

Universidade de Lisboa
Instituto de Geografia e Ordenamento do Território



**Modelo Espacial de Suporte à Decisão:
Criação de Indicadores de Valor Acrescentado para Geomarketing**

Vítor Manuel Jesus Martins

Relatório de Estágio orientado pelo Prof. Doutor Jorge Rocha

Mestrado em Sistemas de Informação Geográfica e Modelação Territorial
Aplicados ao Ordenamento

2020

Universidade de Lisboa
Instituto de Geografia e Ordenamento do Território



**Modelo Espacial de Suporte à Decisão:
Criação de Indicadores de Valor Acrescentado para Geomarketing**

Vítor Manuel Jesus Martins

Relatório de Estágio orientado pelo Prof. Doutor Jorge Rocha

Júri:

Presidente: Doutor Paulo Alexandre Morgado Sousa, Professor Auxiliar do Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa;

Vogais:

- Doutor Jorge Ricardo Costa Ferreira, Professor Auxiliar da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa;
- Doutor Fernando Jorge Pedro da Silva Pinto da Rocha, Professor Auxiliar do Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa.

Índice Geral

Resumo	ix
Abstract	xi
Agradecimentos.....	xiii
Siglas e Acrónimos.....	xv
1. Introdução.....	1
2. Análise de mercado.....	5
2.1 Indicadores.....	5
2.2 Marketing.....	8
2.3 Geomarketing.....	9
2.4 Segmentação dos públicos-alvo.....	12
2.5 Mapidea	14
2.6 Área de estudo	15
3. Geografia e os modelos de localização	17
3.1 Modelos Determinísticos	18
3.1.1 Teoria dos anéis concêntricos	18
3.1.2 Modelo de Hotelling.....	27
3.1.3 Teoria dos Lugares Centrais	32
3.2 Modelos Gravíticos	43
3.2.1 Lei de Reilly.....	43
3.2.2 Modelo de Huff.....	45
3.3 Modelos de Interação Espacial	48
3.3.1 Modelos de localização-alocação.....	48
3.3.1.1 P-Mediana.....	50
3.3.1.2 Maximização da cobertura	55
3.3.1.3 Maximização da participação de mercado (MPM).....	58
3.3.2 Modelo MCI	63
4. Importância dos SIG para a análise espacial	69
4.1 SIG e o Geomarketing	73
4.2 Relevância dos SIG para a criação e análise de indicadores.....	76

5.	Desenvolvimento do modelo	81
5.1	Informação geográfica utilizada	84
5.2	Tratamento da Informação	86
5.3	<i>Clusters</i> de variáveis e indicadores mais relevantes obtidos com os censos ..	90
5.4	Georreferenciação	93
5.4.1	Aplicação da Geocodificação	97
5.5	WebSIG	99
5.5.1	WebSIG da Mapidea	104
5.5.1.1	Utilização do WebSIG da Mapidea	107
6.	Resultados	111
6.1	Análise dos dados mais importantes para a análise da correlação.....	111
6.1.1	População residente	111
6.1.2	Índice de diversidade.....	112
6.2	Regressão.....	113
6.2.1	Aplicação da regressão	115
6.3	Análise Fatorial.....	120
6.4	Desenvolvimento de indicadores com base no modelo fatorial	126
6.5	Análise dos indicadores obtidos com o modelo fatorial	132
6.5.1	Estrutura económica familiar	132
6.5.2	Índice de baixos rendimentos	133
6.5.3	Índice de rendimentos elevados	134
6.5.4	Índice de envelhecimento	135
6.5.5	Diversidade sociocultural	136
6.5.6	Serviços de proximidade	137
6.5.7	Comércio de conveniência	138
6.5.8	Índice de comunicações	139
6.6	Aplicabilidade dos indicadores desenvolvidos	140
7.	Considerações finais.....	145
8.	Bibliografia	149
9.	Anexos.....	157

Índice de Figuras

Figura 2.1 - Relação entre as estratégias de marketing (Adaptado de <i>Chasco-Yrigoyen, 2003</i>).....	9
Figura 2.2 - Evolução das áreas científicas que contribuem para o desenvolvimento do Geomarketing (Adaptado de <i>Cavion & Philips, 2006</i>).....	10
Figura 2.3 - Estados, territórios e áreas de governo local da Austrália.....	16
Figura 3.1 - Rendimento por distância ao mercado (Adaptado de <i>Figueiredo, 2016</i>)...	20
Figura 3.2 - Rendimento obtido de acordo com a intensidade de produção (A) e intensidade de produção face à distância ao mercado (B) (Adaptado de <i>Figueiredo, 2016</i>).....	21
Figura 3.3 - Rendimento de cada produto em função da distância (Adaptado de <i>Figueiredo, 2016</i>).....	22
Figura 3.4 - Anéis de produção agrícola em função da distância à cidade central (Adaptado de <i>Figueiredo, 2016</i>).....	23
Figura 3.5 - Sistemas de anéis de produção (Adaptado de <i>Fischer, 2011</i>).	24
Figura 3.6 - Teoria dos anéis concêntricos (Adaptado de <i>Fischer, 2011</i>).....	25
Figura 3.7 - Influência da cidade pequena e do rio no padrão de produção (A) e efeito da variação do preço do grão no padrão de produção (B) (Adaptado de <i>Figueiredo, 2016</i>).	26
Figura 3.8 - Funções de preço dos dois competidores em um segmento de reta (Adaptado de <i>Eiselt, 2011</i>).	28
Figura 3.9 - Ajustes de preço ao longo do tempo (Adaptado de <i>Eiselt, 2011</i>).....	30
Figura 3.10 - Áreas de mercado hexagonais (Adaptado de <i>Fischer, 2011</i>).....	35
Figura 3.11 - Estrutura hierárquica dos lugares centrais (Adaptado de <i>Christaller, 1966</i>).	36
Figura 3.12 - Princípio de transporte ($K = 4$) (Adaptado de <i>Fischer, 2011</i>).....	37
Figura 3.13 - Princípio administrativo ($K = 7$) (Adaptado de <i>Christaller, 1966</i>).....	38
Figura 3.14 - Dimensão hexagonal das menores áreas de mercado (Adaptado de <i>Fischer, 2011</i>).....	39
Figura 3.15 - Segunda menor área de mercado ($K = 4$) (Adaptado de <i>Fischer, 2011</i>). ...	40
Figura 3.16 - Padrão das áreas de mercado de Lösch (Adaptado de <i>Fischer, 2011</i>).....	41

Figura 3.17 - Relação entre a Lei de Reilly e a Lei da gravitação universal.	44
Figura 3.18 - Localização ideal para as instalações segundo o modelo p-mediana (Adaptado de <i>ReVelle & Swain, 1970</i>).....	51
Figura 3.19 - Localização ideal para as instalações segundo o modelo da maximização da cobertura (Adaptado de <i>Church & ReVelle, 1974</i>)	56
Figura 3.20 - Localização ideal para as instalações segundo o modelo da maximização da participação de mercado (Adaptado de <i>Revelle, 1986</i>).	60
Figura 4.1 - Sistema SIG (Adaptado de <i>Longley, 2008</i>).	70
Figura 4.2 - Combinação dos dados do rendimento semanal por família (em dólares australianos) com as estações de metro, estradas, supermercados e ATM da cidade de Melbourne.	71
Figura 4.3 - Funções de SIG que podem ser aplicadas no âmbito da análise dos indicadores.	78
Figura 5.1 - Estrutura do modelo.....	82
Figura 5.2 - Conjunto de procedimentos aplicados.	83
Figura 5.3 - Exemplificação do processo de geocodificação.	95
Figura 5.4 - Exemplos de camadas de informação que um mapa pode incluir e possível alinhamento.....	96
Figura 5.5 - Relação servidor-clientes estabelecida pelo WebSIG (Adaptado de <i>Fu, 2015</i>).	101
Figura 5.6 - Potencial arquitetura de um WebSIG simples (Adaptado de <i>Fu, 2015</i>). ..	102
Figura 5.7 - Comunicação dos documentos entre a framework e os clientes do WebSIG (Adaptado de <i>Fu, 2015</i>).	104
Figura 5.8 - Interface gráfica do WebSIG da Mapidea.	105
Figura 5.9 - Demonstração das funções disponibilizadas pelo WebSIG da Mapidea. .	106
Figura 5.10 - Apresentação da opção para exportar dados do WebSIG.	107
Figura 5.11 - Grelha com polígonos quadrados de 100 km ² para o conjunto da Austrália.	108
Figura 5.12 - Grelha com polígonos quadrados de 250 m ² para o centro da cidade de Melbourne.	109
Figura 6.1 - População residente nas áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.....	111

Figura 6.2 - Índice de diversidade para as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.....	112
Figura 6.3 - Exemplo de distribuição de observações de forma linear (A) e de distribuição de observações de forma dispersa (B)	118
Figura 6.4 - Exemplo de explicação das variáveis através de fatores comuns e específicos (Adaptado de <i>Quaglino & Pagura, 2014</i>).	121
Figura 6.5 - Estrutura económica familiar consoante as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.	132
Figura 6.6 - Índice de baixos rendimentos para as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.	133
Figura 6.7 - Índice de rendimentos elevados para as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.....	134
Figura 6.8 - Índice de envelhecimento para as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.	135
Figura 6.9 - Diversidade sociocultural consoante as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.	136
Figura 6.10 - Serviços de proximidade consoante as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.....	138
Figura 6.11 - Comércio de conveniência consoante as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.....	139
Figura 6.12 - Índice de comunicações para as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.	140
Figura 6.13 - Instalações escolhidas para a localização de um negócio através do modelo p-mediana.....	142
Figura 6.14 - Instalações escolhidas para a localização de um negócio através do modelo da maximização da cobertura.	143
Figura 6.15 - Instalações escolhidas para a localização de um negócio através do modelo da maximização da participação de mercado.	144

Índice de Quadros

Quadro 5.1 - Informações secundárias das variáveis retiradas dos clusters de variáveis e respectivas exceções.....	87
Quadro 5.2 - Clusters de variáveis dos censos removidos por serem considerados pouco relevantes para o marketing.	89
Quadro 5.3 - Grupos de variáveis removidas por serem consideradas pouco relevantes para o marketing.	89
Quadro 5.4 - Clusters de variáveis mais relevantes obtidos através da seleção de variáveis dos censos.	91
Quadro 5.5 - Organização dos dígitos que compõem os códigos dos censos para as áreas estatísticas SA1 (Adaptado dos censos da Austrália, 2016).....	98
Quadro 5.6 - Organização dos dígitos que compõem os códigos dos censos para as áreas estatísticas SA2 (Adaptado dos censos da Austrália, 2016).....	98
Quadro 6.1 - Variáveis colineares, excluídas através do modelo de regressão.....	117
Quadro 6.2 - Resultado da medida de Kaiser-Meyer-Olki e do teste da esfericidade de Bartlett.....	126
Quadro 6.3 - Representação das variáveis influenciadas pelo fator 1.....	127
Quadro 6.4 - Representação das variáveis influenciadas pelo fator 2.....	128
Quadro 6.5 - Representação das variáveis influenciadas pelo fator 3.....	128
Quadro 6.6 - Representação das variáveis influenciadas pelo fator 4.....	129
Quadro 6.7 - Representação da variável e do indicador influenciado(a) pelo fator 5.	129
Quadro 6.8 - Representação das variáveis influenciadas pelo fator 6.....	130
Quadro 6.9 - Representação das variáveis influenciadas pelo fator 7.....	130
Quadro 6.10 - Representação das variáveis influenciadas pelo fator 8.....	131

Resumo

Após identificarem os segmentos de mercado mais propensos a adquirir/utilizar um produto, os especialistas de marketing têm normalmente dificuldade em delimitar espacialmente os lugares onde predominam esses segmentos. Para o fazer, necessitam de indicadores que tornem o processo cada vez mais célere e eficaz, condições essenciais para decidir onde se deve apostar na oferta ou promoção de um determinado produto.

Assim, este relatório tem como principal objetivo desenvolver um modelo que permita identificar novos indicadores mais adequados para caracterizar uma determinada área de mercado. Como etapas intercalares pretende-se: i) Sintetizar as variáveis/indicadores mais pertinentes para a área de mercado em análise; ii) Transformar as variáveis/indicadores em indicadores de valor acrescentado; iii) Identificar os lugares com maior potencial de negócio; iv) Desenvolver um sistema espacial de apoio à promoção e seleção do conjunto de produtos que mais se adequam a cada lugar, com base em indicadores de valor acrescentado; e v) Explorar o sistema espacial de apoio à decisão para fomentar uma forma de marketing inteligente, i.e., que tenha por base as características específicas de cada área de mercado.

O modelo foi desenvolvido para a área de mercado da Austrália devido à grande disponibilidade de dados recentes para os diferentes níveis geográficos e à constatação de características que fazem com que se verifique uma maior diversidade de preferências dos consumidores, como a existência de uma população étnica ou racial diversa, a distribuição assimétrica da população e o elevado rendimento dos habitantes.

Para o desenvolver, sintetizaram-se as variáveis/indicadores sociodemográficos(as) e económicos(as) mais pertinentes para a área de mercado da Austrália, para que, através da aplicação de procedimentos estatísticos como a regressão e a análise fatorial, se pudesse transformar essas variáveis/indicadores em indicadores de valor acrescentado para o contexto de mercado da Austrália. Desta forma, foi possível obter oito indicadores de valor acrescentado que foram posteriormente representados em mapas, com base na utilização dos Sistemas de Informação Geográfica. Por fim, para a cidade de Melbourne, identificaram-se os lugares com maior potencial de negócio, tendo em conta os oito indicadores de valor acrescentado obtidos.

Este procedimento, por assumir uma estrutura flexível, poderá vir a ser aplicado pelos especialistas de marketing nos mais diversos contextos de negócio.

Palavras chave: Geomarketing, indicadores, análise espacial, segmentação, SIG

Abstract

After identifying the market segments most likely to purchase/use a product, marketing specialists usually have difficulty in defining spatially the places where these segments are located. To do this, they need indicators that make the process faster and more efficient. These are essential conditions to decide where to bet on the offer or promotion of a certain product.

Thus, this report has as a main objective to develop a model that allows the identification of new more appropriate indicators which to characterize a given market area. To achieve this objective, the following steps were established: i) Synthetization of the most relevant variables/indicators for the market area under analysis; ii) Transformation of the variables/indicators into higher value indicators; iii) Identification of the places with the greatest business potential; iv) Development of a space system supporting the promotion and selection of the sets of products that best suit each place, based on higher value indicators; and v) Exploration of the space system that supports the decision, to promote a form of intelligent marketing, i.e., that it is based on the specific characteristics of each market area.

The model was developed for the Australian market area due to the large availability of recent data for different geographical levels and due to the different population characteristics', that allows wider choices in preferences. Some of these characteristics are the existence of an ethnic or racial diversified population, the asymmetric distribution of the population and the high income of the inhabitants.

To develop the model, the most relevant sociodemographic and economic variables/indicators for the Australian market area were synthesized, through the application of statistical procedures such as regression and factor analysis, in order to transform these variables/indicators in higher value indicators for the Australian market context. This way was possible to obtain eight higher value indicators that were later represented on maps, based on the use of Geographic Information Systems. Finally, for the city of Melbourne, the places with the greatest business potential were identified, considering the eight higher value indicators obtained.

As this procedure has an assuming flexible structure, it can be applied by marketing specialists in the most diverse contexts of business.

Keywords: Geomarketing, indicators, spatial analysis, segmentation, GIS

Agradecimentos

Quero, em primeiro lugar, agradecer ao Professor Doutor Jorge Rocha por me ter incentivado a seguir em frente com este trabalho e por ter estado sempre disponível para me apoiar ao longo da realização do mesmo enquanto meu orientador. Sem a sua preciosa ajuda teria tido maior dificuldade em chegar a estes resultados.

Em seguida, quero agradecer à empresa Mapidea por me ter proporcionado este estágio e por me ter disponibilizado a sua aplicação, que foi fundamental para tratar e analisar os dados utilizados. Agradeço também ao professor Miguel Marques e aos colaboradores da empresa Mapidea pela simpatia, pela ajuda e orientação que me prestaram e pelos conhecimentos que me transmitiram, que foram fundamentais para a realização deste trabalho.

Quero também agradecer aos meus pais, Rodrigo Martins e Vanda Vicente, por me terem apoiado nos bons e nos maus momentos, pela preocupação que sempre tiveram para comigo e por me terem dado motivação ao longo da realização deste trabalho.

Por fim, quero agradecer aos meus amigos, Tiago, Mariana e Diogo, pela preocupação e pelo apoio que desde sempre demonstraram, assim como também quero agradecer à minha melhor amiga Margarida, pela ajuda, preocupação e conforto que me dedicou durante este último ano.

Siglas e Acrónimos

API - *Application Programming Interface* [Interface de programação de aplicações]

ATM - *Automated Teller Machine* [Caixa de multibanco]

CBD - *Central Business District* [Área central de negócios de uma cidade]

CSS - *Cascading Style Sheets* [Mecanismo que permite determinar as características de apresentação de uma página da internet]

CSV - *Comma Separated Values* [Ficheiro de texto que permite armazenar dados]

ESRI - *Environmental Systems Research Institute* [empresa de *software* SIG]

GML - *Geographic Markup Language* [Linguagem de marcação geográfica]

GPS - *Global Positioning System* [Sistema de posicionamento global]

HTML - *Hypertext Markup Language* [Linguagem de marcação de hipertexto]

HTTP - *Hypertext Transfer Protocol* [Protocolo de transferência de hipertexto]

KML - *Keyhole Markup Language* [Linguagem de marcação que permite armazenar dados geográficos]

KMO - Medida de *Kaiser-Meyer-Olki*

LGA - *Local Government Areas* [Áreas de governo local]

MCI - *Multiplicative Competitive Interaction* [Modelo multiplicativo de interação competitiva]

MPM - Modelo da maximização da participação de mercado

QGIS - *Quantum GIS* [*software* SIG]

SIG - Sistemas de Informação Geográfica

URL - *Uniform Resource Locator* [localizador uniforme de dados alojados num servidor]

XML - *Extensible Markup Language* [Linguagem de marcação de dados]

1. Introdução

Tendo em conta que o sucesso do negócio de um produto é essencial para a sobrevivência e para o valor da empresa que o produz, cada vez mais empresas desenvolvem e empreendem estratégias de marketing inovadoras com a finalidade de promover e adequar os seus produtos aos potenciais consumidores, por forma a conseguirem aumentar o seu volume de negócios.

Essas estratégias, para serem eficazes, têm de acompanhar os gostos e as preferências dos potenciais consumidores. Por isso, a maioria dos especialistas de marketing das empresas analisam os hábitos de consumo e as principais características dos seus clientes para poderem identificar os segmentos de mercado mais propensos a adquirir um produto. Contudo, após realizarem a identificação desses segmentos, os especialistas de marketing têm normalmente dificuldade em delimitar espacialmente os lugares onde predominam esses segmentos.

Essa dificuldade deve-se principalmente à incapacidade de sintetizarem as informações mais importantes para a caracterização da população de um determinado lugar ou território, de articularem essas informações espaciais com os segmentos de mercado identificados e de estabelecerem posteriormente os limites espaciais onde predominam esses segmentos.

Para contornar estas incapacidades, torna-se relevante desenvolver métodos e técnicas que permitam sintetizar informações (indicadores) de valor acrescentado com base na avaliação do potencial de procura de um produto e que possibilitem a sua representação espacial para cada lugar, por forma a poder-se delimitar de forma simples as áreas de mercado onde predominam os segmentos de mercado de um produto, e onde, por isso, haverá maior potencial de negócio de um produto. Com a determinação dessas áreas, com base nos indicadores de valor acrescentado, é possível fundamentar e apoiar a decisão de localizar a oferta ou promoção de um produto.

Por isso, este relatório de estágio tem como principal objetivo desenvolver um modelo que permita identificar novos indicadores mais adequados para caracterizar uma determinada área de mercado, por forma a facilitar a tarefa de encontrar os lugares com maior concentração dos segmentos de mercado que têm maior potencial para adquirir/utilizar um produto.

Para ser possível cumprir esse objetivo, definiu-se as seguintes etapas intercalares:

- i) Sintetizar as variáveis/indicadores mais pertinentes para a área de mercado em análise;
- ii) Transformar as variáveis/indicadores em indicadores de valor acrescentado;

- iii) Identificar os lugares com maior potencial de negócio;
- iv) Desenvolver um sistema espacial de apoio à promoção e seleção do conjunto de produtos que mais se adequam a cada lugar, com base em indicadores de valor acrescentado;
- v) Explorar o sistema espacial de apoio à decisão para fomentar uma forma de marketing inteligente, i.e., que tenha por base as características específicas de cada área de mercado.

Para o desenvolvimento deste modelo, foi então necessário sintetizar as variáveis/indicadores sociodemográficos(as) e económicos(as) mais pertinentes para a caracterização da área de mercado da Austrália, obtidos(as) com base nos dados dos censos do país, do *OpenStreetMap* e do *OpenCellID*.

Posteriormente, procedeu-se à aplicação de procedimentos estatísticos como o modelo de regressão, que permitiu remover as variáveis/indicadores que poderiam perturbar a análise da correlação, e o modelo fatorial, que permitiu identificar os principais fatores que influenciam a variação dos valores dos(as) variáveis/indicadores em cada unidade estatística. Com base nesses fatores obtidos e na aplicação de uma fórmula, foi possível transformar as variáveis/indicadores obtidos(as) em indicadores de valor acrescentado para o contexto de mercado da Austrália.

Em seguida, com recurso às técnicas de Sistemas de Informação Geográfica, representaram-se em mapas temáticos os indicadores de valor acrescentado obtidos, que permitem caracterizar as populações de cada unidade estatística.

Por fim, com base nos indicadores de valor acrescentado obtidos e na aplicação dos modelos de localização-alocação p-mediana, maximização da cobertura e maximização da participação de mercado, foi possível identificar os lugares da cidade de Melbourne com maior potencial de negócio.

Este relatório de estágio está organizado em sete capítulos. O primeiro desses capítulos corresponde à introdução.

No segundo capítulo, designado de análise de mercado, começa-se por explicar o que são indicadores e qual a sua importância para uma análise de mercado adequada. Em seguida são introduzidos e desenvolvidos os conceitos de marketing e de geomarketing, importantes para compreender o contexto em que se aplica este modelo. Posteriormente é abordada a segmentação dos públicos-alvo, em que se especifica a importância de segmentar os consumidores de um produto para prever o potencial de negócio e de lucro do mesmo dentro de um grupo de consumidores e, conseqüentemente, por lugar considerado. É ainda referida a relevância da aplicação desenvolvida pela empresa Mapidea para a análise de mercado e a área de estudo que serviu de base ao desenvolvimento do modelo.

No terceiro capítulo começa-se por debater a importância da geografia e da informação geográfica para o desenvolvimento dos modelos de localização espacial, e por indicar as categorias de modelos de localização espacial existentes. Depois são abordados, um a um, os principais modelos de cada uma das categorias de modelos referidas e qual a importância de cada um para determinar a melhor localização para realizar a oferta de mercado.

Já no quarto capítulo descreve-se o que são os SIG e qual a importância dos mesmos para a análise espacial, sendo relatados alguns exemplos de funções que os SIG permitem aplicar. É também referida a pertinência dos SIG para o geomarketing e para a criação e análise de indicadores em específico.

No quinto capítulo, que corresponde ao desenvolvimento do modelo, começa-se por apresentar a estrutura do modelo e os dados utilizados na aplicação do mesmo. Posteriormente são explicados os procedimentos a aplicar para sintetizar as variáveis/indicadores mais pertinentes para a área de mercado analisada e procede-se à apresentação das variáveis/indicadores obtidos(as). Em seguida é ilustrada a importância da georreferenciação para articular nos SIG as informações espaciais a uma determinada localização geográfica e é explicado o processo de geocodificação dos dados das variáveis/indicadores obtidos(as), processo efetuado por forma a incluir esses dados na aplicação da Mapidea e para poderem posteriormente ser representados em mapas temáticos. É ainda explicado o que é um WebSIG e quais as componentes que o compõem, para, por fim, se poder apresentar o WebSIG da Mapidea e explicar os procedimentos que foram aplicados no mesmo.

No sexto capítulo, correspondente aos resultados, começa-se por analisar os dados mais importantes para a análise da correlação que será aplicada no modelo de regressão e na análise fatorial. Em seguida, apresentam-se os pressupostos do modelo de regressão e do modelo fatorial, procedendo-se também à explicação dos procedimentos aplicados no âmbito dos modelos para conseguir transformar as variáveis/indicadores obtidos(as) em indicadores de valor acrescentado. Posteriormente, procede-se à apresentação do desenvolvimento dos indicadores de valor acrescentado e analisa-se os dados dos indicadores obtidos através de mapas temáticos que os representam. Por fim, para demonstrar a aplicabilidade do modelo, aplicaram-se os modelos de localização-alocação para identificar os lugares da cidade de Melbourne com maior potencial de negócio.

No último capítulo, o sétimo, são apresentadas as considerações finais deste relatório.

2. Análise de mercado

2.1 Indicadores

A medição do potencial de negócio de um produto é cada vez mais importante para melhorar e incrementar o comércio/utilização de produtos (Franceschini et al., 2007).

Um indicador, segundo a norma UNI 11097 do Organismo Nacional de Unificação de Itália, consiste numa informação acerca de um fenómeno, processo ou resultado em análise, que permite examinar a sua evolução e apurar se os objetivos do mesmo são conseguidos, por forma a facilitar as ações e decisões a tomar, podendo essa informação assumir uma forma qualitativa como quantitativa (ou as duas em simultâneo). Os indicadores permitem então medir o potencial de negócio de um produto.

Assim, a importância dos indicadores no negócio de produtos tem vindo a crescer, principalmente quando utilizados para facilitar a definição de estratégias das empresas (indicadores de previsão) e para medir o efeito das mesmas (indicadores de avaliação) (Franceschini et al., 2007). Esse crescimento deve-se à:

- Necessidade de gerir uma oferta cada vez mais complexa;
- Redução do período de vida dos produtos e serviços;
- Existência de um maior volume de dados;
- Maior número de técnicas de suporte à decisão que utilizam indicadores (Franceschini et al., 2007).

Tendo em conta tais mudanças, se uma empresa não rever os indicadores utilizados, corre o risco de vir a ter uma estratégia de negócio obsoleta (Franceschini et al., 2007). Isto leva ao aumento da importância dos indicadores.

Os indicadores, segundo Franceschini et al. (2007), devem:

- Ser representativos;
- Possibilitar uma utilização e interpretação simples;
- Conseguir evidenciar as tendências temporais;
- Ter em conta as mudanças que ocorrem;
- Implicar uma recolha e processamento de dados acessível;
- Permitir a atualização dos mesmos de forma fácil e rápida.

A forma como é feita a oferta dos produtos tem uma grande importância para o sucesso de uma empresa (Khan, 2006). Para que o negócio de um produto seja bem-sucedido, as empresas devem ter em conta, sempre que possível, as emoções, a personalidade, sentimentos e posição social dos potenciais consumidores para que mais facilmente consigam compreender as suas necessidades, e com isso poderem aumentar o consumo dos produtos da empresa (Dyer, 1982; Hackley, 2005; Khan, 2006). Assim, os indicadores, ao informarem as características e as preferências de grupos de potenciais

consumidores, tornam-se fundamentais para o comércio de produtos, dado permitirem com essas informações adaptar a oferta de produtos em função do lugar e do público-alvo (Khan, 2006).

A criação e análise de indicadores pode então ajudar as empresas a definir estratégias que permitam:

- Potenciar a procura pelos produtos (atrair novos clientes);
- Promover a empresa que disponibiliza os produtos;
- Informar o preço e as características dos produtos aos potenciais consumidores;
- Divulgar um produto de forma criativa e convincente;
- Convencer os potenciais consumidores das vantagens dos produtos;
- Tornar os seus produtos competitivos face aos produtos disponibilizados por outras empresas;
- Garantir a distribuição dos produtos;
- Aumentar a compra/utilização dos produtos;
- Expandir a área de mercado dos produtos (Hackley, 2005; Gallopel, 2006; Khan, 2006).

Analisar os consumidores de um produto através de indicadores, para além de poder ajudar a potenciar um produto, pode, com a atração dos consumidores ao local de negócio, contribuir para o aumento do consumo dos restantes produtos do local de negócio (Hackley, 2005; Khan, 2006).

Normalmente, as empresas analisam os potenciais consumidores de um produto quando:

- Surge um novo produto ou uma nova funcionalidade do mesmo;
- Querem tornar um produto competitivo;
- Existe uma mudança no nome da empresa que disponibiliza o produto;
- É aberto um novo local de negócio;
- Se faz uma promoção ou atualização do produto (Khan, 2006).

Ao analisarem as características e as preferências dos grupos de potenciais consumidores, os indicadores permitem também avaliar o potencial de concretização dos objetivos de negócio de uma empresa (Franceschini et al., 2007).

Se uma empresa define uma estratégia com objetivos de negócio sem ter em conta indicadores, essa estratégia perde a utilidade, dado não ser possível verificar se a estratégia permitiu obter resultados positivos (ou não) na concretização dos objetivos (Franceschini et al., 2007). Assim, ter em conta os indicadores na estratégia de negócio e avaliar o seu desempenho, é fundamental para se obter a estratégia que permite atingir melhores resultados (Franceschini et al., 2007).

Uma análise bem-sucedida dos potenciais clientes de uma empresa, essencial em qualquer estratégia de negócio, implica então que não se fique apenas pela simples escolha de alguns indicadores, mas que se identifiquem os indicadores que refletem o potencial dos consumidores comprar ou utilizar um produto, os chamados indicadores de valor acrescentado (Franceschini et al., 2007; Vitanov, 2016).

Assim, os indicadores de valor acrescentado devem contribuir para:

- Comunicar o efeito que um dado fator tem para o negócio do produto;
- Avaliar e controlar o desempenho atual ou potencial de negócio de um produto;
- Melhorar o desempenho de negócio de um produto e da empresa como um todo, facilitando a identificação das falhas existentes (Franceschini et al., 2007).

A identificação dos indicadores de valor acrescentado mais adequados para analisar os clientes de uma empresa ou consumidores de um produto vai depender das:

- Características dos produtos de oferta;
- Público-alvo em que se quer potenciar o produto;
- Número de pessoas a que se quer disponibilizar o produto;
- Objetivos de negócio definidos para cada produto e para o conjunto da empresa;
- Marketing adotado (Hackley, 2005; Khan, 2006).

Isto pressupõem a necessidade de uma forte articulação na empresa entre quem oferece o produto e quem analisa o respetivo mercado do mesmo (Khan, 2006).

Os indicadores de valor acrescentado podem variar consoante as características do lugar (área urbana ou rural), a idade das pessoas (jovens ou idosos), sexo e o rendimento das pessoas a que se quer potenciar o produto (Khan, 2006).

A criação e análise de indicadores de valor acrescentado deve ter em conta as preferências dos potenciais consumidores de um produto (Gallopel, 2006). Para identificar essas preferências é então essencial efetuar uma análise das características sociodemográficas e económicas dos atuais clientes e da sua respetiva localização, por forma a encontrar um padrão de eventuais consumidores de um produto que permita determinar um conjunto de preferências dos mesmos.

Com as preferências dos potenciais consumidores é possível analisar quem compra/utiliza um produto, o que compram/utilizam e como compram/utilizam os produtos, o que é essencial para a determinação do padrão desses possíveis consumidores e para a formulação de indicadores de valor acrescentado que descrevam o respetivo padrão (Khan, 2006). Com esse padrão, as empresas podem selecionar as áreas onde é mais conveniente efetuar a oferta de um produto (Gallopel, 2006).

O processo de criação e análise de indicadores de valor acrescentado, essencial para os especialistas de marketing de uma empresa poderem realizar uma adequada

potencialização de um produto, será demonstrado ao longo da representação do modelo desenvolvido.

2.2 Marketing

Os comportamentos e hábitos de consumo estão sempre em constante mudança (Kotler, 2000). A oferta e as estratégias de negócio de produtos devem acompanhar essas mudanças, identificando e satisfazendo as necessidades humanas em cada momento, para que as empresas continuem a ser lucrativas (Kotler, 2000).

O marketing pode então ajudar a definir essas estratégias para as empresas, dado este consistir num processo em que as empresas tentam compreender os comportamentos dos consumidores, por forma a facilitar a criação e a comunicação da oferta de produtos que lhes sejam adequados (Kotler, 2000; Kotler & Armstrong, 2018).

O objetivo do marketing é então obter conhecimentos consideráveis acerca dos potenciais consumidores de um mercado-alvo, para que o produto disponibilizado por uma empresa lhes seja adequado (Kotler, 2000). No processo de gestão de marketing analisam-se os comportamentos e o perfil de alguns clientes atuais (por exemplo: idade, sexo, rendimento) para se poder escolher o público-alvo (através da distribuição dos clientes por segmentos consoante as suas características e comportamentos) que possa manter ou aumentar o número de clientes de uma dada empresa (Kotler, 2000; Cliquet, 2006).

Com base no conhecimento do público-alvo escolhido e na identificação da sua distribuição espacial, o marketing define então estratégias que têm em conta a adequação do produto em oferta, a forma como deve ser feita e comunicada a oferta do produto, o preço que o produto deve ter (tendo em conta os custos de produção, o segmento de mercado, a procura e os objetivos de diferenciação do produto) e a localização da oferta (figura 2.1) (Kotler, 2000; Chasco-Yrigoyen, 2003).

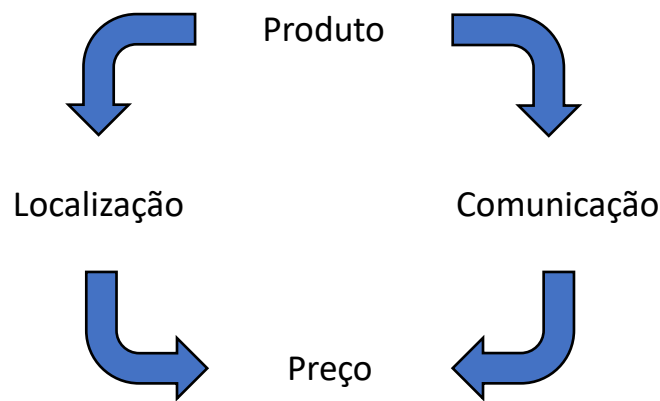


Figura 2.1 - Relação entre as estratégias de marketing (Adaptado de *Chasco-Yrigoyen, 2003*).

Os decisores de marketing, com base na localização dos consumidores e nas características do público-alvo definido, poderão, para aumentar o consumo dos produtos, escolher os territórios a realizar a oferta do produto, consoante o potencial de consumo (Cliquet, 2006). Desta forma, torna-se relevante para os especialistas de marketing medir o potencial de consumo das diversas áreas de mercado. Para medir esse potencial é preciso determinar os indicadores mais adequados para o efeito. Assim, por forma a facilitar essa tarefa, este relatório vai tentar simplificar o processo de determinação e análise dos indicadores mais adequados para medir o potencial consumo dos produtos.

2.3 Geomarketing

Os consumidores de um produto estão localizados num espaço geográfico preciso (Cliquet, 2006). Assim, uma empresa, para aumentar o consumo dos seus produtos, deve conhecer a sua área de mercado porque:

- A procura por um produto varia consoante o espaço analisado;
- A existência de locais de negócio e a disponibilidade dos produtos variam consoante o local;
- A procura e a oferta de um produto podem se encontrar em locais distintos (Cliquet, 2006).

Estes factos fazem com que uma empresa deva analisar a distribuição espacial dos consumidores, para poder-se aproximar dos locais de procura do produto que disponibiliza, e assim poder aumentar o consumo dos mesmos (Dion & Cliquet, 2006). A informação espacial acerca dos consumidores pode então contribuir para a definição das estratégias de negócio das empresas (Cliquet, 2006). Neste contexto, as empresas

terão como objetivo conquistar novos locais para disponibilizarem os seus produtos e tentarem cobrir o maior número de potenciais consumidores (Cliquet, 2006).

Deste modo, com o desenvolvimento da relação entre o marketing e a geografia, com o surgimento dos modelos de localização espacial - essenciais para compreender como os consumidores escolhem um dado local de negócio para comprar os produtos, e com o aparecimento dos Sistemas de Informação Geográfica foi possível, na década de 1980, o nascimento do conceito de geomarketing (figura 2.2) (Cavion & Philips, 2006; Cliquet, 2006).

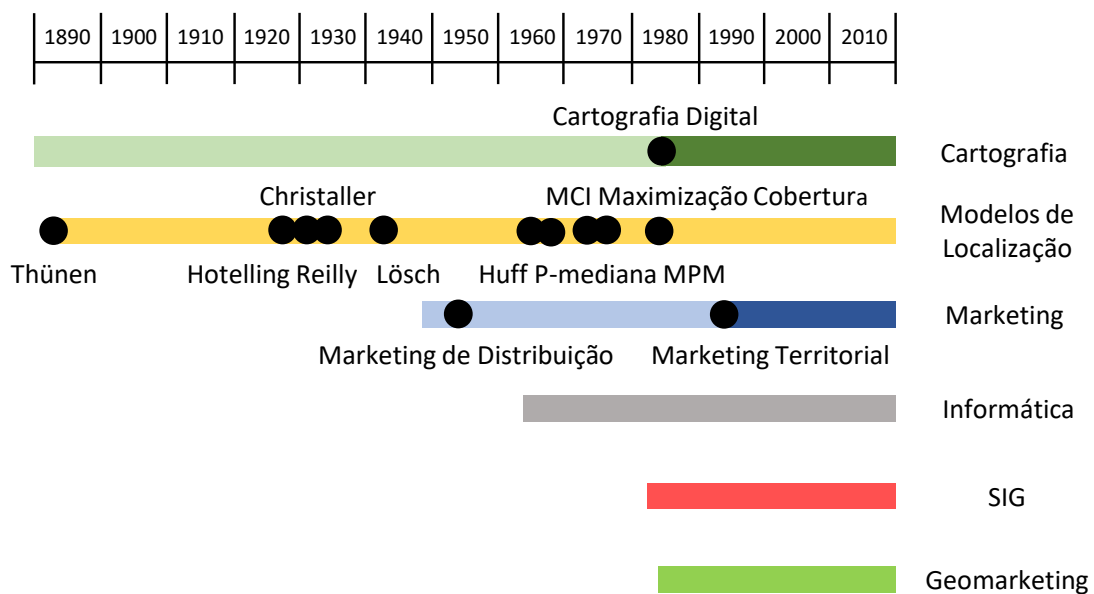


Figura 2.2 - Evolução das áreas científicas que contribuem para o desenvolvimento do Geomarketing (Adaptado de Cavion & Philips, 2006).

O geomarketing consiste num conjunto de técnicas que permitem a manipulação de dados cartográficos e estatísticos para a análise do contexto sociodemográfico e económico de uma área geográfica, com o objetivo de obter informações úteis para o marketing de uma empresa ou produto (Chasco-Yrigoyen, 2003; Cliquet, 2006).

Este pressupõe a relação de características do marketing como o produto, a comunicação, distribuição e o preço com a perspetiva espacial associada aos mesmos, por forma a entender melhor o impacto de uma possível decisão de marketing (Chasco-Yrigoyen, 2003).

A utilização do geomarketing tem como objetivo gerir múltiplas informações dos consumidores e dos lugares em análise, dominar os territórios alvo da estratégia de

marketing, adequar os produtos das empresas aos clientes e prever a evolução dos diversos fatores que influenciam o consumo do produto (Douard, 2006).

Inicialmente, o geomarketing começou por fazer representações cartográficas de informações existentes de censos demográficos, rendimento e infraestruturas (Douard, 2006). A crescente disponibilidade de bases de dados socioeconómicas de cariz espacial (obtidas através de censos populacionais mais completos, de trabalhos de investigação ou de informações disponibilizadas pelas empresas ou entidades públicas), assim como também a necessidade de compreender melhor a área de mercado e os segmentos de consumidores, possibilitaram um maior desenvolvimento da aplicação do geomarketing (Chasco-Yrigoyen, 2003; Douard, 2006). Esse desenvolvimento do geomarketing fez surgir o marketing relacional, que permitiu o desenvolvimento de estratégias inovadoras para as empresas, tendo em conta o cruzamento de diferentes informações que caracterizam os consumidores (Douard, 2006).

As técnicas de geomarketing permitem que a informação geográfica possa ser adaptada para se obterem dados relativos a domínios de marketing como o comportamento dos consumidores, localização da oferta dos produtos e gestão do marketing (Cliquet, 2006).

Segundo Chasco-Yrigoyen (2003) e Douard (2006), o geomarketing pode então ser aplicado nas seguintes áreas:

- Criação de novos locais de negócio;
- Avaliação dos locais de negócio existentes;
- Impacto da localização da concorrência no negócio de produtos;
- Adaptação da oferta comercial;
- Micromercados;
- Fixação de objetivos em relação ao potencial de negócio;
- Adaptação das campanhas publicitárias por forma a atrair os potenciais consumidores;
- Melhorar os investimentos em marketing direto;
- Ajudar na potencialização do negócio de produtos.

Com a utilização das técnicas do geomarketing é então facilitada a análise dos perfis de potenciais consumidores e a sua respetiva segmentação, dado possibilitarem a combinação das informações económicas e sociodemográficas com dados do consumo de produtos, poder de compra e localização dos consumidores (Douard, 2006).

A aplicação do geomarketing na estratégia de negócio pode ser útil para a adequação da oferta de produtos ao público-alvo, dado o potencial de consumo poder variar consoante o local e as características do público-alvo do produto (Chasco-Yrigoyen, 2003). Por isso, este relatório vai desenvolver um modelo que, ao permitir identificar mais facilmente a presença de públicos-alvo de um produto num dado território, possa tornar mais simples a aplicação das técnicas de geomarketing na estratégia de negócio.

2.4 Segmentação dos públicos-alvo

A procura para a maior parte dos produtos varia consoante as características e comportamentos dos consumidores, ou seja, varia consoante o seu rendimento, hábitos de consumo, necessidades, desejos, estilo de vida, atitudes, valores culturais, entre outros (Khan, 2006; Petr, 2006; O'Kelly, 2009; Widaningrum, 2015). Essas características permitem distinguir os consumidores e agrupá-los em diferentes segmentos consoante o seu potencial de procura (Cooper & Nakanishi, 1988).

Se uma empresa não segmentar os consumidores de um produto poderá estar a produzir um produto e a utilizar uma estratégia de marketing que tente agradar a todos os consumidores, o que pode levar a que ninguém depois goste do produto por não se identificar com o mesmo (Khan, 2006).

A segmentação dos consumidores permite às empresas desenvolver os produtos e definir estratégias de marketing que melhor consigam atrair as pessoas que têm maior probabilidade de adquirir/utilizar um produto, dado a forma do produto e as estratégias de marketing poderem ter efeitos diferentes consoante os vários segmentos de consumidores (Cooper & Nakanishi, 1988; Khan, 2006; Petr, 2006; O'Kelly, 2009). Por exemplo, uma estratégia de marketing vocacionada para os jovens aplicada num produto destinado a pessoas jovens teria melhores resultados do que se a mesma estratégia fosse aplicada a um produto destinado a pessoas idosas. Assim, a segmentação dos consumidores é essencial no desenvolvimento dos produtos de uma empresa e em qualquer estratégia de marketing, dado assim poderem-se satisfazer melhor as necessidades ou desejos específicos dos diferentes consumidores (Khan, 2006; Petr, 2006).

Por isso, as empresas, com base nas características e comportamentos dos consumidores, podem segmentar os consumidores de cada produto consoante grupos de idade, sexo, rendimento, nacionalidade, etnia, escolaridade, profissão, estado civil etc., aos quais estão associadas determinadas características e comportamentos que definem os consumidores, e escolher um ou mais segmentos que melhor caracterizam o potencial de compra/utilização de um produto, com base na distribuição da procura existente nos vários segmentos de consumidores, nas necessidades a que o produto pode responder e no lucro que pode gerar (Birkin et al., 1999; Douard, 2006; Khan, 2006; O'Kelly, 2009; Berman & Evans, 2013; Montejano & Bello, 2018).

Devem ser escolhidos os segmentos de consumidores que permitam uma maior distinção entre os consumidores que mais consomem e os que menos consomem um dado produto, em que o produto pode satisfazer as necessidades dos segmentos de consumidores considerados e que levem a um maior lucro potencial (Khan, 2006; Montejano & Bello, 2018). Dentro de cada segmento, a variação do potencial de compra/utilização deve ser menor do que a variação do potencial de compra/utilização

para o conjunto dos diferentes segmentos, para que se possa prever a procura e o lucro potencial em cada segmento (Khan, 2006).

Para a escolha dos segmentos de consumidores, as empresas recolhem e analisam os dados demográficos e socioeconómicos dos seus clientes, consoante os produtos adquiridos por estes e a frequência com que o fazem, para poderem obter os segmentos de consumidores que mais consomem ou poderão consumir um dado produto, e assim conseguirem definir a forma do produto e as estratégias de negócio mais adequadas para os segmentos com maior procura potencial (Birkin et al., 1999; Petr, 2006; O'Kelly, 2009; Montejano & Bello, 2018). Os dados de compra e de frequência dos clientes podem ser recolhidos pelas empresas através de faturas, questionários, dados do cartão de crédito ou de dados do cartão de cliente (Birkin et al., 1999; Khan, 2006; Petr, 2006).

Com base nos dados demográficos/socioeconómicos (recolhidos dos segmentos de clientes e de uma dada área considerada) e nas técnicas de geomarketing, as empresas podem também prever o volume de negócio num lugar por segmento que um dado produto poderá ter e identificar os lugares com maior concentração dos segmentos de consumidores que mais consomem um produto, o que pode ajudar as empresas a definir a sua localização, as áreas de mercado onde é mais favorável potenciar a comercialização e a utilização dos seus produtos e as estratégias de marketing mais adequadas a um dado lugar, por forma a aumentar de maneira eficaz o negócio dos produtos (Cooper & Nakanishi, 1988; Birkin et al., 1999; Douard, 2006; O'Kelly, 2009; Berman & Evans, 2013).

A previsão do volume de negócio de um produto por segmento pode ser feita pelas empresas através do cruzamento dos dados de consumo de um produto por segmento (recolhidos pela empresa) com a análise de variáveis/indicadores socioeconómicos e demográficos de uma dada área considerada (cruzamento feito através da sobreposição de camadas), por forma a encontrar as pessoas que se inserem no perfil de consumidores do segmento definido e assim prever a utilização do produto (Desmet & Zollinger, 2006; Douard, 2006; Petr, 2006; Berman & Evans, 2013). Este cruzamento dos dados assume que a procura existente num segmento de consumidores por um produto, numa dada área, é proporcional à dimensão populacional do respetivo segmento de consumidores (O'Kelly, 2009). Na previsão do volume de negócio deve também ser considerado que o lugar onde os consumidores residem (urbano, rural, região, país, bairro, rua) e a forma geográfica do lugar em que residem (ilha, litoral, montanha) influencia o seu consumo, dado os consumidores que residem no mesmo espaço tendem a comportar-se da mesma forma e a serem influenciados pelas características dos lugares (Douard, 2006; Petr, 2006).

Os dados das variáveis socioeconómicas e demográficas podem ser obtidos para pequenos setores geográficos, freguesias, municípios e regiões, através dos censos populacionais que estão normalmente disponíveis em bases de dados públicas, da

recolha de informação ou da compra de informação de outras empresas (Birkin et al., 1999; Gallopel, 2006; Berman & Evans, 2013). Estes dados permitem comparar a presença dos segmentos de consumidores nos diferentes níveis geográficos (Desmet & Zollinger, 2006). Normalmente estes são recolhidos para pequenos setores geográficos, representados pela menor unidade de área do censo (Longley, 2007).

Contudo, muitos dos dados disponíveis são apenas dados socioeconómicos e demográficos, o que dificulta a compreensão da complexidade do comportamento dos consumidores (Dion & Cliquet, 2006). Assim, no âmbito do geomarketing, têm vindo a ser desenvolvidos indicadores de valor acrescentado que integram dados comportamentais, obtidos através da interação entre inúmeras variáveis e indicadores que caracterizam os consumidores, para se poder compreender os seus comportamentos (Douard, 2006).

A caracterização e a segmentação dos consumidores facilitam então a escolha dos locais onde a oferta do produto deve ser realizada, para que a empresa possa aumentar o potencial de negócio de um dado produto. Por isso, o modelo que vai ser apresentado pretende desenvolver indicadores de valor acrescentado que, com base na caracterização dos consumidores efetuada pelos mesmos, permitam identificar espacialmente os locais com uma concentração superior dos segmentos de consumidores com maior potencial para consumir um produto.

2.5 Mapidea

Cada vez mais empresas utilizam aplicações geodemográficas avançadas onde são integrados dados de variáveis/indicadores sociodemográficos e económicos recolhidos para uma dada área, por forma a facilitarem a identificação dos lugares com maior concentração dos segmentos de consumidores pretendidos (Doyle, 2001).

Para responder a essa procura, a empresa Mapidea desenvolveu uma aplicação geodemográfica que promove a utilização da capacidade dos mapas e da análise da localização para os mais diversos contextos de negócio que necessitem de identificar lugares consoante a concentração de segmentos de potenciais consumidores.

Fundada em 2014, a empresa Mapidea pretende então ajudar outras empresas e organizações a analisar os dados de variáveis/indicadores socioeconómicos e/ou demográficos mais relevantes em diversos contextos, de forma simples e interativa, para assim poderem melhorar as suas decisões.

Através da aplicação da Mapidea pode-se empreender funções de:

- Geomarketing, dado poderem-se identificar as melhores áreas de mercado para o comércio e serviços, quais as características dos potenciais consumidores e qual a influência que poderão ter os concorrentes;
- Marketing, em que se podem definir estratégias de negócio consoante as características do território onde se pretende potenciar um produto;
- Organização e produção de dados, em que qualquer utilizador da aplicação pode inserir e analisar dados de clientes, locais de negócio, concentração de pessoas, concorrentes, entre outros;
- SIG, dado a aplicação fornecer funções para filtrar dados, calcular novos dados e analisar a proximidade (*buffer*);
- WebSIG, dado ser possível, através da aplicação, partilhar mapas com dados espaciais na internet para outros membros de uma empresa ou organização, por forma a que estes possam utilizar os dados.

A aplicação da Mapidea destaca-se ainda por permitir que qualquer pessoa sem formação em SIG possa utilizar e analisar dados espaciais, dado a aplicação ser de fácil utilização e intuitiva. Por isso, a aplicação pode ser utilizada pelos especialistas de marketing para analisarem e explorarem os dados dos indicadores de valor acrescentado obtidos e para articularem os mesmos com outros dados específicos que caracterizam as áreas de mercado e/ou com outros dados concretos das empresas.

2.6 Área de estudo

A criação dos indicadores de valor acrescentado para o geomarketing foi feita para o conjunto da Austrália, um país que, segundo os censos da Austrália de 2016, tem 23.401.892 habitantes e se subdivide em 12 estados ou territórios e 562 LGA (figura 2.3). Este país destaca-se por ter um grande território, dado contar com uma área de 7.688.126,32 km^2 , e por ter os seus habitantes distribuídos de forma desigual, em que se concentram sobretudo nas zonas mais próximas do litoral.

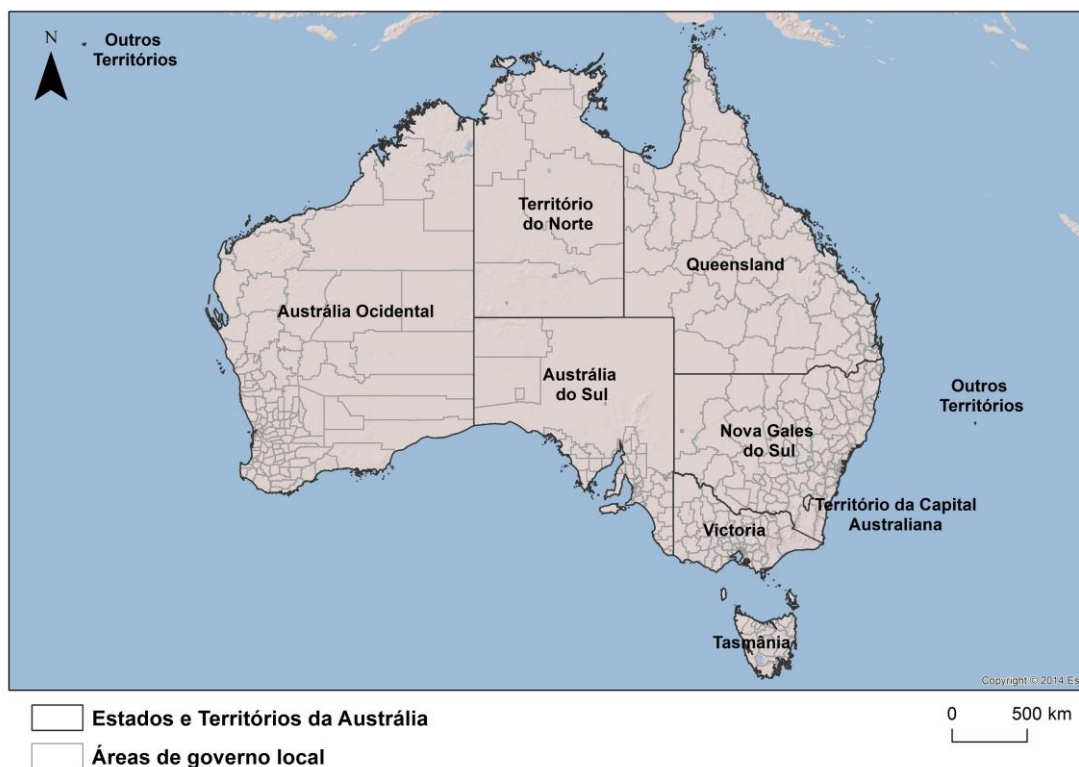


Figura 2.3 - Estados, territórios e áreas de governo local da Austrália.

A grande disponibilidade de dados recentes (de 2016 ou mais atualizados) para os diferentes níveis geográficos (pequenos setores geográficos, áreas locais, regionais, estaduais e para o conjunto do país), assim como também algumas das características do país que fazem com que se verifique uma maior diversificação das preferências dos consumidores, como a existência de uma considerável população étnica ou racial diversa, as grandes assimetrias na distribuição da população (em que a densidade populacional pode ser muito reduzida/nula em muitas áreas do país ou chegar às 15800 pessoas por km^2) e o elevado rendimento individual das pessoas (11º país do mundo com maior PIB *per capita* em 2019 segundo o Fundo Monetário Internacional), levaram a que este país fosse escolhido para desenvolver o modelo que permite identificar novos indicadores mais adequados para caracterizar os consumidores de uma determinada área de mercado.

3. Geografia e os modelos de localização

A Geografia preocupa-se com o estudo dos vários lugares de um dado território e com as relações estabelecidas entre as pessoas e os dados lugares, tentando identificar as características dos lugares e as razões de localização e de distribuição de um ou vários elementos num determinado território (Johnston, 2018).

A informação geográfica de um dado território permite fazer a representação cartográfica e a caracterização dos lugares/elementos existentes nesse território (Cliquet, 2006). Ao permitir representar e caracterizar os elementos de um território, a informação geográfica pode também contribuir para detetar as razões que estão na base da localização desses elementos (Cliquet, 2006).

Deste modo, a geografia, tendo em conta a informação geográfica, desenvolveu um conjunto de modelos de localização espacial para ajudar as empresas e as entidades públicas a encontrarem a melhor localização para a oferta de produtos (Fischer, 2011).

Os vários modelos de localização espacial têm em conta um espaço (região geográfica) onde é identificado(a) o conjunto de lugares existentes (caso da localização da procura no estudo das áreas de mercado), uma medida de distância entre os diversos elementos e os lugares candidatos para a localização de um ou vários elementos em estudo (caso da determinação dos locais mais adequados para realizar a oferta de produtos) (Eiselt & Marianov, 2011).

Os modelos de localização podem então ser utilizados para determinarem os melhores lugares para a localização da oferta de produtos, tendo por base os locais candidatos para localizar a oferta, a localização da procura (potenciais consumidores), a distância entre a procura e os locais candidatos definidos e os objetivos específicos definidos para a oferta do produto (se o produto pretende suprir e/ou potenciar a procura do mesmo) (Eiselt & Marianov, 2011; Douard et al., 2015).

A procura a considerar no modelo de localização deve ser o público-alvo do produto, definido através da segmentação da população (O'Kelly, 2009). Assim, os modelos podem identificar os lugares com maior concentração de pessoas de um dado segmento da população.

Os primeiros modelos de localização espacial debruçaram-se sobre o estudo da localização e da aglomeração das atividades económicas, das populações e/ou dos locais de comércio em específico, focando-se nas razões que levaram a localização e/ou aglomeração destes em certos lugares (Fischer, 2011). A Teoria dos Anéis Concêntricos, o Modelo de Hotelling e a Teoria dos Lugares Centrais, analisados no próximo capítulo, são exemplos desses primeiros modelos de localização, considerados de carácter determinístico.

Posteriormente surgiram os modelos de localização gravíticos, que se focam no estudo da localização do comércio, definindo os limites das áreas de mercado com base na ideia de que os consumidores não se deslocam ao local de negócio mais próximo, mas sim ao que está situado na cidade com maior atratividade (medida em função da população da cidade e da distância a que a mesma se situa) ou, no caso do outro modelo, ao que tem uma maior atratividade (medida em função da área do local de negócio e do tempo de deslocamento até ao mesmo) (Joseph & Kuby, 2011). A lei da gravitação do comércio e o modelo de Huff, analisados neste relatório, são exemplos de modelos gravíticos.

Os modelos de localização baseados na teoria de utilidade aleatória têm em conta que a escolha de um local de negócio está relacionada com a sua utilidade, definida em função de um conjunto de razões, com base no pressuposto de que a preferência entre opções alternativas é obtida através da soma das várias razões de escolha de forma aleatória (Montejano & Bello, 2018). Os modelos de localização-alocação e o modelo MCI, analisados neste relatório, são exemplos de modelos de localização que têm por base a teoria de utilidade aleatória.

No âmbito da apresentação do modelo espacial que pretende facilitar o desenvolvimento dos indicadores de valor acrescentado para o geomarketing, serão aplicados os modelos de localização-alocação mais relevantes para identificar os lugares com maior potencial de negócio, e onde, por isso, se aconselha a localização das novas instalações de uma empresa.

Os outros modelos referidos são também analisados porque, no caso do modelo de Huff e do modelo MCI, poderão ser igualmente adequados para a identificação dos lugares com maior potencial de negócio (sendo a adequação do modelo determinada consoante a estratégia de marketing da empresa) e porque, no caso dos restantes modelos, são importantes para determinar os fatores que influenciam o potencial de negócio e/ou a delimitação adequada das áreas de mercado, que podem ser tidos em conta nos modelos a aplicar.

3.1 Modelos Determinísticos

3.1.1 Teoria dos anéis concêntricos

Johann Heinrich von Thünen (1826) foi um agricultor que analisou os fatores que geraram os diferentes usos do terreno em diferentes áreas e os que determinavam o preço dos produtos cultivados. Essa análise permitiu-lhe criar um modelo de localização da produção agrícola, publicado no seu livro: *Der isolierte Staat* (O estado isolado) que teve por base três fatores: custo de transporte, valor do terreno e o preço dos produtos.

No seu modelo, von Thünen considera um estado isolado composto por uma superfície agrícola plana e homogénea, sem rios, em que todos os solos têm inicialmente igual fertilidade, existindo no centro desse estado uma única cidade (mercado), que desencadeia um padrão de produção. Pressupõe-se que esse padrão de produção ocorre devido aos agricultores transportarem os produtos apenas para a cidade central e por estes tentarem sempre aumentar o lucro.

Assim, se para uma cultura considerar-se que:

- Y é o rendimento por unidade de terreno;
- p é o preço do produto por unidade de peso do mesmo;
- c é o custo de produção por unidade de peso do produto;
- t é o custo de transporte por unidade de peso do produto e por unidade de distância;
- d é a distância à cidade;
- Yp é o valor que se pode obter por unidade de terreno;
- Yc é o custo de produção por unidade de terreno;
- Ytd é o custo de transporte por unidade de terreno a uma distância d ;

a soma do valor do terreno, do custo de produção e do custo de transporte permite obter o excedente E para o agricultor, da seguinte forma:

$$E = Y(p - c - td) \quad (\text{Eq. 3.1})$$

Von Thünen assume então que:

- O custo de produção c (influenciado pela presunção da igual fertilidade dos solos) e do transporte t (por unidade de distância) são uniformes;
- O preço de mercado p final é fixo para cada produto, independentemente da localização onde é produzido;
- A procura da cidade consegue absorver toda a produção existente.

Por isso, a distância à cidade d é o único fator que diferencia o excedente E entre diferentes terrenos agrícolas (dado a distância influenciar os custos de transporte e de produção), o que faz com que a equação 3.1 possa ser assim reescrita:

$$E = a - bd \quad (\text{Eq. 3.2})$$

onde:

$$a = Y(p - c) \quad (\text{Eq. 3.3})$$

e:

$$b = Yt \quad (\text{Eq. 3.4})$$

sendo a e b elementos constantes.

Esse excedente acaba por ser o lucro potencial que resulta do cultivo do terreno agrícola após a dedução de todos os custos e do desgaste dos equipamentos (von Thünen, 1966). O excedente é também designado por renda da terra dado este representar o valor do terreno, ou seja, o valor máximo pelo qual um agricultor estaria disposto a pagar pelo uso de um terreno agrícola, que é influenciado pela qualidade do solo e pela respetiva localização. Esta renda resulta então da vantagem que um melhor terreno agrícola tem sobre o pior terreno.

Assim, von Thünen concluiu que, quanto maior for a distância do terreno agrícola à cidade (custo de transporte total), menor será o lucro potencial do produto (figura 3.1), o que faz com que exista um limite, designado d_0 , que impede um produto de ser lucrativo (ter excedente). Esse limite, que varia em função das características do produto, permite definir limites entre áreas de produção.

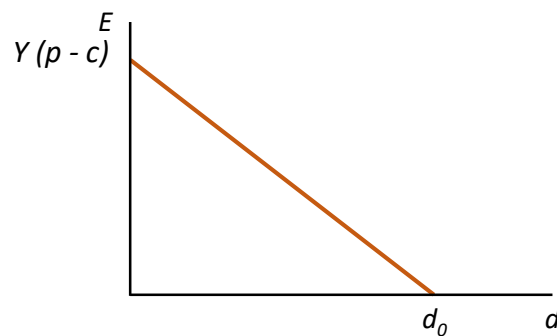


Figura 3.1 - Rendimento por distância ao mercado (Adaptado de *Figueiredo, 2016*).

Por outro lado, von Thünen reconheceu que os solos não têm todos igual fertilidade, e que por isso esta influência a produção de uma dada cultura. Deste modo, se um agricultor quiser aumentar a produtividade do seu terreno agrícola, tem de usar esterco nos solos agrícolas ou aplicar um sistema de rotação de culturas, o que aumentará o custo de produção c (que faz com que c deixe de ser uma constante). Esse aumento do custo de produção será compensado pelo conseqüente aumento do rendimento Y , devido à maior produtividade, através da seguinte expressão:

$$Y = Y(c) \quad (\text{Eq. 3.5})$$

No entanto, a partir de um dado aumento do custo de produção, o rendimento tende a estabilizar (figura 3.2 - A), dado atingir-se uma menor diferença entre o preço de mercado fixo do produto e os custos de produção e transporte, que faz diminuir a margem de rendimento. Assim, as áreas mais próximas do mercado vão ter uma maior intensidade de cultivo face às áreas mais distantes (figura 3.2 - B), pois ao terem menores custos de transporte associados, permitem que o agricultor possa fazer um maior investimento na produção.

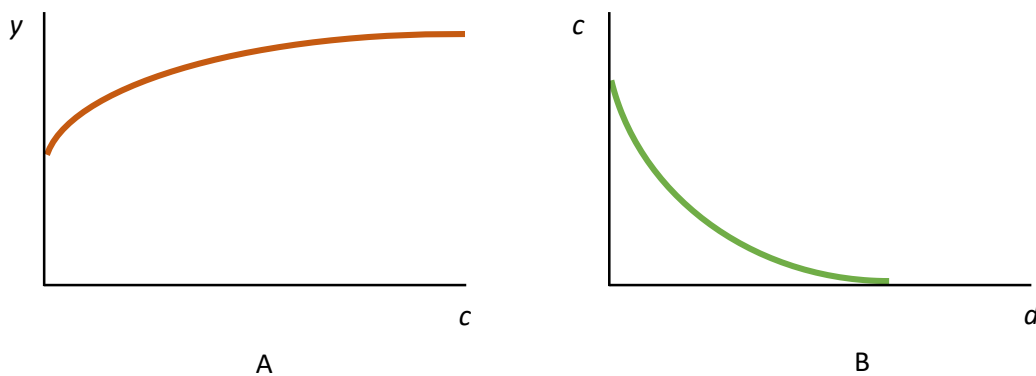


Figura 3.2 - Rendimento obtido de acordo com a intensidade de produção (A) e intensidade de produção face à distância ao mercado (B) (Adaptado de *Figueiredo, 2016*).

Pode-se então concluir que a intensidade de cultivo de um produto varia de acordo com o investimento na produção (que o custo de transporte possibilita). Porém, no caso dos produtos em que se obtém um menor excedente, devido ao seu reduzido preço de mercado, vai ser necessário uma maior intensidade de produção para que se consiga algum lucro. Isto leva a que esses produtos tenham de ser produzidos próximo do mercado.

O modelo de von Thünen pressupõe então que o agricultor escolhe a localização do terreno para produzir um dado produto em função da intensidade de produção que é adequada para o dado produto dar lucro. Assim, se os agricultores aplicarem diferentes investimentos a cada local de produção dos diferentes produtos, cada local terá um único custo de produção c e um único rendimento Y . Esses diferentes custos de produção de cada produto, em função da distância a que cada terreno está do mercado, pode ser demonstrada na seguinte figura:

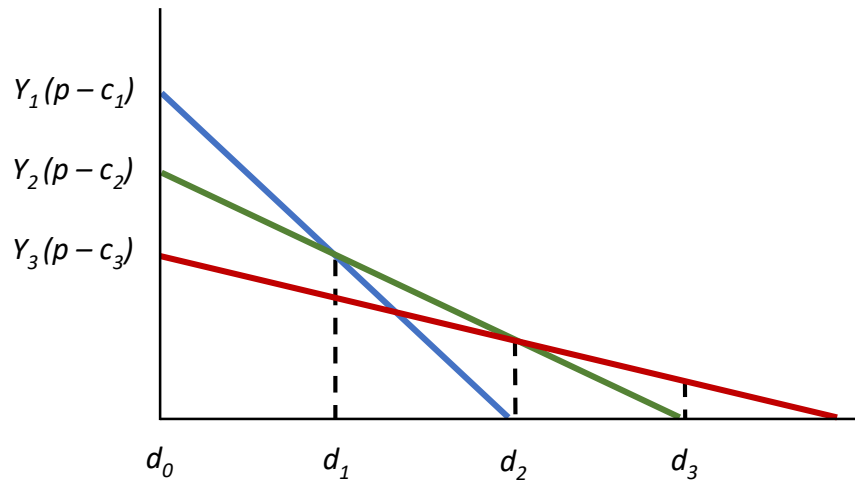


Figura 3.3 - Rendimento de cada produto em função da distância (Adaptado de *Figueiredo, 2016*).

A relação entre os diferentes custos de produção dos três produtos da figura (c_1 , c_2 e c_3), e os rendimentos obtidos por cada um deles (Y_1 , Y_2 e Y_3), no local de mercado, pode ser expressa da seguinte forma:

$$Y_1(p - c_1) > Y_2(p - c_2) > Y_3(p - c_3) \quad (\text{Eq. 3.6})$$

De acordo com a relação estabelecida para os três produtos considerados, o que necessita de uma maior renda deve ser produzido na área mais próxima da cidade, de d_0 a d_1 , o que necessita de uma renda intermédia deve ser produzido na área de d_1 a d_2 , e o que necessita de uma menor renda deve ser produzido na área de d_2 a d_3 , para que todos os produtos possam dar lucro ao agricultor. Em $d > d_3$ considera-se que não haverá qualquer cultivo. Assim, com a definição das melhores áreas de produção agrícola para cada produto, é possível definir um padrão de produção baseado em três anéis produtivos sob os quais a produção deve-se organizar (figura 3.4). Essa definição pressupõe que para cada anel, o rendimento Y_i ($i = 1, 2, 3$) aumenta de forma mais rápida que o custo c_i .

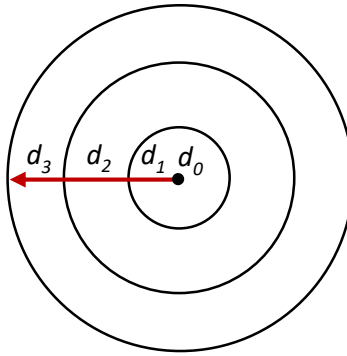


Figura 3.4 - Anéis de produção agrícola em função da distância à cidade central (Adaptado de Figueiredo, 2016).

Este padrão de produção de von Thünen demonstra que o uso do terreno agrícola é feito sob a forma de anéis circulares à volta de uma cidade, em função da distância e da procura gerada pela mesma, tendo em conta o volume dos produtos nas deslocações.

Com base nessas características do padrão de produção, von Thünen concluiu (figura 3.5) que no primeiro anel era feita a produção de lacticínios e de culturas de produção intensiva (morangos, batatas), dado esses produtos necessitarem de ser produzidos mais perto do mercado para não se estragarem, e porque caso fossem produzidos nos terrenos mais distantes o preço dos mesmos teria de aumentar para um dado valor que ninguém os aceitaria consumir. No segundo anel iriam-se localizar as áreas florestais para a produção de lenha, devido ao difícil transporte da lenha a grandes distâncias por causa do seu peso, que faz com que tenha de ser produzida próximo do mercado. Deve-se ter em conta que na época de von Thünen a lenha era muito utilizada para o aquecimento das habitações e para cozinhar os alimentos (Fischer, 2011). No terceiro anel iriam-se produzir culturas como o grão, por serem produtos perecíveis e de fácil transporte. Neste anel supunha-se que era feito um sistema de rotação de culturas. O último anel seria então dedicado à pecuária, devido ao maior custo de produção nas áreas mais próximas da cidade, que ao não ser compensado pelo menor custo de transporte, impede a sua produção, e à produção de manteiga, que não precisa de ser produzida a uma distância menor. Assim, com a exceção do segundo anel, a produção agrícola quanto mais próxima da cidade está, mais esta se intensifica.

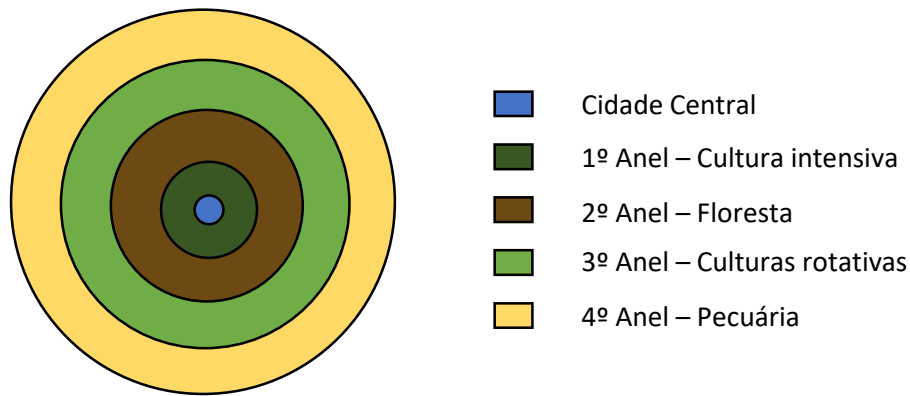


Figura 3.5 - Sistemas de anéis de produção (Adaptado de *Fischer, 2011*).

Tendo em conta o dado padrão de produção definido, von Thünen enunciou os seguintes princípios para a localização da produção agrícola:

- Quanto maior o custo de produção, mais distante da cidade o produto deve ser produzido;
- Quanto maior for o rendimento de um produto, mais próximo da cidade este deve ser produzido;
- O produto que extrai uma maior fertilidade dos solos deve ser cultivado mais longe da cidade;

Pode-se também concluir, segundo este padrão de produção, que quanto maior for distância ao mercado, maior vai ser o custo de transporte, e conseqüentemente menor é o valor do terreno agrícola, e que se o rendimento de um produto for constante, a renda do terreno vai diminuir com a maior distância.

Este padrão de produção, baseado em anéis de produção agrícola, é designado como teoria dos anéis concêntricos (figura 3.6) que se dispõe da seguinte forma:

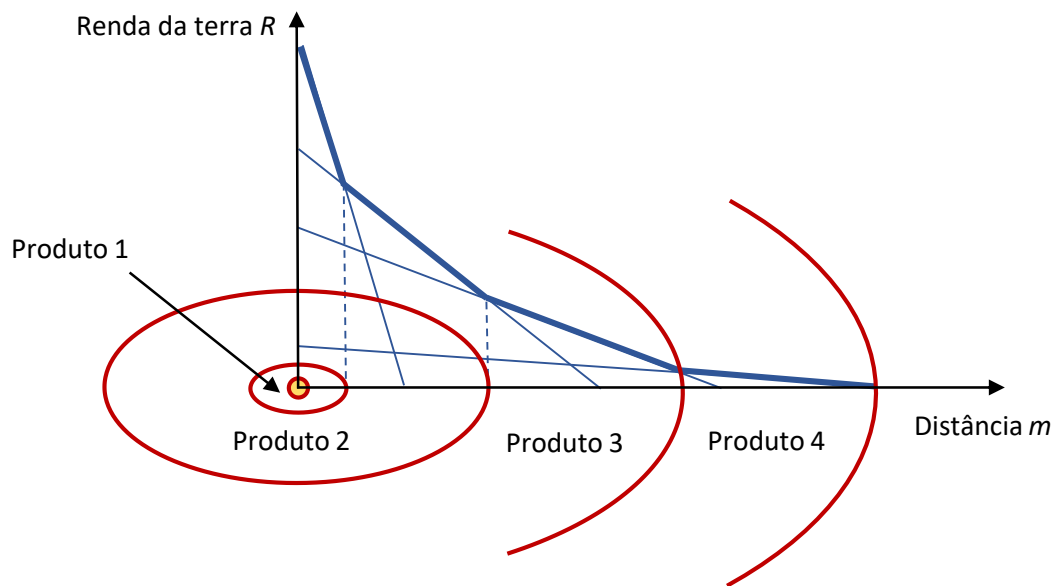


Figura 3.6 - Teoria dos anéis concêntricos (Adaptado de Fischer, 2011).

Esta teoria de von Thünen também permite ser adaptada em função das características do território, dado a fórmula da renda do terreno (equação inicial) permitir a flexibilização do valor das variáveis, e conseqüentemente dos princípios assumidos. Deste modo, von Thünen considerou como distorções na sua teoria: a existência de cidades menores no meio da área do padrão de produção que possuem os seus próprios sistemas de anéis, de menor dimensão, designados como estados independentes (figura 3.7 - A); a influência que um rio pode ter no padrão de produção, dado os rios levarem a áreas de custos de transporte igual à volta deles (devido ao custo do transporte fluvial ser mais reduzido do que o custo do transporte por via terrestre) (figura 3.7 - A); e o efeito da variação do preço do grão (figura 3.7 - B).

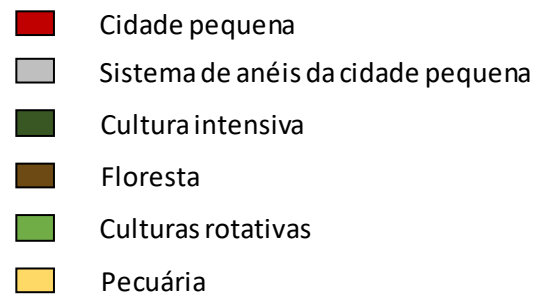
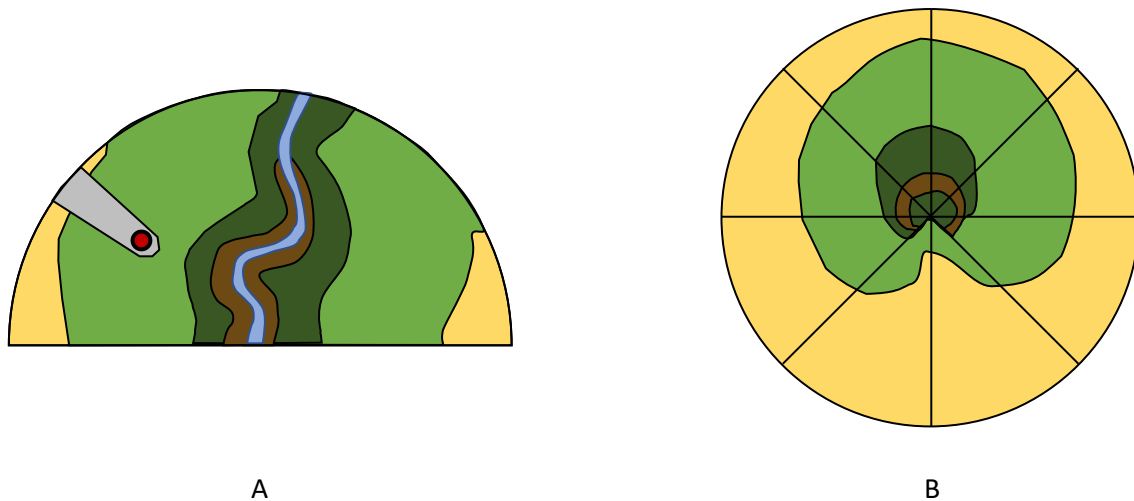


Figura 3.7 - Influência da cidade pequena e do rio no padrão de produção (A) e efeito da variação do preço do grão no padrão de produção (B) (Adaptado de *Figueiredo, 2016*).

Com a teoria dos anéis concêntricos, von Thünen contribuiu para o desenvolvimento dos conceitos de rendimento do terreno agrícola e de equilíbrio da utilização dos solos agrícolas, tendo em conta onde se deve localizar cada atividade agrícola (Fischer, 2011). Mas as suas maiores contribuições foram a introdução dos custos de transporte na teoria económica e as suas respetivas consequências na organização espacial (uso do solo agrícola) e nos preços dos produtos, e por ter sido das primeiras pessoas a aplicar a dimensão espacial na modelação das atividades económicas, marcando assim este modelo o início da análise teórica da localização geográfica (Fischer, 2011).

Contudo, o modelo de von Thünen foi criticado por:

- Ser considerado demasiado simplista, devido ao modelo ter em conta que o espaço é homogéneo;
- Supor que só existe uma única cidade principal;
- Ser estático;

- Considerar um preço fixo para os produtos;
- Não levar em conta os recursos naturais;
- Pressupor que existe uma procura ilimitada, em que qualquer produto pode ser inserido no mercado;
- Não contemplar o comércio entre diferentes mercados;
- Não ter em consideração a especialização da produção em culturas onde a área de mercado da cidade teria vantagem competitiva, dado o modelo presumir que a área de mercado deve produzir os produtos todos, independentemente do facto de os solos serem bons ou não para a produção (Fischer, 2011; Figueiredo, 2016).

Apesar de ser um modelo clássico, continua a ser importante analisá-lo porque desenvolveu alguns dos fatores que podem influenciar o potencial de negócio e a delimitação das áreas de mercado dos produtos, como a distância na qual um produto é rentável (que permite determinar a distância máxima que a área de mercado pode atingir), o princípio de que diversos produtos poderão gerar diferentes rendimentos (o que faz com que as características do produto influenciem a definição da sua área de mercado) e a existência ou não de vias de comunicação que podem levar a redução dos custos de transporte (o que pode influenciar a atração dos consumidores e, conseqüentemente, a delimitação da área de mercado).

3.1.2 Modelo de Hotelling

Harold Hotelling (1929) foi um economista que introduziu pela primeira vez a concorrência nos modelos de localização. Ao contrário dos modelos anteriores, que abordavam a concorrência numa economia sem espaço, o modelo de Hotelling destaca-se por introduzir a competição no espaço, com base no princípio de que a decisão da localização é feita tendo em conta as decisões dos vários decisores. No seu modelo, Hotelling aborda principalmente as conseqüências da aglomeração de instalações de diferentes empresas.

O modelo de Hotelling pressupõe um espaço que é fechado ao meio por um segmento de reta com comprimento l , que representa uma rua principal ou uma parte de uma ferrovia. Ao longo dessa linha a procura, considerada estática no modelo, é distribuída de forma uniforme em uma dada densidade, que faz com que a procura total seja igual a l .

Nesse espaço, segundo Hotelling, duas empresas concorrentes vão querer localizar uma instalação de negócio simultaneamente e posteriormente definir um único preço de mercado para um produto, que será o preço de fábrica. Dado os consumidores terem de se deslocar à instalação da empresa escolhida para comprar o produto, o preço final

do produto vai incluir o preço de fábrica e o custo de transporte dos clientes. A escolha da localização, segundo o modelo, é feita antes do preço do produto devido à localização ser permanente, enquanto que os preços podem ser ajustados a qualquer altura, e devido à localização escolhida também poder influenciar os preços que a empresa pratica.

Outro pressuposto do modelo de Hotelling é que os consumidores não têm preferência por nenhuma empresa em concreto para a compra de um produto, o que faz com que estes venham a escolher a empresa que oferece o preço mais baixo pelo produto, independentemente da distância. É ainda assumido no modelo que os custos de transporte são proporcionais à distância percorrida, e que as duas empresas concorrentes têm custos iguais, que são normalizados como zero.

Se considerarmos que:

- O segmento de reta de comprimento l representa o mercado de um produto;
- Uma empresa A está localizada no lado esquerdo desse segmento de reta e uma empresa B no lado direito;
- a e b representa a distância a que as empresas estão separadas da extremidade esquerda e direita da sua área de mercado;
- p_A e p_B representam o preço de fábrica que as instalações das empresas A e B respetivamente pedem pelos produtos;
- c corresponde aos custos de transporte;
- X representa o consumidor que lhe é indiferente comprar na empresa A ou B ;

o modelo de Hotelling pode ser representado pela seguinte figura:

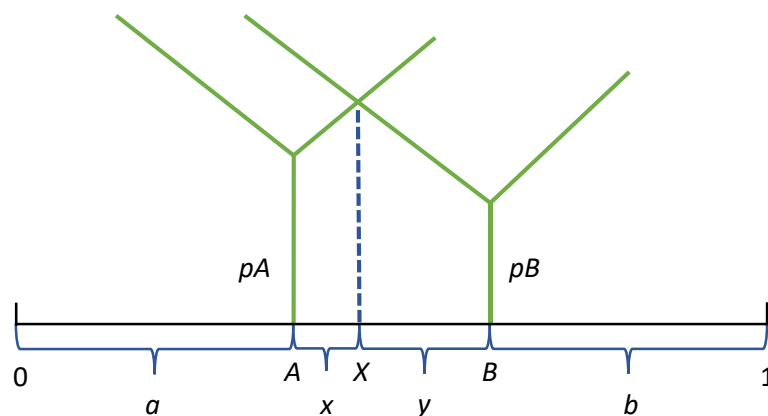


Figura 3.8 - Funções de preço dos dois competidores em um segmento de reta (Adaptado de *Eiselt, 2011*).

As formas de Y na figura 3.8 representam o preço total dos produtos nas empresas A e B , em que a base da forma de Y é o preço de fábrica e os ramos do Y é o custo do transporte c .

Tendo em conta que os produtos são iguais em ambas as empresas, os consumidores vão comprar o produto na empresa em que fica mais barato. Desta maneira, os consumidores vão comprar o produto na empresa que esteja representada abaixo dos ramos Y , dado ser a empresa que permite obter um menor preço total pelo produto, tal como os consumidores que estão à esquerda de X vão comprar na empresa A e os que estão à direita de X vão comprar na empresa B pela mesma razão.

A forma de Y na figura permite então, com base nestes pressupostos, representar a área de mercado da empresa A , que vai da extremidade esquerda (0) até ao ponto X e a área de mercado da empresa B , que vai do ponto X até à extremidade direita (1). Assim, a área de mercado à esquerda da empresa A e a área de mercado à direita da empresa B podem ser consideradas como o interior do mercado da empresa A e B respetivamente, com um comprimento a e b e a área de mercado entre a empresa A e B como a área de mercado competitiva, em que x é a área competitiva mais próxima da empresa A e y a área competitiva mais próxima da empresa B (Eiselt, 2011).

Pode-se também representar o modelo pela seguinte fórmula:

$$a + x + y + b = l \quad (\text{Eq. 3.7})$$

O preço total para o consumidor em X é igual em ambas as empresas, ou seja:

$$p_A + cx = cy + p_B \quad (\text{Eq. 3.8})$$

em que:

$$x = \frac{1}{2} \left[l - a - b + \frac{1}{c} (p_B - p_A) \right] \quad (\text{Eq. 3.9})$$

e

$$y = \frac{1}{2} \left[l - a - b + \frac{1}{c} (p_A - p_B) \right] \quad (\text{Eq. 3.10})$$

representam a distância competitiva do mercado A e B respetivamente.

Se considerarmos que uma empresa B ao estabelecer um preço p_B por um dado produto, a empresa A reage estabelecendo um preço p_A para o mesmo produto consoante uma relação, então os valores dos preços p_A e p_B podem ser obtidos através do equilíbrio do modelo pelas seguintes fórmulas:

$$p_A = c \left(l + \frac{a - b}{3} \right) \quad (\text{Eq. 3.11})$$

$$p_B = c \left(l - \frac{a - b}{3} \right) \quad (\text{Eq. 3.12})$$

Hotelling no seu modelo descreve ainda que duas empresas começam com preços desequilibrados, e que os vão ajustando entre si de forma sequencial. Desta forma, se tivermos em conta uma dada função A de reação para a otimização da variação do preço A (empresa A) e outra função B de reação para a otimização da variação do preço B (empresa B), vai-se obter um ponto de equilíbrio E , representado na seguinte figura:

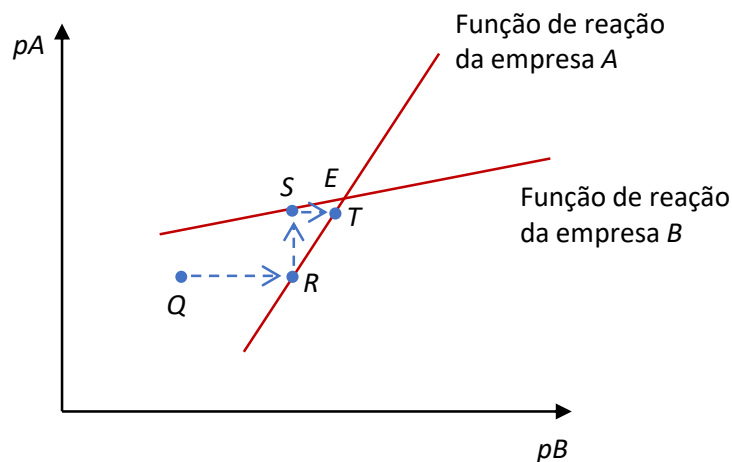


Figura 3.9 - Ajustes de preço ao longo do tempo (Adaptado de *Eiselt, 2011*).

Se considerarmos que duas empresas querem alcançar o preço do ponto Q , tendo em conta que os preços do produto em ambas as empresas estão longe do ponto de equilíbrio, qualquer uma das empresas vai ter de elevar o seu preço para poder alcançar o ponto. Com isto, se a empresa A reagir primeiro de forma imprevista, mudando o preço do ponto Q para o ponto R (que está na sua função da reação), esta assume que

a empresa concorrente não vai reagir de imediato. No entanto, após a mudança do preço efetuada pela empresa *A*, a empresa *B* vai também reagir e mudar o preço para o ponto *S* (inserido na sua função de reação), e conseqüentemente a empresa *A* vai reagir novamente, mudando o preço para o ponto *T*, e assim sucessivamente.

Estes fatores, segundo Hotelling, só se verificam se a diferença do preço entre empresas não for superior ao custo de envio de um produto da empresa *A* para a empresa *B* ou vice-versa. Se não se verificarem, o ponto de equilíbrio deixa de ser o ponto *E*.

Segundo o modelo, em lugares onde o custo de transporte é mais elevado, as empresas obteriam mais lucros dado a maior dificuldade de deslocação entre ambas as empresas potenciar que cada empresa se comportasse como um monopolista, dado poderem cobrar os preços que queriam sem ter impacto no negócio do produto. Desta maneira, quanto mais longe estiverem duas instalações de diferentes empresas, maior será o lucro potencial de ambas por produto vendido.

No entanto, se as empresas quiserem atrair um maior número de consumidores, para obterem um maior lucro pelo total de produtos vendidos, é importante antes diminuir os custos de transporte. Deste modo, segundo o modelo, no caso da empresa *A*, o custo de transporte no mercado entre 0 e *A* (*CA*) (representado na figura 3.8) pode ser obtido através da seguinte fórmula:

$$\int_{t=0}^a CA = \frac{1}{2}ca^2 \quad (\text{Eq. 3.13})$$

Para o conjunto da área de mercado das duas empresas, o custo de transporte total (*CT*) é obtido tendo em conta a equação 3.14:

$$CT = \frac{1}{2}c(a^2 + b^2 + x^2 + y^2) \quad (\text{Eq. 3.14})$$

Assim, se *a* for maior que *b*, o preço de *pA* vai ser também maior que *pB*, o que fará com que consumidores na área de mercado competitiva mais próximos da instalação da empresa *A* venham a efetuar as compras na instalação da empresa *B*, facto este que levará ao aumento dos custos de transporte. Isto deve-se ao princípio do modelo de que os consumidores comprem o produto na empresa que oferece o menor preço total pelo mesmo.

Como analisado na figura 3.8, as duas empresas *A* e *B* tendem a estar relativamente próximas uma da outra. Isto ocorre devido ao facto de a proximidade entre ambas as empresas permitir o aumento do lucro obtido pelas mesmas. Assim sendo, quando a

área de mercado tiver três empresas concorrentes, o modelo de Hotelling pressupõe que a terceira empresa, para que possa aumentar o lucro, se vá localizar próxima da empresa A e da empresa B, mas não entre elas.

Normalmente a procura para a maioria dos produtos é elástica (à exceção dos bens essenciais em que a procura pode ser estática). Contudo, segundo o modelo de Hotelling, a procura para todos os tipos de produto é elástica em vez de estática, o que faz com que essa mudança (passar de duas para três empresas) não tenham impacto nos resultados do modelo, embora exista uma menor tendência de aglomeração das empresas.

Tendo em conta os pressupostos do modelo, pode-se concluir que uma empresa que quer entrar no mercado com um novo produto deve fazer com que o mesmo seja ligeiramente diferente dos já existentes, para que possa maximizar o lucro obtido pelo mesmo, assim como também deve localizar a oferta do produto na proximidade das empresas já existentes (Eiselt, 2011).

O modelo de Hotelling contribuiu assim para a introdução da competição nos modelos de localização, descrevendo o impacto que as empresas competidoras terão na escolha da localização ideal e dos preços efetuada pelos decisores, a forma como as empresas devem reagir perante uma mudança, nos preços de um produto ou da localização, efetuada pelas empresas competidoras, e como devem antecipar a reação das empresas competidoras face as decisões que tomam (Eiselt, 2011).

A análise deste modelo é então importante porque desenvolveu outros fatores que podem influenciar o potencial de negócio e a delimitação das áreas de mercado dos produtos, como a localização de negócio (que pode influenciar os preços dos produtos e a atração dos consumidores, elementos que permitem determinar o potencial de negócio e as áreas de mercado), a concorrência entre empresas (que poderá levar a que uma empresa atraia mais consumidores que outra, em função do preço que cobra pelos produtos e do custo de transporte dos consumidores, fazendo com que essa empresa tenha uma maior área de mercado face à empresa concorrente) e o princípio da aglomeração das empresas como fator de lucro (que poderá significar que, se uma empresa localizar as suas instalações na proximidade das empresas já instaladas, possivelmente será capaz de atingir um maior volume de negócio).

3.1.3 Teoria dos Lugares Centrais

Walter Christaller (1933) foi o primeiro investigador a estudar a organização espacial dos sistemas urbanos e das respetivas áreas de mercado. Com base nesse estudo desenvolveu no seu livro: *“Die zentralen Orte in Süddeutschland”* (Os lugares centrais no sul da Alemanha) a teoria dos lugares centrais, em que demonstra que as cidades

(lugares centrais) organizam-se segundo uma hierarquia, que permite disponibilizar todos os produtos a toda a população, e que é possível definir regras que determinem o número, ordem e distribuição de lugares centrais numa dada região.

Ao verificar a tendência natural para a centralização das atividades em torno de um centro ou núcleo, Christaller conclui que o principal objetivo da cidade (mercado) é estar no centro de uma área agrícola. Com base nesse pressuposto afirma então que os lugares centrais (cidades) constituem áreas urbanas importantes para as áreas envolventes porque disponibilizam os bens centrais a estas. Os bens e serviços centrais, apesar de serem consumidos nos mais variados lugares, são apenas produzidos em certos lugares centrais. As escolas, lojas ou postos médicos são alguns exemplos desses bens e serviços centrais considerados por Christaller. Essa centralidade das cidades é então explicada pelo facto de ser nesses locais que se faz a oferta de certos produtos, por necessitarem de elevadas quantias de capital (em maior abundância nas cidades) para serem disponibilizados, e não tanto pela produção dos dados bens.

Como princípios da teoria dos lugares centrais, Christaller considerou que:

- A superfície é plana e homogénea;
- Todos os consumidores têm os mesmos rendimentos e poder de compra;
- Os custos de transporte por unidade de distância são uniformes;
- Os produtos em oferta são iguais em todos os lugares;
- Todos os consumidores procuram um produto ou um conjunto de produtos semelhantes;
- Os consumidores preferem sempre a cidade e o local de negócio mais próximo para a compra de produtos de menor valor;

Segundo Christaller o que determina então a centralidade de um lugar é a distância económica, que é obtida através do custo de transporte e do tempo de deslocação. Essa distância económica leva à definição do alcance de um produto, que consiste na distância máxima que um dado consumidor está disposto a percorrer para comprar um produto, que varia de indivíduo para indivíduo. O alcance de um produto, de acordo com Christaller, permite determinar a área de mercado que um lugar central serve, que varia de acordo com o preço do produto e com o custo de transporte. Assim, uma maior acessibilidade, ao facilitar o transporte de produtos e ao reduzir o seu custo, permite que os produtos tenham maior alcance e maior consumo.

Christaller considera que a área de mercado de um lugar central varia também entre os centros de maior e menor ordem, dado nos centros de maior ordem (caso de cidades com universidade) serem oferecidos todos os bens e serviços de menor ordem (como uma escola), mas nos centros de menor ordem não são oferecidos os de maior ordem. A diferença da área de mercado entre centros de maior e menor ordem leva à hierarquia dos lugares centrais, em que a importância de uma cidade (centro) não depende do

número de habitantes, mas sim do número de produtos centrais oferecidos e seu respectivo alcance.

No entanto, o número de habitantes e a ordem de uma cidade influencia o alcance e o consumo dos produtos, dado que, quanto maior for a ordem e dimensão do lugar central, mais produtos esse lugar vai oferecer, o que por sua vez vai tornar o lugar central mais atrativo e assim levar ao aumento do alcance e do consumo dos produtos. O alcance de um produto também é influenciado pelas suas características, dado que um produto que pode ser substituído, como é o caso do leite que pode ser comprado em qualquer local de negócio, terá um alcance menor que um produto que é mais difícil substituir, como por exemplo um fato de casamento, que tem de ser comprado num local de negócio específico (Fischer, 2011).

Desta forma, o modelo pressupõe que os produtos de maior valor são comprados pelos consumidores de uma cidade, como pelos consumidores das cidades vizinhas mais pequenas, dado os consumidores estarem dispostos a percorrer maiores distâncias para os comprarem, pois as cidades maiores podem oferecer um preço mais reduzido e uma melhor qualidade pelos produtos de maior valor.

Segundo Christaller também existe uma distância mínima que as pessoas têm de percorrer para comprar um produto, de forma a que este seja rentável, que se designa por limiar. Quando esse limiar e o alcance do produto são maiores, o lugar é classificado como um lugar de maior ordem, quando são menores, o lugar é classificado como um lugar de menor ordem. Se o limiar for menor e o alcance maior significa que o produto pode ser disponibilizado em muitos lugares, o que faz dele um produto de menor ordem.

O modelo de Christaller tem ainda em conta que:

- Um lugar central vai tentar aumentar a área de mercado tanto quanto possível;
- Os locais de negócio têm de ter uma quantidade suficiente de clientes para poderem ser rentáveis;
- A procura de toda a população por um produto deve ser satisfeita com um número mínimo de locais de negócio, para que tenham o máximo lucro possível.

Com base nas premissas enunciadas, Christaller assume que um produto de um lugar A tem um alcance de intervalo x que determina a área de mercado e os respetivos clientes que um produto serve, e que para todas as áreas exteriores à área de mercado considerada sejam servidas pelo produto devem ser acrescentados lugares centrais A para as servir (figura 3.10). Para evitar que partes das áreas de mercado dos lugares centrais fiquem sobrepostas, dado todas as áreas terem de ser servidas, deve-se assumir para as áreas de mercado uma forma hexagonal, determinada com base na disposição dos lugares centrais num padrão de triângulos (figura 3.10).

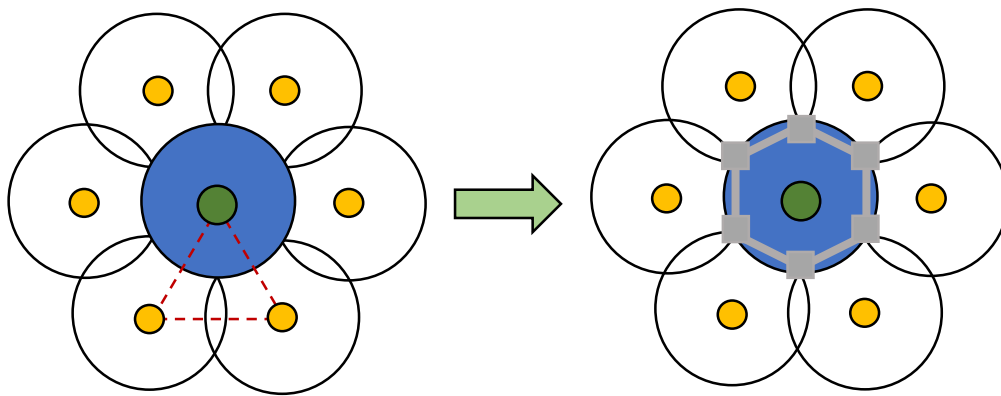


Figura 3.10 - Áreas de mercado hexagonais (Adaptado de *Fischer, 2011*).

Visto que os lugares centrais do modelo espacial têm ordens e áreas de mercado diferentes, fazem com que se obtenha um sistema de hexágonos de tamanhos diferentes que representam as áreas de mercado de cada lugar central (figura 3.11). No sistema de hexágonos da figura 3.11 os lugares centrais de maior ordem, que oferecem produtos de ordem superior, são designados por *C* e os lugares de menor ordem, que oferecem produtos de ordem inferior, por *B*. Os lugares de maior e menor ordem são equidistantes.

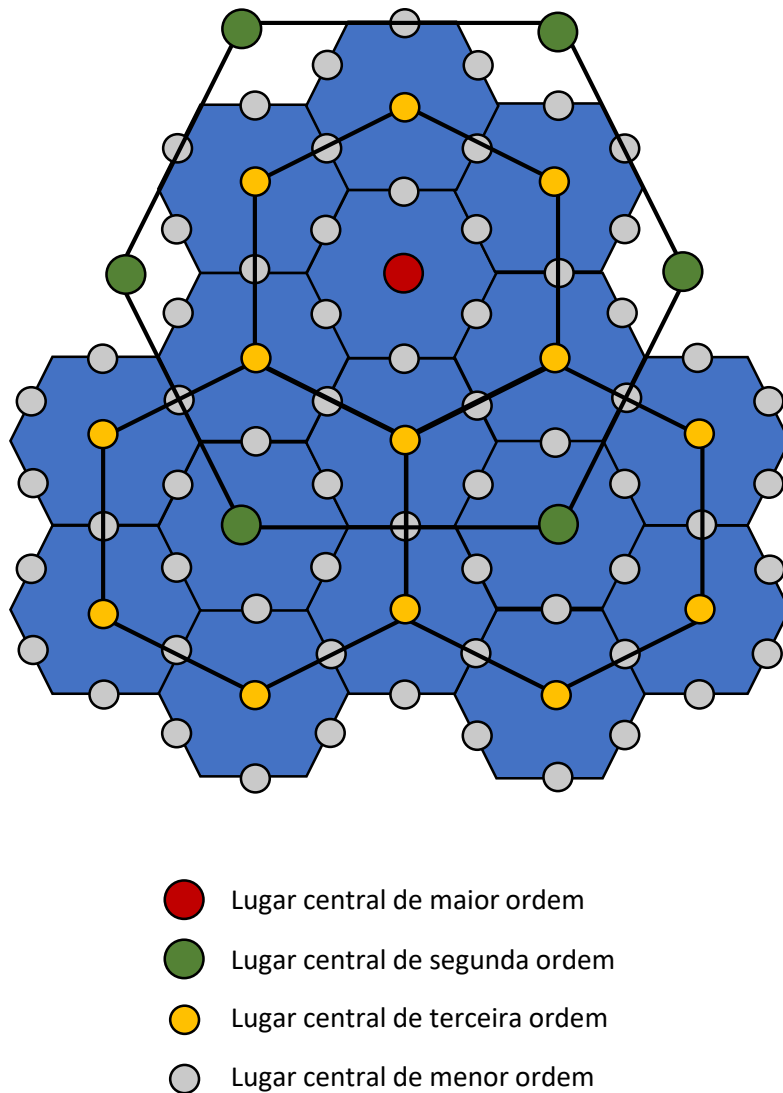


Figura 3.11 - Estrutura hierárquica dos lugares centrais (Adaptado de *Christaller, 1966*).

Este padrão de localização, é designado de princípio de mercado ou sistema $K = 3$, devido à área de mercado de um lugar de maior ordem corresponder ao mercado local de menor ordem e a um terço das seis áreas de mercado circundantes, o que faz com que o mercado do lugar de maior ordem seja três vezes maior que o de menor ordem (Christaller, 1966). Com este padrão de localização de Christaller conclui-se que quanto menor for a ordem de um lugar central, maior será o número de lugares dessa ordem e que quanto maior for a ordem do lugar central, maior será a sua área de mercado e o número de produtos que oferece.

Contudo, neste padrão de localização não são abordadas as condições de deslocação existentes, porque Christaller supõem que elas resultam do sistema de lugares centrais, uma vez que as pessoas, ao terem de se deslocar aos lugares centrais para adquirirem os produtos, leva a que exista o movimento de produtos e que se crie a infraestrutura

de transportes. Isto faz com que as estruturas de fornecimento de produtos sejam más, porque se considerarmos que dois lugares de ordem superior estão ligados, apenas dois lugares de ordem inferior serão encontrados porque os outros lugares centrais não estão em linha reta.

Deste modo, Christaller, com o objetivo de ligar um maior número de lugares centrais em rede, tendo em conta o menor custo possível, considera outro princípio de padrão de localização, que é designado por princípio de transporte ou sistema $K = 4$, devido à área servida por um lugar de maior ordem corresponder à área local de menor ordem e a metade das seis áreas de mercado circundantes, o que faz com que o local de ordem superior sirva quatro lugares de ordem inferior. Estas mudanças face ao princípio anterior devem-se a uma nova organização dos lugares centrais, em que os lugares de menor ordem passam a estar noutra posição (figura 3.12).

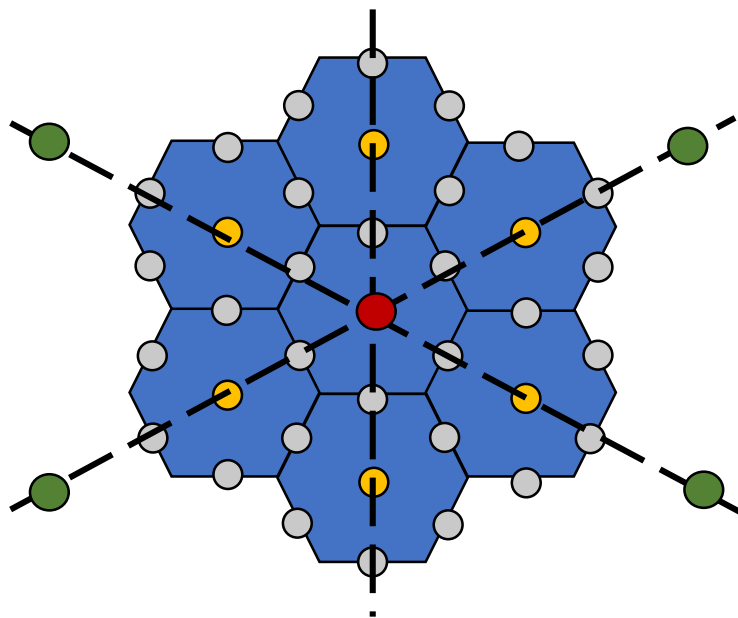


Figura 3.12 - Princípio de transporte ($K = 4$) (Adaptado de Fischer, 2011).

Christaller ainda considera um terceiro princípio com o objetivo de criar unidades administrativas para os lugares centrais, designado princípio administrativo ou sistema $K = 7$, devido a colocar um lugar de ordem superior e seis de ordem inferior juntos numa área por forma a criar a dada unidade administrativa (figura 3.13).

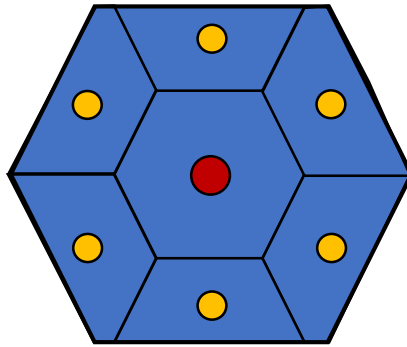


Figura 3.13 - Princípio administrativo ($K = 7$) (Adaptado de *Christaller, 1966*).

Destes princípios enunciados, Christaller considera o princípio de mercado o mais importante para a distribuição dos lugares centrais.

Com o livro *“Die räumliche Ordnung der Wirtschaft”* (A economia da localização) August Lösch (1940) também contribuiu para a formulação da Teoria dos Lugares Centrais. Teve em conta, tal como Christaller, as relações entre lugares de produção e consumo, e a distribuição da área de mercado segundo uma forma hexagonal. Contudo, enquanto que Christaller se foca no menor número de lugares centrais para fornecer um produto, Lösch foca-se no efeito que a concorrência tem para a menor área de mercado possível, através da consideração de um maior número de localizações.

Lösch supõe também que cada local é escolhido de forma a ter a maior utilidade possível, ou seja, que tenha um maior lucro e que seja o lugar com os mínimos custos de transporte e de produção (tendo em conta uma procura estática). Tem-se em conta também que a procura depende do preço do produto e da localização do local de negócio, dado a mudança de localização de um local de negócio levar a uma nova área de mercado, que pode fazer mudar a procura.

Com base na maximização da utilidade e do número de localizações, foi possível a Lösch apresentar um modelo espacial que assenta em cinco condições para a área de mercado:

1. A localização do local de negócio deve ser a mais vantajosa possível tendo em conta a procura existente;
2. Deve haver um considerável número de locais de negócio que permitam a cobertura do espaço;
3. Os preços devem ser iguais aos custos;
4. As áreas de mercados devem ser o menor possíveis para evitar a entrada de concorrentes;
5. Nos limites das áreas de mercado é indiferente qual o local de negócio a que a área pertence.

No seu modelo, Lössch tem também em conta que:

- As matérias-primas estão distribuídas uniformemente;
- O espaço é homogéneo;
- Os espaços de produção estão igualmente distribuídos;
- Quanto maior for a distância, menor será a procura;
- Só as pessoas que vivem na distância de transporte necessária compram o produto.

Assim, pode-se constatar que Lössch não tem em conta a distância máxima que as pessoas estão dispostas a percorrer para comprar o produto como Christaller, mas sim a área de mercado mínima que tem de ser coberta para tornar a produção útil. Lössch também tem em conta uma população distribuída de forma descontínua e considera a existência de um conjunto de quintas localizadas nos vértices na matriz de triângulos, em vez de pequenos lugares centrais como Christaller.

O tamanho do hexágono da área de mercado de qualquer produto depende do custo de produção e da procura. Os hexágonos dispõem-se de acordo com a seguinte figura:

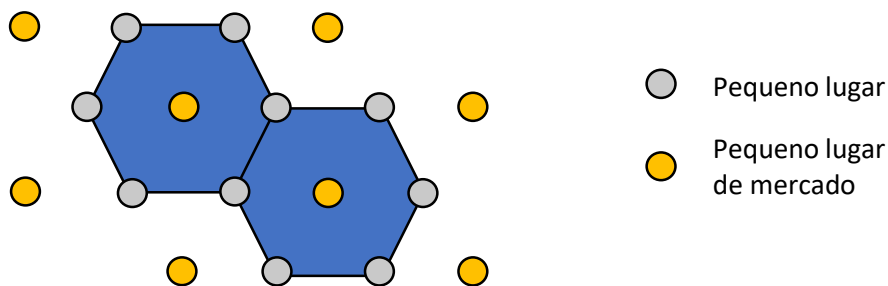


Figura 3.14 - Dimensão hexagonal das menores áreas de mercado (Adaptado de *Fischer, 2011*).

Lössch também tem em conta um limiar de procura no modelo, tal como Christaller, porém ele assume que pode haver produtos que serviam quatro lugares ($K = 4$) em vez de três ($K = 3$). Isto deve-se ao facto de cada pequeno lugar poder ser servido por dois mercados, em que o mercado maior corresponde aos produtos de maior ordem, tendo este, contudo, uma mesma distância de alcance que o mercado mais pequeno, mudando apenas a área servida (figura 3.15).

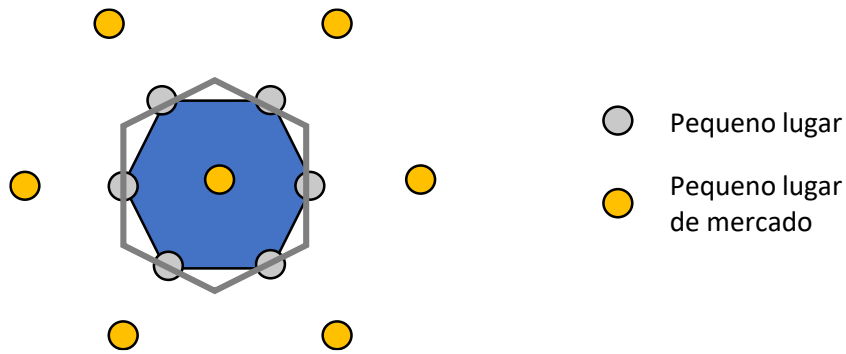
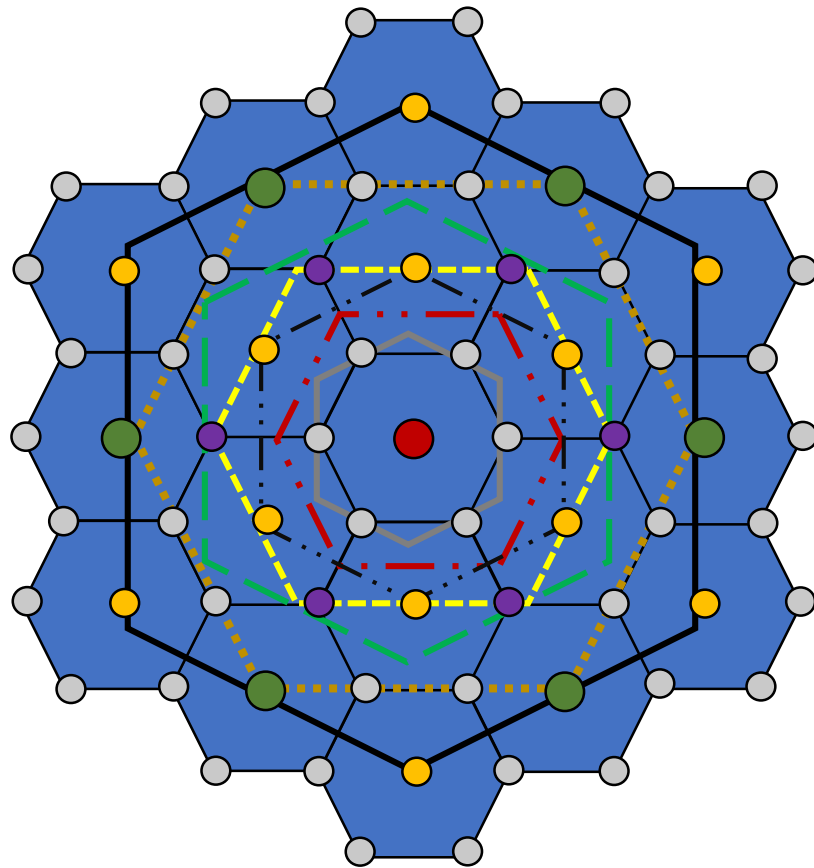


Figura 3.15 - Segunda menor área de mercado ($K = 4$) (Adaptado de Fischer, 2011).

Segundo Lösch, cada produto necessita de uma área de mercado específica. Por isso, ele considera que devem-se calcular áreas de mercado sucessivas, $K = 3, 4, 7...$ (sendo K a área de mercado), que poderão corresponder ao alcance real de um ou mais produtos. Estas áreas de mercado são consideradas de acordo com a ordem dos produtos, em que, com uma maior ordem destes, a área de mercado e o número de lugares servidos vão aumentando gradualmente, com base na sucessão anterior. Isto faz com que o modelo considere que um lugar único de ordem superior, ao oferecer todos os produtos, vá tornar-se o centro das diferentes redes de lugares que o produto serve (Lösch, 1954). Essas redes assumem uma forma hexagonal.

Ao contrário de Christaller, Lösch não tem em conta que cada produto de ordem inferior seja oferecido nos lugares de ordem superior, considerando apenas que tal só acontece na cidade central de maior ordem. Isto deve-se aos centros de menor ordem se especializarem num dado produto. Assim, verifica-se que Lösch, ao contrário de Christaller, não considera a existência de relações hierárquicas, tendo apenas em conta um padrão resultante da coincidência dos menores centros com centros de maior ordem (figura 3.16). Esse padrão faz com que quantos mais centros forem coincidentes, menor será o custo de transporte associado.



- Lugar mais central onde estão disponíveis todos os produtos (1-10)
- Lugar onde está disponível o 3º produto
- Lugar onde está disponível o 2º produto (e o 1º produto)
- Lugar onde está disponível o 1º produto
- Pequeno lugar

Figura 3.16 - Padrão das áreas de mercado de Lösch (Adaptado de *Fischer, 2011*).

Também assume que nem todas as áreas de mercado definidas precisam de existir de facto (dado que a área de mercado pode ser demasiado pequena para tornar a oferta do produto útil), que uma maior área de mercado pode não englobar mais nenhum lugar central e que o preço, a diferenciação do produto, a produtividade e a acessibilidade podem levar a que as áreas de mercado tenham uma forma irregular e/ou que sejam sobrepostas.

O contributo de Lösch permite aumentar a abrangência do modelo porque transforma a estrutura hierárquica dos lugares apresentada por Christaller num sistema de lugares centrais mais flexível e ao mesmo tempo mais complexo, considera que os lugares podem-se especializar em produtos diferentes, permite que os menores lugares possam abastecer os lugares maiores (em contradição com Christaller) e porque tem por base o princípio da maximização do lucro do comércio (Fischer, 2011).

Segundo Fischer (2011), pode-se então concluir que a Teoria dos Lugares Centrais tem como aspetos relevantes:

- A introdução das relações existentes entre as diferentes atividades económicas, das localizações dessas atividades e do respetivo alcance das mesmas (dado esta teoria não considerar os lugares de forma isolada);
- Assumir que os consumidores têm um valor máximo para gastar;
- Presumir que os consumidores compram essencialmente os produtos próximos do local em que residem;
- Deduzir que um local de negócio tem de vender (ou ser utilizado) o suficiente para valer a pena estar aberto;
- Considerar o abastecimento de produtos a todos os consumidores com um custo mínimo.

Como críticas frequentes do modelo tem-se:

- A suposição de que o espaço é homogéneo (o que faz com que este não tenha em conta deformações na estrutura urbana como o relevo ou os rios na definição dos lugares centrais e das respetivas áreas de mercado);
- A consideração de que, no sistema de Christaller, os princípios representam leis para a distribuição dos lugares centrais e de que, no sistema de Lösch, a posição da cidade principal define os restantes elementos, faz com que a teoria, em ambos os sistemas, seja rígida e estática (devido a essas assunções simplificadas da realidade);
- A dedução de que o consumidor compra apenas um produto, não tendo em conta a possibilidade de comprar vários produtos (Fischer, 2011; Rocha, 2012).

No caso do sistema de Christaller critica-se também o facto de este assumir que a população está distribuída de forma uniforme (o que não acontece na realidade, devido à população, por razões históricas, se aglomerar nos centros de maior dimensão) e de a hierarquia dos lugares centrais ser demasiado restritiva.

Já no caso concreto do sistema de Lösch critica-se ainda, por um lado, a não existência de razões para que os lugares centrais assumam o padrão de localização demonstrado, e por outro, o facto de este padrão ser menos regular, o que dificulta o ajuste do modelo a uma situação concreta, e conseqüentemente a sua aplicabilidade.

Assim, a análise deste modelo continua a ser relevante porque desenvolveu fatores que podem influenciar o potencial de negócio e a delimitação das áreas de mercado dos produtos, como a oferta e o consumo de produtos (que permite determinar os locais onde existe ou poderá existir uma maior oferta e/ou um maior consumo de cada produto, essencial para definir as áreas de mercado onde se deve realizar a oferta dos produtos), a distância entre os locais de oferta e consumo dos produtos (que permite delimitar as áreas de mercado dos produtos e avaliar o potencial de negócio, dado os consumidores comprarem os produtos na proximidade do local onde residem, exceto se forem produtos mais valiosos) e os consumidores atuais ou potenciais (que permitem determinar a área de mercado mínima que torna a oferta de um produto rentável).

Com o sistema de Christaller em concreto foi possível obter outros fatores, como a distância que cada consumidor está disposto a percorrer para adquirir um produto (que, ao poder variar de indivíduo para indivíduo, faz com que se deva ter em conta a distância máxima que os consumidores de cada segmento estão dispostos a percorrer para adquirir um produto na delimitação das áreas de mercado), a importância de cada centro urbano (que faz com que os maiores centros urbanos tenham maior potencial para o negócio de determinados produtos que os centros mais pequenos, dado alguns produtos necessitem de um maior rendimento para serem rentáveis) e a existência de uma distância mínima que os consumidores têm de percorrer para que a oferta do produto seja rentável (que influencia o potencial de negócio e a área de mercado).

Já com o sistema de Lösch em concreto foi possível identificar fatores como as características dos produtos (que fazem com que cada produto tenha a sua própria área de mercado, pelo que, tem de ser tidas em conta na delimitação das áreas de mercado), o rendimento dos consumidores/preço dos produtos (que influenciam o potencial de negócio e, conseqüentemente, a delimitação da área de mercado) e o efeito da concorrência (que poderá ter conseqüências no lucro que se obtém de um produto, e assim, influenciar o potencial de negócio e a área de mercado).

3.2 Modelos Gravíticos

3.2.1 Lei de Reilly

William Reilly (1931) introduziu, na sua obra *"The Law of Retail Gravitation"* a lei da gravitação do comércio, que é uma analogia da lei de gravitação de Newton, com o objetivo de identificar pontos de rutura das áreas de mercado de duas cidades concorrentes (tendo em conta o tamanho e as distâncias às mesmas), que permitam calcular a área de mercado de cada cidade.

Esta lei baseia-se no pressuposto de que as cidades com maior população atraem mais consumidores do que as cidades com menor população, em que essa atração é

proporcional à população da dada cidade, e inversamente proporcional ao quadrado da distância que separa a maior cidade da cidade mais pequena. Assim, o modelo define que os consumidores fazem compras na cidade (centro de comércio) que mais os atrai, e que quanto mais próxima for a cidade de um dado lugar, maior é a atração dos consumidores desse lugar (Reilly, 1931).

Na formulação do modelo, Reilly trocou os elementos correspondentes às massas dos corpos pelas populações de dois centros de comércio, tendo por base a lei da gravitação universal. A troca desses elementos por centros de comércio é demonstrada na figura 3.17 por C_1 e C_2 . Sendo o ponto de rutura das áreas de mercado de cada cidade representado por Y , os consumidores que se localizam entre a cidade C_1 e o lugar Y vão ser atraídos para a cidade C_1 e os que se localizam entre a cidade C_2 e o lugar Y vão ser atraídos para a cidade C_2 .

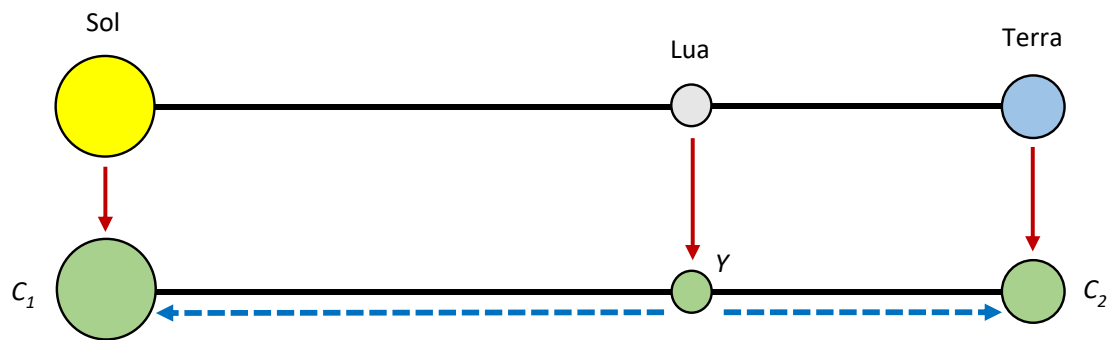


Figura 3.17 - Relação entre a Lei de Reilly e a Lei da gravitação universal.

Por isso, se tivermos em conta que P é a população do centro de comércio, que F são as forças de atração que cada centro de comércio faz sobre o consumidor, que x é a localização do consumidor, que A é o centro de comércio C_1 , que B é o centro de comércio C_2 e que d é a distância, podemos obter a força de atração de cada centro de comércio pela seguinte expressão:

$$F_{Ax} = \frac{P_A}{d_{Ax}^2} \text{ e } F_{Bx} = \frac{P_B}{d_{Bx}^2} \quad (\text{Eq. 3.15})$$

e a distância de cada cidade ao ponto de rutura:

$$d_A = \frac{d}{1 + \sqrt{\frac{P_B}{P_A}}} \text{ e } d_B = \frac{d}{1 + \sqrt{\frac{P_A}{P_B}}} \quad (\text{Eq. 3.16})$$

em que:

$$d = d_A + d_B \quad (\text{Eq. 3.17})$$

A lei de Reilly tem sido criticada pela sua simplicidade, dado apenas considerar a dimensão da população e a distância no cálculo do ponto de rutura das áreas de mercado, quando poderia também considerar outros fatores, como as barreiras existentes na estrutura urbana, as condições de acessibilidade ou variáveis de segmentação da população para uma melhor definição das áreas de mercado de um produto (Cliquet, 2006; Montejano & Bello, 2018). É também criticada por dar a entender que um consumidor é obrigado a fazer as suas compras numa cidade específica (Cliquet, 2006).

Contudo, a análise desta lei continua a ser relevante porque contribuiu para comprovar que fatores como a distância aos centros de comércio e a dimensão desses centros influenciam a atração dos consumidores, o que fazem com que tenham de ser tidos em conta na análise do potencial de negócio e na definição dos limites das áreas de mercado dos produtos (Reilly, 1931; Cliquet, 2006).

3.2.2 Modelo de Huff

David Huff (1964), no seu artigo designado “*Defining and Estimating a Trading Area*” (Definição e estimação de uma área de mercado), criou um modelo gravitacional que pressupõem que a atração que dado local de negócio efetua sobre os consumidores tem por base fatores probabilísticos, contradizendo assim o modelo de Reilly que se baseia em fatores determinísticos.

Com base em trabalhos relacionados com as áreas de mercado, Huff demonstrou que:

- A proporção de clientes de um dado local de negócio muda consoante a distância a que estão desse mesmo local;
- A proporção de clientes de vários locais de negócio varia com base nos produtos oferecidos por cada local;

- As distâncias que os clientes percorrem para os diversos locais de negócio são diferentes para as distintas categorias de produtos que se pode comprar;
- A atração de um local de negócio é influenciada pela iminência de locais de negócio concorrentes.

Huff verificou que muitos autores consideram que a área de mercado de um dado produto está dependente de uma distância fixa, o que considera errado dado as áreas de mercado poderem variar com as características de cada região, como a densidade populacional e a localização dos locais de negócio concorrentes.

Dos estudos analisados das áreas de mercado, Huff destacou a lei gravitacional de Reilly pela observação dos movimentos dos consumidores entre cidades. Nesta lei, Huff identificou como limitações a incapacidade de prever a procura específica de cada local de negócio, a definição inconsistente dos limites da área de mercado onde a competição de cada área é igual, e a possibilidade de existirem locais que estão fora de qualquer área de mercado.

Com base nas evidências demonstradas e nas limitações da Lei de Reilly, Huff desenvolveu um novo modelo onde o foco é o consumidor e não o local de negócio, dado o consumidor ser o fator que influencia a área de mercado desse local. Este modelo evidencia a probabilidade dos consumidores escolherem um local de negócio para obterem determinados bens e serviços, com base num conjunto de locais alternativos (Huff, 1964).

Assim, se tivermos em conta que:

- P_{ij} é a probabilidade de um consumidor localizado no lugar i se deslocar para o local de negócio j ;
- A_j é a área em metros quadrados do local de negócio;
- T_{ij} é o tempo de deslocação entre o lugar i e o local de negócio j ;
- λ é o parâmetro a aferir de maneira empírica que transmite a importância do tempo de deslocação com base na categoria do produto a comprar;

pode-se obter a probabilidade de um consumidor se deslocar ao local de negócio pela seguinte equação:

$$P_{ij} = \frac{\frac{A_j}{T_{ij}^\lambda}}{\sum_{j=1}^n \frac{A_j}{T_{ij}^\lambda}} \quad (\text{Eq. 3.18})$$

Huff pressupõe assim que o poder de atração de um local de negócio sobre o consumidor é equivalente à área do respetivo local e inversamente proporcional ao tempo de deslocação para o local de negócio.

Com este modelo, Huff evidenciou as seguintes conclusões gerais:

- A área de mercado corresponde aos lugares de procura que contêm potenciais consumidores para um dado produto de um local de negócio;
- Um local de negócio pode conter uma única empresa ou inúmeras empresas;
- Os lugares de procura podem ser caracterizados por um conjunto de medidas que retratam o potencial de negócio.
- As medidas que avaliam a procura são de origem probabilística, alternando estas entre zero e um, em que o valor de um representa o domínio da dada medida;
- O conjunto de potenciais consumidores que se localizam numa área de mercado de um local de negócio correspondem à soma do número esperado de consumidores em cada uma das medidas que avaliam a procura;
- As medidas que avaliam a procura de cada uma das empresas concorrentes sobrepõem-se, fazendo com que nos locais onde as medidas de igual probabilidade se cruzam seja atingida uma posição de equilíbrio competitivo.

Estas conclusões levaram a que Huff descrevesse a área de mercado de um produto como um território definido onde se encontram os potenciais consumidores que têm uma probabilidade superior a zero para comprar um produto disponibilizado por uma empresa ou conjunto de empresas.

Tendo em conta que:

- T_j corresponde à área de mercado de uma empresa ou conjunto de empresas j , isto é, ao número de consumidores num território que poderão procurar no local j um determinado tipo de produtos;
- P_{ij} é a probabilidade de um consumidor, que se localiza numa medida i , de realizar compras no local j ;
- C_i representa o número de consumidores que se localiza numa dada medida i ;

pode-se expressar a área de mercado pela seguinte fórmula:

$$T_j = \sum_{i=1}^n (P_{ij} * C_i) \quad (\text{Eq. 3.19})$$

A análise deste modelo é importante porque, para além de introduzir a área do local de negócio e o tempo de deslocação como fatores importantes para a avaliação do potencial de negócio e para a definição dos limites das áreas de mercado, pode ser

aplicado na identificação dos lugares com maior potencial de negócio (e onde, por isso, se aconselha a localização das novas instalações de uma empresa), quando o objetivo for realizar uma análise simples da melhor localização de negócio com base nos comportamentos dos consumidores.

3.3 Modelos de Interação Espacial

3.3.1 Modelos de localização-alocação

Os modelos de localização-alocação são modelos que permitem selecionar lugares ótimos para a localização das instalações de negócio, tendo em conta a alocação (atribuição) de potenciais clientes (procura) a essas instalações, para que estas possam atingir os objetivos definidos pelas empresas (Church, 2008; O'Kelly, 2009; Montejano & Bello, 2018). Assim, estes modelos podem ajudar as empresas a localizar as suas instalações em lugares que permitam atingir um maior número de clientes e/ou um maior lucro potencial (O'Kelly, 2009).

Ao modelar várias instalações em simultâneo, os modelos de localização-alocação permitem também analisar o impacto da localização de múltiplas instalações e otimizar a alocação de bens ou outras atividades à procura existente, feita por cada instalação (Church, 2008; Montejano & Bello, 2018). A alocação de bens ou outras atividades nos modelos de localização-alocação tem por base o princípio de que os consumidores fazem a viagem até ao local de negócio com um único propósito (O'Kelly, 2009).

A localização ótima para as instalações é obtida nos modelos de localização-alocação através de uma fórmula matemática que inclui variáveis quantitativas que permitem analisar as interações entre os potenciais clientes e as possíveis localizações das instalações dentro de uma área de mercado, consoante os vários objetivos definidos (Church, 2008; O'Kelly, 2009). Algumas das variáveis que permitem encontrar a localização ótima das instalações, utilizadas normalmente nos modelos de localização-alocação são a distância aos clientes (face aos seus locais de residência, de trabalho e das vias de acesso), a procura espacial a cobrir e a medição da equidade espacial das instalações (Church, 2008; Moreno, 2012, citado por Montejano & Bello, 2018). Assim, quando aplicados no geomarketing, os modelos podem encontrar a localização mais eficiente e a que permite melhorar a cobertura espacial da área de mercado (Montejano & Bello, 2018).

A existência de diferentes modelos de localização-alocação permite avaliar diferentes alternativas de localização para as instalações de uma empresa, escolhidas pelos mesmos (Montejano & Bello, 2018).

A fórmula matemática do dado modelo pode também incluir fatores que influenciam a atração dos consumidores como o comportamento dos comerciantes, dos consumidores no geral e as características do local de negócio, o que torna os modelos de localização-alocação bastante flexíveis (Church, 2008; O'Kelly, 2009). Estas características fazem com que os modelos de localização-alocação sejam úteis para a decisão da localização de várias instalações (Church, 2008).

Segundo Cliquet (2006), os modelos de localização-alocação devem assim incluir:

- Uma função objetiva a ser otimizada de acordo com as diferentes localizações possíveis;
- Áreas de procura reduzidas a um lugar central onde a procura por bens e serviços está concentrada (designado por lugar de procura);
- Os lugares possíveis para a localização das instalações, baseados na acessibilidade e na infraestrutura;
- A matriz de distância ou de tempo de deslocação;
- As regras da atribuição da procura aos locais de oferta ou a forma como os consumidores escolhem entre os locais de oferta propostos.

Os modelos de localização-alocação foram inicialmente utilizados para determinar a localização industrial, dado o surgimento destes modelos ter como fundamento a resposta ao problema de Weber (1909) de identificar onde se deve localizar o centro de produção, de forma a que a distância entre o centro de produção e a fonte das matérias-primas fosse minimizada (Montejano & Bello, 2018).

Contudo, os modelos de localização-alocação têm sido utilizados principalmente para encontrar as localizações ideais para instalações comerciais. (O'Kelly, 2009). Isto deve-se ao facto de as empresas comerciais, para desenvolverem redes de instalações que cubram vastas áreas de mercado de forma adequada (para aumentarem a venda dos seus produtos), precisem de modelos de localização-alocação que otimizem o planeamento da localização das instalações, consoante a localização da procura (Cliquet, 2006). A localização ideal pode assim maximizar os lucros de uma empresa, dado os modelos poderem ter em conta as vantagens da localização escolhida face às localizações alternativas e à despesa de localizar uma instalação (Cliquet, 2006; O'Kelly, 2009).

Os modelos de localização-alocação podem também ser úteis para evitar que uma área de mercado de uma instalação *A* se sobreponha à área de mercado de uma instalação *B*, o que, em caso de sobreposição, poderia fazer com que as duas instalações competissem pelos mesmos consumidores e assim levar a que os lucros de cada instalação fossem menores (Church, 2008).

Para uma aplicação adequada dos vários modelos de localização-alocação deve-se representar espacialmente os lugares de procura, verificar como a procura está

distribuída pelo território, definir os lugares possíveis para a localização das instalações e medir o tempo de deslocação entre os diversos pontos de procura e de instalações a localizar (Church, 2008).

Os modelos de localização-alocação devem ter em conta uma certa minimização dos custos de deslocação entre a localização dos consumidores e as novas instalações, e o número ideal de locais de negócio a instalar, por forma a satisfazerem a procura da área de mercado (Azarmand e Neishabouri, 2009, citado por Montejano & Bello, 2018).

Tendo em conta estes princípios, serão apresentados em seguida os modelos de localização-alocação p-mediana (que minimiza a distância dos clientes face às instalações), maximização da cobertura (que maximiza as áreas de mercado das instalações) e maximização da participação de mercado (que maximiza a participação de mercado de uma instalação face à presença de instalações concorrentes).

3.3.1.1 P-Mediana

Muitas das empresas têm os seus clientes distribuídos de forma dispersa. Isto exige que as empresas tenham de otimizar a localização das instalações que permitem a oferta dos produtos, por forma a facilitar a atração e a deslocação dos potenciais clientes (Marianov & Serra, 2011).

O modelo p-mediana procura então encontrar as melhores localizações para as instalações, face à localização dos consumidores, através da identificação do conjunto de deslocações que permitem minimizar a distância média percorrida pelos consumidores entre o lugar de residência e o lugar da instalação proposta (Church, 2008; Alcaide et al., 2012 e Moreno, 2012 citados por Montejano & Bello, 2018). Assim, o modelo permite diminuir a distância entre a procura e a oferta de um produto, por forma a maximizar a acessibilidade aos lugares que oferecem o produto e minimizar os custos de transporte associados à deslocação (Church, 2005; Marianov & Serra, 2011).

Este modelo pressupõe a determinação da localização ideal de um número fixo de instalações (Church, 2008). Se esses lugares servirem cada um dos lugares de procura alocados aos mesmos na menor distância possível, então o resultado do modelo indicará as instalações que minimizam a distância percorrida pelos consumidores (Marianov & Serra, 2011). Isso mesmo pode ser demonstrado pela seguinte figura:

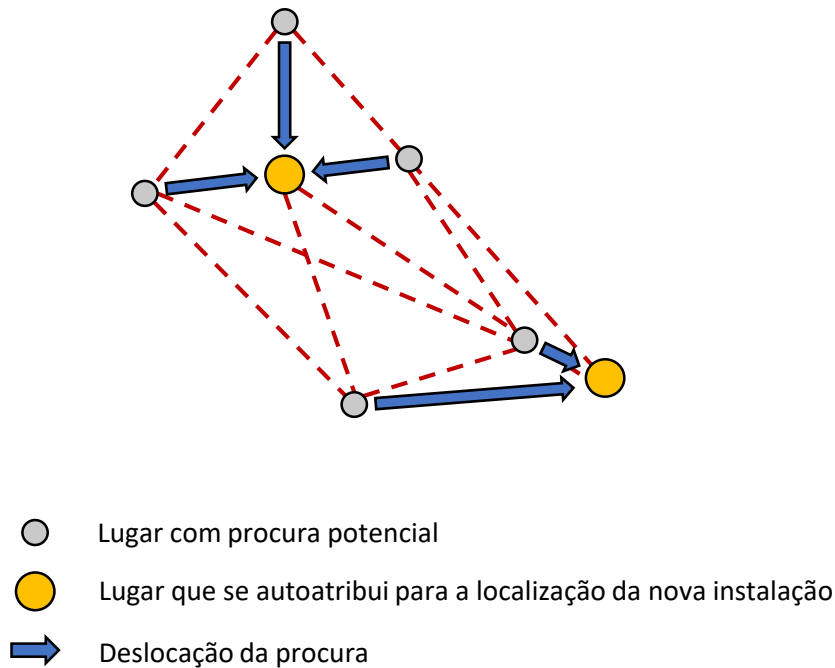


Figura 3.18 - Localização ideal para as instalações segundo o modelo p-mediana (Adaptado de *ReVelle & Swain, 1970*).

A localização ideal das instalações é então encontrada no modelo p-mediana através da minimização da soma das distâncias dos caminhos mais curtos entre os lugares de procura e os lugares mais próximos onde é possível localizar as instalações, ponderando a dimensão da procura em cada lugar de procura (Marianov & Serra, 2011). Ao dividir-se a soma dessas distâncias pelos consumidores pode-se obter a distância mínima, em média, entre consumidores e instalações (Marianov & Serra, 2011).

O modelo p-mediana surgiu com Louis Hakimi em 1964, com o objetivo de identificar a localização para um centro de transferência de uma rede de telecomunicações que permitisse minimizar a extensão dos cabos e os custos de operação (Montejano & Bello, 2018).

Este modelo advém do princípio de que o vértice mediana é o vértice de uma rede em que a soma dos comprimentos dos caminhos mais curtos para todos os restantes vértices é menor (Marianov & Serra, 2011). Deste modo, as instalações podem localizar-se em qualquer vértice de uma rede, desde que seja um vértice mediana (Marianov & Serra, 2011).

No modelo devem então ser representados os lugares de procura, os possíveis lugares para a localização das instalações e a distância entre ambos (Church, 2008). A distância pode ser estimada através da distância euclidiana (Church, 2008).

Se considerarmos que:

- i é o índice de lugares com procura potencial;
- j é o índice dos lugares candidatos para a localização das instalações;
- a_i é a procura existente no lugar i ;
- d_{ij} representa a menor distância entre o lugar de procura potencial i e o lugar candidato j ;
- p é o número total de instalações a serem localizadas;
- m é o número total de lugares candidatos para a localização das instalações;
- n é o número total de lugares de procura;
- x_{ij} assume o valor de 1 se o lugar de procura i for atribuído a um lugar j e o valor de 0 caso contrário;
- x_{jj} assume o valor de 1 se uma instalação for colocada no lugar candidato j e o valor de 0 caso contrário;

o modelo p-mediana, segundo ReVelle & Swain (1970), pode ser representado pela seguinte equação:

$$\text{Minimizar } Z = \sum_{i=1}^n \sum_{j=1}^m a_i d_{ij} x_{ij} \quad (\text{Eq. 3.20})$$

Esta equação tem as seguintes condições:

$$1 \quad \sum_{j=1}^m x_{ij} = 1, \quad i = 1, 2, \dots, n$$

$$2 \quad x_{ij} \leq x_{jj}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m$$

$$3 \quad \sum_{j=1}^m x_{jj} = p$$

$$4 \quad x_{ij}, x_{jj} \in \{0, 1\}; \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m$$

em que a condição 1 faz com que cada lugar de procura seja atribuído a uma única instalação, a condição 2 faz com que o lugar de procura i seja alocado a uma instalação

j apenas se houver uma instalação a criar naquele lugar, a condição 3 define o número de instalações a ser colocadas e a condição 4 faz com que todas as variáveis assumam números inteiros (Church, 2005; Marianov & Serra, 2011). Se um lugar j atribuir a si mesmo $x_{jj} = 1$, deve ser criada uma instalação nesse lugar (Marianov & Serra, 2011). Se todos os x_{ij} forem 0 ou 1, então o resultado do modelo inclui a atribuição da respetiva instalação a todos os lugares de procura, o que faz com que exista sempre a alocação de cada consumidor a uma única instalação (Marianov & Serra, 2011).

A equação do modelo (3.20) tem como objetivo encontrar um valor mínimo de Z que consiga uma maior otimização das deslocações entre cada lugar de procura i e cada lugar de instalação p mais próxima (Moreno, 2012, citado por Montejano & Bello, 2018). Essa otimização das deslocações é feita tendo em conta uma matriz de origem-destino dos consumidores que é obtida através da fórmula, e que permite a programação linear das deslocações (Montejano & Bello, 2018).

A programação linear da matriz de deslocações no modelo p -mediana pode ser feita através dos algoritmos (quando envolvem um reduzido número de lugares) ou através de métodos heurísticos (quando o tamanho do modelo impossibilita o uso de algoritmos), dado um maior número de lugares de procura e de restrições dificultar a resolução do modelo p -mediana (Church, 2005; Church, 2008; Montejano & Bello, 2018).

Os métodos heurísticos, ao basearem-se no conhecimento que o decisor tem da localização, podem resolver os problemas de localização de forma mais simples e rápida, o que leva a que se reduzam os custos de execução do modelo (Church, 2005).

No modelo pode-se também adicionar um vértice caso se verifique que a localização a meio de uma estrada seja melhor do que dentro de um lugar de procura, o que contribui para uma melhor definição da localização das instalações (Marianov & Serra, 2011). A procura local desse lugar seria zero.

Segundo Church (2005) e Marianov & Serra (2011), o modelo do p -mediana supõe assim que:

- O custo da localização de uma instalação é igual em qualquer uma das possíveis localizações;
- Cada lugar pode ser um lugar com procura potencial como um lugar candidato para localizar uma instalação;
- Os consumidores deslocam-se pelos caminhos mais curtos entre os locais onde se encontram e as instalações a localizar;
- Um lugar de procura tem de ser totalmente alocado a uma única instalação.

No entanto, se o decisor souber o caminho mais utilizado pelos consumidores, o princípio do caminho mais próximo pode não ser aplicado (Marianov & Serra, 2011). As

várias soluções do modelo devem ser avaliadas para evitar erros de abordagem (Church, 2008).

O decisor, com base no modelo, pode então decidir a localização das instalações, o caminho de distância mínima que os clientes terão de percorrer até chegar à instalação e as áreas com consumidores alocados a cada instalação (área de mercado) (Church, 2005; Marianov & Serra, 2011).

O modelo p-mediana pode ser adaptado para que o decisor possa identificar subconjuntos específicos de lugares de procura ou para que possa limitar a distância de cobertura de mercado (Church, 2005).

No caso da limitação da distância de cobertura, apenas as distâncias d_{ij} que excedam a distância-limite de cobertura, definida como S , não são consideradas (mantendo-se os restantes pressupostos do modelo). Assim, para a equação do modelo (3.20) não considerar as distâncias superior a S (Montejano & Bello, 2018):

$$\begin{cases} x_{ij} = 1, & \text{se } d_{ij} \leq S \\ x_{ij} = 0, & \text{se } d_{ij} > S \end{cases}$$

Isto pode ser útil para reforçar a minimização da distância a que os novos locais de oferta estarão dos potenciais clientes, tornando assim mais eficiente a localização da oferta de produtos.

Assim conclui-se que o modelo, ao incluir fatores que influenciam o potencial de negócio e a delimitação das áreas de mercado como o número de potenciais consumidores em cada lugar considerado e as vias de comunicação que permitiram ligar os lugares com consumidores aos lugares candidatos a receber os locais de negócio, pode ser aplicado na identificação dos lugares onde se aconselha a localização das novas instalações de uma empresa quando o objetivo for a minimização da distância entre os novos locais de oferta e os seus potenciais clientes.

Tendo em conta estas características, o modelo pode então ser aplicado em diversos problemas de localização, desde a localização de escolas e de clínicas de saúde, até à localização de garagens de manutenção de estradas e de veículos de resposta a emergências (Church, 2005).

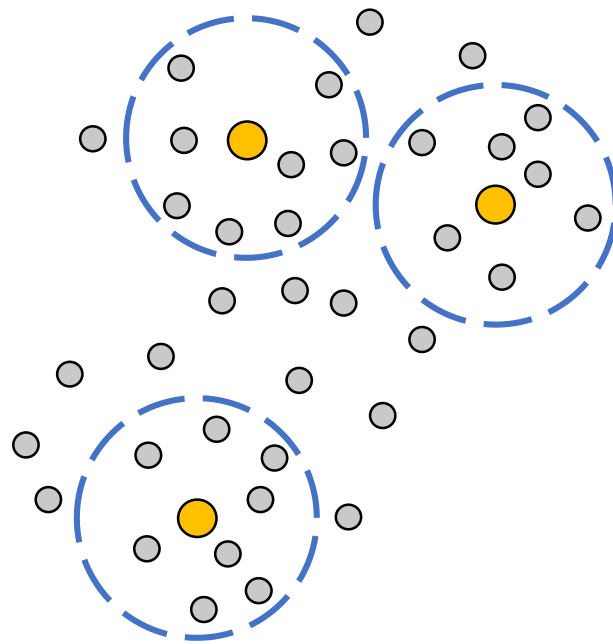
3.3.1.2 Maximização da cobertura

O modelo da maximização da cobertura, desenvolvido por Church e ReVelle em 1974, pretende determinar a melhor localização para uma instalação, tendo por base os objetivos de cobrir o maior número de lugares de procura com potenciais clientes dentro de uma distância pré-definida e de atribuir essa procura ao lugar central mais próximo (Church & ReVelle, 1974; Church, 2005; Montejano & Bello, 2018).

Assim, tendo em conta um contexto em que uma dada empresa ou instituição não consegue abrir o número de instalações necessárias para cobrir a procura total de uma dada área de mercado numa distância de atendimento considerada razoável, este modelo pode então contribuir para escolher as localizações que permitam às novas instalações cobrir o conjunto de lugares com maior volume de procura, por forma a minimizar a procura não coberta (Church & ReVelle, 1974; Church, 2005; Church, 2008; Sendra et al., 2011).

Desta forma, o modelo pode levar à minimização das instalações necessárias para cobrir o total da procura existente num território e à maximização da procura atendida por cada instalação (Church, 2005; Church, 2008).

Tendo por base estes pressupostos e o princípio de que é previamente determinado o número de instalações que se pretende abrir, a localização ideal para as instalações, segundo o modelo, pode ser assim demonstrada pela seguinte figura:



- Lugar com procura potencial
- Lugar que se autoatribui para a localização da nova instalação
- Limite da área de mercado atribuída a cada nova instalação

Figura 3.19 - Localização ideal para as instalações segundo o modelo da maximização da cobertura (Adaptado de Church & ReVelle, 1974)

Se considerarmos que:

- i é o índice de lugares de procura potencial;
- j é o índice dos lugares candidatos para a localização das instalações;
- a_i é a procura existente no lugar i ;
- V representa o conjunto de lugares de procura i ;
- p é o número total de instalações a serem localizadas;
- x_j assume o valor de 1 se uma instalação for colocada no lugar candidato j e o valor de 0 caso contrário;
- y_i assume o valor de 1 se o lugar de procura i estiver na área de cobertura adequada de alguma instalação e o valor de 0 caso contrário.

o modelo da maximização da procura, segundo Church & ReVelle (1974), pode ser representado pela seguinte equação:

$$\text{Maximizar } Z = \sum_{i \in V} a_i y_i \quad (\text{Eq. 3.21})$$

Esta equação tem as seguintes condições:

$$1 \quad \sum_j x_j \geq y_i \quad \forall i \in V$$

$$2 \quad \sum_{j \in V} x_j = p$$

$$3 \quad x_j \in \{0, 1\}, \quad \forall j \in V$$

$$4 \quad y_i \in \{0, 1\}, \quad \forall i \in V$$

A condição 1 impede que um potencial cliente seja considerado coberto caso não tenha sido considerado para nenhuma instalação a criar, a condição 2 define o número de instalações a serem criadas e as condições 3 e 4 indicam que os valores de x_j e y_i assumem apenas os valores de 0 e 1 (Snyder, 2011).

Esta equação do modelo (3.21) tem como objetivo encontrar um valor máximo de Z que consiga maximizar a procura coberta por uma instalação (Snyder, 2011).

O valor de y_i é determinado com base num raio de distância R que se considera ser a distância-limite que consegue garantir a qualidade de atendimento de uma instalação (Church & ReVelle, 1974; Sendra et al., 2011). Por isso, em todos os lugares de procura que estiverem a uma distância da instalação superior a R deve ser atribuído o valor de 0 a y_i (Church & ReVelle, 1974; Sendra et al., 2011).

O decisor, com base no modelo, pode então decidir o número de instalações que quer criar e a procura máxima que pretende que uma instalação sirva (Snyder, 2011).

Assim, segundo Church & ReVelle (1974), Church (2005), Snyder (2011) e Montejano & Bello (2018), este modelo supõe que:

- O atendimento efetuado por uma instalação tende a ter uma menor qualidade a partir de uma dada distância;
- A procura é servida se estiver dentro da distância máxima definida a que a procura deve estar da dada instalação;

- A cobertura dos lugares de procura é ponderada pela procura que geram, por isso, este favorece a cobertura de lugares com maior procura.

Os métodos heurísticos também podem ajudar a resolver o modelo da maximização da cobertura. Um exemplo é o método heurístico *Greedy Adding*, que cria instalações em sequência, escolhendo em cada repetição a instalação que permite aumentar mais a procura coberta (Snyder, 2011). Com este método torna-se mais simples a identificação da localização que mais procura permite cobrir.

O modelo da máxima cobertura pode também ser adaptado pelo decisor, que pode considerar uma nova distância-limite de atendimento da instalação T para impedir que haja qualquer lugar de procura de uma área de mercado mais distante do que essa nova distancia-limite (Church & ReVelle, 1974; Sendra et al., 2011).

Assim, com essa nova distância-limite, se algum lugar de procura permanecer mais distante do que a mesma, o decisor deve abrir um maior número de instalações para impedir que qualquer lugar de procura continue mais distante do que a nova distancia-limite definida (R. Church & ReVelle, 1974).

Esta restrição pode ser importante para poder-se garantir que se mantém uma certa qualidade de atendimento aos lugares de procura que ficaram mais distantes do que a distancia-limite inicial R que garante a qualidade de atendimento da instalação (Church & ReVelle, 1974).

Assim conclui-se que o modelo, ao basear-se em fatores que influenciam o potencial de negócio e a delimitação das áreas de mercado como o número de potenciais consumidores em cada lugar considerado e a distância que se considera que cada consumidor está disposto a percorrer para adquirir um produto, pode ser aplicado na identificação dos lugares onde se aconselha a localização das novas instalações de uma empresa quando o objetivo for maximizar a procura atendida por cada instalação (ou seja, a maximização da área de mercado).

Tendo em conta estas características, o modelo pode então ser aplicado em diversos problemas de localização, desde instalações públicas, até a instalações de negócio como pizzarias (Church & ReVelle, 1974; Montejano & Bello, 2018).

3.3.1.3 Maximização da participação de mercado (MPM)

O modelo da maximização da participação de mercado, elaborado por Michael Goodchild em 1984, pretende determinar a melhor localização para as instalações de uma empresa, tendo por base os objetivos de maximizar a procura potencial de produtos dessa empresa face à presença de instalações concorrentes que oferecem o

mesmo produto, e de atribuir essa procura ao lugar central mais próximo (Revelle, 1986; Church, 2008; Sendra et al., 2011; Montejano & Bello, 2018).

Por isso, tendo em conta a seleção de um conjunto de lugares de uma dada área de mercado (nos quais uma empresa poderá ou não abrir novas instalações) e um contexto de competição entre diferentes empresas que leve uma empresa a tentar atrair o maior número de clientes possíveis das outras empresas já existentes, este modelo procura então identificar quais desses lugares selecionados permitem captar a totalidade ou a maior parte da procura de uma dada área de mercado, por forma a se poder localizar nesses lugares as novas instalações e assim se conseguir maximizar a participação de mercado da empresa (Revelle, 1986; Sendra et al., 2011).

Tendo por base estes pressupostos e o princípio de que é previamente determinado o número de instalações que se pretende abrir, a localização ideal para as instalações, segundo o modelo, pode ser assim demonstrada pela seguinte figura:

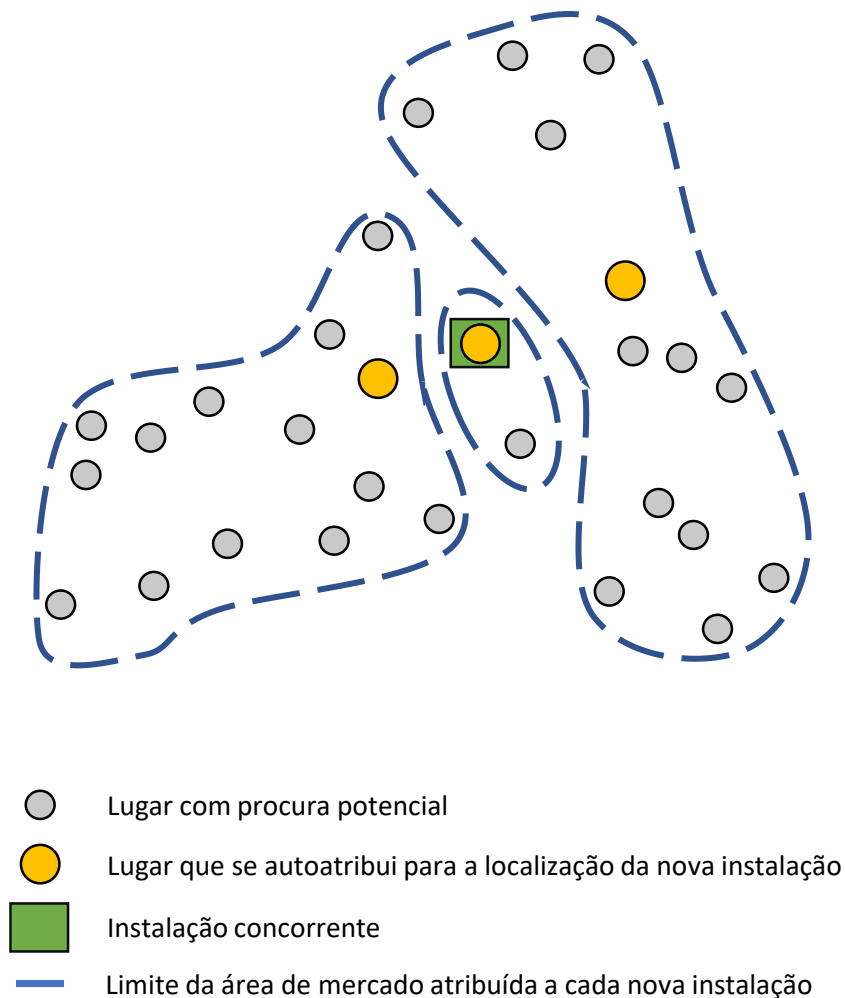


Figura 3.20 - Localização ideal para as instalações segundo o modelo da maximização da participação de mercado (Adaptado de *Revelle, 1986*).

O desenvolvimento deste modelo foi feito com base no modelo de Hotelling, dado este modelo basear-se na competição entre duas instalações de comércio evidenciada por Hotelling para formular a entrada de várias instalações numa área de mercado com instalações concorrentes já existentes (ReVelle, 1986).

Segundo Revelle (1986), se considerarmos que:

- i é o índice de lugares de potencial procura;
- j é o índice de lugares candidatos para a localização das instalações;
- I constitui o número total de lugares de procura;
- J constitui o número total de lugares candidatos para a localização das instalações;
- J_0 representa o conjunto de lugares candidatos que estão atualmente ocupados;

- J_N representa o conjunto de lugares candidatos que não estão ocupados;
- a_i é a procura existente no lugar i ;
- d_{ij} representa a menor distância entre o lugar de potencial procura i e o lugar candidato j ;
- S_i constitui a distância entre o lugar de potencial procura i e a instalação concorrente atual mais próxima;
- p é o número total de instalações a serem localizadas;
- x_{ij} assume o valor de 1 se o lugar de procura i for atribuído a um lugar j e o valor de 0 caso contrário;
- x_{jj} assume o valor de 1 se uma instalação for colocada no lugar candidato j e o valor de 0 caso contrário;

o modelo da maximização da participação de mercado pode ser representado pela seguinte equação:

$$\text{Maximizar } Z = \sum_{i=1}^I \sum_{j \in J_N \cup J_0}^J c_{ij} x_{ij} \quad (\text{Eq. 3.22})$$

em que:

$$c_{ij} = \begin{cases} a_i & \text{se } d_{ij} < S_i \\ a_i/2 & \text{se } d_{ij} = S_i \\ 0 & \text{se } d_{ij} > S_i \end{cases} \quad (\text{Eq. 3.23})$$

A equação do modelo (3.22) tem as seguintes condições:

$$1 \quad \sum_{j \in J_N \cup J_0} x_{ij} = 1, \quad i = 1, 2, \dots, I$$

$$2 \quad x_{jj} \geq x_{ij}, \quad i = 1, 2, \dots, I, \quad j = 1, 2, \dots, J$$

$$3 \quad \sum_{j \in J_N \cup J_0} x_{jj} = p$$

A condição 1 faz com que a procura existente no lugar i seja alocada de forma única, ou seja, leva a que o conjunto da procura apenas possa ser captada por uma instalação j mais próxima ou, quando a distância entre duas instalações concorrentes é igual, apenas possa ser dividida entre as mesmas (Revelle, 1986; Marianov & Serra, 2011). Já a condição 2 faz com que a procura do lugar i seja alocada a uma instalação j ou dividida entre duas instalações (quando a distância de um lugar de procura a ambas as instalações concorrentes for igual) apenas se houver uma instalação a criar na proximidade do lugar de procura (Revelle, 1986; Church, 2005; Marianov & Serra, 2011). Por fim, a condição 3 limita o número de novas instalações a serem colocadas (Revelle, 1986; Church, 2005; Marianov & Serra, 2011).

Através da equação 3.23 verifica-se que c_{ij} pode assumir diferentes valores (Sendra et al., 2011). O primeiro desses valores significa que, se a distância entre o lugar de procura e o lugar da nova instalação for menor do que a distância entre esse lugar de procura e a instalação concorrente já existente, então a nova instalação cobre toda a procura desse lugar (Sendra et al., 2011). Já o segundo valor indica que, se ambas as distâncias anteriores forem iguais, então a procura existente nesse lugar é dividida entre a nova instalação e a instalação concorrente (Sendra et al., 2011). Por fim, o terceiro valor demonstra que, caso o lugar de procura estiver mais próximo da instalação concorrente do que da nova instalação, então não será atribuída qualquer procura à nova instalação (Sendra et al., 2011).

No modelo podem também ser consideradas mais do que duas instalações a igual distância de um único lugar de procura, desde que a procura a alocar a cada instalação resulte da divisão da procura total pelo número de instalações a igual distância (ReVelle, 1986).

Constata-se também, através da equação do modelo (3.22), que não há limitação ou até mesmo indicação do número de instalações já existentes antes da criação das novas instalações, nem se refere o facto de as instalações pré-existentes serem só de uma empresa ou de diversas empresas que concorrem entre si (ReVelle, 1986). Isto deve-se ao facto de ser mais relevante, neste modelo, a comparação da distância entre o lugar de procura e a instalação pré-existente com a distância entre esse lugar e a nova instalação a criar para determinar que proporção da procura desse lugar deve ser alocada a cada instalação (ReVelle, 1986).

No entanto, se for necessário incluir na formulação do modelo as instalações da dada empresa que já existem numa área de mercado, tem de se retirar todos os lugares de procura atribuídos inicialmente às instalações da empresa já existentes, para não se considerar a procura captada por essas instalações (ReVelle, 1986).

Assim, de acordo com Revelle (1986), Sendra et al. (2011) e Montejano & Bello (2018), este modelo supõe que:

- Já existem instalações concorrentes e/ou instalações da empresa fixadas na área de mercado em que se pretende abrir as novas instalações;
- O volume de procura não varia com a distância;
- A procura de produtos pode ser deslocada das instalações concorrentes para as novas instalações, caso as novas instalações estejam mais próximas dos consumidores;
- A alocação da procura a cada instalação é feita apenas em função do custo mínimo de deslocação, não sendo consideradas as particularidades das respetivas instalações.

No entanto, o decisor pode adaptar o modelo para que este considere que a procura de um lugar i a alocar a uma instalação j seja menor quanto maior for a distância face à instalação, através do cálculo do valor de c_{ij} :

$$c_{ij} = \begin{cases} a_i (1 - d_{ij}/L) & \text{se } d_{ij} \leq L, d_{ij} < S_i \\ (a_i/2) (1 - d_{ij}/L) & \text{se } d_{ij} \leq L, d_{ij} = S_i \\ 0 & \text{se } d_{ij} > S_i \end{cases} \quad (\text{Eq. 3.24})$$

em que L representa a distância-limite após a qual a procura é nula (ReVelle, 1986).

Assim conclui-se que o modelo, ao basear-se em fatores que influenciam o potencial de negócio e a delimitação das áreas de mercado como o número de potenciais consumidores em cada lugar considerado, a concorrência entre empresas e o princípio da aglomeração das empresas como fator de lucro, pode ser aplicado na identificação dos lugares onde se aconselha a localização das novas instalações de uma empresa quando o objetivo for maximizar a participação de mercado de cada nova instalação face à presença de instalações de empresas concorrentes.

Tendo em conta as suas características, este modelo pode então ser aplicado, por exemplo, na expansão de uma cadeia de instalações comerciais ou da restauração (Revelle, 1986; Montejano & Bello, 2018).

3.3.2 Modelo MCI

Nakanishi e Cooper (1974) desenvolveram no seu artigo "*Parameter Estimation for a Multiplicative Competitive Interaction Model - Least Squares Approach*" um modelo multiplicativo de interação competitiva (MCI) que consiste numa evolução do modelo

de Huff que tem em conta um maior conjunto de variáveis, para além da distância e do tamanho do local de negócio, por forma a se poderem considerar os comportamentos espaciais dos consumidores e a competitividade que existe entre as diferentes localizações como fatores determinantes para a localização de um negócio (Nakanishi & Cooper, 1974; Cliquet, 2006; Baviera-Puig et al., 2013; Montejano & Bello, 2018).

O modelo MCI foi desenvolvido com o objetivo de considerar um mercado que é segmentado em pequenas áreas de mercado, determinadas segundo um grupo de potenciais consumidores ou uma dada localização (Wieland, 2017).

Este modelo tem por base o princípio de que a escolha da localização entre várias alternativas é feita através da soma das variáveis (Alcaide et al, 2012, citado por Montejano & Bello, 2018). O modelo assume assim que cada consumidor prefere o local de negócio com maior utilidade, em que essa utilidade tem por base um conjunto vasto de razões (representadas como variáveis), que podem mudar entre diferentes locais de negócio e consoante os consumidores (utilidade aleatória) (Montejano & Bello, 2018).

Nakanishi e Cooper, no seu modelo, também pressupõem que:

- O número de variáveis a considerar no modelo pode ser teoricamente ilimitado;
- Todas as variáveis consideradas têm o mesmo peso para a atração do negócio;
- As variáveis a incluir no modelo podem ser obtidas de forma objetiva e subjetiva (Cliquet, 2006).

Ao considerarem o princípio da independência das alternativas irrelevantes no modelo, Nakanishi e Cooper pressupõem que um novo local de negócio irá competir de forma igual com outros locais de negócio existentes, segundo as suas características (Baviera-Puig et al., 2013). Isto deve-se ao facto da definição da utilidade de um local de negócio ser independente da competição existente (Luce, 1959, citado por Baviera-Puig et al., 2013).

A elaboração do modelo teve por base o cruzamento dos modelos de interação espacial gravíticos com os modelos de participação do mercado baseados no Teorema fundamental de Kotler (Nakanishi & Cooper, 1974). O teorema fundamental de Kotler pressupõe uma relação proporcional entre a participação de mercado de uma empresa e as suas ações de marketing, em que quantas mais forem as ações de marketing da empresa, maior é a sua participação de mercado (Cooper & Nakanishi, 1988).

Segundo Nakanishi & Cooper (1974), se considerarmos que:

- P_{ij} é a probabilidade de que um consumidor residente em i escolha o local de negócio j ;
- X_{kij} é a variável k que descreve o espaço j num lugar i ;
- β_k é o parâmetro de sensibilidade relativo à variável k ;
- m é o número de locais de negócio alternativos para escolha;

- q é o número de variáveis;
- ζ_{ij} corresponde ao valor do erro específico associado ao modelo;

pode-se obter a probabilidade de um consumidor escolher um dado local de negócio através da seguinte equação:

$$P_{ij} = \frac{(\prod_{k=1}^q (X_{kij} \beta_k)) \zeta_{ij}}{\sum_{j=1}^m [(\prod_{k=1}^q (X_{kij} \beta_k)) \zeta_{ij}]}$$
 (Eq. 3.25)

Se tivermos ainda em conta que P_{ij} é a variável dependente, que \overline{P}_i e \overline{X}_{ki} correspondem a média dos valores de P_{ij} e X_{kij} respetivamente e que ε representa o valor de erro que pode advir tanto da determinação do valor de ζ_{ij} , como das diferentes amostras de valores observados nas variáveis que determinam P_{ij} , a equação 3.25, segundo Nakanishi & Cooper (1974), pode ser transformada de forma logarítmica na seguinte equação de regressão múltipla:

$$y_{ij} = \sum_{k=1}^q \beta_k z_{kij} + \varepsilon_{ij}$$
 (Eq. 3.26)

tendo como condições:

$$1 \quad y_{ij} = \log \left(\frac{P_{ij}}{\overline{P}_i} \right)$$

$$2 \quad z_{kij} = \log \left(\frac{X_{kij}}{\overline{X}_{ki}} \right)$$

$$3 \quad i = 1, 2 \dots M ; j = 1, 2 \dots m$$

Esta transformação implica que as variáveis binárias tenham de ser modificadas para que não assumam valores zero, dado estes impossibilitarem o cálculo logarítmico (Nakanishi & Cooper, 1974). A transformação do modelo permite ao decisor testar a influência das diferentes variáveis para a determinação da área de mercado de uma

empresa, dado permitir obter os coeficientes de regressão de cada variável, em que o valor do coeficiente determina a influência da variável (Nakanishi & Cooper, 1974; Huff & McCallum, 2008, citados por Wieland, 2017).

De acordo com Nakanishi & Cooper (1974), para estimar-se o valor de β_k tem de se primeiro transformar a equação 3.26 em notação matricial:

$$y = Z\beta + \varepsilon \quad (\text{Eq. 3.27})$$

tendo em conta as seguintes condições:

$$1 \quad y'_i = (y_{i1} \quad y_{i2} \quad \dots \quad y_{im})$$

$$2 \quad Z_i = \begin{pmatrix} z_{1i1} & z_{2i1} & \dots & z_{qi1} \\ z_{1i2} & z_{2i2} & \dots & z_{qi2} \\ \dots & \dots & \dots & \dots \\ z_{1im} & z_{2im} & \dots & z_{qim} \end{pmatrix}$$

$$3 \quad \varepsilon'_i = (\varepsilon_{i1} \quad \varepsilon_{i2} \quad \dots \quad \varepsilon_{im})$$

$$4 \quad y = \begin{pmatrix} y_1 \\ \dots \\ y_2 \\ \dots \\ \vdots \\ \dots \\ y_M \end{pmatrix} \quad Z = \begin{pmatrix} Z_1 \\ \dots \\ Z_2 \\ \dots \\ \vdots \\ \dots \\ Z_M \end{pmatrix} \quad \varepsilon = \begin{pmatrix} e_1 \\ \dots \\ e_2 \\ \dots \\ \vdots \\ \dots \\ e_M \end{pmatrix}$$

e depois estimar-se o valor médio do β para cada parâmetro, através da seguinte equação:

$$\bar{\beta} = (Z'Z)^{-1} (Z'y) \quad (\text{Eq. 3.28})$$

Considerando que só podem ser introduzidas na equação de regressão variáveis com valores que estejam numa única escala de intervalo (dado no modelo MCI todas as variáveis terem de ter o mesmo peso), tem de se efetuar ainda uma transformação dos valores das variáveis em zeta ao quadrado, para que os valores das variáveis medidos

em diferentes escalas de intervalo possam dispor-se numa única escala (Cooper & Nakanishi, 1983; Montejano & Bello, 2018).

Desta forma, segundo Cooper & Nakanishi (1983), se considerarmos que:

- ζ_{kij}^2 é o valor da variável obtido com a transformação;
- z_{kij} é o valor de padronizado de X_{kij} ;
- $I_{ki0}^{1/2}$ é o momento central de inércia;

a transformação dos valores das variáveis em zeta ao quadrado pode ser feita da seguinte forma:

$$\zeta_{kij}^2 = \begin{cases} (z_{kij})^2 & \text{se } X_{kij} \geq \bar{X}_{ki} \\ 1 / (z_{kij})^2 & \text{se } X_{kij} \leq \bar{X}_{ki} \end{cases} \quad (\text{Eq. 3.29})$$

em que:

$$z_{kij} = \frac{(X_{kij} - \bar{X}_{ki})}{I_{ki0}^{1/2}} \quad (\text{Eq. 3.30})$$

e

$$I_{ki0}^{1/2} = \frac{1}{m_i} \sum_{j=1}^{m_i} (X_{kij} - \bar{X}_{ki}) \quad (\text{Eq. 3.31})$$

O modelo MCI possibilita que a distância não seja considerada para a explicação da atração do local de negócio. Isto deve-se ao facto de a espacialidade no modelo ser implícita, dado a variável i representar o espaço geográfico, o que faz com que a variável i seja o único elemento gravitacional do modelo (Cliquet, 2006). No entanto, a distância entre o local onde reside o consumidor e o local onde poderá ser localizada uma nova instalação de negócio é normalmente incluída no modelo (Cliquet, 2006).

Este modelo é criticado por ser de difícil implementação, devido à complexidade de medição e integração das variáveis subjetivas, da determinação das escalas de medida das variáveis (dado terem de estar numa única escala) e à dificuldade de efetuar uma divisão geográfica adequada através do modelo (Cliquet, 2006). Também é limitado por não poderem ser utilizados valores zero ou negativos para as variáveis do modelo (Baviera-Puig et al., 2013).

Desta forma pode-se concluir que o modelo, ao basear-se em fatores que influenciam o potencial de negócio e a delimitação das áreas de mercado como os comportamentos dos consumidores e a localização de negócio (dado o modelo escolher a melhor localização para o local de negócio através da competitividade entre diferentes localizações), pode ser aplicado na identificação dos lugares onde se aconselha a localização das novas instalações de uma empresa quando o objetivo for encontrar as localizações com maior utilidade para os consumidores, podendo essa utilidade ser determinada com base em diversas variáveis que poderão ser livremente escolhidas pelos especialistas de marketing das empresas.

Dadas as suas características, o modelo tem também sido utilizado para estimar as áreas de mercado, para prever o impacto da entrada de empresas concorrentes e para prever o volume de negócio de um estabelecimento comercial (Garrocho, 2003, citado por Baviera-Puig et al., 2013).

4. Importância dos SIG para a análise espacial

Quase tudo acontece num certo lugar, dado a maioria das atividades humanas desenrolarem-se num lugar específico ou em vários lugares da superfície da Terra (Longley et al., 2015; Bolstad, 2016). Desta forma, a análise da localização de diferentes atividades e a consequente perceção de como essas atividades se desenrolam e se relacionam no espaço é fundamental para a resolução dos problemas da sociedade (Longley et al., 2015; Bolstad, 2016).

Tendo em conta que a informação espacial permite descrever e caracterizar os lugares (dado esta indicar a localização e as características daquilo que representa), torna-se útil utilizar técnicas que permitam representar e modificar as informações espaciais de um dado território, para poder-se efetuar uma análise adequada das atividades alvo de estudo (Longley et al., 2015; Bolstad, 2016; Chang, 2019). Essas técnicas estão associadas aos Sistemas de Informação Geográfica (Bolstad, 2016).

Um Sistema de Informação Geográfica (SIG) consiste num sistema que utiliza procedimentos computacionais para facilitar a recolha, armazenamento, representação, consulta e análise das informações georreferenciadas de um dado território (dados espaciais), por forma a tornar mais simples o processo de decisão (Longley et al., 2015; Bolstad, 2016; Chang, 2019).

Os SIG têm como objetivo representar o mundo real de maneira prática e robusta, por forma a facilitar a abordagem e a resolução dos problemas da sociedade e do meio ambiente (Longley, 2008; Bolstad, 2016).

Eles permitem fazer a representação espacial dos lugares e das várias características associadas aos mesmos, com base nos dados espaciais sobre esses lugares, assim como também permitem descrever como os vários elementos espaciais funcionam e se relacionam, através da combinação dos dados espaciais relativos a esses elementos (Longley, 2008; Bolstad, 2016).

A combinação dos dados espaciais nos SIG também permite desenvolver novos dados espaciais (Bolstad, 2016). Desta forma, a informação geográfica recolhida para um dado território pode ser transformada em novos dados mais completos e adequados aos problemas em análise, o que faz com que a análise espacial e a tomada de decisão se tornem mais simples.

As funções de SIG permitem então melhorar e tornar mais rápido o processo de tratamento e análise de dados espaciais (Longley et al., 2015).

Segundo Longley (2008), Bolstad (2016) e Chang (2019), para que os SIG possam ser utilizados de forma adequada, estes devem incluir sempre as seguintes componentes (figura 4.1):

- Um *software* de SIG para executar o conjunto das funções espaciais;
- O *hardware*;
- Os dados espaciais (inseridos nas bases de dados);
- O conjunto de pessoas e instituições que realizam operações com dados espaciais;
- Os procedimentos de SIG aplicados.

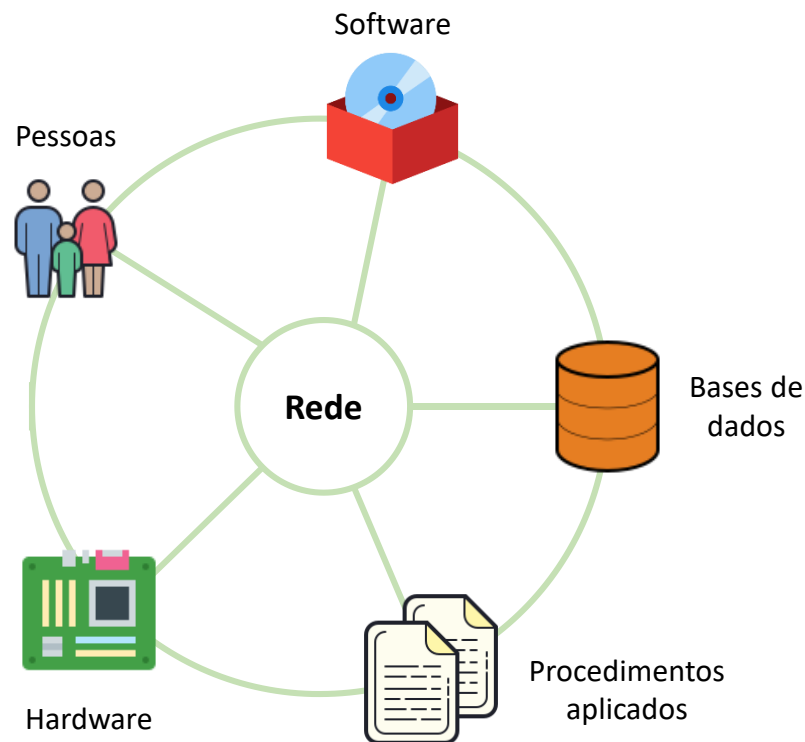


Figura 4.1 - Sistema SIG (Adaptado de Longley, 2008).

O *software* de SIG, tanto comercial como livre, fornece um ambiente de trabalho que permite, através das funções disponibilizadas, criar representações digitais que simplificam a complexidade das várias componentes espaciais de um território representado (Galati, 2006). A este estão associados os programas e aplicações que são utilizados num computador para executar as funções espaciais (de representação, análise e visualização de informações espaciais), assim como também as aplicações adicionais para análises mais específicas, desenvolvidas através de linguagem *Python* ou *JavaScript* (Bolstad, 2016; Chang, 2019). O *ArcGIS* (de âmbito comercial, desenvolvido pela ESRI), o *QGIS* (de acesso livre, desenvolvido pela *Open Source Geospatial Foundation*) e o *MapInfo* (de âmbito comercial, desenvolvido pela *MapInfo Corporation*) constituem alguns exemplos de *software* SIG (Longley et al., 2015; Bolstad, 2016).

O *hardware* de SIG são os dispositivos que interagem com a aplicação das operações de SIG (Longley et al., 2015). Os computadores utilizados para o processamento e armazenamento de dados espaciais, as impressoras que permitem imprimir os mapas obtidos, os *scanners* que possibilitam a digitalização dos dados espaciais e o GPS ou outros aparelhos de recolha dos dados constituem alguns exemplos desses dispositivos (Bolstad, 2016; Chang, 2019).

Os dados espaciais permitem representar a localização dos elementos espaciais nos SIG (Longley et al., 2015; Bolstad, 2016; Chang, 2019). A estes estão associados um sistema de coordenadas geográficas (expressas através da latitude e da longitude) ou planas (expressas através de coordenadas x e y), que permitem localizar os elementos na superfície da Terra (Longley et al., 2015; Bolstad, 2016; Chang, 2019). Ao terem um único sistema de coordenadas e ao poderem ser sobrepostos em camadas ou organizados por temas, quando se representa nos SIG diferentes dados espaciais de um mesmo lugar (figura 4.2), eles aparecem automaticamente sobrepostos, o que facilita a integração dos diversos dados, e, conseqüentemente, a análise desse lugar (Galati, 2006; Chang, 2019).

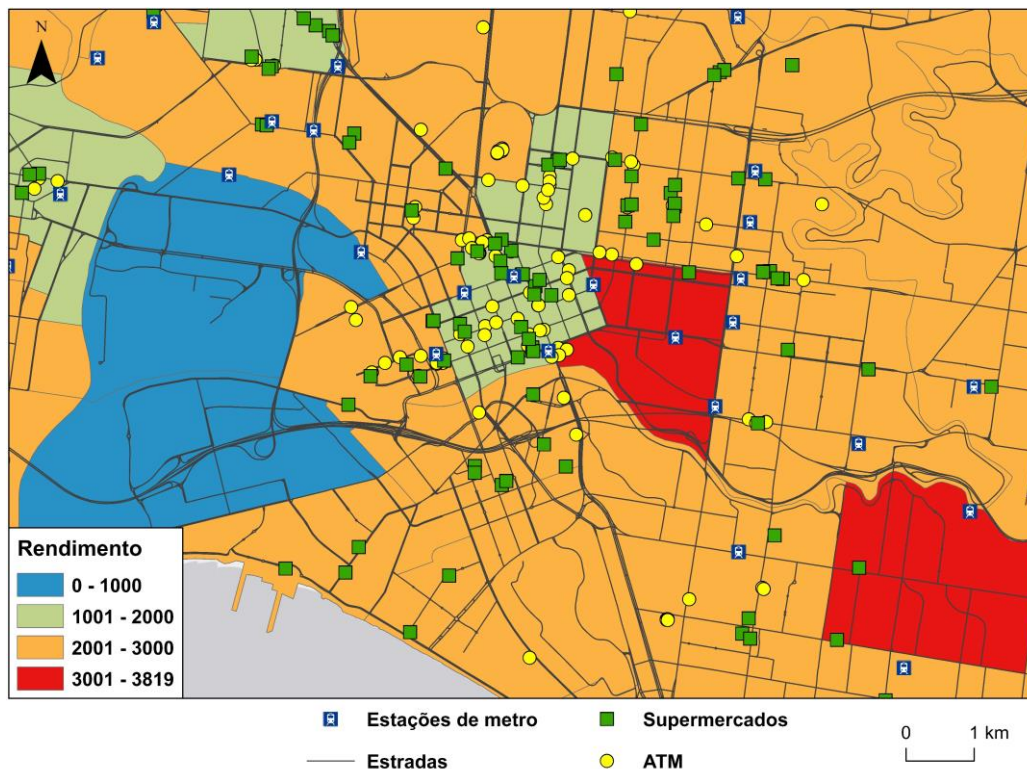


Figura 4.2 - Combinação dos dados do rendimento semanal por família (em dólares australianos) com as estações de metro, estradas, supermercados e ATM da cidade de Melbourne.

Para corresponder os diferentes dados de forma correta e para ter-se noção da dimensão do que é representado, os dados espaciais nos SIG são organizados segundo uma escala, que faz com que elementos de igual dimensão no mundo real sejam representados nos SIG com a mesma dimensão (Longley et al., 2015).

Os dados espaciais podem ser de tipo vetorial ou raster (Chang, 2019). Os dados de tipo vetorial podem assumir a forma de pontos, linhas ou polígonos, para que se possa representar as características dos elementos espaciais corretamente (Chang, 2019). Já os dados de tipo raster assumem a forma de uma grelha de células que representa os elementos espaciais (Chang, 2019).

As pessoas e as instituições são também incluídas nos SIG porque, para além de terem de ter conhecimentos em SIG, são estas quem definem as funções e os objetivos da utilização dos SIG, quem recolhe os dados de base e quem analisa, seleciona e apresenta as informações que são úteis para a tomada de decisão (Longley et al., 2015; Bolstad, 2016; Chang, 2019).

Os procedimentos aplicados no âmbito dos SIG são a recolha, o armazenamento, a representação, a consulta e a análise dos dados espaciais (Longley et al., 2015; Bolstad, 2016; Chang, 2019).

O primeiro procedimento aplicado é, normalmente, a recolha dos dados espaciais, por ser necessário obter as informações que se quer analisar (Chang, 2019). Esses dados, quando introduzidos pela primeira vez num SIG, poderão ter de ser georreferenciados ou corrigidos se tiverem erros de digitalização (Chang, 2019).

Depois, é efetuado o armazenamento dos dados em bases de dados espaciais que estão associadas aos SIG, e que, para além de armazenarem os dados espaciais, também permitem modificar, acrescentar, eliminar e organizar as informações (atributos) associadas aos mesmos (Longley et al., 2015; Bolstad, 2016; Chang, 2019).

A representação dos dados espaciais em mapas (figura 4.2) é um processo fundamental para se poder consultar e analisar os dados que estão a ser tratados (Chang, 2019). Esses mapas podem ser de cariz informal, quando se quer apenas visualizar as informações, ou formal quando têm como finalidade serem apresentados em relatórios ou palestras (Chang, 2019).

Com a consulta de dados nos SIG, feita através de mapas, tabelas ou gráficos que descrevem as características dos dados, é possível não só visualizar, mas também manipular esses dados representados para se obter novas informações, dado os *softwares* SIG possibilitarem a realização de tarefas como a classificação das informações, junção de dados, seleção de dados por atributos e localização, entre outras, o que facilita a análise de dados (Bolstad, 2016; Chang, 2019).

O processo de análise de dados é também facilitado por funções disponibilizadas nos *softwares* SIG como o *buffer* (que permite criar áreas envolventes a determinado elemento), a sobreposição de dados (que permite combinar a forma geométrica e os atributos de diferentes dados), a medição de distâncias e áreas e a criação, divisão e unificação de dados geográficos (Longley, 2008; Longley et al., 2015; Bolstad, 2016; Chang, 2019).

Com base nestes pressupostos, os SIG podem ser aplicados nas mais diversas áreas onde o espaço exerça influência, dado permitirem representar de forma mais simples a localização da população, a forma física do território, a distribuição dos recursos económicos, entre outras variáveis (Longley, 2008).

O marketing é uma dessas áreas onde os SIG podem ser aplicados, dado grande parte dos dados que são necessários analisar no âmbito do marketing estarem relacionados com um determinado espaço ou lugar (Chang, 2019). Exemplo disso é o processo de análise das áreas de mercado, procedimento essencial no marketing abordado neste relatório que pode ser facilitado pelos SIG, dado poder-se efetuar nos mesmos a conjugação da representação espacial da população, das condições económicas e de outros indicadores, para se poderem determinar os lugares onde predominam determinados segmentos da população mais propensos a adquirir um produto, e onde, por isso, é mais favorável potenciar a oferta e o consumo de um produto (Chang, 2019).

4.1 SIG e o Geomarketing

Os SIG têm vindo a adquirir um papel fundamental no mundo dos negócios, dado cada vez mais os dados utilizados na conceção das opções de marketing das empresas serem de carácter espacial (Birkin et al., 1999; Doyle, 2001; Berman & Evans, 2013). Os limites das áreas de mercado, os indicadores demográficos e socioeconómicos (que permitem identificar os potenciais consumidores dessas áreas de mercado) e as localizações atuais e potenciais dos locais de negócio constituem exemplos desses dados (Doyle, 2001; Douard, 2006; Suárez-Vega et al., 2012; Montejano & Bello, 2018).

Isto faz com que se tenha, muitas vezes, de adaptar e representar sob diversas formas os dados espaciais, para poder-se efetuar uma correta análise dos mesmos. Através dos SIG é possível realizar a adaptação dos dados espaciais (para obterem-se novos dados mais apropriados) e a representação dos mesmos em mapas (para ser mais simples verificar como estes variam no espaço) (Doyle, 2001; Chasco-Yrigoyen, 2003; Douard, 2006; Suárez-Vega et al., 2012; Widaningrum, 2015).

Os dados espaciais relativos à procura e aos locais de negócio são representados nos SIG por pontos, as áreas de mercado por linhas ou polígonos e as vias de comunicação por linhas (Doyle, 2001; Suárez-Vega et al., 2012; Montejano & Bello, 2018). A cada

elemento representado pode ser associado um conjunto de informações (atributos) como a população (total, por faixa etária, sexo, etc.), rendimento médio ou outras características que o mesmo representa, o que torna mais simples a modelação dos dados utilizados no marketing (Doyle, 2001; Clarke & Hayes, 2006; Montejano & Bello, 2018).

Devido à crescente utilização de informações espaciais, torna-se necessário proceder à sua georreferenciação e ao seu armazenamento. Com os SIG é possível realizar a georreferenciação e o armazenamento (em bases de dados) de um grande volume de informações espaciais, o que reforça a utilidade dos SIG para o geomarketing (Chasco-Yrigoyen, 2003; Clarke & Hayes, 2006; Suárez-Vega et al., 2012).

É ainda possível, nos SIG, seleccionar determinados dados por localização e por atributo, para mais facilmente encontrar esses dados, o que facilita a realização das tarefas de geomarketing como a seleção das áreas onde predominam determinados segmentos de consumidores ou a identificação de instalações ou equipamentos importantes para a análise da área de mercado (Birkin et al., 1999; Doyle, 2001).

A avaliação do potencial de mercado (procura e lucro) e a consequente definição dos limites da área de mercado de um produto, procedimentos bastante importantes e regularmente efetuados pelo marketing das empresas, podem ser realizados nos SIG, dado estes permitirem identificar as áreas com maior número de potenciais consumidores de um produto (Birkin et al., 1999; Doyle, 2001; Berman & Evans, 2013; Montejano & Bello, 2018).

Para se poder fazer a identificação dessas áreas é necessário primeiro analisar a frequência e os valores de compra dos atuais clientes de um produto, para se poder determinar os segmentos de consumidores com maior potencial para adquirir um produto (Birkin et al., 1999; Doyle, 2001; Cliquet, 2006; Berman & Evans, 2013).

Após a determinação desses segmentos de consumidores, é preciso ainda identificar informações que caracterizam esses consumidores, para se poder seleccionar as variáveis ou indicadores sociodemográficos/económicos (como a população residente, rendimento, entre outros) que permitem determinar os lugares onde se concentram os segmentos de consumidores mais favoráveis a adquirir um produto (Doyle, 2001; Clarke & Hayes, 2006; Berman & Evans, 2013).

Posteriormente, essas variáveis/indicadores obtidos podem ser processados e combinados nos SIG para identificarem-se espacialmente as áreas com maior presença desses segmentos de consumidores, que poderão constituir a área de mercado do produto (Birkin et al., 1999; Doyle, 2001; Berman & Evans, 2013).

Depois de identificadas as áreas com maior número de potenciais consumidores de um produto, os decisores do marketing podem definir nos SIG os limites da área de mercado desse produto tendo em conta:

- A distância que os potenciais consumidores do produto estão dispostos a percorrer até ao local de negócio;
- A estimativa da procura pelo produto que poderá ser gerada;
- A previsão do lucro do produto;
- A estratégia de negócio a adotar (Birkin et al., 1999; Doyle, 2001).

No processo de definição das áreas de mercado podem-se identificar áreas de mercado principais e secundárias, que podem variar consoante as características desse produto (Berman & Evans, 2013).

No âmbito da análise da área de mercado, pode ser então realizado nos SIG a:

- Representação dos dados espaciais e dos seus atributos;
- Localização de determinados elementos;
- Criação de *buffers* em torno de elementos para representar a área de mercado;
- Medição de distâncias num mapa;
- Alteração da extensão do mapa (Doyle, 2001).

Ao saber a localização dos potenciais clientes de uma empresa, através da análise da área de mercado, os especialistas de marketing podem ainda propor a criação de novos locais de negócio que permitam colmatar as lacunas de cobertura da área de mercado, dado os SIG possibilitarem a comparação da área de mercado dos locais de negócio existentes com a localização dos potenciais clientes (Doyle, 2001).

A identificação dos lugares mais atrativos para a localização de novas instalações de uma empresa, outro procedimento efetuado pelos decisores de marketing, pode ser igualmente realizado nos SIG, dado estes permitirem aplicar de forma simples a maioria dos modelos de localização existentes (Cliquet, 2006; Murray, 2010; Suárez-Vega et al., 2012).

Os SIG simplificam a aplicação dos modelos de localização porque permitem:

- Delimitar a área de mercado a ter em conta no modelo (de acordo com a distância que os consumidores estão dispostos a percorrer até ao local de negócio e com a procura coberta);
- Considerar as variáveis/indicadores que caracterizam os potenciais clientes existentes na área de mercado das novas instalações;
- Identificar a movimentação das pessoas e os transportes que utilizam;
- Representar as vias de comunicação que os potenciais clientes terão de percorrer até ao local de negócio;
- Calcular a distância que os potenciais clientes terão de percorrer;
- Aplicar funções para ajustar a modelação dos dados de acordo com os princípios associados a cada modelo;

- Executar os modelos de localização e visualizar as possíveis soluções obtidas com a execução desses modelos (Birkin et al., 1999; Doyle, 2001; Cliquet, 2006; Murray, 2010; Suárez-Vega et al., 2012; Widaningrum, 2015; Montejano & Bello, 2018).

Para além das vias de comunicação, os SIG também permitem considerar barreiras no modelo de localização, como rios ou montanhas, que podem impedir o acesso dos clientes ao novo local de negócio, o que reforça a importância da utilização dos SIG no âmbito dos modelos de localização (Suárez-Vega et al., 2012).

A aplicação de modelos de localização nos SIG pode assim ajudar a criar sistemas de suporte à decisão que permitam considerar e relacionar diversos dados espaciais, adaptar os dados espaciais aos objetivos de localização, delimitar as áreas de mercado de forma mais precisa, analisar um maior número de problemas e de estratégias de localização e verificar se, com o modelo executado, as localizações escolhidas poderão ajudar as novas instalações a cumprir os objetivos definidos para as mesmas, tarefas que dificilmente poderiam ser bem conseguidas com o simples mapeamento dos dados (Birkin et al., 1999; Cliquet, 2006; Murray, 2010; Suárez-Vega et al., 2012; Bolstad, 2016).

Os SIG são também utilizados no geomarketing para representar informações de uma empresa e das suas instalações através de mapas (como a localização, o caminho que os clientes de um lugar têm de percorrer para chegar à instalação e outras informações descritivas) e para o marketing baseado na localização, que faz a verificação do local onde se encontra o cliente, através de, por exemplo, um cartão-cliente, para informar o mesmo das ofertas e dos descontos que estão disponíveis no estabelecimento comercial mais próximo (Doyle, 2001).

Neste relatório, a aplicação dos SIG irá contribuir para que o modelo de geomarketing apresentado possa facilmente desenvolver novos indicadores mais adequados para a caracterização dos consumidores da área de mercado (dado estes indicadores serem desenvolvidos com base nas informações espaciais dessa área), por forma a tornar mais simples a tarefa dos especialistas de marketing de identificar os lugares onde predominam os segmentos de consumidores com maior potencial para consumir um produto. Também irá contribuir para que, no âmbito do modelo apresentado, se possa aplicar de forma simples os modelos de localização na seleção dos lugares mais atrativos para instalar um local de negócio.

4.2 Relevância dos SIG para a criação e análise de indicadores

Os indicadores, tal como a maior parte das representações efetuadas nos SIG, tentam identificar as informações mais importantes para serem representadas, por forma a se obter uma análise mais simplificada de um dado problema (Longley, 2007). Tal objetivo

comum tem contribuído para o sucesso da aplicação dos indicadores e dos SIG nas tarefas do marketing, onde é necessário fazer uma análise eficiente e pormenorizada da área de mercado, para que, através da compreensão das características da população (residente ou transeunte), das suas atividades e do território que ocupam, se possa descobrir as suas necessidades (Longley, 2007; Lassarre et al., 2012; Kalábová et al., 2013).

Desta forma, no âmbito do marketing, os indicadores representam muitas vezes informações espaciais relacionadas com as áreas de mercado, que precisam, por vezes, de ser adaptadas e modificadas para se poder destacar as informações mais importantes para a caracterização dessas áreas (Longley, 2007; Lassarre et al., 2012; Socha et al., 2012; Greenhalgh & King, 2013).

Os SIG, ao permitirem realizar a recolha, armazenamento, adaptação, representação, visualização e a análise das informações espaciais dos indicadores, podem facilitar a utilização dos indicadores no processo de caracterização das áreas de mercado efetuado pelo marketing (Longley, 2007; Lassarre et al., 2012; Socha et al., 2012; Greenhalgh & King, 2013; Janowski et al., 2014).

Ao possibilitarem o armazenamento e a organização de um grande volume de informações georreferenciadas relativas aos indicadores em bases de dados associadas, os SIG podem contribuir para um adequado enquadramento espacial da informação dos indicadores (Lassarre et al., 2012; Socha et al., 2012).

Dado poder-se sobrepor as informações dos diferentes indicadores por camadas, calcular os dados dos diferentes indicadores e combinar os dados das diferentes camadas, os SIG podem contribuir para a análise espacial dos indicadores e para a obtenção de novas informações que poderão servir de base ao desenvolvimento de novos indicadores (Lassarre et al., 2012; Socha et al., 2012; Greenhalgh & King, 2013; Kalábová et al., 2013).

Com os SIG é também possível representar as informações espaciais dos indicadores através de mapas que classificam os territórios de acordo com essas informações, o que facilita a apresentação e o destaque de determinadas características do território, descritas pelos indicadores, que poderão ajudar a caracterizar a área de mercado e a verificar as relações espaciais existentes entre os diferentes indicadores (Socha et al., 2012; Greenhalgh & King, 2013; Kalábová et al., 2013).

A visualização das informações espaciais nos SIG permite ainda identificar mais facilmente as principais características do território descritas pelos indicadores, dado ser possível efetuar nos SIG a seleção das áreas onde o indicador tem um valor específico ou onde o valor representado pelo indicador é mais baixo ou mais elevado, o que poderá ser útil para identificar os lugares onde existe maior potencial de mercado (Lassarre et al., 2012; Socha et al., 2012; Greenhalgh & King, 2013; Kalábová et al., 2013).

O processo de análise dos indicadores é também facilitado pelos SIG porque estes permitem efetuar funções como a visualização de tabelas de atributos com os valores dos indicadores por área considerada, a mudança de escala, a comparação dos valores verificados em diferentes períodos temporais, o cálculo da área a que corresponde o valor espacial representado pelo indicador e a ponderação dos valores obtidos nos indicadores em cada área considerada para áreas com diferentes dimensões (figura 4.3), o que faz com que se torne mais simples a caracterização das áreas de mercado (Lassarre et al., 2012; Greenhalgh & King, 2013).

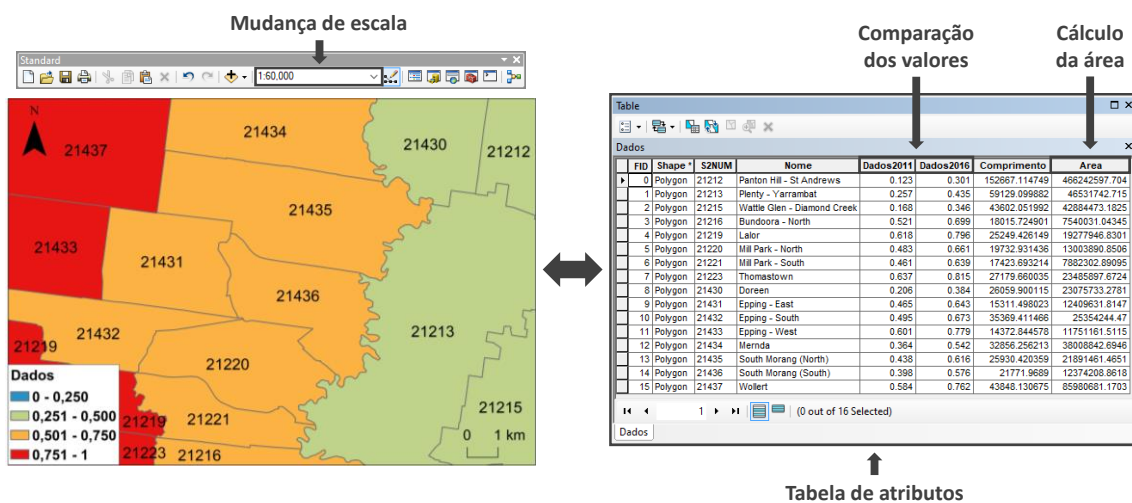


Figura 4.3 - Funções de SIG que podem ser aplicadas no âmbito da análise dos indicadores.

Através dos SIG é também possível a contextualização espacial dos valores obtidos nos indicadores em cada lugar, dado ser possível nos SIG comparar a informação espacial dos indicadores com dados de base, como por exemplo o relevo, os rios ou outras barreiras físicas do território, e com dados associados ao funcionamento da sociedade, como por exemplo, a rede rodoviária, o tráfego de veículos, o movimento de pessoas ou os transportes públicos, que poderão influenciar o potencial de negócio (Lassarre et al., 2012; Greenhalgh & King, 2013; Janowski et al., 2014). Esses dados podem ser obtidos, por exemplo, no *OpenStreetMap*, que consiste num servidor de dados abertos que disponibiliza um vasto conjunto de dados que podem ser processadas nos SIG (Janowski et al., 2014).

Desta forma, com a utilização dos SIG, torna-se então mais simples a seleção dos indicadores e das respetivas informações mais relevantes para a caracterização das áreas de mercado, assim como o desenvolvimento de indicadores mais vocacionados

para a análise dessas áreas (Lassarre et al., 2012; Socha et al., 2012; Kalábová et al., 2013; Janowski et al., 2014).

Com o processamento espacial dos indicadores que permitem antever os comportamentos dos consumidores nos SIG, pode ser assim facilitada a escolha das áreas onde é mais conveniente oferecer um produto, a seleção dos novos locais para localizar o negócio de produtos e a previsão das vendas futuras de um produto num dado lugar (Doyle, 2001).

No âmbito do modelo de geomarketing apresentado, os SIG serão utilizados para facilitar o desenvolvimento de novos indicadores de valor acrescentado, através da transformação dos dados de carácter espacial em informações por área estatística que, conjuntamente com os dados dos censos, servirão de base ao desenvolvimento de novos indicadores, e para a representação dos indicadores obtidos em mapas temáticos que poderão tornar mais simples a tarefa de analisar a área de mercado.

5. Desenvolvimento do modelo

O modelo espacial desenvolvido pretende que seja possível a sua aplicação no contexto de qualquer território. Por isso, definiu-se uma estrutura única para o modelo (figura 5.1), por forma a que a aplicação do modelo seja o mais flexível possível. Neste relatório, o modelo desenvolvido será aplicado para o contexto da área de mercado da Austrália, tendo em conta a aplicação de um conjunto de procedimentos (figura 5.2).

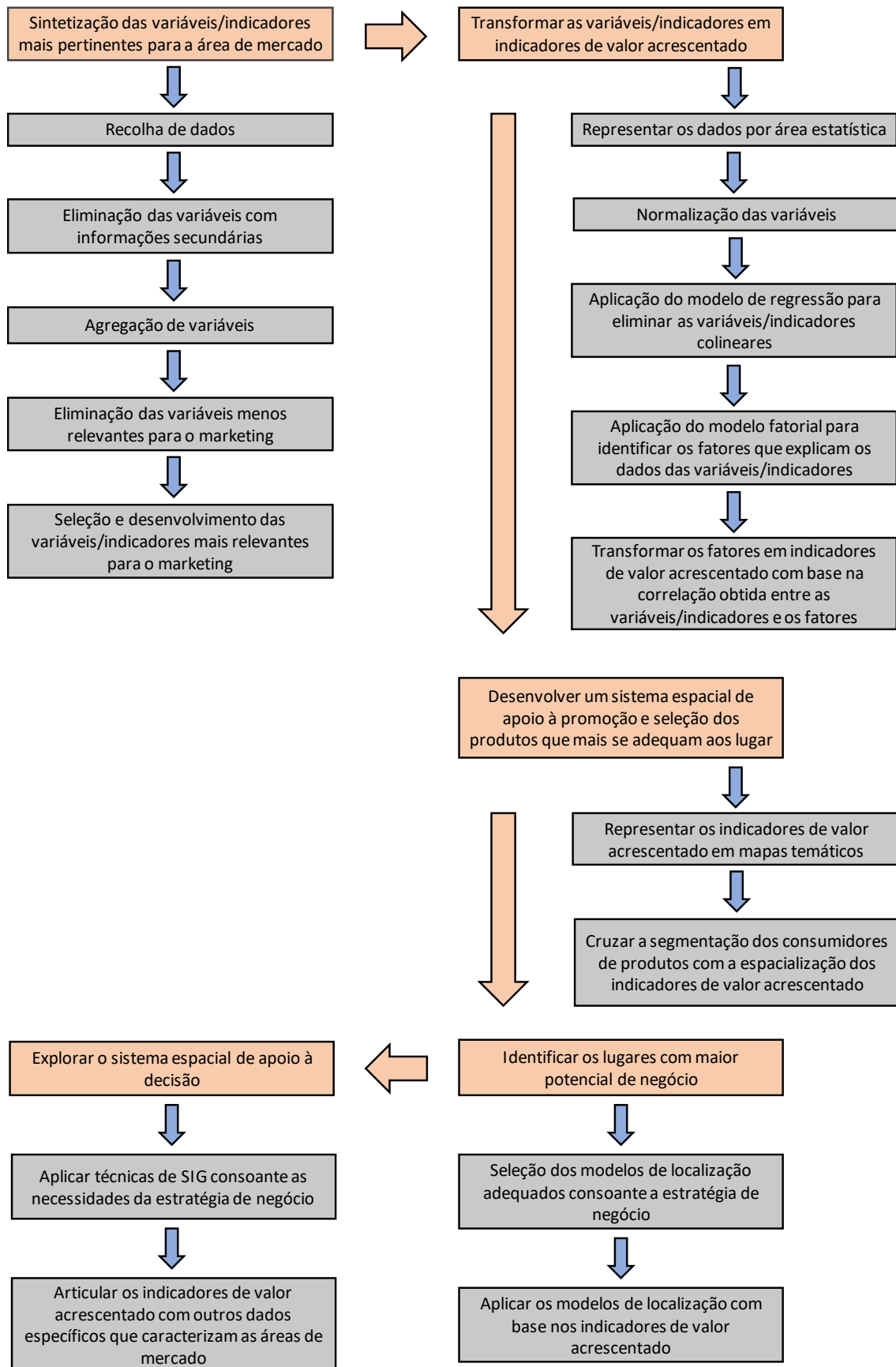


Figura 5.1 - Estrutura do modelo.

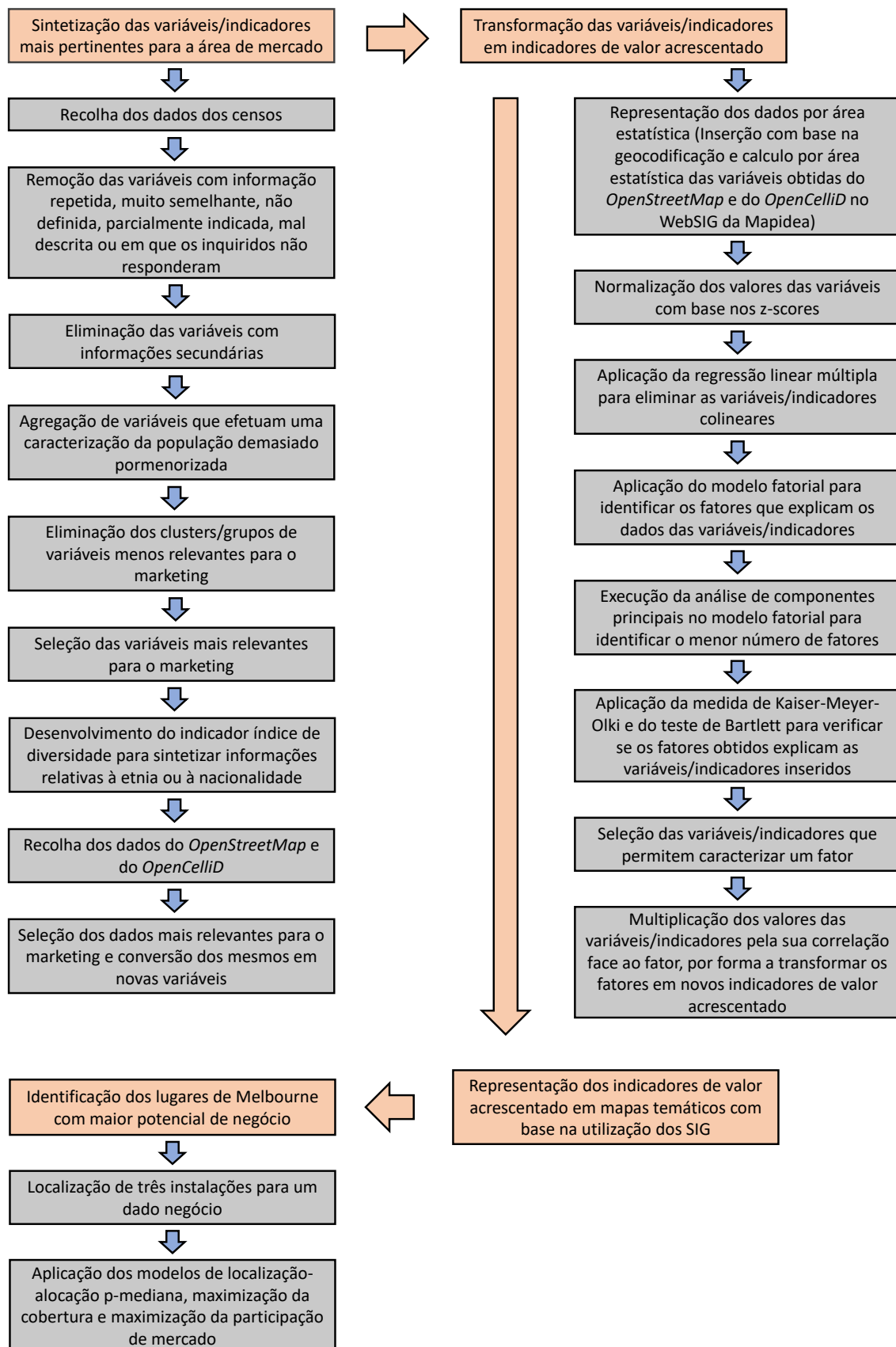


Figura 5.2 - Conjunto de procedimentos aplicados.

5.1 Informação geográfica utilizada

Para a aplicação deste modelo, foram utilizados apenas dados abertos. Segundo a Open Knowledge Foundation (2005), são considerados dados abertos os dados que qualquer pessoa, no lugar que queira e para a aplicação que quiser pode utilizar, construir, modificar e compartilhar de forma livre.

É relevante a utilização de dados abertos devido à:

- Disponibilidade do total do conjunto de dados;
- Possibilidade de serem utilizados e modificados de forma simples;
- Reutilização e redistribuição dos dados que pode ser efetuada;
- Poderem ser combinados com outros conjuntos de dados (Monino & Sedkaoui, 2016).

A utilização de dados abertos possibilitou então a análise e modificação de um conjunto vasto de dados relativos à Austrália, que, caso não fossem de acesso livre, seria muito difícil proceder à sua recolha.

Esses dados abertos recolhidos para a Austrália foram obtidos através dos:

- Censos;
- *OpenStreetMap*;
- *OpenCellID*.

Dos censos da Austrália foram recolhidos dados de variáveis demográficas e socioeconómicas que se encontravam numa base de dados disponibilizada pelo *Australian Bureau of Statistics*, que tem todos os dados de acesso livre relativos aos censos de 2016 (os censos mais recentes efetuados). Desses dados disponibilizados foram selecionados todos os que eram relativos às áreas estatísticas de nível 1 (SA1), que são as unidades territoriais mais pequenas de acesso livre, e os que eram relativos às áreas estatísticas de nível 2 (SA2), que constituem uma unidade territorial de tamanho médio dos censos (com um tamanho superior face às SA1). As áreas de governo local da Austrália, por não coincidirem com as áreas estatísticas selecionadas, não foram consideradas.

Segundo os censos da Austrália as unidades territoriais SA1 têm geralmente, cada uma delas, uma população entre as 200 e as 800 pessoas (à exceção das áreas não habitadas), com uma população média de 400 pessoas cada, e cobrem todo o território da Austrália sem lacunas nem sobreposições. Estas pretendem representar partes de territórios contínuos predominantemente rurais (com densidade populacional inferior a 200 pessoas por km^2) ou urbanos (com densidade populacional superior a 200 pessoas por km^2).

Já as unidades territoriais SA2, que são constituídas com base nas SA1, têm geralmente cada uma delas uma população entre as 3000 e as 25000 pessoas (à exceção das áreas não habitadas), com uma população média de 10000 pessoas cada, e cobrem também todo o território sem lacunas nem sobreposições. Estas pretendem representar uma área funcional, de uma cidade pequena ou um subúrbio de uma cidade grande, que demonstre as interações sociais e económicas de uma comunidade.

Foram também recolhidas, dos censos, as subdivisões das SA1 e SA2, para se poder georreferenciar os diferentes dados recolhidos em cada uma das SA1 e SA2.

Do *OpenStreetMap*, que constitui um servidor de dados abertos desenvolvido por uma comunidade de voluntários que insere e atualiza dados de pontos de interesse nos mapas do servidor, foram recolhidos para a Austrália dados de bancos, transportes públicos e centros comerciais. Este servidor destaca-se pelo facto de qualquer utilizador poder visualizar, carregar e atualizar dados no mesmo, o que faz aumentar o volume de dados disponíveis, e por poder-se utilizar esses dados para aquilo que se quiser. Para proceder à recolha dos dados do *OpenStreetMap* utilizou-se o *Geofabrik*¹, que constitui um servidor de acesso livre que fornece os arquivos de dados atualizados do *OpenStreetMap*.

Por fim, do *OpenCellID*², que representa a maior base de dados com a posição GPS das antenas telefónicas, foram recolhidos os dados das antenas telefónicas para a Austrália. Esta base de dados de livre acesso é também desenvolvida por uma comunidade de voluntários e os seus utilizadores podem usar os dados das antenas para aquilo que quiserem.

¹ O *Geofabrik* consiste num servidor de dados geográficos que procede diariamente à recolha, seleção e processamento dos dados do *OpenStreetMap* com vista à sua disponibilização livre. Este servidor permite efetuar o *download* gratuito de dados geográficos do *OpenStreetMap* para todo o mundo em formato *shapefile* (formato que possibilita o processamento dos dados em qualquer *software* SIG), podendo o utilizador do servidor escolher o país (ou a região em alguns países) sobre o(a) qual pretende recolher os dados.

² O *OpenCellID* corresponde a uma base de dados que contém as coordenadas GPS das antenas telefónicas recolhidas para poderem ser utilizadas livremente em qualquer contexto. Através do *Website* do *OpenCellID* é possível realizar o *download* gratuito da posição das antenas telefónicas, tendo o utilizador apenas de criar uma conta gratuita e fazer, no *Website*, o pedido da API relativa à localização da base de dados para conseguir fazer o *download* do ficheiro CSV com as coordenadas latitude e longitude das antenas telefónicas. Os utilizadores podem escolher efetuar a recolha da posição de todas as antenas telefónicas do mundo ou de um país em específico.

5.2 Tratamento da Informação

Após recolhidos os dados dos censos da Austrália para as unidades territoriais SA1 e SA2, foi criada uma tabela em *Excel* com a descrição de todas as variáveis de informação disponibilizadas pelos censos. Devido ao número elevado de variáveis fornecidas pelos censos, 15535, numa primeira fase foram removidas todas as variáveis com informação repetida, muito semelhante, não definida, parcialmente indicada, descrita de forma inadequada ou em que os inquiridos pelos censos optaram por não responder. Com a remoção das mesmas, ainda sobraram muitas variáveis (acima de 10000).

Por isso, procedeu-se à seleção e criação das variáveis com relevância para o marketing, consoante a seguinte metodologia:

- Eliminação de variáveis com informações secundárias;
- Agregação de variáveis;
- Eliminação de *clusters* de variáveis e grupos de variáveis menos relevantes para o marketing.

Assim, primeiro optou-se por retirar dos clusters de variáveis fornecidos pelos censos (anexo 1) as variáveis com informações secundárias. Foram consideradas informações secundárias aquelas que descrevem outras informações associadas às pessoas caracterizadas num dado *cluster* de variáveis. As variáveis com essas informações foram selecionadas e posteriormente eliminadas em todos os clusters de variáveis com base nas informações referidas no quadro 5.1 (salvo exceções referidas no mesmo quadro).

Quadro 5.1 - Informações secundárias das variáveis retiradas dos *clusters* de variáveis e respectivas exceções.

Informações secundárias das variáveis retiradas dos <i>clusters</i> de variáveis	<i>Clusters</i> de variáveis onde essas informações não foram retiradas
Sexo	<ul style="list-style-type: none"> • Sexo • Idade • Rendimento individual total semanal • Relação do agregado familiar (apenas nas variáveis solteiros(as) com filhos) • Percentagem de desemprego
Idade	<ul style="list-style-type: none"> • Idade • Rendimento individual total semanal • Necessidade de assistência • Situação da força de trabalho (apenas nas variáveis do desemprego)
Ano de chegada à Austrália	
Crianças dependentes	<ul style="list-style-type: none"> • Cuidados infantis não remunerados • Número de Filhos
Composição das famílias	<ul style="list-style-type: none"> • Relação do agregado familiar (apenas nas variáveis solteiros(as) com filhos) • Famílias e a sua composição • Filhos dependentes
Ascendência dos Pais	<ul style="list-style-type: none"> • Ascendência
Presença em agregados familiares	<ul style="list-style-type: none"> • Rendimento do agregado familiar total semanal (apenas nas variáveis das famílias por agregado familiar)
Estrutura da Habitação	<ul style="list-style-type: none"> • Estrutura da Habitação; • Situação de posse da habitação
Tipo de Senhorio	
Ocupação dos trabalhadores	<ul style="list-style-type: none"> • Ocupação dos trabalhadores
Atividade dos trabalhadores empregados	<ul style="list-style-type: none"> • Atividade dos trabalhadores empregados

Depois de retiradas as variáveis com informações secundárias da tabela com as variáveis dos censos, ficou-se com 1681 variáveis, o que ainda é um número elevado.

Desta forma, optou-se, em seguida, por agregar variáveis de alguns *clusters* de variáveis em novas variáveis, por forma a simplificar a apresentação da informação e a aumentar o valor da informação das variáveis. As variáveis finais com grupos de informação mais relevantes para um dado *cluster*, obtidas algumas delas através da agregação de variáveis, foram representadas no Anexo 2.

O critério utilizado para a agregação das variáveis, nos casos dos *clusters* da ascendência, país de nascimento e capacidade de falar inglês e/ou outras línguas (anexo 2), foi manter as dez variáveis com maior número de pessoas reportadas, e agrupar as restantes variáveis numa variável única designada outros(as) países/lugares/línguas, consoante o *cluster* (à exceção da ascendência Aboriginal, das pessoas que falam línguas indígenas Australianas e dos totais dos *clusters* que se mantiveram fora da agregação). No caso do método de viagem para o trabalho agregou-se quem utiliza o automóvel numa única variável. Nos restantes *clusters* com variáveis agregadas, a agregação das classes foi feita de forma a normalizar essas informações face às variáveis de outros *clusters*.

Para os *clusters* de variáveis apresentados no anexo 2 foram eliminadas todas as restantes variáveis com informação não indicada em nenhuma das variáveis referidas como variáveis finais.

Após a agregação das variáveis de alguns *clusters*, a tabela com as variáveis dos censos passou a ter 586 variáveis. Como ainda é um número elevado de variáveis, procedeu-se à remoção de *clusters* de variáveis e de grupos de variáveis (de um *cluster*) que se considerou, tendo em conta o conjunto de *clusters* de variáveis normalmente utilizados pelas empresas de marketing para segmentar os públicos-alvo, menos relevantes para o marketing. Os *clusters* de variáveis removidos são apresentados no quadro 5.2 e os grupos de variáveis removidas no quadro 5.3.

Quadro 5.2 - *Clusters* de variáveis dos censos removidos por serem considerados pouco relevantes para o marketing.

<i>Clusters</i> de variáveis removidos	
<ul style="list-style-type: none"> • Pessoas na noite do censo em casa • Caracterização dos Indígenas • Pessoas com cidadania australiana • Pessoas na noite do censo consoante a propriedade de residência • Lugar de residência habitual na noite do censo • Estado civil perante a sociedade • Estrutura da habitação - quartos 	<ul style="list-style-type: none"> • Maior ano de escolaridade concluído (à exceção da variável com as pessoas que nunca foram à escola) • Assistência não remunerada a uma pessoa com deficiência • Rendimento dos agregados familiar total semanal (à exceção das variáveis com o total de presença em agregados familiares) • Prestação do empréstimo mensal - habitações

Quadro 5.3 - Grupos de variáveis removidas por serem consideradas pouco relevantes para o marketing.

Grupos de variáveis removidas	
<ul style="list-style-type: none"> • Várias religiões cristãs, indígenas ou outras crenças seculares ou espirituais • Propriedade da instituição de ensino • Estudantes em regime de full-time/part-time • Composição das famílias consoante filhos e/ou estudantes dependentes • Pessoas e famílias em outro tipo de famílias • Habitações de outro tipo de propriedade consoante a estrutura 	<ul style="list-style-type: none"> • Lugar em que residiam as pessoas com morada diferente à 1 e/ou à 5 anos atrás • Pessoas empregadas longe de casa • Desempregados à procura de trabalho • Percentagem de força de trabalho • Percentagem de pessoas empregadas • Tipo de emprego que os desempregados procuram • Pessoas com formação escolar mista

Com a remoção dos *clusters* de variáveis e das variáveis consideradas menos relevantes para o marketing, pôde-se obter as 381 variáveis provenientes dos censos que se considerou serem relevantes para o marketing. Estas variáveis obtidas irão servir de base à seleção das variáveis/indicadores mais relevantes para o marketing (para serem inseridos/as na máquina da Mapidea e utilizados/as no modelo fatorial aplicado que pretende desenvolver indicadores de valor acrescentado) e à georreferenciação da informação (necessária para incluir as variáveis/indicadores na máquina da Mapidea).

No caso dos dados do *OpenStreetMap*, depois de serem recolhidos, foram adicionados ao *ArcGIS* todos os dados de pontos provenientes do *Geofabrik* para poderem ser visualizados. Depois de serem adicionados, foram selecionados os dados relativos aos transportes, bancos, ATM, postos de combustível, centros comerciais e supermercados, por constituírem dados relevantes para o marketing, devido a representarem algumas das variáveis que mais influenciam a área de mercado e o consumo dos produtos, e por nesses lugares ser mais provável encontrar potenciais consumidores de um produto. No caso dos supermercados, estes foram divididos em supermercados Coles, Coles Express, Woolworths e outros supermercados, para se obter uma análise mais pormenorizada. Os dados das antenas do *OpenCellID* não precisaram de ser tratados. Estas 10 variáveis obtidas irão igualmente servir de base à georreferenciação da informação, para poderem ser inseridas na máquina da Mapidea e utilizadas no modelo fatorial aplicado.

5.3 Clusters de variáveis e indicadores mais relevantes obtidos com os censos

Depois de obterem-se as variáveis provenientes dos censos com relevância para o marketing, foram selecionadas as variáveis dos *clusters* de variáveis que se considerou terem maior relevância para o marketing (por constituírem dados referentes à idade, sexo, rendimento, escolaridade e profissão, considerados importantes segundo a bibliografia) e foi desenvolvido um indicador para sintetizar informações relativas à etnia ou à nacionalidade (por serem também consideradas informações importantes pela bibliografia), para se poder inserir as respetivas variáveis e o dado indicador na máquina da Mapidea (por forma a facilitarem a análise das áreas de mercado) e utilizar os mesmos no modelo fatorial aplicado (para permitirem o desenvolvimento dos indicadores de valor acrescentado).

Os *clusters* de variáveis considerados mais importantes para o marketing, obtidos através da seleção de 42 variáveis dos censos, encontram-se referidos no quadro 5.4. Antes de obterem-se as variáveis dos *clusters* considerados foi, no entanto, necessário efetuar alguns procedimentos:

- No *cluster* de variáveis dos residentes consoante a estrutura da habitação foi preciso agregar previamente duas das variáveis (vivendas e moradias);

- Já no *cluster* de variáveis relativas às famílias foi preciso agregar previamente as variáveis das famílias monoparentais e das famílias de casais que têm filhos com menos de 15 anos, para obter-se uma variável única relativa às famílias com filhos com menos de 15 anos, e as famílias monoparentais e as famílias de casais sem filhos com menos de 15 anos, para se obter uma variável única com as famílias sem filhos com menos de 15 anos, de forma a se poder adicionar estas variáveis e o total de famílias;
- No caso do *cluster* de variáveis relativas à escolaridade não se incluiu as variáveis com o tipo de certificado;
- Por fim, no *cluster* de variáveis do método de viagem para o trabalho, foi preciso agregar previamente numa única variável quem usa transportes públicos para ir para o trabalho, noutra variável única quem usa automóvel ou mota, e noutra variável única quem usa bicicleta, vai a pé ou outros.

Quadro 5.4 - *Clusters* de variáveis mais relevantes obtidos através da seleção de variáveis dos censos.

<i>Clusters</i> de variáveis selecionados	
<ul style="list-style-type: none"> • Sexo • Idade (dos jovens, pessoas em idade ativa e idosos) • Mediana de idade das pessoas • Mediana do rendimento semanal por pessoa • Mediana da renda semanal da habitação • Mediana do rendimento semanal por família • Média do tamanho do agregado familiar • Estudantes (total) • Famílias (total e consoante a existência de filhos) • Filhos dependentes (total) • Rendimento familiar total semanal 	<ul style="list-style-type: none"> • Veículos a motor (nenhum veículo e total de habitações com veículos) • Habitações por ocupação • Residentes consoante a estrutura da habitação • Prestação do empréstimo mensal (total de famílias) • Prestação de renda semanal (total de habitações arrendadas) • Situação da força de trabalho (total de pessoas empregadas e desempregadas) • Escolaridade • Método de viagem para o trabalho

O indicador desenvolvido para sintetizar as informações relativas à etnia ou à nacionalidade designou-se por índice de diversidade.

O índice de diversidade criado consiste na probabilidade de uma pessoa escolhida aleatoriamente numa SA1 ou SA2 de ser uma pessoa indígena, de ter um dos pais que nasceu ou ambos os pais nasceram fora da Austrália ou de ser uma pessoa que nasceu fora do país (Reese-Cassal, 2014). Este índice é útil porque permite medir a diversidade étnica e racial de cada uma das unidades territoriais SA1 e SA2, fator que muitas vezes influencia o consumo de produtos.

Os valores do índice de diversidade que foi desenvolvido variam entre 0 (sem diversidade étnica ou racial), em que todas as pessoas nasceram na Austrália e são descendentes de australianos não indígenas, e 1 (diversidade étnica ou racial total), em que todas as pessoas não nasceram na Austrália e não são descendente de australianos não indígenas (Reese-Cassal, 2014).

Assim, ao considerarem-se as seguintes variáveis *i* dos *clusters* de variáveis no índice de diversidade:

1. Pessoas Indígenas (Indígenas);
2. Ambos os pais nasceram fora do país (Ascendência);
3. Só o Pai nasceu fora do país (Ascendência);
4. Só a Mãe nasceu fora do país (Ascendência);
5. Pessoas que nasceram fora do país (País de nascimento)

e que se deve retirar da soma dos valores das variáveis 2, 3 e 4 o valor da variável 5 (por forma a excluir as pessoas que nasceram fora da Austrália e que têm um dos pais ou ambos os pais que nasceram fora da Austrália), o índice de diversidade para cada área estatística considerada pode ser calculado com base na seguinte equação:

$$D = \frac{\sum n_i}{p} \quad (\text{Eq. 5.1})$$

em que:

- *D* é o índice de diversidade;
- *n_i* consiste no número de pessoas que é caracterizada pela variável *i*;
- *p* é a população total da área estatística considerada.

Desta forma, foi possível obter as 42 variáveis dos 19 *clusters* de variáveis e o indicador desenvolvido com informações provenientes dos censos (anexo 3), que foram incluídas na máquina da Mapidea. Para incluir estas variáveis/indicadores e as 10 variáveis

obtidas com os dados do *OpenStreetMap* e do *OpenCellID* (anexo 3) na base de dados da máquina da Mapidea foi necessário proceder à georreferenciação das mesmas.

5.4 Georreferenciação

A georreferenciação consiste no processo de atribuição da localização geográfica a determinados dados, através do nome dos lugares, da morada, de códigos desenvolvidos ou do sistema de coordenadas (latitude e longitude) (Hill, 2006; Yao, 2009).

Com a georreferenciação é então possível articular conjuntos de informações a uma determinada localização geográfica (Hill, 2006). Essa articulação é útil, pois permite verificar como as informações dos dados se relacionam entre si e com as divisões administrativas, rios, relevo, ou qualquer outro ponto geográfico de interesse (Hill, 2006). Através das localizações obtidas com a georreferenciação dos dados é também possível determinar a distância e a direção de um lugar georreferenciado face a outros lugares representados, e verificar se existe sobreposição da área desse lugar face a outros lugares (Hill, 2006).

A aplicação da georreferenciação deve ter em conta a individualidade da referenciação, em que deve-se atribuir a cada lugar uma única referência (coordenada de latitude e longitude, código ou nome) exclusiva para esse lugar, a resolução espacial, ou seja, a dimensão da área que é coberta pela figura ou ponto espacial a georreferenciar, a escala e o sistema de projeção dos dados (Hill, 2006; Yao, 2009).

Assim, a georreferenciação é essencial para a visualização, adaptação, armazenamento e recuperação de informações em que a distribuição geográfica é importante (Hill, 2006). A visualização de informação associada aos lugares num território é importante, pois permite mais facilmente identificar os lugares que têm mais ou menos importância para uma dada análise (Hill, 2006).

Como tipos de georreferenciação existe a georreferenciação direta, que georreferencia os locais através da colocação das coordenadas geográficas da figura ou ponto espacial num sistema de coordenadas geográficas (caso da georreferenciação através da latitude e longitude), e a georreferenciação indireta, que georreferencia um lugar através da sua relação com elementos geográficos, sendo a relação determinada pelos nomes dos lugares, moradas e códigos (caso da geocodificação) (Hill, 2006; Yao, 2009; Li, 2018).

Essa relação na georreferenciação indireta pode ter em conta o facto de o lugar corresponder a uma figura espacial já georreferenciada, de este estar a uma distância de uma figura já georreferenciada e do mesmo estar relacionado com uma figura georreferenciada (Yao, 2009).

A georreferenciação direta, por constituir uma área contínua, não tem um limite de resolução espacial, enquanto que a georreferenciação indireta, por localizar locais únicos de um território como áreas estatísticas, administrativas ou moradas, tem um limite de resolução de acordo com aquilo que é georreferenciado (Yao, 2009).

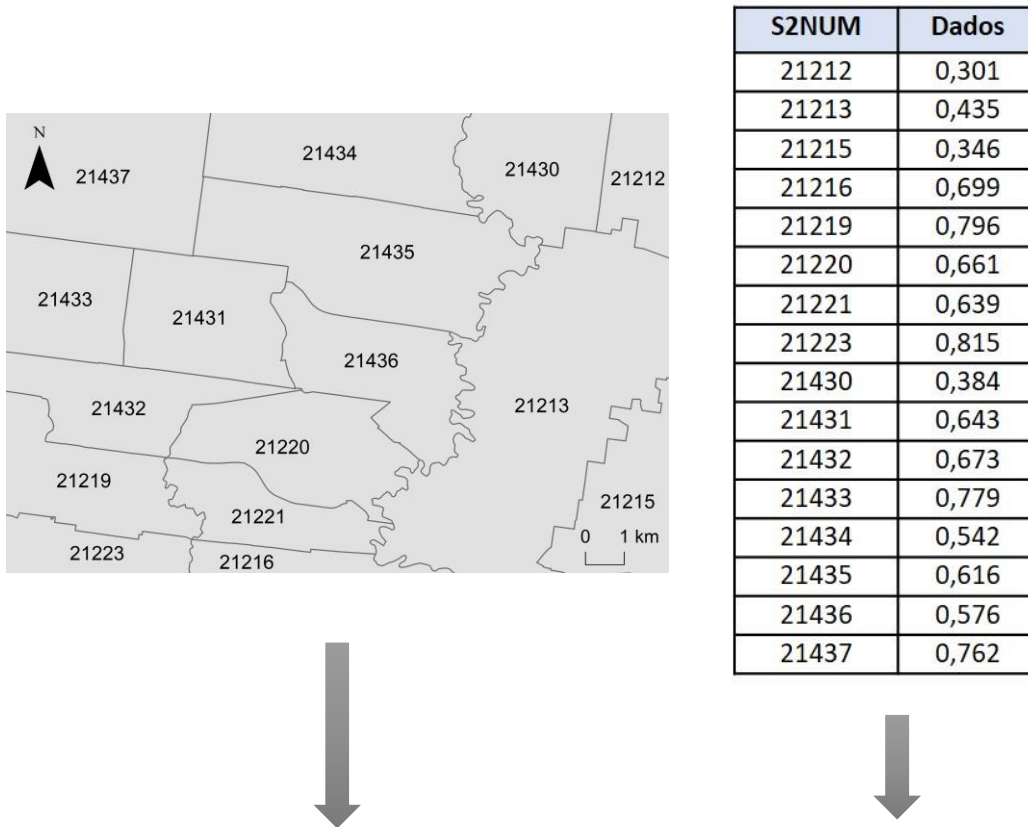
Como forma de georreferenciação direta, a georreferenciação através das coordenadas da latitude e longitude permite localizar espacialmente as informações com coordenadas associadas, tendo por base sistemas de posicionamento, navegação e de informação geográfica (Hill, 2006; Yao, 2009).

Uma forma de georreferenciação indireta é a geocodificação, que consiste no processo de atribuição de um código único (código postal, código estatístico, código de cadastro ou outro tipo de código) para cada uma das figuras georreferenciadas previamente, por forma a articular (referenciar espacialmente) potenciais dados e informações associadas a essas figuras (Jackson, 1997, citado por Hill, 2006; Goldberg et al., 2007; Yao, 2009; Bolstad, 2016). A geocodificação da informação pode ser feita para figuras de pontos, linhas ou áreas (Goldberg et al., 2007).

A atribuição do código para as figuras georreferenciadas, ou seja, com coordenadas de latitude e longitude associadas, é feita tendo em conta que um território considerado é dividido em áreas administrativas ou estatísticas, segundo uma hierarquia, e que para cada nível dessa hierarquia as áreas administrativas ou estatísticas têm um código único que permite identificar cada uma dessas áreas na geocodificação (Hill, 2006; Yao, 2009; Li, 2018). Esse código único atribuído, para cada área administrativa ou estatística, deve ser colocado, tanto nas figuras georreferenciadas, como nas informações das áreas a relacionar com os dados espaciais (figura 5.3) (Hill, 2006; Li, 2018). Numa geocodificação das figuras previamente realizada, se existirem códigos duplicados, estes devem ser modificados para que se atinja uma correta geocodificação que permita realizar a correspondência adequada das informações (Goldberg et al., 2007; Li, 2018).

Depois de atribuídos os códigos, procede-se ao *input* dos dados, em que são incluídos os dados espaciais que representam o lugar em que se quer geocodificar a informação e os dados com as novas informações a geocodificar (figura 5.3). Em seguida, procede-se ao processamento do algoritmo, em que é efetuada a correspondência entre os dados espaciais e as informações a geocodificar, através da ligação dos códigos correspondentes existentes nos dados espaciais e nas informações a geocodificar. Por fim, obtêm-se os dados espaciais com a informação inserida (Goldberg et al., 2007; Li, 2018).

Atribuição de um código único para cada área administrativa ou estatística:



Junção dos dados espaciais com os dados que contêm as novas informações, através do código atribuído:

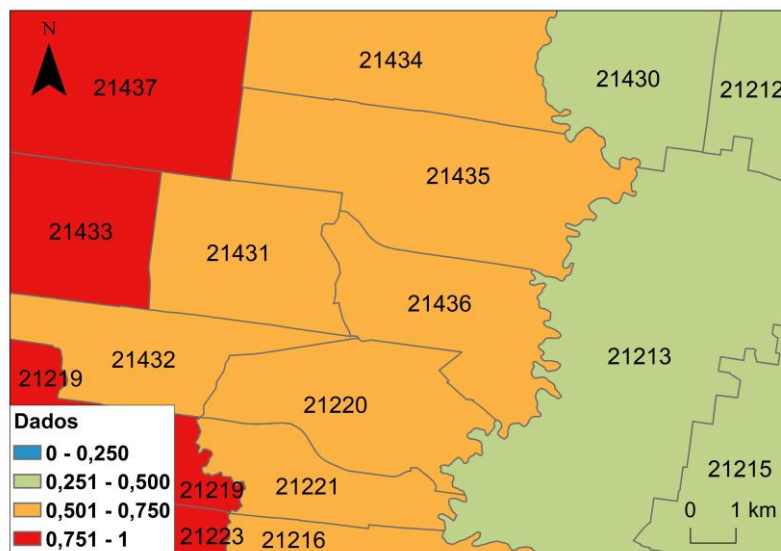


Figura 5.3 - Exemplificação do processo de geocodificação.

O processo de geocodificação permite, desta forma, que qualquer conjunto de dados espaciais codificados possam ser inseridos num ambiente SIG e/ou numa base de dados espacial para serem mapeados (Longley et al., 2015). A geocodificação é também essencial em qualquer base de dados espacial porque permite organizar as informações armazenadas por hierarquias de localização, o que facilita a procura da informação armazenada (Hill, 2006; Yao, 2009).

Através dos códigos geográficos e da georreferenciação por coordenadas, é então possível referenciar corretamente as informações num SIG, o que facilita o alinhamento das diferentes camadas de informação referenciadas espacialmente e dos respetivos dados (figura 5.4) que permitem criar mapas e análises novas da informação (Hill, 2006).

Como camadas de informação, um mapa pode incluir (figura 5.4):

- Cartografia de base (limites administrativos, limites estatísticos, relevo, rede de estradas, parques verdes, entre outras informações de base);
- Dados de análise (informações sociodemográficas e/ou económicas) (Hill, 2006).

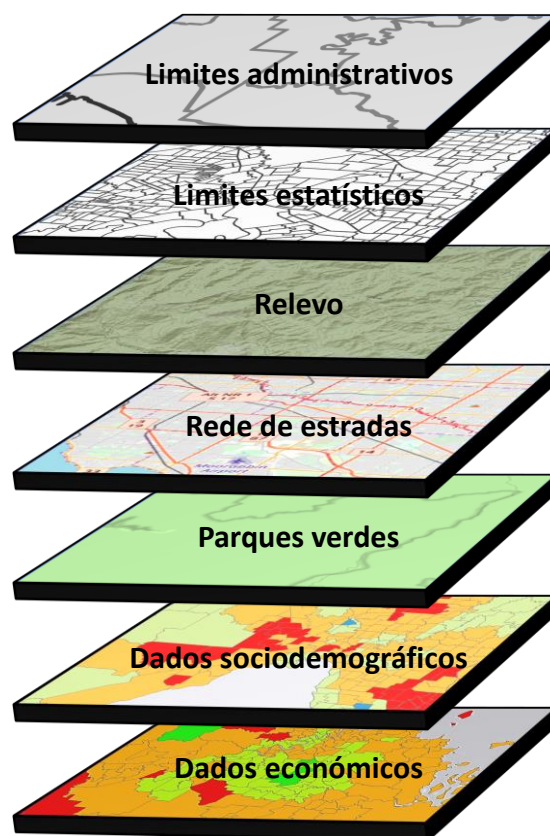


Figura 5.4 - Exemplos de camadas de informação que um mapa pode incluir e possível alinhamento.

Assim, a inclusão e a representação de informações nos conjuntos de dados espaciais existentes é facilitada pela geocodificação e pelas georreferenciação por coordenadas, o que é importante para encontrar novos fatores que influenciam a variação das informações e para identificar novas informações (Li, 2018). Por isso, a geocodificação e a georreferenciação por coordenadas são bastante aplicadas no marketing das empresas, porque tornam mais simples a identificação da localização de potenciais clientes, dos locais de negócio e consequentemente dos padrões de compra/utilização dos produtos e dos impactos das campanhas de marketing (Hill, 2006; Bolstad, 2016).

Para facilitar a interoperabilidade entre a geocodificação e os dados que são geocodificados é importante representar as informações em metadados (Hill, 2006; Yao, 2009).

Os metadados constituem um conjunto de informações detalhadas e organizadas que descrevem uma ou mais informações em análise, sendo estes desenvolvidos para facilitar o uso das informações, e que sem eles seria mais complicado encontrar as informações específicas a ter em conta na geocodificação (Hill, 2006). Estes devem incluir, para cada um dos dados representados, um código de geocodificação (para poder-se relacionar as informações dos dados à sua localização geográfica), um título, uma descrição, a data de recolha e a fonte dos dados (Hill, 2006). Estas informações descritivas simplificam a análise dos dados, pois permitem aceder, descobrir e recuperar as informações, avaliar a possível utilização dos dados, ver os critérios para a utilização e ver como os dados estão organizados (Hill, 2006).

A tabela de *Excel* onde foram colocadas as variáveis provenientes dos *clusters* de variáveis recolhidos dos censos e o indicador desenvolvido constitui uma tabela de metadados.

5.4.1 Aplicação da Geocodificação

Para incluir as informações das variáveis finais selecionadas e o indicador desenvolvido na base de dados da Mapidea foi necessário proceder à geocodificação das informações correspondentes às áreas estatísticas SA1 e SA2, através de códigos estatísticos únicos fornecidos pelos censos da Austrália que identificam cada uma das áreas estatísticas.

Esses códigos foram elaborados pelos censos tendo por base um esquema hierárquico. De acordo com esse esquema, no caso das áreas estatísticas SA1, os códigos são compostos por sete dígitos, em que nos dígitos é identificado primeiro, por ordem alfabética, o estado ou território a que pertence a área estatística SA1, depois são indicadas as áreas estatísticas SA2 em que se inserem as áreas estatísticas SA1 e por fim as áreas estatísticas SA1 (quadro 5.5).

Quadro 5.5 - Organização dos dígitos que compõem os códigos dos censos para as áreas estatísticas SA1 (Adaptado dos censos da Austrália, 2016).

Estado/Território	SA2	SA1
5	1041	18

Para as áreas estatísticas SA2, os códigos são compostos por cinco dígitos, em que nos dígitos é identificado, por ordem alfabética, o estado ou território a que pertence a área estatística SA2 e as áreas estatísticas SA2 (quadro 5.6).

Quadro 5.6 - Organização dos dígitos que compõem os códigos dos censos para as áreas estatísticas SA2 (Adaptado dos censos da Austrália, 2016).

Estado/Território	SA2
5	1041

Com esta hierarquização dos códigos, definida pelos censos, é garantida a unicidade dos códigos atribuídos para cada área estatística, o que é essencial para o processo de geocodificação das informações das variáveis/indicadores nos dados espaciais que representam as áreas estatísticas SA1 e SA2.

Estes códigos dos censos vêm inseridos, tanto na base de dados com as variáveis recolhidas, como nos dados espaciais que representam as áreas estatísticas SA1 e SA2, o que permite realizar a geocodificação logo após se inserir os dados espaciais.

Contudo, para geocodificar as informações na base de dados da Mapidea, foi ainda necessário atribuir um código único que substituísse o nome de cada uma das 387 variáveis dos censos consideradas relevantes para o marketing, o indicador desenvolvido e o nome do campo com os geocódigos das áreas estatísticas, devido a ser preciso codificar toda a informação quando inserida na base de dados no âmbito da geocodificação.

Esse código único atribuído, que substitui o nome de cada variável dos censos e o indicador desenvolvido, varia entre V001 (primeira variável/indicador da tabela) e V388

(última variável/indicador da tabela), para ser possível identificar posteriormente, de forma mais fácil, cada uma das variáveis/indicadores codificados(as) na base de dados. Foram codificadas mais sete variáveis/indicadores face as variáveis que foram inicialmente obtidas, devido a se ter acrescentado as variáveis criadas com a agregação de variáveis (efetuada durante o processo de seleção dos *clusters* de variáveis mais relevantes) e o novo indicador desenvolvido. Para substituir o nome do campo com os geocódigos das áreas estatísticas por um código, atribuiu-se o código S1NUM como nome dos geocódigos das SA1 e o código S2NUM como nome dos geocódigos das SA2.

Para incluir os dados selecionados do *OpenStreetMap* e os dados do *OpenCellID* na base de dados da Mapidea procedeu-se então à georreferenciação dos mesmos através das coordenadas da latitude e da longitude de cada lugar representado nos dados, que são fornecidas pelos dados. O nome das variáveis, que compõem estes dados, foi também codificado, de forma a poder-se identificar mais facilmente as variáveis (V390 a V399 - anexo 3).

O processo de codificação de toda a informação inserida teve por base uma formatação *Unicode*. Esta formatação constitui um padrão internacional que apresenta várias formas para codificar um texto de um documento digital, em que cada uma dessas formas de codificação atribui um código único para cada carácter (Hill, 2006). A codificação padrão de texto *Unicode* é o UTF-8, por isso foi utilizado na codificação de toda a informação (Hill, 2006).

O WebSIG da Mapidea, ao qual a base de dados da Mapidea está associada, permitiu carregar as *layers* (camadas) das áreas estatísticas na base de dados e realizar as operações de geocodificação e de georreferenciação das coordenadas que possibilitaram o carregamento dos dados na base de dados.

A inserção dos dados na base de dados da Mapidea foi então importante pois permitiu que os dados carregados pudessem ser visualizados, analisados e adaptados de forma intuitiva no WebSIG, devido às funções que este disponibiliza.

5.5 WebSIG

O aparecimento e o desenvolvimento das tecnologias associadas à internet e a necessidade de novas formas de análise de dados geográficos levaram ao desenvolvimento dos WebSIGs (Li et al., 2011).

Um WebSIG consiste numa página de um navegador da internet ou de outro tipo de aplicação cliente que disponibiliza dados georreferenciados, funções de SIG e/ou outros serviços, por forma a permitir que diversos dados geográficos de determinados temas sejam partilhados, visualizados, utilizados, criados, adaptados e analisados sem ser

preciso um *software* SIG (Peterson, 2008; Kulawiak et al., 2010; Li et al., 2011; Yang et al., 2011; Zavala-Romero et al., 2014).

Os dados georreferenciados, provenientes de bases de dados e/ou de serviços da internet, são ou podem ser representados num WebSIG através de mapas de base que os permitem organizar em camadas de dados e evidenciar o contexto geográfico dos dados (Peterson, 2008; Kulawiak et al., 2010; Fu, 2015). Os mapas de base são inseridos num WebSIG através de serviços de mapas da internet (Fu, 2015).

Para se adaptarem às necessidades dos diferentes grupos de utilizadores, os mapas do WebSIG normalmente oferecem a possibilidade de escolher as camadas de dados que se quer ver e como se quer ver, consultar as informações associadas aos dados e escolher os formatos de informação que se quer visualizar (Kulawiak et al., 2010; Li et al., 2011; Zavala-Romero et al., 2014; Fu, 2015). A esses mapas estão normalmente associadas funções SIG (como a geocodificação, a georreferenciação, a identificação de uma localização, as operações de cálculo, a seleção de dados, o *buffer*, entre outras) que permitem analisar e modificar os dados a apresentar no WebSIG, de acordo com as preferências do utilizador (Peterson, 2008; Kulawiak et al., 2010; Li et al., 2011; Fu, 2015). Assim, um WebSIG permite representar os dados geográficos de forma interativa e funcional (Li et al., 2011). Alguns WebSIGs fornecem plataformas que permitem aos utilizadores descarregar os dados disponibilizados ou obtidos através do WebSIG, o que faz aumentar a funcionalidade do WebSIG (Yang et al., 2011).

O WebSIG e os respetivos dados e funções podem ser disponibilizados para as pessoas em geral (quando público) ou apenas para as empresas, instituições ou a um número limitado de utilizadores (Peterson, 2008).

Através de um WebSIG é então possível:

- Partilhar facilmente dados geográficos para pessoas e organizações de todo o mundo;
- Que um elevado número de pessoas utilize determinados dados geográficos georreferenciados;
- Reduzir os custos associados à partilha de dados;
- Tornar mais fácil a utilização de funções SIG e dos dados geográficos, dado um WebSIG apresentar os mesmos numa interface gráfica simples e intuitiva;
- Apresentar funções que adaptam a apresentação dos dados e que analisam os dados consoante as necessidades dos utilizadores;
- Facilitar a criação de aplicações para as mais diversas áreas (Kulawiak et al., 2010; Li et al., 2011; Yang et al., 2011; Fu, 2015).

Para que um WebSIG possa ser partilhado para inúmeros utilizadores de diferentes locais, este deve ter por base uma estrutura que se baseie na relação servidores-clientes (figura 5.5), em que as máquinas clientes (aplicações ou páginas da internet), através da

ligação a máquinas servidor, requerem os serviços disponibilizados pelos servidores, que são essenciais para o funcionamento de um WebSIG (Peterson, 2008; Kulawiak et al., 2010; Li et al., 2011). Esta estrutura permite ainda que o WebSIG tenha um ambiente aberto e adaptável, no qual diversas funções e opções podem ser disponibilizadas (Peterson, 2008).



Figura 5.5 - Relação servidor-clientes estabelecida pelo WebSIG (Adaptado de Fu, 2015).

Assim, um WebSIG é normalmente composto por um ou vários servidores de dados geográficos, pelos servidores de mapas, por um servidor de configuração da apresentação do WebSIG e pelo navegador da internet (Yang et al., 2011).

Para que a ligação dos vários servidores às máquinas cliente permita uma correta visualização do WebSIG, este, na sua forma mais simples, tem de ser organizado pelas seguintes camadas (figura 5.6):

- A camada dos clientes do WebSIG, que serve de apresentação do WebSIG. Esta contém normalmente os mapas, as funções, os dados geográficos e os resultados das análises de dados previamente efetuadas a disponibilizar no WebSIG;
- A camada dos servidores do WebSIG, que contém os elementos que servem de base ao WebSIG. É onde são definidas as funções, os conteúdos e a capacidade que o WebSIG deve conter;

- A camada dos servidores da base de dados SIG, que contém os dados necessários para suportar as camadas dos servidores e dos clientes do WebSIG (Fu, 2015).

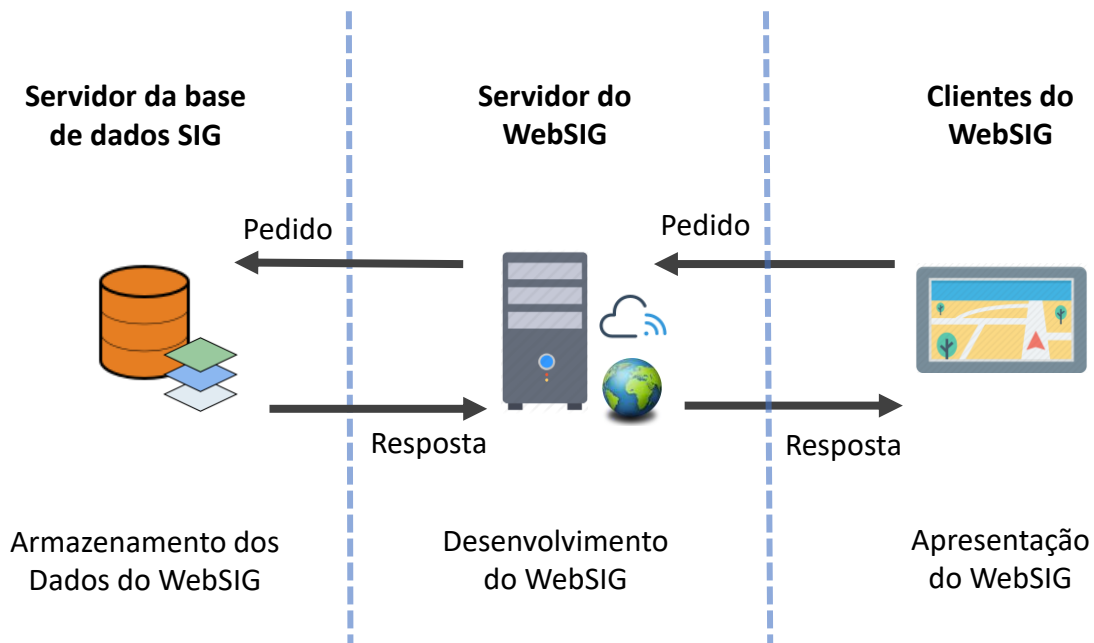


Figura 5.6 - Potencial arquitetura de um WebSIG simples (Adaptado de Fu, 2015).

Tendo em conta a organização por camadas de um WebSIG e para que este seja desenvolvido de forma correta, em primeiro lugar, na camada da base de dados, deve-se inserir os dados espaciais a apresentar no WebSIG, para ser possível a sua visualização na camada dos servidores do WebSIG (Kulawiak et al., 2010; Zavala-Romero et al., 2014; Fu, 2015). Os dados a inserir podem ser de formato CSV, *GeoTIFF* ou *shapefile* quando são provenientes de bases de dados associadas ao WebSIG e de formato KML quando são provenientes da internet (Yang et al., 2011; Fu, 2015).

Em seguida, na camada dos servidores do WebSIG, primeiro devem ser configurados, no servidor de mapas, os dados provenientes da camada da base de dados ou de servidores da internet para se poder definir quais os dados a apresentar e como deve ser feita a sua apresentação (formato, especificações, cor e simbologia) e depois na *framework* deve ser configurada a apresentação do WebSIG, através de linguagem de código ou de um *software* interoperável (caso do *Mapbender3*) que utiliza uma ou várias API da internet que permitem aceder aos dados dos diferentes servidores de mapas, criar ou escolher (quando disponíveis) as funções de SIG a apresentar, organizar as camadas de dados a apresentar nos mapas, a execução simples das funções do WebSIG pelos utilizadores, apresentar os resultados das funções executadas pelos utilizadores e definir as características da aplicação ou página de apresentação do WebSIG (posição

dos elementos representados, cor, formato, entre outras) (Peterson, 2008; Tsou, 2008; Kulawiak et al., 2010; Zavala-Romero et al., 2014; Fu, 2015).

No servidor de mapas do WebSIG podem ser configurados servidores de dados da internet, de dados vetoriais (para incluir os dados vetoriais da camada da base de dados no WebSIG), de dados matriciais (para incluir os dados raster da camada da base de dados), entre outros (Zavala-Romero et al., 2014; Fu, 2015). A esses servidores deve ser associado um endereço URL que permita ligar os servidores à *framework* (Zavala-Romero et al., 2014). Os servidores de mapas mais utilizados para a disponibilização de dados geográficos são o *MapServer* e o *GeoServer* (Peterson, 2008; Kulawiak et al., 2010; Zavala-Romero et al., 2014).

Para aceder aos dados espaciais dos servidores de mapas do WebSIG a *framework* pode também ligar-se aos servidores de mapas através de formatos do *Open Geospatial Consortium* (OGC), como o WMS (*Web Map Service*) que representa mapas com uma ou mais camadas de dados geográficos para uma determinada área, o WFS (*Web Feature Service*) que representa dados geográficos vetoriais e o WCS (*Web Coverage Service*) que representa dados multidimensionais que variam no tempo e no espaço (Peterson, 2008; Tsou, 2008; Kulawiak et al., 2010; Yang et al., 2011; Zavala-Romero et al., 2014). A *framework* pode ainda inserir dados dos servidores de mapas nos formatos XML e GML que facilitam a comunicação e organização da informação, dado poderem especificar a ligação ao servidor de mapas, o nome, atributos e a hierarquia dos dados (Tsou, 2008; Yang et al., 2011; Zavala-Romero et al., 2014; Davis & Alves, 2017).

Para se poder incluir no WebSIG mapas de base com informações georreferenciadas de qualquer lugar, a *framework* pode ser associada a API do *Google Maps*, do *Bing Maps* ou a uma outra API de mapas (Kulawiak et al., 2010; Li et al., 2011; Yang et al., 2011; Fu, 2015).

Por fim, na camada cliente, é efetuada a apresentação do WebSIG. A apresentação deste é programada no âmbito da *framework*, através de documentos HTML que programam os dados e as funções a apresentar no WebSIG e de documentos CSS que formatam o estilo dos elementos representados (Tsou, 2008; Fu, 2015). Esses documentos, que estão na base da programação do WebSIG, utilizam normalmente uma linguagem de programação *JavaScript* porque permite que o WebSIG possa ser visualizado em computadores, tablets e smartphones quando disponibilizado no navegador da internet (Tsou, 2008; Kulawiak et al., 2010; Zavala-Romero et al., 2014; Fu, 2015).

A disponibilização do WebSIG no navegador da internet pode ser feita através do HTTP, que faz a transferência dos documentos HTML e CSS da *framework* para o navegador da internet (figura 5.7) com base num endereço URL que permite ao navegador localizar esses documentos (Yang et al., 2011; Fu, 2015; Davis & Alves, 2017).

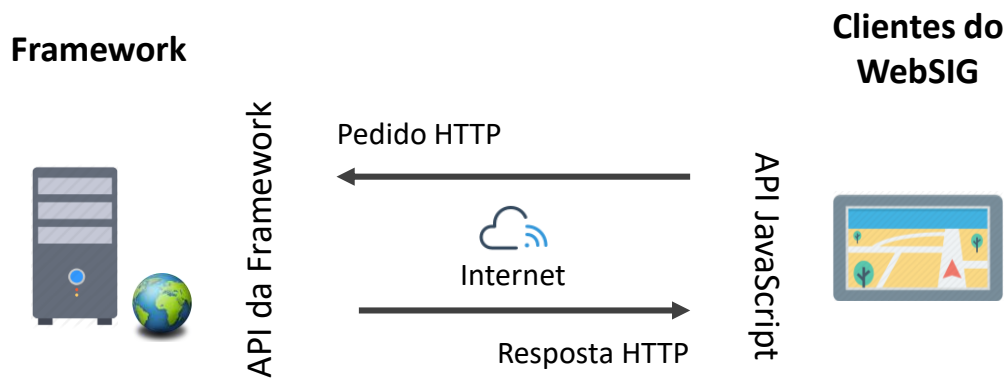


Figura 5.7 - Comunicação dos documentos entre a *framework* e os clientes do WebSIG
(Adaptado de Fu, 2015).

5.5.1 WebSIG da Mapidea

O WebSIG da Mapidea Location Analytics constitui um WebSIG desenvolvido totalmente pela Mapidea com o objetivo de simplificar e melhorar as decisões das empresas através da análise da localização. Este WebSIG destaca-se pela sua interface gráfica intuitiva e de fácil utilização que permite incluir, visualizar e analisar dados geográficos de forma acessível e rápida, sem ser preciso aplicar linguagem de programação ou utilizar um *software* SIG (figura 5.8). O WebSIG desenvolvido pode ser utilizado por qualquer pessoa de uma empresa ou organização, através da internet, desde que possua uma conta de utilizador. A simplicidade do WebSIG permite que os seus utilizadores possam desenvolver e criar dados geográficos sem ser preciso serem especialistas em SIG ou aplicar procedimentos complexos.

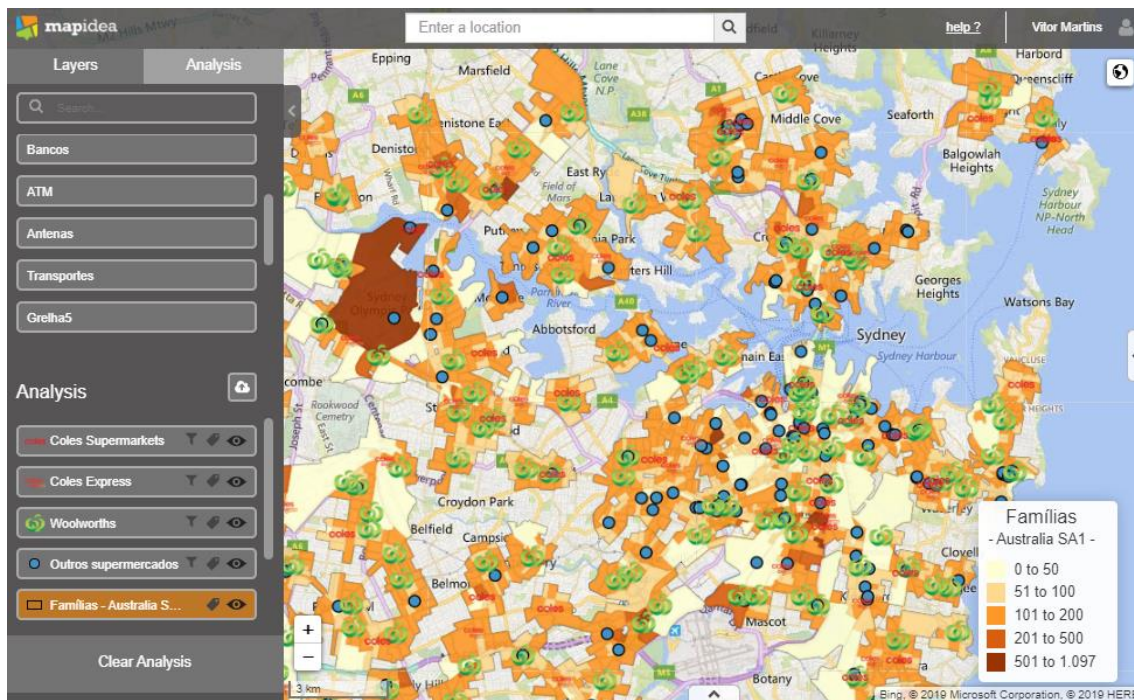


Figura 5.8 - Interface gráfica do WebSIG da Mapidea.

A visualização e análise dos dados geográficos no WebSIG da Mapidea pode ser feita através dos mapas que os utilizadores podem produzir, com base nos dados geográficos inseridos (que podem ser de formato *GeoJSON*, *MapInfo*, *ESRI Shapefile*, *CSV*, *Excel* ou *Google Earth - KML*) e nas funções disponibilizadas. Na produção do mapa os utilizadores podem escolher visualizar o padrão de distribuição dos dados de uma variável ou podem escolher comparar duas variáveis do mesmo ou de diferentes *clusters* de variáveis. Através dos mapas produzidos, os utilizadores podem ainda visualizar dados de uma unidade territorial, através de quadros de dados e de gráficos que os utilizadores podem configurar livremente.

Algumas das funções disponibilizadas por este WebSIG (figura 5.9) são a possibilidade de realizar a filtragem de dados (por atributos de um conjunto de dados ou espacial através da relação com outras camadas de dados), de criar novos dados a partir de operações de cálculo (com os atributos dos dados ou através das relações espaciais entre camadas) e de calcular a proximidade face a um elemento representado numa camada de dados (*buffer*).

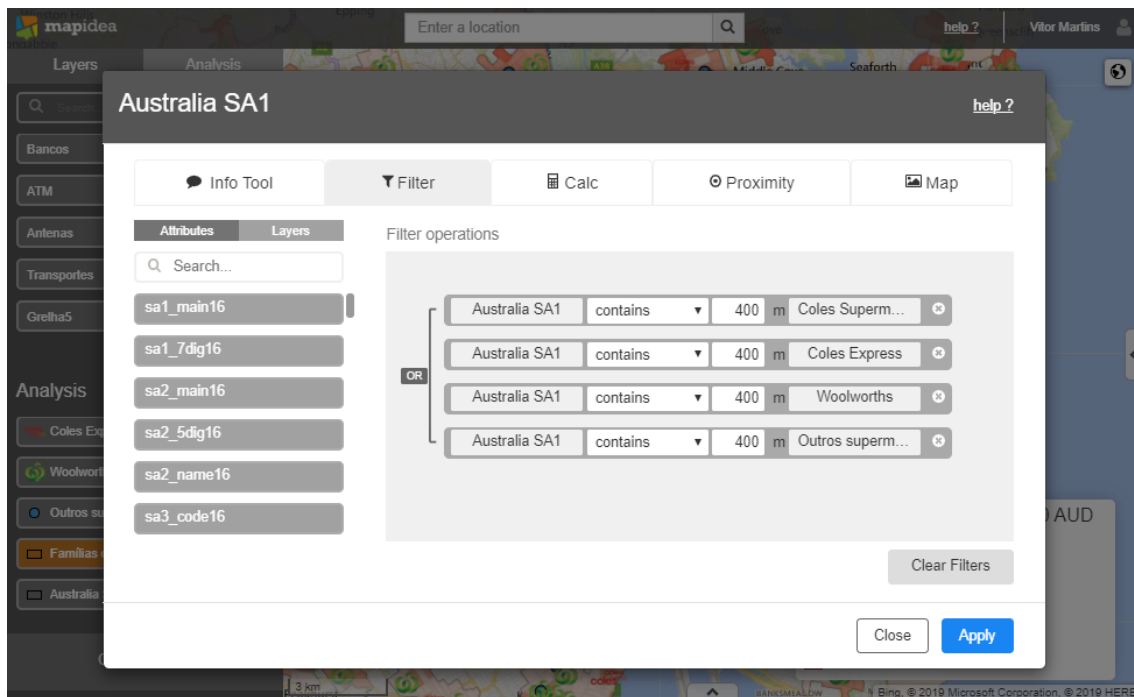


Figura 5.9 - Demonstração das funções disponibilizadas pelo WebSIG da Mapidea.

Os novos dados desenvolvidos pelos utilizadores, através da aplicação das funções do WebSIG, podem ser partilhados com outros utilizadores, guardados para poderem ser visualizados posteriormente e descarregados em formato *Excel*, *CSV*, *Mapinfo*, *Shapefile*, *Google Earth* (KML) e *GeoJSON* (figura 5.10), o que permite aumentar a utilidade do WebSIG e das suas funções.

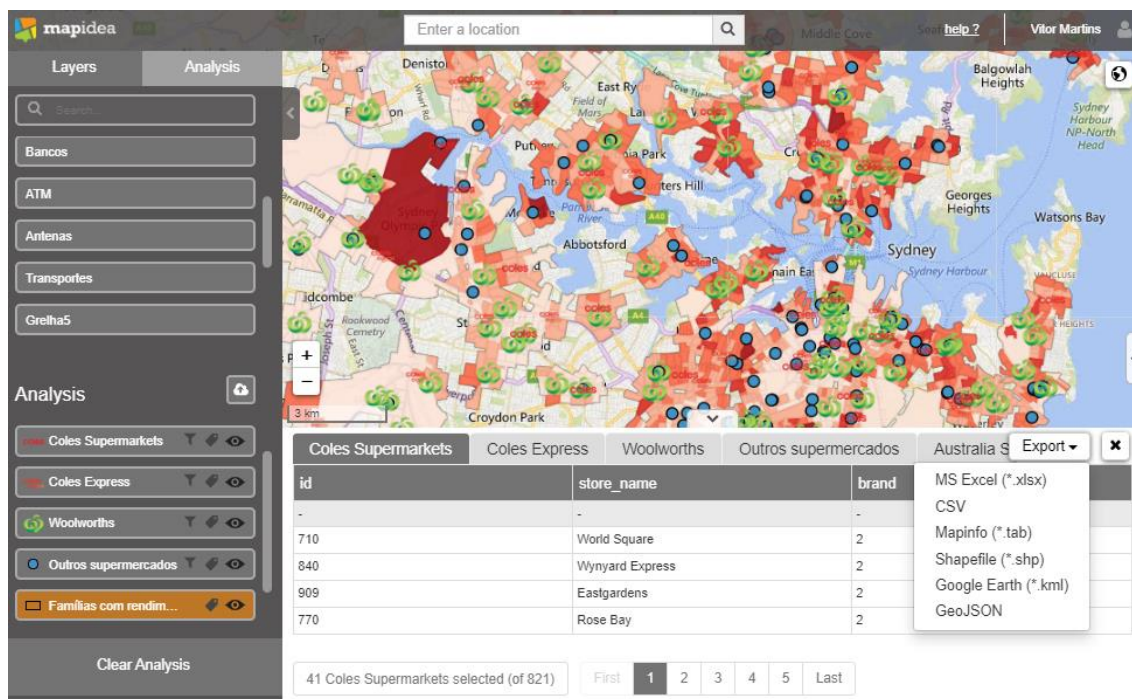


Figura 5.10 - Apresentação da opção para exportar dados do WebSIG.

5.5.1.1 Utilização do WebSIG da Mapidea

Após o carregamento dos dados dos censos, do *OpenStreetMap* e do *OpenCellID* na base de dados da Mapidea, procedeu-se à criação de duas grelhas com estes dados (fora do WebSIG), que foram posteriormente inseridas no WebSIG da Mapidea para ser possível realizar no mesmo análises mais adequadas da área de mercado da Austrália.

Uma dessas grelhas desenvolvidas (figura 5.11) dividiu a Austrália em polígonos quadrados de 100 km^2 , para ser mais fácil analisar as informações sociodemográficas e económicas das áreas rurais da Austrália, que ao serem caracterizadas pela sua baixa densidade populacional, torna-se pouco útil utilizar unidades territoriais de menor dimensão.

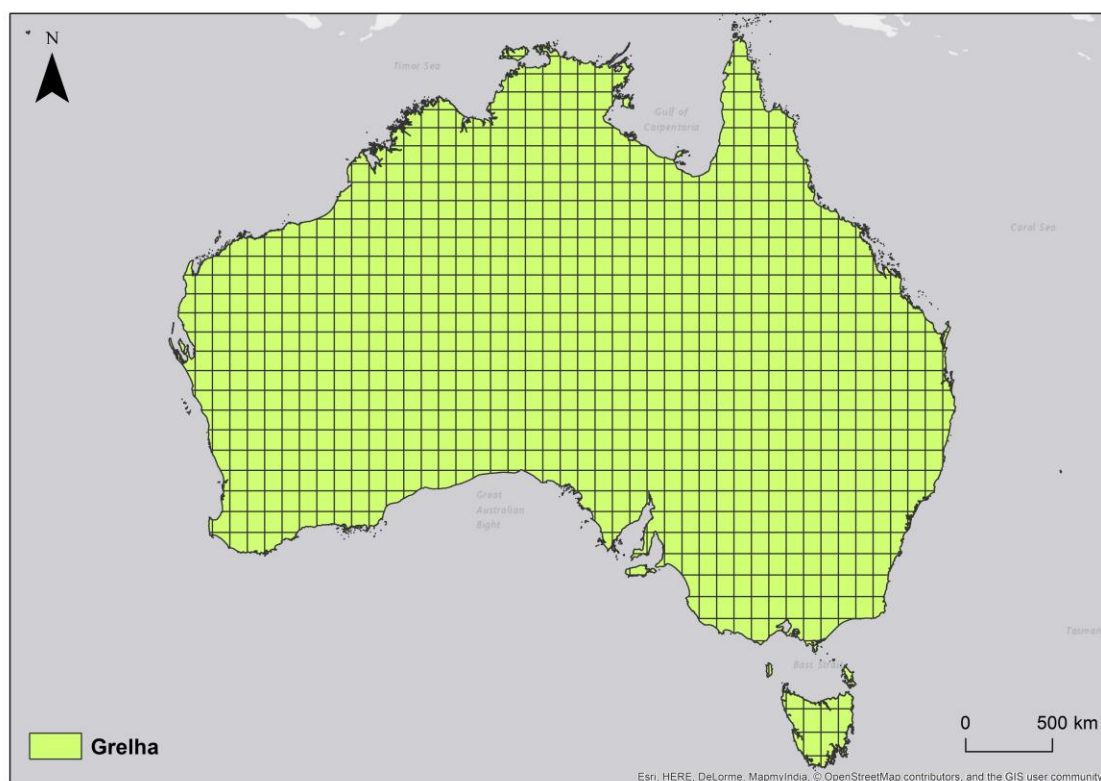


Figura 5.11 - Grelha com polígonos quadrados de 100 km^2 para o conjunto da Austrália.

A outra grelha, que foi apenas elaborada para as principais áreas urbanas da Austrália (figura 5.12), dividiu estas áreas em polígonos quadrados de 250 m^2 , para ser possível analisar em pormenor as informações sociodemográficas e económicas das áreas urbanas mais habitadas, e assim se poder determinar de forma adequada as áreas de mercado, que tendem a ter uma dimensão espacial bastante reduzida nestas áreas.

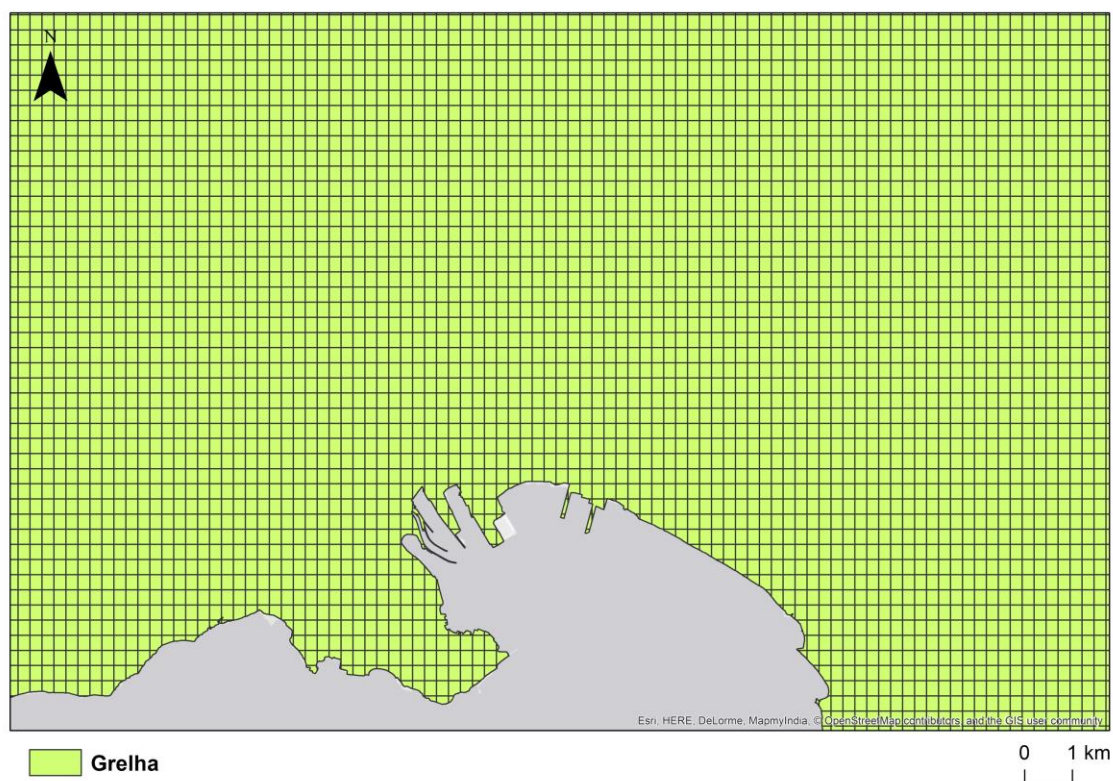


Figura 5.12 - Grelha com polígonos quadrados de 250 m^2 para o centro da cidade de Melbourne.

Com os dados todos inseridos, foi possível calcular no WebSIG da Mapidea o total de paragens de transportes, bancos, ATM, postos de combustível, centros comerciais, supermercados (Coles, Coles Express, Woolworths, e outros supermercados) e antenas por áreas estatísticas SA1 e SA2. O cálculo destes dados por cada área estatística SA1 e SA2 é essencial para se poder relacionar os mesmos com as informações dos censos (já representadas por áreas estatísticas), procedimento que será efetuado pelos modelos de regressão e fatorial aplicados no âmbito deste modelo espacial demonstrado.

Estes dados inseridos no WebSIG, após realizado o trabalho, ficaram disponibilizados na área de administração do WebSIG da Mapidea para poderem vir a ser utilizados em futuras análises de mercado efetuadas pela empresa.

6. Resultados

6.1 Análise dos dados mais importantes para a análise da correlação

Para se compreender melhor a importância dos dados que serão essenciais para a análise da correlação que será aplicada no modelo de regressão e na análise fatorial, decidiu-se representar a população residente e o índice de diversidade para as áreas estatísticas do conjunto da Austrália e das três principais cidades (Sydney, Melbourne e Brisbane).

6.1.1 População residente

Representar a população residente nas diversas áreas estatísticas da Austrália (figura 6.1) pode ser importante para verificar-se de que forma a população se encontra distribuída pelo território do país, ou seja, em que zonas a população das áreas estatísticas tende a ser maior ou menor.

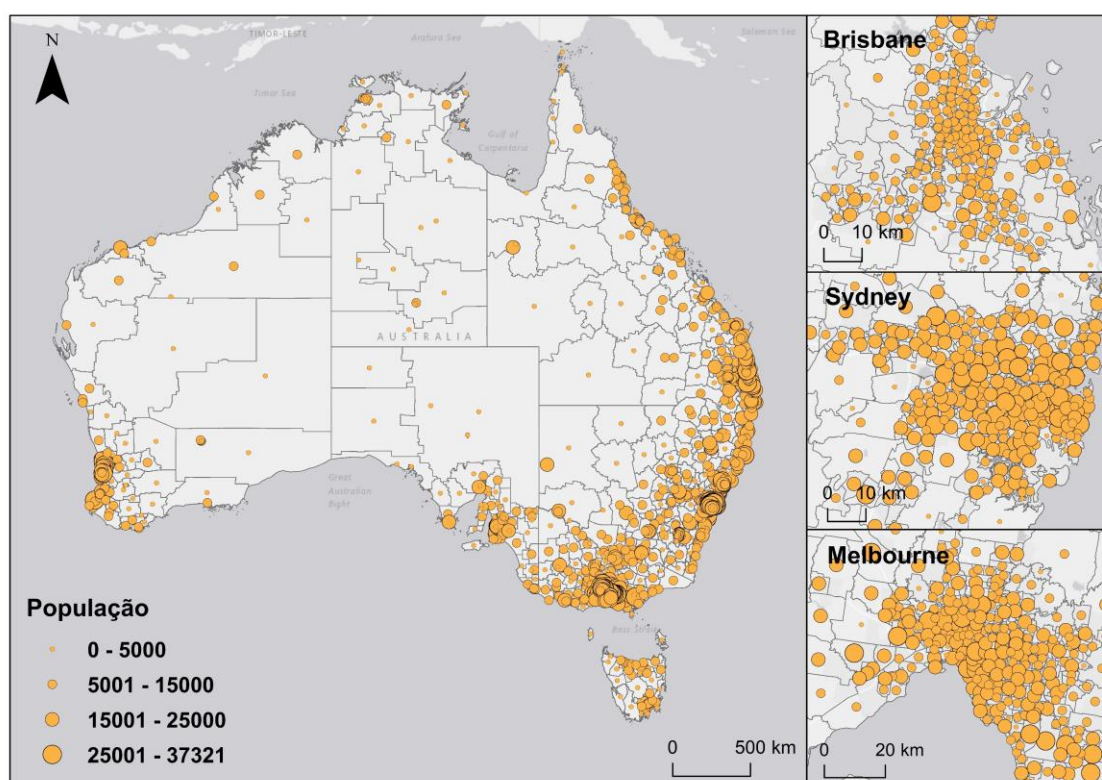


Figura 6.1 - População residente nas áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.

A representação da variação espacial da população permite constatar que a maioria das áreas estatísticas onde a população é maior e a sua dimensão espacial é normalmente mais reduzida estão situadas no litoral da Austrália, sobretudo no centro das principais cidades (à exceção da cidade de Brisbane em que a população é maior apenas em determinadas áreas mais periféricas por as áreas estatísticas do centro da cidade terem uma menor dimensão). Nas restantes zonas do país as áreas estatísticas tendem a ter menos população e uma maior dimensão espacial.

Introduzir a população residente na análise da correlação é importante porque permite comparar a população de cada área estatística (procura potencial) com o valor que as outras variáveis/indicadores assumem.

6.1.2 Índice de diversidade

Com o índice de diversidade (figura 6.2) pode-se demonstrar em que zonas as áreas estatísticas têm uma maior ou menor diversidade, ou seja, em que zonas existe um maior ou menor número de pessoas indígenas ou oriundas de outros países.

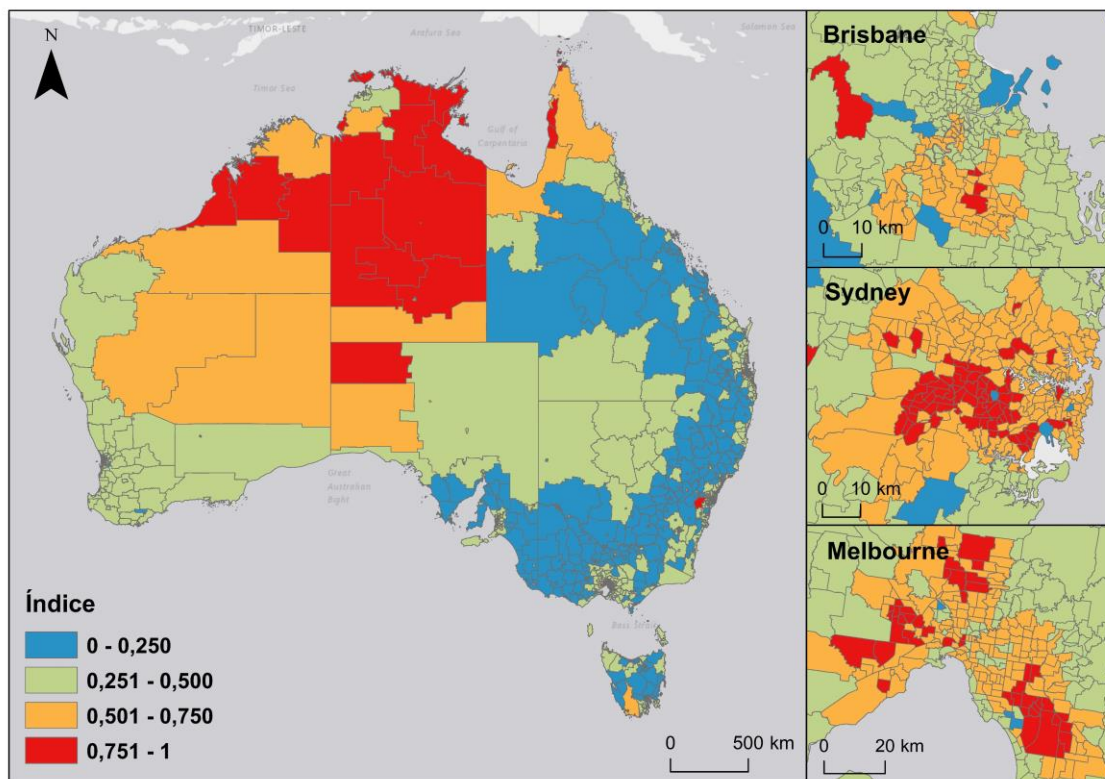


Figura 6.2 - Índice de diversidade para as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.

Através da representação do índice de diversidade pode-se verificar que as áreas estatísticas com uma população mais diversa localizam-se essencialmente na periferia das principais cidades (na sua maioria constituem pessoas que nasceram fora do país ou com um ou ambos os pais a nascerem fora do país) e no interior da Austrália (que constituem na sua maioria população indígena). Nas restantes zonas do país as áreas estatísticas tendem a ter uma população menos diversa.

Dado os hábitos culturais de cada pessoa influenciar o tipo de produtos que consomem e a forma como o fazem, torna-se importante incluir o índice de diversidade na análise da correlação que pretende desenvolver indicadores de valor acrescentado para a análise da área de mercado.

6.2 Regressão

As decisões de gestão e do marketing têm normalmente em conta a forma como duas ou mais variáveis/indicadores com informações descritivas da sociedade se relacionam, para que essas decisões considerem os fatores que poderão tornar mais adequada uma dada localização para um local de negócio ou que podem influenciar a compra/utilização de um produto (Anderson et al., 2007; Bingham & Fry, 2010). Uma técnica estatística que permite verificar como se relacionam as diferentes variáveis/indicadores é a análise de regressão (Anderson et al., 2007; Bingham & Fry, 2010).

Uma análise de regressão consiste num procedimento estatístico que tenta prever, através de uma equação, a relação que existe entre as variáveis/indicadores, ou seja, de que forma uma variável dependente y se relaciona com uma ou mais variáveis/indicadores independentes x , por forma a facilitar a identificação dos fatores que levam a que elas se relacionem (Anderson et al., 2007; Aljandali, 2016).

Analisar a regressão é útil, porque se conhecermos uma dada variável/indicador x , e se descobirmos que essa variável/indicador é bastante correlacionada com a variável y que queremos conhecer, podemos descobrir novas informações sobre a variável y com base na variável/indicador x que conhecemos (Bingham & Fry, 2010). Com essa análise é então possível explicar o porquê da variação dos valores de uma variável (Goodwin, 2018). Desta forma, a variável dependente y é a que queremos prever, enquanto que a(s) variável(is)/indicador(es) independente(s) x é(são) a(s)/o(s) que permite(m) prever o valor da variável dependente y (Anderson et al., 2007; Bingham & Fry, 2010; Goodwin, 2018).

Tendo em conta os vários modelos de análise de regressão, o modelo de regressão linear é o mais aplicado (Aljandali, 2016). Este consiste num método que demonstra como os valores médios de uma variável/indicador variam em conjuntos de valores definidos sob forma linear (Gelman & Hill, 2006). Neste método, a relação entre as

variáveis/indicadores y e x é maior quanto maior for a aproximação da posição dos valores a uma linha reta (Aljandali, 2016).

A regressão linear divide-se em dois tipos de modelo: regressão linear simples e regressão linear múltipla (Anderson et al., 2007). A regressão linear simples é quando se relaciona uma variável dependente y com uma variável/indicador independente x (Anderson et al., 2007; Keith, 2019). Já a regressão linear múltipla relaciona uma variável dependente y com duas ou mais variáveis/indicadores independentes x (Anderson et al., 2007). Neste relatório será aplicada a última, pois permite comparar uma variável com várias outras variáveis/indicadores, o que facilita a identificação das relações que existem entre as variáveis/indicadores, para posteriormente ser possível identificar os fatores que provocam essa relação.

Assim, se considerarmos que:

- y é a variável dependente;
- β_0 é o intercepto;
- $\beta_1, \beta_2, \dots, \beta_n$ representam o declive;
- ε representa o erro;
- x_1, x_2, \dots, x_n representam as variáveis/indicadores independentes;

a regressão linear múltipla pode ser demonstrada pela seguinte equação (Anderson et al., 2007; Larose & Larose, 2019):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (\text{Eq. 6.1})$$

O intercepto é o valor previsto para a variável dependente quando o valor equivalente nas variáveis/indicadores independentes é zero e o erro indica as vezes em que y não pode ser explicado pela relação entre as variáveis/indicadores x e y (Anderson et al., 2007; Keith, 2019).

Caso considerarmos que não existe erro, para que os valores médio ou esperados de y tenham em conta a relação entre a variável dependente e as variáveis/indicadores independentes, é preciso calcular a seguinte equação (Anderson et al., 2007):

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (\text{Eq. 6.2})$$

Com a substituição dos valores de $\beta_0, \beta_1, \dots, \beta_n$ por valores estimados b_0, b_1, \dots, b_n a partir de uma amostra (devido aos primeiros valores não serem normalmente conhecidos),

devemos considerar a seguinte equação de regressão estimada múltipla (Anderson et al., 2007; Larose & Larose, 2019):

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (\text{Eq. 6.3})$$

O valor b da equação 6.3 representa o coeficiente de regressão, que corresponde ao valor multiplicado à variável/indicador independente que permite que o valor dessa variável/indicador seja equivalente ao da variável dependente (Keith, 2019). Com este valor é então possível comparar os dados das diferentes variáveis/indicadores (Gelman & Hill, 2006).

6.2.1 Aplicação da regressão

Para se poder desenvolver novos indicadores de valor acrescentado para o contexto da área de mercado da Austrália foi preciso realizar inicialmente uma regressão que permita determinar as relações que existem entre as variáveis/indicadores, de forma a se poderem identificar padrões de relação que levem à definição dos fatores que explicam a variação dos dados das variáveis/indicadores, e com esses fatores se poderem criar novos indicadores.

A regressão aplicada teve por base os valores das variáveis/indicadores relativos às unidades estatísticas SA1, porque estas contêm um maior número de observações, o que torna a análise da regressão mais precisa.

Antes de efetuar a regressão, é necessário garantir que cada uma das diferentes variáveis/indicadores assumem a mesma importância, para assim poderem ser relacionados(as) na regressão (Larose & Larose, 2019). Por isso, foi preciso normalizar os valores das variáveis/indicadores selecionados (anexo 3), para que tenham uma variação de unidades de valor equivalente. Essa normalização foi feita com base nos *z-scores*, em que se determinou uma variação de unidades equivalentes para as diferentes variáveis/indicadores através do cálculo da diferença proporcional entre os valores das observações de uma variável/indicador e a média do conjunto de valores dessa variável/indicador (Gelman & Hill, 2006; Anderson et al., 2007). Assim, se considerarmos que para cada valor de observação x_i de uma variável/indicador existe outro valor designado *z-score* (valor padronizado), a normalização dos valores das variáveis/indicadores pode ser feita da seguinte forma (Gelman & Hill, 2006; Anderson et al., 2007):

$$z_i = \frac{x_i - \bar{x}}{s} \quad (\text{Eq. 6.4})$$

em que:

- z_i é o *z-score* para x_i ;
- \bar{x} é a média da amostra;
- s é o desvio padrão da amostra.

Os *z-scores* podem ser interpretados como a diferença em unidades de desvio padrão entre x_i e a média, equivalente para todas as variáveis/indicadores (Gelman & Hill, 2006; Anderson et al., 2007). Um valor de *z-score* maior que zero significa que o valor de x_i é maior que a média de valores e um *z-score* menor que zero significa que x_i assume um valor mais baixo que a média (Anderson et al., 2007).

Depois de normalizados os valores das variáveis/indicadores, foi possível efetuar no SPSS uma regressão linear múltipla com as variáveis/indicadores para identificar as variáveis/indicadores colineares, ou seja, as variáveis/indicadores que têm elevada correlação (relação) entre si, para poderem ser retirados(as), dado perturbarem o resultado da análise fatorial a ser realizada posteriormente. Na regressão realizada foi considerado como variável dependente o número de pessoas, as restantes variáveis/indicadores (anexo 3) como variáveis independentes e foi utilizado o método enter para ter em conta todas as variáveis/indicadores independentes em análise.

O cálculo da correlação na regressão foi feito através do coeficiente de Pearson, um coeficiente que varia entre -1 e 1 e que calcula a correlação de cada variável/indicador independente face à variável dependente y (Wackerly et al., 2008).

Tendo em conta que uma correlação de Pearson acima de 0.8 e inferior a -0.8 é considerada uma correlação elevada, foram retiradas as variáveis referidas no quadro 6.1 por serem consideradas variáveis colineares, dado terem elevadas correlações entre si e/ou com outras variáveis/indicadores inseridos(as) no modelo (Schober et al., 2018).

Quadro 6.1 - Variáveis colineares, excluídas através do modelo de regressão.

Variáveis Colineares	
<ul style="list-style-type: none"> • Homens V001 • Mulheres V002 • Mediana do rendimento semanal por pessoa V027 • Pessoas em idade ativa V039 • Estudantes V108 • Famílias que têm filhos com menos de 15 anos V215 • Famílias V219 • Crianças V223 • Habitações ocupadas com veículos a motor V239 	<ul style="list-style-type: none"> • Habitações ocupadas V240 • Residentes em vivendas ou moradias V251 • Famílias com prestação mensal de empréstimo V284 • Pessoas empregadas V311 • Pessoas com bacharelato V324 • Pessoas com diploma avançado V325 • Pessoas empregadas que se deslocam para o trabalho apenas de automóvel ou mota V387

Com a remoção das variáveis colineares obteve-se 37 variáveis/indicadores não colineares. Para medir a proporção de variação da variável dependente (número de pessoas) que é explicada pelas restantes variáveis/indicadores independentes (variáveis/indicadores não colineares obtidos(as)) pode-se calcular o coeficiente de determinação, que corresponde ao valor do coeficiente de Pearson elevado ao quadrado (Anderson et al., 2007; Aljandali, 2016).

Os valores do coeficiente de determinação (r^2) variam entre 0 e 1 (ou entre 0 e 100%), em que um valor do coeficiente próximo ou igual a 1 (figura 6.3 - A) significa que existe um elevado grau de correlação (dispondo-se as observações sob a forma linear) e um valor próximo ou igual a 0 (figura 6.3 - B) significa que existe uma reduzida correlação entre as variáveis/indicadores (dispondo-se as observações de forma dispersa) (Aljandali, 2016; Goodwin, 2018). Através do valor do coeficiente de determinação é então possível verificar a percentagem de valores observados que são explicados pela equação da regressão (Anderson et al., 2007).

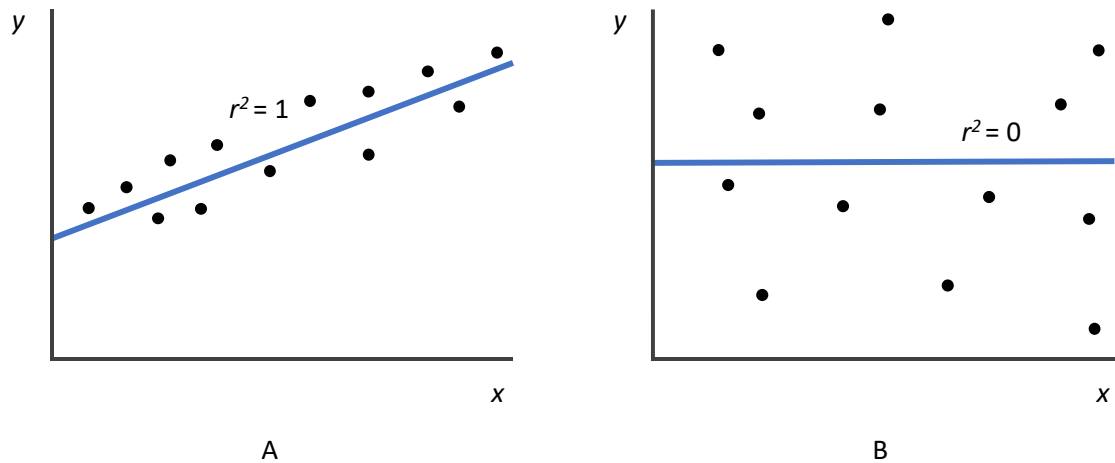


Figura 6.3 - Exemplo de distribuição de observações de forma linear (A) e de distribuição de observações de forma dispersa (B)

Se tivermos em conta que:

- STQ é a soma total dos quadrados;
- SQR é a soma dos quadrados da regressão (fornecidos pela variável/indicador independente);
- SQE é a soma dos quadrados do erro (diferença da variável dependente face a variável/indicador independente);
- \hat{y}_i é o valor estimado da variável y para a observação i ;
- y_i é o valor observado da variável y para a observação i ;
- \bar{y} representa a média dos valores observados de y ;

o cálculo do coeficiente de determinação para uma variável dependente y pode também ser feito da seguinte forma (Anderson et al., 2007):

$$r^2 = \frac{SQR}{STQ} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \quad (\text{Eq. 6.5})$$

em que:

$$STQ = SQR + SQE = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2 \quad (\text{Eq. 6.6})$$

Assim, a medição do coeficiente de determinação é importante para encontrar a equação que melhor se adequa a regressão analisada (Anderson et al., 2007). Na nova regressão efetuada, após retirar as variáveis colineares, obteve-se um coeficiente de determinação de 0,959.

Na análise das variáveis/indicadores independentes que melhor explicam uma variável dependente, o coeficiente de determinação tende a aumentar à medida que se adicionam novas variáveis/indicadores independentes não colineares ao modelo (Anderson et al., 2007; Keith, 2019). Para que o valor do coeficiente de determinação que explica a amplitude de valores da equação da regressão não seja influenciado, ou mais concretamente, superestimado pela possível inclusão de novas variáveis/indicadores independentes, deve-se ajustar o valor de r^2 da seguinte forma (Anderson et al., 2007; Keith, 2019):

$$r_a^2 = 1 - (1 - r^2) \frac{t - 1}{t - n - 1} \quad (\text{Eq. 6.7})$$

em que:

- r_a^2 corresponde ao coeficiente de determinação ajustado;
- t é o número total de observações;
- n é o número de variáveis/indicadores independentes.

Na nova regressão, o valor do coeficiente de determinação quando ajustado é de 0,959, o que significa que as variáveis/indicadores independentes não colineares inseridas explicam 95,9% da variação da população, independentemente da entrada ou não de novas variáveis/indicadores no modelo. Assim, este valor do coeficiente de determinação ajustado indica que as variáveis/indicadores independentes não colineares permitem explicar a variação da população quase toda, o que possibilita a identificação de características da população através da variação das variáveis/indicadores e dos padrões de relação que existem entre essas variáveis/indicadores (o que é essencial para que os novos indicadores desenvolvidos sejam de valor acrescentado).

Tendo em conta os resultados obtidos com o modelo de regressão, foi então possível realizar uma análise fatorial com as variáveis/indicadores não colineares, para se poder, com base na correlação que existe entre as variáveis/indicadores, identificar os fatores que explicam a variação dos dados das variáveis/indicadores e que permitam a criação de novos indicadores de valor acrescentado para o contexto da área de mercado da Austrália.

6.3 Análise Fatorial

Um dos procedimentos estatísticos mais utilizados, a análise fatorial, permite criar modelos concisos que, com base nos dados de um conjunto de variáveis/indicadores, possibilitam a identificação e caracterização dos fatores que são responsáveis pela variação do conjunto dos dados das variáveis/indicadores (Fabrigar & Wegener, 2012; Quaglino & Pagura, 2014; Brown, 2015).

Um fator na análise fatorial consiste numa variável não determinada que influencia mais uma dada realidade do que um conjunto de variáveis/indicadores explicativos(as) conhecidos(as) e que é responsável pela variação dos dados dessas variáveis/indicadores e pela correlação que existe entre as variáveis/indicadores (dado assumir-se que existem correlações entre as variáveis/indicadores explicativos(as)) (Fabrigar & Wegener, 2012; Quaglino & Pagura, 2014; Brown, 2015).

Através da análise fatorial, é então possível determinar um conjunto menor de variáveis (fatores) que permitem explicar uma dada realidade e fornecer informações acerca da mesma, face às variáveis/indicadores indicados(as) inicialmente no modelo como explicativos(as) (Fabrigar & Wegener, 2012; Brown, 2015). Para determinar os fatores que explicam uma dada realidade, é preciso no modelo analisar de que forma se correlacionam as diferentes variáveis/indicadores (para se poder agrupar as variáveis/indicadores conhecidos(as) consoante fatores de acordo com a maior ou menor correlação que possa existir entre eles(as)) e identificar quais as causas que produzem as correlações entre as variáveis/indicadores (Fabrigar & Wegener, 2012; Quaglino & Pagura, 2014; Brown, 2015).

A análise fatorial pode ser de tipo exploratória ou confirmatória (Fabrigar & Wegener, 2012; Brown, 2015; Loehlin & Beaujean, 2017).

A análise fatorial de tipo exploratória tem como principal foco os dados das variáveis/indicadores, pelo que não tem em conta à priori o número de fatores nem o padrão da relação entre os fatores e as variáveis/indicadores (o peso fatorial) (Fabrigar & Wegener, 2012; Brown, 2015; Loehlin & Beaujean, 2017). Este tipo de análise tem como objetivos identificar o menor número de fatores que explicam o conjunto dos dados e as correlações que existem entre as variáveis/indicadores, e quais as variáveis/indicadores que são mais explicados(as) por cada fator (através da dimensão do peso fatorial) (Fabrigar & Wegener, 2012; Brown, 2015; Loehlin & Beaujean, 2017).

Já na análise fatorial de tipo confirmatória, tem de ser indicado à priori o número de fatores que explica o conjunto dos dados, o padrão da importância das variáveis/indicadores e outras especificações que influenciam a variação e a independência dos fatores (Fabrigar & Wegener, 2012; Brown, 2015). Esta análise tem então como objetivo medir a correlação dos fatores previamente identificados com as variáveis/indicadores medidos(as) (Brown, 2015).

Por isso, neste relatório, foi aplicada, através do SPSS, a análise de tipo exploratória, por não se saber à partida o menor número de fatores que permitem explicar os dados das variáveis/indicadores e as correlações que existem entre as mesmas.

O peso fatorial constitui uma estimativa padronizada para prever a variação da variável/indicador que é explicada pelo fator, sendo interpretado como uma correlação face ao fator (Fabrigar & Wegener, 2012; Brown, 2015). Este, quando elevado ao quadrado, representa a proporção de variação da variável/indicador que o dado fator explica (Brown, 2015). Com o peso fatorial é então possível analisar a influência que cada fator exerce nas variáveis/indicadores em análise (Fabrigar & Wegener, 2012). Este pode também ser calculado para identificar a importância de cada fator para a explicação do conjunto das variáveis/indicadores do modelo (Loehlin & Beaujean, 2017).

Os dados de cada variável/indicador na análise fatorial são explicados pela variação do fator (também designada variação comum) e pela variação única (Fabrigar & Wegener, 2012; Brown, 2015; Loehlin & Beaujean, 2017). A variação do fator consiste na variação dos dados das variáveis/indicadores que é explicada por um fator ou por um conjunto de fatores que explicam várias variáveis/indicadores, designados fatores comuns (figura 6.4) (Fabrigar & Wegener, 2012; Brown, 2015; Loehlin & Beaujean, 2017). Já a variação única consiste na variação dos dados que é explicada pelo fator específico da variável/indicador, que influencia apenas uma única variável/indicador (figura 6.4) (Fabrigar & Wegener, 2012; Brown, 2015; Loehlin & Beaujean, 2017).

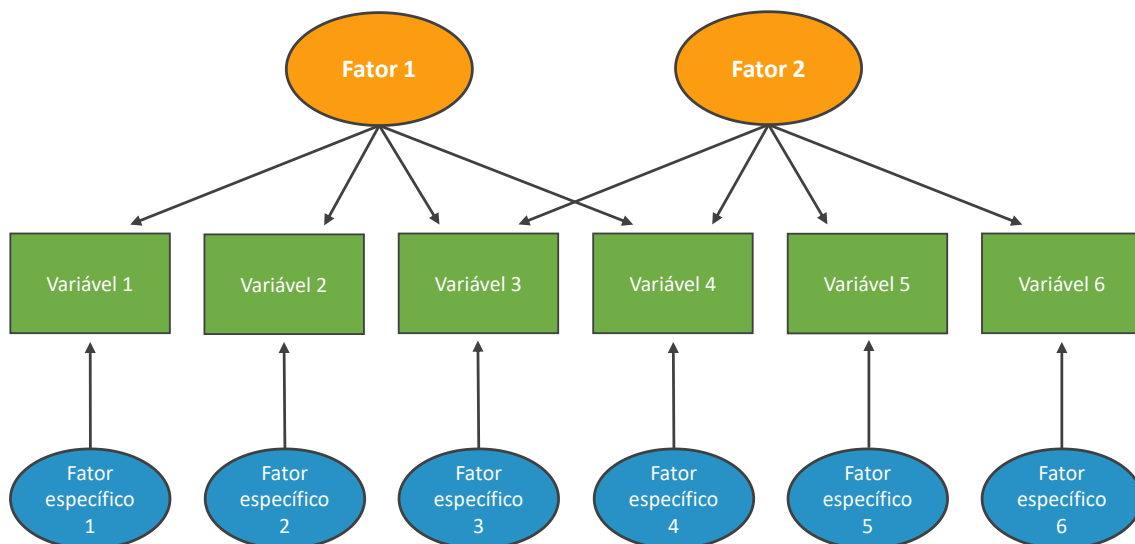


Figura 6.4 - Exemplo de explicação das variáveis através de fatores comuns e específicos

(Adaptado de Quaglino & Pagura, 2014).

O fator específico, por poder influenciar apenas uma única variável/indicador, não permite explicar a existência de correlações entre as variáveis/indicadores (Fabrigar & Wegener, 2012; Brown, 2015).

O modelo da análise fatorial pode ser assim representado pela seguinte equação:

$$y_j = \lambda_{j1}\eta_1 + \lambda_{j2}\eta_2 + \dots + \lambda_{jm}\eta_m + \varepsilon_j \quad (\text{Eq. 6.8})$$

em que:

- y_j constitui cada uma das variáveis/indicadores y obtidos(as) da amostra de dados n ;
- η_m representa cada um dos fatores η considerados;
- λ_{jm} é o peso fatorial que relaciona a variável/indicador j ao fator m ;
- ε_j é a variação única da variável/indicador j (Quaglino & Pagura, 2014; Brown, 2015).

Para o conjunto das variáveis/indicadores, a explicação da variação do fator e da variação única pode ser representada da seguinte forma:

$$y = \Lambda_y \eta + \varepsilon \quad (\text{Eq. 6.9})$$

em que:

- Λ_y representa o peso fatorial das variáveis/indicadores;
- ε constitui a variação única das variáveis/indicadores (Quaglino & Pagura, 2014; Brown, 2015).

Essa explicação pode também ser representada sob a forma de matriz:

$$\Sigma = \Lambda_y \Psi \Lambda_y' + \theta \varepsilon \quad (\text{Eq. 6.10})$$

em que:

- Σ é a matriz de correlação simétrica das variáveis/indicadores ($p \times p$);
- Λ_y representa a matriz $p \times m$ do peso fatorial λ ;
- Ψ constitui a matriz de correlação simétrica dos fatores ($m \times m$);

- $\Theta\epsilon$ é a matriz diagonal $p \times p$ da variação única (ϵ) (Quaglino & Pagura, 2014; Brown, 2015).

Os dados das variáveis/indicadores a inserir no modelo fatorial devem ser padronizados(as) previamente, para que o peso fatorial seja interpretado como uma correlação de regressão padronizada (Quaglino & Pagura, 2014; Brown, 2015; Loehlin & Beaujean, 2017). Tal procedimento foi efetuado antes de aplicar a análise de regressão.

Quando aplicada uma análise fatorial de tipo exploratória, deve ser primeiro escolhido o método para estimar o modelo fatorial com o menor número de fatores que explicam adequadamente as correlações e depois deve ser escolhida a técnica para rotacionar a matriz fatorial inicial, para perceber as variáveis/indicadores que estão relacionados(as) aos determinados fatores (Quaglino & Pagura, 2014; Brown, 2015; Loehlin & Beaujean, 2017).

Para estimar o modelo fatorial existem inúmeros métodos: máxima verosimilhança, componentes principais, mínimos quadrados não ponderados, mínimos quadrados generalizados (Quaglino & Pagura, 2014; Brown, 2015). Destes foi utilizado no SPSS o método de componentes principais, porque permite identificar os principais fatores que explicam a variação dos dados observados nas variáveis/indicadores, através da análise das semelhanças e diferenças que existem entre os dados das variáveis/indicadores para uma mesma área, ao contrário dos restantes métodos de análise fatorial que se baseiam apenas na relação que existe entre os conjuntos de dados das variáveis/indicadores (Quaglino & Pagura, 2014; Brown, 2015). Assim, com este método, é possível reduzir o número de fatores que explicam um dado problema, o que permite explicar a variação dos dados de forma sintetizada (Quaglino & Pagura, 2014; Brown, 2015).

O número de fatores identificados depende do número de variáveis/indicadores introduzidos(as) no modelo, ou seja, quantas mais variáveis/indicadores são introduzidos(as), mais fatores são identificados (Brown, 2015).

Após a seleção automática do número de fatores, quando selecionados dois ou mais fatores estes devem ser rotacionados porque a rotação, ao maximizar os valores de correlação entre variáveis/indicadores que estão próximos de 1 e ao minimizar os valores de correlação mais próximos de 0, permite evitar que o modelo fatorial produza um número infinito de soluções de variáveis/indicadores relacionados(as) aos fatores igualmente adequadas, o que dificulta a interpretação dos resultados (Quaglino & Pagura, 2014; Brown, 2015; Loehlin & Beaujean, 2017).

A rotação a aplicar pode ser do tipo ortogonal (em que se impede que os fatores estejam correlacionados entre si), caso das rotações *varimax* e *quartimax*, ou oblíqua (em que os fatores podem correlacionarem-se entre si), caso da rotação *direct oblimin* (Quaglino & Pagura, 2014; Brown, 2015; Loehlin & Beaujean, 2017). Neste caso, foi aplicada no SPSS uma rotação de tipo ortogonal (rotação *varimax*) para que a variação entre os

valores de correlação mais próximos de 1 e os valores de correlação mais próximos de 0 seja a maior possível (por forma a identificar as variáveis/indicadores com maior correlação com cada fator) e para que o peso fatorial das variáveis/indicadores represente apenas a correlação que existe entre as variáveis/indicadores e os fatores, e não tenha em conta a correlação que possa existir entre os vários fatores (por forma a diferenciar os fatores) (Quaglino & Pagura, 2014; Brown, 2015; Loehlin & Beaujean, 2017).

Esta rotação consegue cumprir tais objetivos porque os eixos das coordenadas dos fatores são colocados em ângulos de 90 graus (para os fatores não se correlacionarem entre si) e são girados para se poder maximizar os pesos fatoriais mais elevados e minimizar os pesos fatoriais mais reduzidos (Brown, 2015; Loehlin & Beaujean, 2017).

Um peso fatorial é considerado relevante se for igual ou maior que 0,30 (Brown, 2015). Por isso, no *SPSS* definiu-se ainda que no resultado da rotação não deviam aparecer correlações inferiores a 0,30, para facilitar a interpretação dos resultados.

Com a definição das especificações anteriormente referidas foi possível executar o modelo fatorial. O resultado deste apenas é adequado se os fatores identificados tiverem, cada um, mais de duas variáveis/indicadores associados(as) com forte peso fatorial e se num fator as variáveis/indicadores tiverem um peso fatorial forte e nos restantes fatores tiverem um peso fatorial muito reduzido (Fabrigar & Wegener, 2012; Brown, 2015).

O modelo fatorial executado identificou oito fatores, que cumpriram as recomendações anteriores. Com a soma dos pesos fatoriais de cada fator em cada variável/indicador elevados ao quadrado, para o conjunto dos fatores, obteve-se o valor de 23,707. Este valor, quando comparado com o total de variáveis/indicadores inseridos(as) (37), indica que 64,070 % da variação das variáveis/indicadores é explicada pelo conjunto dos fatores identificados.

Para verificar se o número de fatores identificados pelo modelo são adequados, pode ser efetuada a regra de *Kaiser-Guttman*, o teste do *Scree Plot* e/ou a análise paralela (Fabrigar & Wegener, 2012; Brown, 2015; Loehlin & Beaujean, 2017).

A regra de *Kaiser-Guttman* pressupõe que o número de fatores corresponde aos fatores que assumem um autovalor (*eigenvalue*) maior que 1 (Fabrigar & Wegener, 2012; Loehlin & Beaujean, 2017). Isto deve-se ao facto de se assumir que os fatores que têm um autovalor inferior a 1 são constituídos por variáveis/indicadores em que a variação explicada pelo dado fator é menor do que a variação única da variável/indicador (Brown, 2015).

O teste do *Scree Plot* determina o número de fatores com base num gráfico que cruza no eixo vertical os autovalores (obtidos na matriz de entrada) com o número de fatores no eixo horizontal, e em que através do qual se determina o ponto onde a linha que

crusa os valores muda a inclinação de um declínio rápido para uma inclinação mais plana, que corresponde ao número de fatores ideal (Fabrigar & Wegener, 2012; Brown, 2015; Loehlin & Beaujean, 2017). A determinação dos fatores é possível através deste método, porque os autovalores ao representarem a variação de um dado fator leva a que os autovalores de valor elevado correspondam aos principais fatores comuns (Fabrigar & Wegener, 2012).

A análise paralela determina o número de fatores com base na comparação dos autovalores obtidos através do conjunto dos dados das variáveis/indicadores com os autovalores obtidos através de dados aleatórios das variáveis/indicadores (Fabrigar & Wegener, 2012). Assim, o número de fatores neste método vai corresponder aos autovalores do conjunto dos dados das variáveis/indicadores que são maiores que os autovalores dos dados aleatórios das variáveis/indicadores (Fabrigar & Wegener, 2012).

Dos métodos indicados, foi aplicada a regra de *Kaiser-Guttman* e o teste do *Scree Plot*, em que se obteve em ambos o mesmo número de fatores que foram identificados pelo modelo, o que indica que o número de fatores é adequado.

Para verificar se o modelo de análise fatorial aplicado permite explicar, de forma adequada, as variáveis/indicadores que foram inseridos(as), deve ser aplicada a medida de *Kaiser-Meyer-Olki* e o teste da esfericidade de Bartlett (Dziuban & Shirkey, 1974; Arsham & Lovric, 2011; Jr. et al., 2014).

A medida de *Kaiser-Meyer-Olki* (KMO) consiste na medição da proporção de correlação que não é composta por correlação parcial, ou seja, pela correlação que exista entre duas variáveis/indicadores apenas (após retirar a correlação que tenham com outras variáveis/indicadores), por forma a determinar se um conjunto de variáveis/indicadores podem ser explicadas de maneira adequada pelos fatores indicados no modelo de análise fatorial (Kaiser, 1970; Cerny & Kaiser, 1977). A medida KMO varia entre 0 e 1. Se conjuntos de várias variáveis/indicadores partilharem a influência de determinados fatores, a correlação parcial será menor e o valor de KMO será próximo de 1, valor que indica a ausência de correlação parcial. Um valor próximo de 0 indica que a maioria da correlação é parcial (Dziuban & Shirkey, 1974). Por isso, os fatores explicam a variação das variáveis/indicadores de forma adequada se o valor de KMO for maior que 0,8 (Kaiser, 1970; Dziuban & Shirkey, 1974).

O teste da esfericidade de Bartlett é um procedimento que permite verificar se existe uma variação de dados igual ou semelhante em diferentes variáveis/indicadores, dado este testar a hipótese das correlações entre as variáveis/indicadores serem nulas (Arsham & Lovric, 2011; Jr. et al., 2014). Tendo em conta que um modelo de análise fatorial adequado deve ter algumas variáveis/indicadores com uma variação de dados semelhante, ou seja, com correlações significativas (não nulas) entre si, o resultado da significância das correlações nulas no teste de Bartlett não deve ser maior que 0,05 (Arsham & Lovric, 2011; Jr. et al., 2014). Tal como pode ser observado no quadro 6.2,

em que na medida de *Kaiser-Meyer-Olki* obteve-se um valor de 0,839, o que significa que os fatores explicam a variação das variáveis/indicadores de forma adequada, dado o valor da medida ser maior que 0,8. Já no teste da esfericidade de Bartlett obteve-se um valor de significância de 0,000, o que quer dizer que não existe nenhuma variável/indicador que não tenha correlação com outra variável/indicador.

Quadro 6.2 - Resultado da medida de *Kaiser-Meyer-Olki* e do teste da esfericidade de Bartlett

Medida de <i>Kaiser-Meyer-Olki</i>		0,839
Teste da esfericidade de Bartlett	Qui-quadrado aproximado	1360823,877
	Grau de liberdade	666
	Significância	0,000

Dado ter-se constatado que o número de fatores identificados pelo modelo é adequado, que os fatores explicam a maior parte da variação das variáveis/indicadores de forma adequada e que qualquer variável/indicador do modelo se correlaciona com pelo menos uma outra, conclui-se que o modelo de análise fatorial executado permite a criação de indicadores com base nos fatores identificados.

6.4 Desenvolvimento de indicadores com base no modelo fatorial

O modelo fatorial aplicado no *SPSS*, para além de ter identificado os fatores com base nas correlações que existem entre as variáveis/indicadores, também produziu as correlações que existem entre as variáveis/indicadores e os fatores identificados, que representam a variação da variável/indicador que é explicado(a) por um fator.

Com o resultado das correlações, foi então possível determinar e analisar as variáveis/indicadores que têm maior correlação com os dados fatores, para se poder, através da informação destas, identificar as características de cada fator, por forma a transformar os fatores identificados em indicadores de valor acrescentado para o contexto da Austrália. Assim, determinou-se que as variáveis/indicadores que têm uma correlação com peso fatorial aproximado de/ou superior a 0,7 face a um fator, são consideradas as variáveis/indicadores que permitem caracterizar o fator (dado as variáveis/indicadores serem mais influenciados(as) por este).

Por isso, para o primeiro fator identificado, determinou-se que sete variáveis permitem caracterizar o fator (indicadas no quadro 6.3). Ao analisar as informações que estas variáveis representam, verificou-se que estas ao incluírem as diferentes categorias de rendimento familiar e os potenciais membros das famílias, faz com que demonstrem sobretudo o peso do rendimento familiar em cada lugar. Assim, concluiu-se que este fator vai permitir o desenvolvimento de um novo indicador designado estrutura económica familiar.

Quadro 6.3 - Representação das variáveis influenciadas pelo fator 1

Fator 1		
Código	Nome da variável	Peso fatorial obtido
V038	Pessoas entre os 0 e os 17 anos	0,871
V328	Pessoas com certificado	0,865
V228	Famílias com rendimento semanal entre 1500 e 1999 dólares australianos	0,834
V003	Pessoas	0,821
V227	Famílias com rendimento semanal entre 1000 e 1449 dólares australianos	0,801
V229	Famílias com rendimento semanal entre 2000 e 2999 dólares australianos	0,753
V226	Famílias com rendimento semanal entre 500 e 999 dólares australianos	0,644

Para o segundo fator identificado, obtiveram-se seis variáveis que permitem caracterizar o fator (quadro 6.4). Tendo em conta as informações que estas variáveis representam, verificou-se que estas traduzem no geral aspetos de uma qualidade de vida mais baixa, dado indicar pessoas que vivem em apartamentos, potencialmente arrendados e que não têm qualquer veículo a motor, o que faz com que tenham de se deslocar de transportes públicos, a pé ou de bicicleta. Por isso, este fator vai consistir num novo indicador designado índice de baixos rendimentos.

Quadro 6.4 - Representação das variáveis influenciadas pelo fator 2

Fator 2		
Código	Nome da variável	Peso fatorial obtido
V253	Residentes em apartamentos	0,864
V234	Habitações ocupadas com nenhum veículo a motor	0,845
V299	Habitações arrendadas	0,812
V386	Pessoas que se deslocam para o trabalho apenas de transportes públicos	0,739
V388	Pessoas que se deslocam para o trabalho apenas de bicicleta ou a pé ou outros	0,684
V322	Pessoas com pós-graduação	0,677

Já no terceiro fator identificado, foram definidas quatro variáveis que permitem caracterizar o fator (quadro 6.5). Com a análise das informações que estas variáveis representam, verificou-se que estas estão associadas essencialmente ao rendimento familiar e, mais concretamente, aos rendimentos mais elevados. Tendo em conta estes aspetos, determinou-se que este fator vai representar um novo indicador designado índice de rendimentos elevados.

Quadro 6.5 - Representação das variáveis influenciadas pelo fator 3

Fator 3		
Código	Nome da variável	Peso fatorial obtido
V029	Mediana do rendimento semanal por família	0,886
V230	Famílias com rendimento semanal superior a 3000 dólares australianos	0,857
V323	Pessoas com o diploma de graduação e certificado de graduação	0,669
V028	Mediana da renda semanal da habitação	0,655

No quarto fator identificado, determinou-se que apenas duas variáveis possibilitam a caracterização do fator (quadro 6.6), sendo que estas indicam que os lugares com maior número de pessoas com mais de 65 anos têm uma mediana de idade das pessoas maior. Tendo em conta as informações transmitidas pelas variáveis, decidiu-se que este fator vai corresponder ao novo indicador designado índice de envelhecimento.

Quadro 6.6 - Representação das variáveis influenciadas pelo fator 4

Fator 4		
Código	Nome da variável	Peso fatorial obtido
V025	Mediana de idade das pessoas	0,865
V040	Pessoas com 65 anos ou mais	0,836

No que se refere ao quinto fator, foram identificadas uma variável e um indicador que permitem descrever o fator (quadro 6.7). Através da análise das informações que representam, percebe-se que as famílias que têm membros indígenas ou oriundos de outros países tem tendência a ter um tamanho de agregado familiar maior. Desta forma, considera-se que este fator pode corresponder ao indicador que se vai designar diversidade sociocultural, dado o tamanho do agregado familiar estar muitas vezes ligado a questões culturais.

Quadro 6.7 - Representação da variável e do indicador influenciado(a) pelo fator 5

Fator 5		
Código	Nome da variável/indicador	Peso fatorial obtido
V076	Índice de diversidade	0,772
V031	Média do tamanho do agregado familiar	0,655

No sexto fator identificado, dado apenas uma variável ter uma correlação com peso fatorial superior a 0,7, consideraram-se mais duas variáveis com peso fatorial superior a 0.6. Assim, pôde-se identificar três variáveis que permitem caracterizar o fator (quadro 6.8). Ao examinar as informações que estas variáveis transmitem, verificou-se que estas relacionam os supermercados a serviços que são importantes para a atividade comercial

(dado as pessoas muitas vezes precisarem de ir aos bancos ou ao ATM para levantar ou consultar o dinheiro para depois poderem comprar os produtos). Deste modo, entendeu-se que este fator pode corresponder ao indicador que se vai designar serviços de proximidade.

Quadro 6.8 - Representação das variáveis influenciadas pelo fator 6

Fator 6		
Código	Nome da variável	Peso fatorial obtido
V391	Bancos	0,738
V395	Outros supermercados	0,619
V392	ATM	0,612

Para o sétimo fator identificado, verificou-se que duas variáveis proporcionam a descrição do fator (quadro 6.9). Com a análise das informações que descrevem, observa-se que estas relacionam as bombas de gasolina aos Coles Express, que constituem supermercados de pequena dimensão que estão junto a um número considerável de bombas de gasolina. Por isso, entendeu-se que este fator pode corresponder a um indicador que se vai chamar comércio de conveniência, por retratar locais de comércio que podem ser úteis para os utilizadores das bombas de gasolina.

Quadro 6.9 - Representação das variáveis influenciadas pelo fator 7

Fator 7		
Código	Nome da variável	Peso fatorial obtido
V396	Bombas de gasolina	0,785
V397	Coles Express	0,707

Por fim, no oitavo fator identificado, foram também determinadas duas variáveis que permitem descrever o fator (quadro 6.10). Através da análise das informações que as variáveis representam, verifica-se que estas retratam locais de comunicação, seja a ligação com outros lugares (Transportes) ou a transmissão de informação (Antenas). Por

isso, concluiu-se que este fator pode representar um novo indicador, designado índice de comunicações.

Quadro 6.10 - Representação das variáveis influenciadas pelo fator 8

Fator 8		
Código	Nome da variável	Peso fatorial obtido
V394	Transportes	0,768
V393	Antenas	0,757

Para desenvolver os indicadores de valor acrescentado referidos, foi necessário determinar para cada indicador um valor único referente a cada área estatística SA1 e SA2. Assim, para se obter esse valor único para cada indicador foi preciso realizar os seguintes cálculos:

$$y_a = \frac{\sum n_{ia}f}{\sum f} \quad (\text{Eq. 6.11})$$

$$x_a = \frac{y_a}{m} \quad (\text{Eq. 6.12})$$

em que:

- n_{ia} consiste no valor que cada uma das variáveis/indicadores i do fator assume em cada área a considerada;
- f representa o peso fatorial de cada variável/indicador i do fator;
- x_a é o valor final de cada indicador de valor acrescentado para cada área a considerada;
- m corresponde ao valor máximo de y_a .

Através da análise das correlações que existem entre as variáveis/indicadores e os fatores identificados e do cálculo do valor único que cada novo indicador vai assumir nas áreas estatísticas SA1 e SA2, foi então possível desenvolver oito indicadores de valor acrescentado para o contexto da área de mercado da Austrália.

6.5 Análise dos indicadores obtidos com o modelo fatorial

Para se poder verificar a importância dos indicadores desenvolvidos com o modelo fatorial para o estudo da área de mercado da Austrália, procedeu-se à representação e análise desses indicadores para o conjunto da Austrália e para as três maiores cidades do país.

6.5.1 Estrutura económica familiar

Assim, com o indicador da estrutura económica familiar (figura 6.5) podem ser identificadas as áreas estatísticas onde o rendimento familiar tem maior ou menor importância económica.

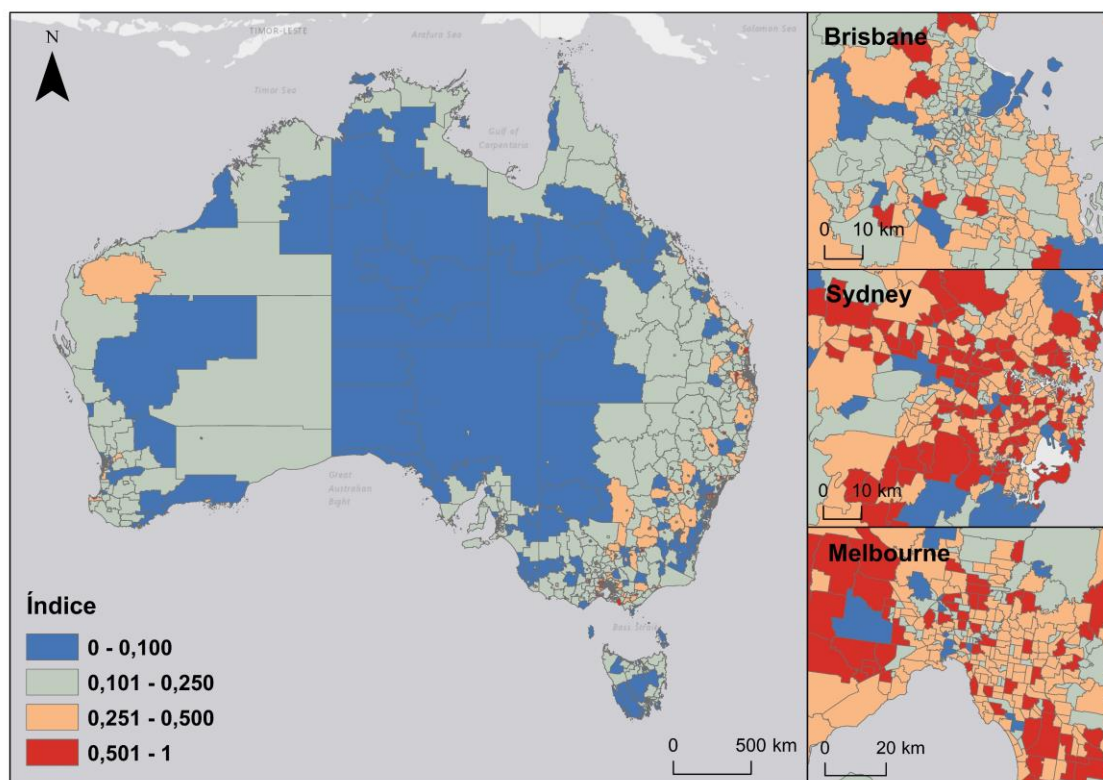


Figura 6.5 - Estrutura económica familiar consoante as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.

Através da análise do indicador, verificou-se que as áreas estatísticas onde o rendimento familiar tem maior importância, por terem um valor do índice mais elevado, localizam-se nas zonas urbanas, sobretudo nas zonas mais afastadas do centro das cidades (à

exceção de Sydney, onde o rendimento familiar tem um peso considerável em quase todas as áreas da cidade). Já as áreas onde o rendimento familiar tem menor importância localizam-se principalmente no interior da Austrália e no centro das maiores cidades (com a exceção de Sydney).

Analisar este indicador pode ser importante para o estudo da área de mercado, porque ao identificar as áreas estatísticas onde o rendimento familiar tem maior importância, permite determinar os lugares onde os produtos principalmente destinados às famílias poderão ter maior potencial de negócio, o que facilita a definição da área de mercado em que as empresas devem potenciar um produto com estas características.

6.5.2 Índice de baixos rendimentos

No caso do índice de baixos rendimentos (figura 6.6) pode-se identificar as áreas estatísticas onde um maior ou menor número de pessoas vive com dificuldades económicas.

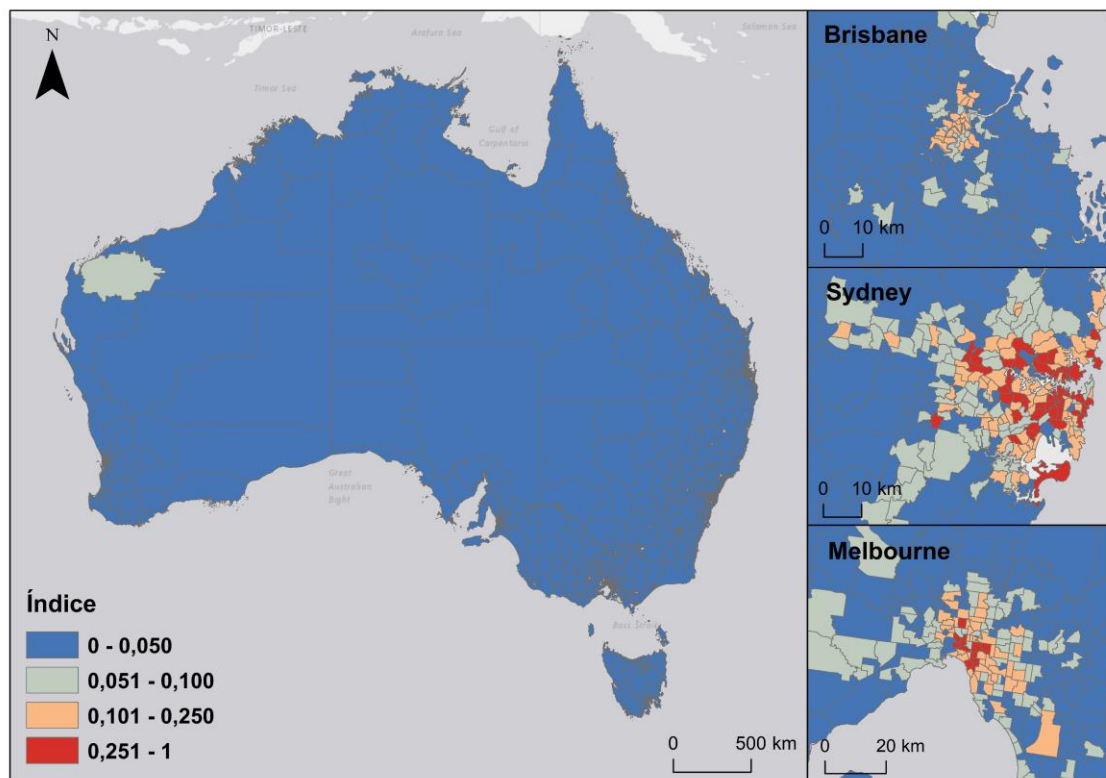


Figura 6.6 - Índice de baixos rendimentos para as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.

Com este índice, constatou-se que as áreas estatísticas onde um maior número de pessoas vive com baixos rendimentos localizam-se essencialmente no centro das principais cidades. Nas restantes partes do país, o número de pessoas que vive com baixos rendimentos é bastante reduzido.

Desta forma, os especialistas de marketing, com base nestes dados, podem selecionar como área de mercado o centro das principais cidades para disponibilizar ou promover produtos de valor reduzido ou verificar se um dado produto tem potencial de negócio nesse lugar.

6.5.3 Índice de rendimentos elevados

Para o índice de rendimentos elevados (figura 6.7) pode-se identificar as áreas estatísticas onde um maior ou menor número de pessoas vive sem problemas económicos e com um poder de compra superior à média.

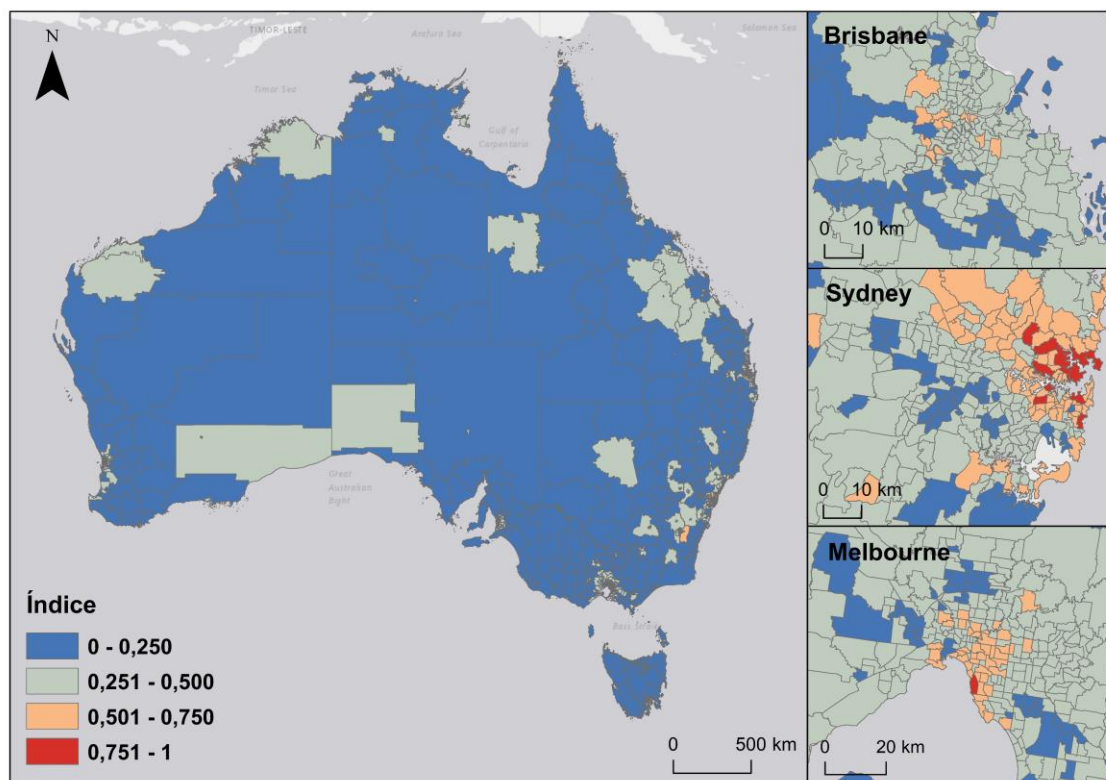


Figura 6.7 - Índice de rendimentos elevados para as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.

Com a análise deste índice, constatou-se que as áreas estatísticas onde um maior número de pessoas tem um rendimento elevado localizam-se em zonas específicas das principais cidades (a sul do centro de Melbourne, a norte do centro de Sydney e a oeste do centro de Brisbane) e na periferia a leste de Camberra. No resto do país, tendencialmente um menor número de pessoas vive com rendimentos elevados

Tendo por base estas informações, os especialistas de marketing podem selecionar as áreas referidas que têm um maior número de pessoas com rendimento elevado como área de mercado para produtos mais personalizados e de valor mais elevado.

6.5.4 Índice de envelhecimento

No que se refere ao índice de envelhecimento (figura 6.8) pode-se determinar as áreas estatísticas em que as pessoas são tendencialmente mais idosas ou mais jovens.

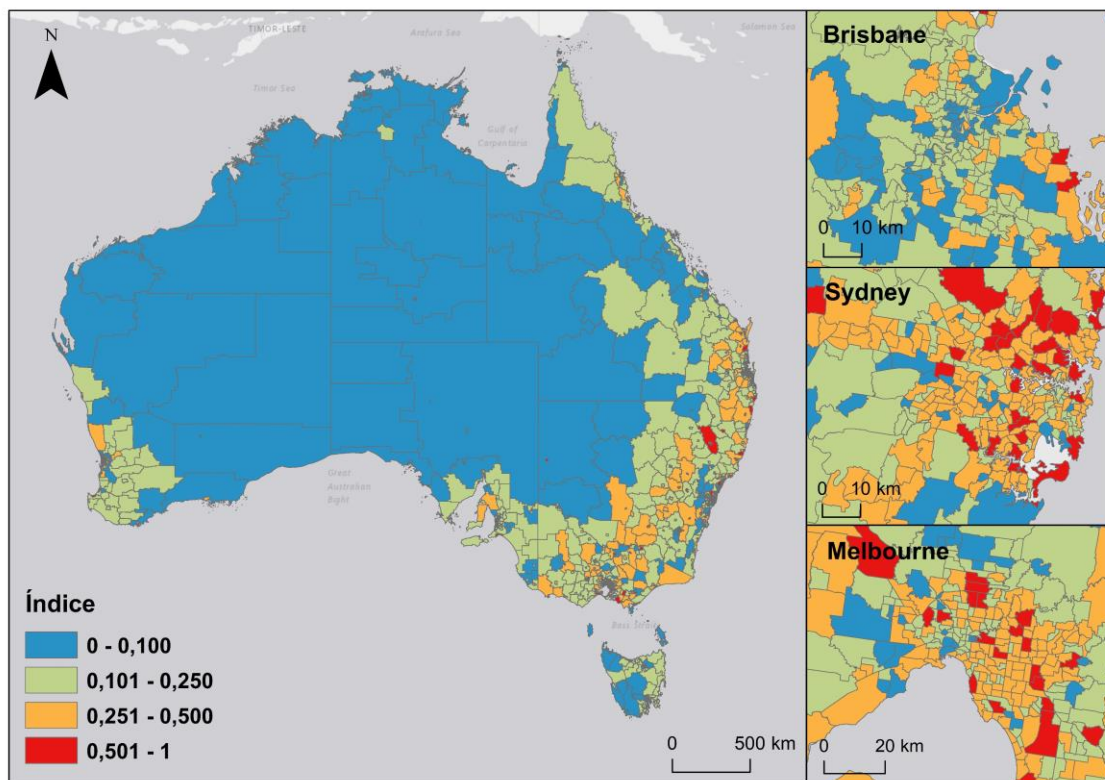


Figura 6.8 - Índice de envelhecimento para as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.

Ao analisar-se este indicador, verificou-se que as áreas estatísticas onde o índice de envelhecimento é maior, ou seja, que têm um maior número de idosos, situam-se nas zonas urbanas, sobretudo em zonas centrais à volta do CBD das principais cidades, e junto à costa leste da Austrália. Nas restantes áreas do país tende a existir pessoas mais jovens, dado o índice de envelhecimento assumir um valor mais baixo nessas áreas.

Com estes dados, as empresas podem adaptar a oferta de produtos a cada área, disponibilizando produtos destinados principalmente a pessoas com mais idade nas áreas onde o índice de envelhecimento é maior, e produtos destinados a pessoas mais novas onde o índice de envelhecimento é menor.

6.5.5 Diversidade sociocultural

Com o indicador da diversidade sociocultural (figura 6.9) pode-se especificar em que áreas estatísticas as famílias tendem a ter uma maior ou menor diversidade (ou seja, se têm ou não membros indígenas ou oriundos de outros países) e ao mesmo tempo se têm um tamanho de agregado familiar maior ou menor.

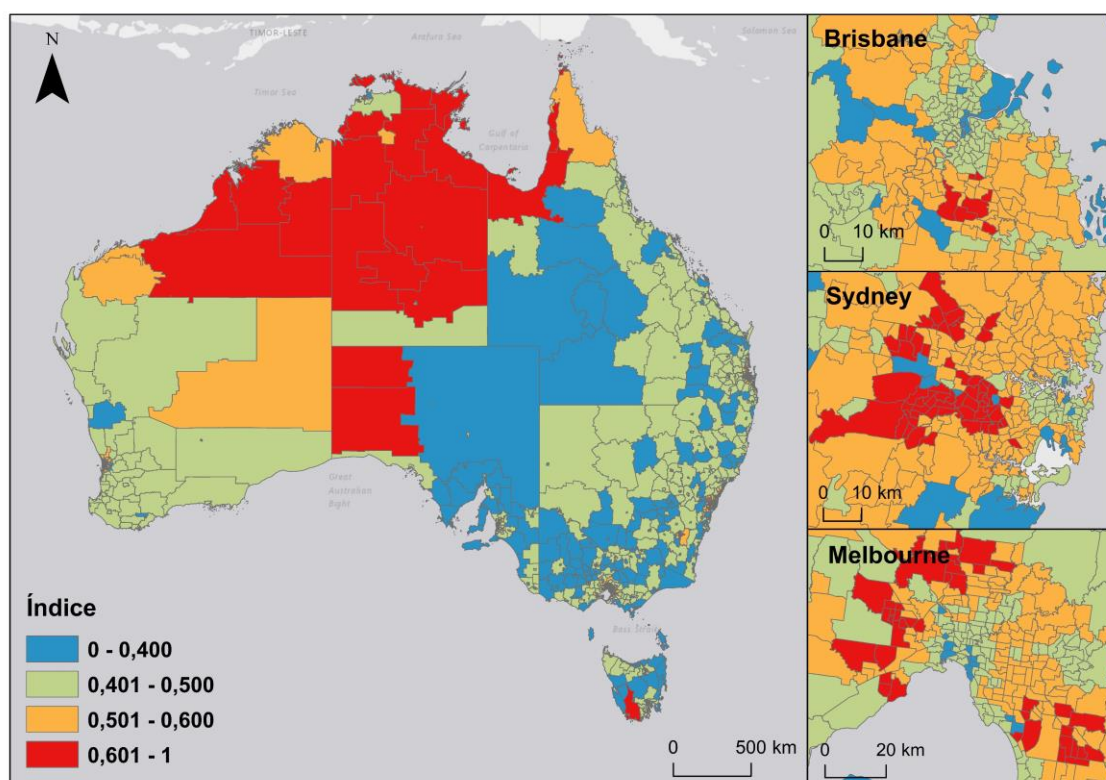


Figura 6.9 - Diversidade sociocultural consoante as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.

A variação espacial do índice permite constatar que as áreas estatísticas onde a diversidade sociocultural é maior, ou seja, onde tendencialmente as famílias têm membros indígenas ou oriundos de outros países e um tamanho do agregado familiar maior, localizam-se na periferia das principais cidades, no interior da Austrália (sobretudo nas zonas de interior mais a norte do país) e a sul da ilha da Tasmânia. Nos restantes territórios do país, a maioria das famílias não têm membros indígenas ou oriundos de outros países e tendem a ter um tamanho do agregado familiar menor.

Assim, nas áreas onde o índice de diversidade sociocultural é maior, as empresas podem disponibilizar produtos em que o principal público-alvo são as pessoas de outras culturas ou as famílias com um tamanho do agregado familiar maior.

6.5.6 Serviços de proximidade

No caso do indicador que analisa os serviços de proximidade (figura 6.10) pode-se determinar onde existe um maior ou menor número de serviços na proximidade de supermercados, ou seja, onde existe mais bancos ou caixas de multibanco (ATM) junto a supermercados.

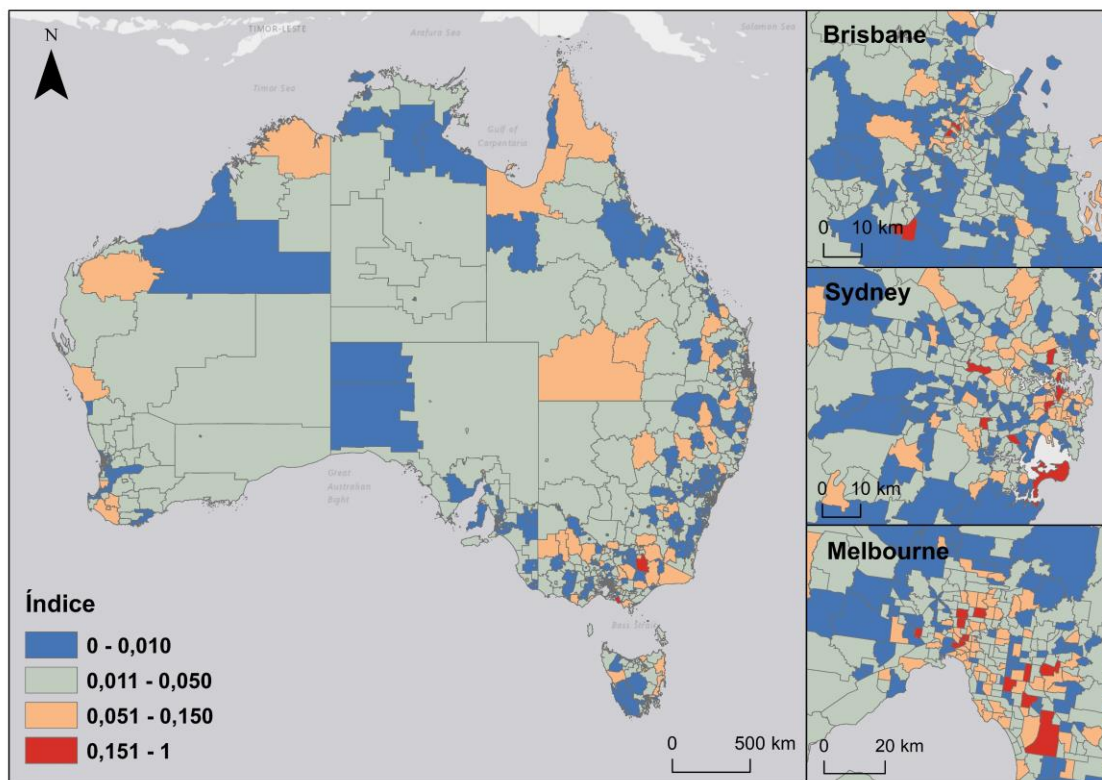


Figura 6.10 - Serviços de proximidade consoante as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.

Com a variação espacial do indicador, verificou-se que as áreas estatísticas onde existem mais serviços de proximidade estão situadas principalmente no centro das cidades e em lugares específicos da periferia das áreas urbanas. Nas restantes zonas do país os serviços de proximidade tendem a existir em menor número.

Determinar os lugares onde os serviços de proximidade estão em maior número pode ser importante para ajudar as empresas a escolherem a melhor localização para um potencial local de negócio de um produto.

6.5.7 Comércio de conveniência

No indicador que analisa o comércio de conveniência (figura 6.11) é possível identificar as áreas estatísticas onde existe um maior ou menor número de locais de comércio que podem ser úteis para os utilizadores de um equipamento (neste caso maioritariamente as bombas de gasolina).

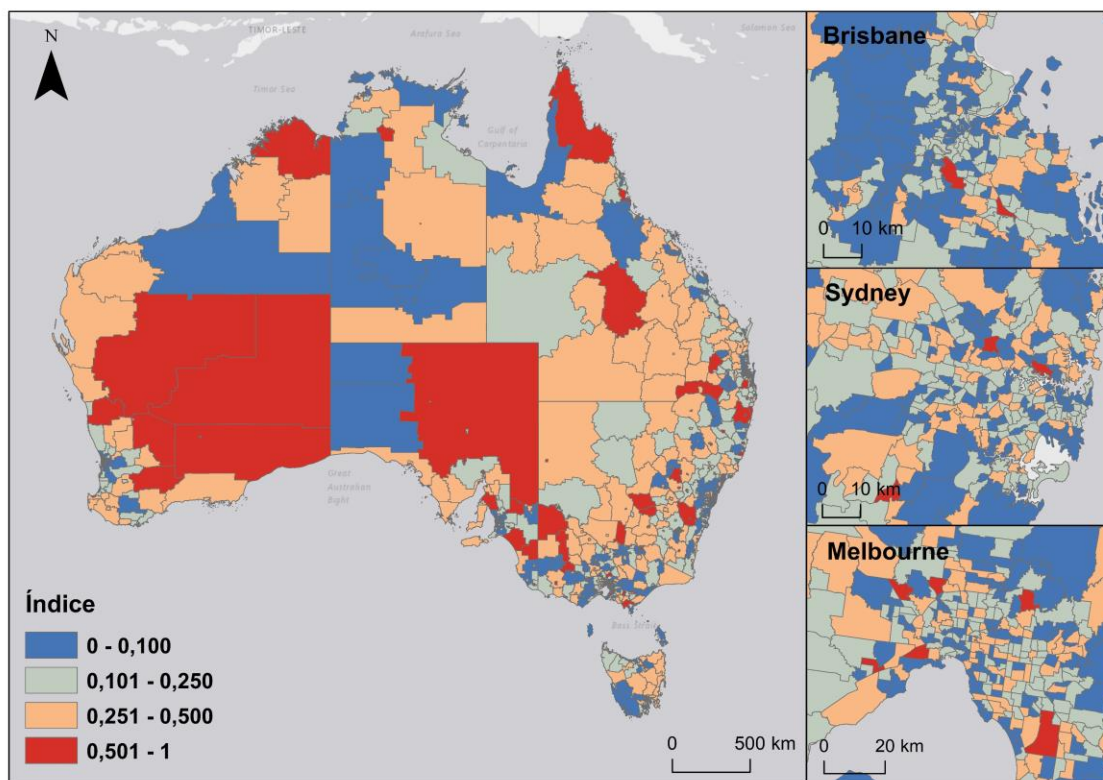


Figura 6.11 - Comércio de conveniência consoante as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.

Através da análise da distribuição espacial dos valores do índice, constata-se que as áreas estatísticas onde existe mais comércio de conveniência localizam-se sobretudo no interior da Austrália e nas zonas mais afastadas dos centros urbanos. Nos centros urbanos e nas restantes zonas do país o comércio de conveniência tende a ter uma presença reduzida.

Verificar os lugares onde existe mais comércio de conveniência pode ser útil para as empresas identificarem áreas onde, apesar de terem poucos residentes, existe um considerável fluxo de pessoas que poderá justificar a criação de um novo local para a oferta ou divulgação de um produto.

6.5.8 Índice de comunicações

No caso do índice de comunicações (figura 6.12), este pode servir para encontrar as áreas estatísticas onde existe um maior ou menor número de locais de comunicação, sejam esses locais de ligação a outros lugares (transportes) ou de transmissão de informação (antenas).

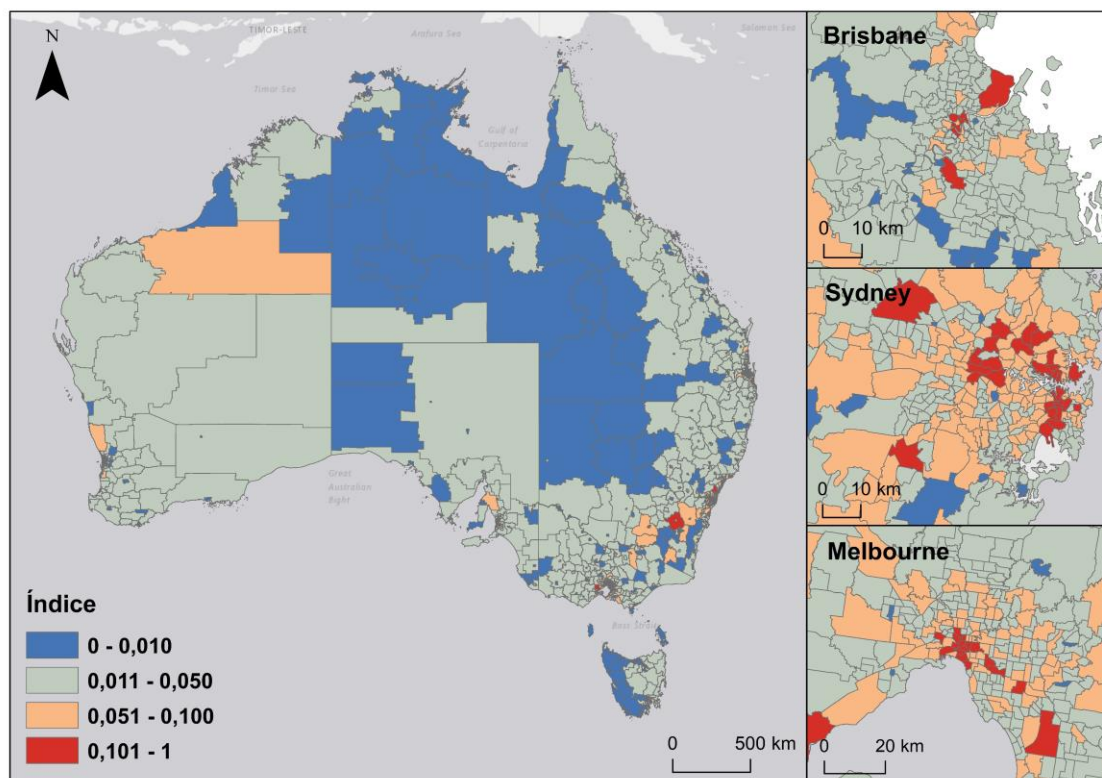


Figura 6.12 - Índice de comunicações para as áreas estatísticas SA2 do conjunto da Austrália e das três maiores cidades do país.

Quando analisada a variação espacial do indicador, verifica-se que as áreas estatísticas onde existem mais locais de comunicação estão situadas principalmente no centro das cidades e em pontos específicos das áreas urbanas, caracterizados por serem locais de transporte intermodal, ou seja, que permitem a mudança de um meio de transporte para outro, e/ou por serem locais onde existe uma maior circulação de pessoas que tendem a comunicar com outras pessoas. No interior da Austrália, na Tasmânia e em algumas zonas periféricas das áreas urbanas os locais de comunicação tendem a ter uma presença reduzida.

Identificar as áreas estatísticas onde existem mais locais de comunicação pode ajudar as empresas a identificarem as zonas onde existe um grande fluxo de pessoas, que podem justificar a oferta ou a divulgação de um produto nesses locais.

6.6 Aplicabilidade dos indicadores desenvolvidos

Para demonstrar como os indicadores desenvolvidos podem ser utilizados na tarefa de encontrar as melhores localizações para um local de negócio, aplicaram-se os modelos

de localização-alocação p-mediana, maximização da cobertura e maximização da participação de mercado, com o objetivo de localizar três instalações de um dado negócio, tendo por base os conceitos teóricos desenvolvidos no capítulo 3.3.1.

Aplicou-se o modelo p-mediana, porque ao permitir a redução da distância dos potenciais clientes face aos locais de negócio a localizar, este pode ser útil para localizar equipamentos sociais, culturais, entre outros. Já a aplicação do modelo da maximização da cobertura foi feita, porque ao permitir a maximização das áreas de mercado dos locais de negócio a localizar, pode ser útil para localizar equipamentos públicos e locais de negócio como pizzarias, entre outros. No caso do modelo da maximização da participação de mercado este foi aplicado porque, ao permitir a maximização da participação de mercado do local de negócio a localizar face à presença de instalações concorrentes, pode ser útil para localizar estabelecimentos que procurem dominar uma dada área de mercado, como supermercados, restaurantes, entre outros.

A aplicação destes modelos teve por base a área de mercado da cidade de Melbourne, por esta ter uma população elevada (o que pode proporcionar uma maior procura por um dado produto), uma ampla rede de estabelecimentos comerciais e uma boa rede de estradas que conecta os lugares de residência da população aos dados estabelecimentos comerciais.

No processo de aplicação dos modelos, foi selecionado, para todos os modelos a aplicar, os dados da variável população (V003) ou de cada um dos indicadores de valor acrescentado desenvolvidos relativos às áreas estatísticas SA1 como lugares de procura potencial.

Para todos os modelos, consideraram-se também os supermercados Coles e Woolworths como instalações de negócio concorrentes e os outros supermercados como instalações candidatas (lugares candidatos) que uma dada empresa pode comprar para localizar um determinado negócio.

Definiu-se ainda, nos vários modelos, uma distância-limite de cobertura da procura (área de mercado) de 3 km por cada instalação a localizar, dado se tratar de uma área urbana. Desta forma, a procura que está a mais de 3 km de cada instalação a localizar não é alocada a nenhuma instalação e a procura que está a uma distância inferior a 3 km é totalmente alocada à instalação a localizar.

Com a análise dos resultados verificou-se que, para cada tipo de modelo aplicado, as instalações escolhidas para localizar um negócio foram as mesmas consoante fosse selecionado no modelo como lugares de procura potencial a variável população ou cada um dos indicadores de valor acrescentado desenvolvidos. Este resultado permite concluir que os indicadores de valor acrescentado podem contribuir para reforçar a centralidade das localizações escolhidas pelos diferentes tipos de modelos aplicados.

Já quando comparados os resultados dos diferentes tipos de modelos, constatou-se que tanto no modelo p-mediana (figura 6.13), como no modelo da maximização da cobertura (figura 6.14) as instalações escolhidas para um dado negócio foram as mesmas. Essas instalações escolhidas situam-se na zona central da cidade, onde a procura potencial (medida pela população e pelo valor dos indicadores desenvolvidos) é maior, estando duas das instalações próximas das extremidades opostas do centro da cidade e a outra no meio da zona central da cidade, a igual distância das outras duas instalações, o que permite reduzir a distância dos locais de negócio face à maioria da procura potencial e, ao mesmo tempo, aumentar a área de mercado do negócio.

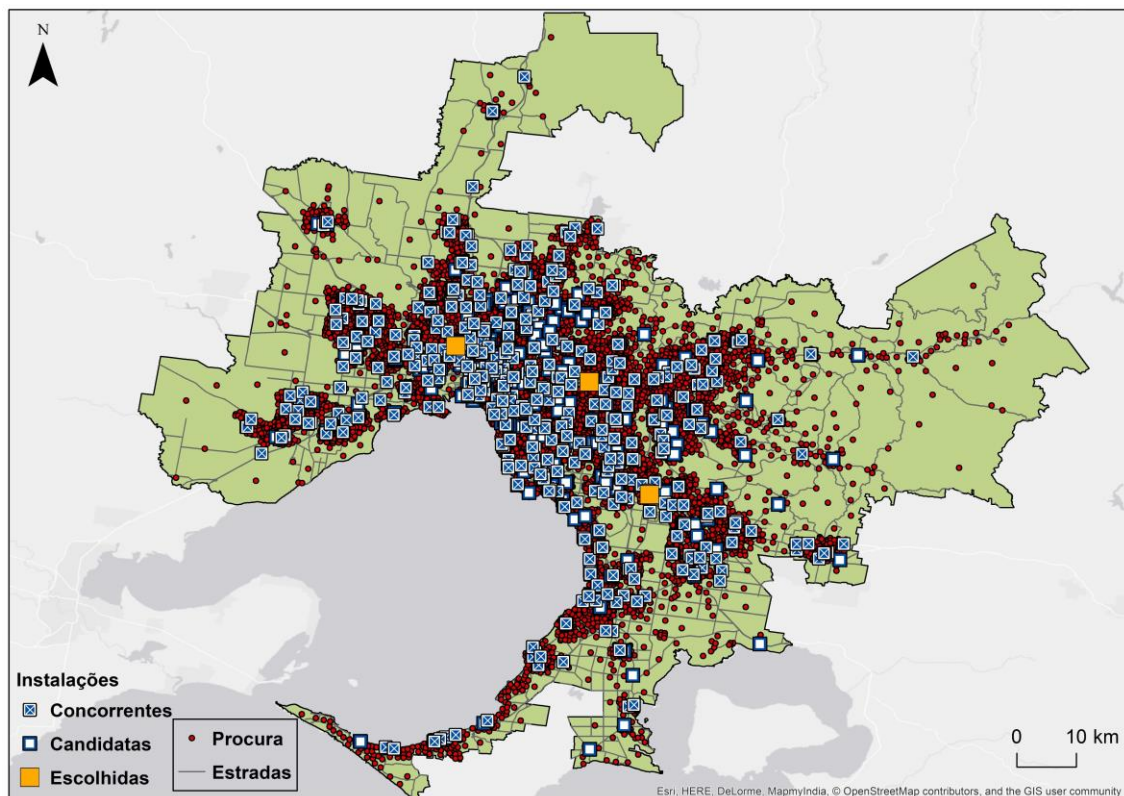


Figura 6.13 - Instalações escolhidas para a localização de um negócio através do modelo p-mediana.

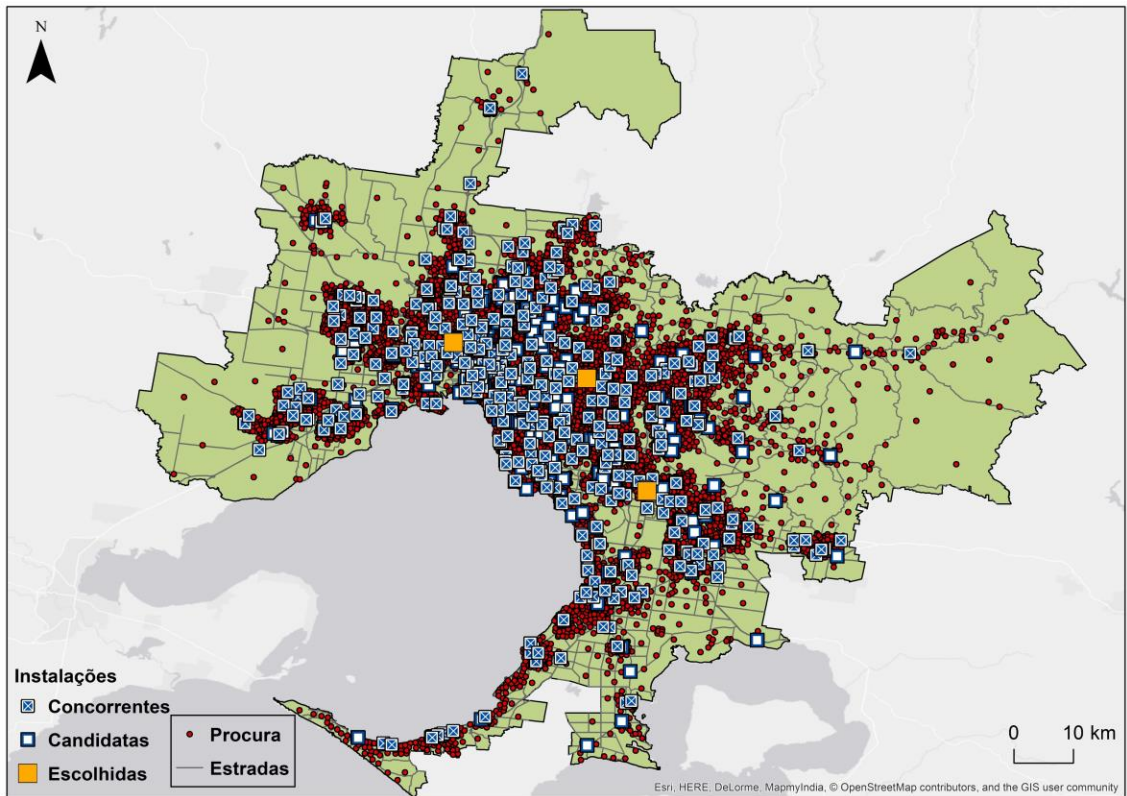


Figura 6.14 - Instalações escolhidas para a localização de um negócio através do modelo da maximização da cobertura.

Contudo, no caso do modelo da maximização da participação de mercado (figura 6.15), foram escolhidas outras instalações para um dado negócio. Estas instalações escolhidas situam-se na zona central da cidade, estando as três instalações aglomeradas na proximidade do CBD para que cada instalação nesta área possa cobrir um maior número de potenciais consumidores (procura) e para que a dada empresa consiga dominar a comercialização de produtos nesta área da cidade.

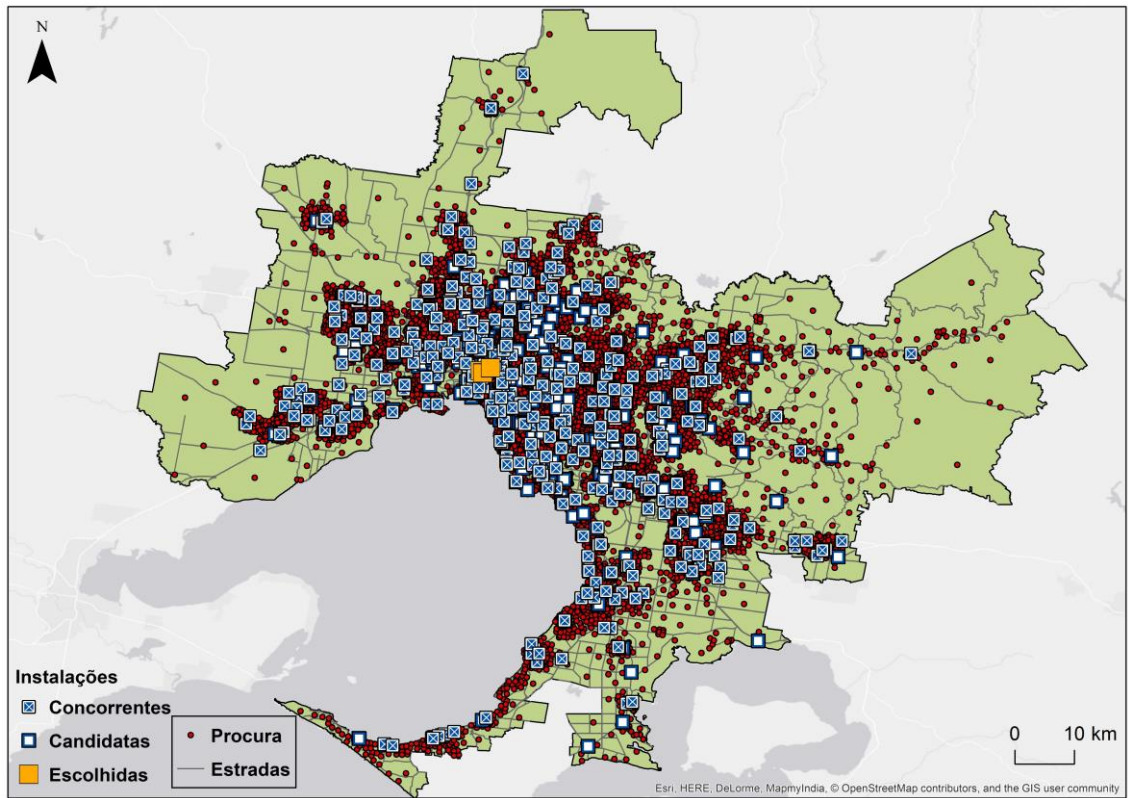


Figura 6.15 - Instalações escolhidas para a localização de um negócio através do modelo da maximização da participação de mercado.

7. Considerações finais

Tendo em conta a dificuldade que os especialistas de marketing normalmente revelam na delimitação dos lugares onde predominam os segmentos de mercado de um produto, considerou-se que seria importante desenvolver um modelo espacial que proceda à criação de indicadores de valor acrescentado para o contexto de uma área de mercado, para que esses indicadores possam tornar o processo de delimitação cada vez mais célere e eficaz, condições que são essenciais para decidir onde se deve apostar na oferta de um determinado produto.

A aplicação do modelo espacial desenvolvido no contexto da Austrália, ao ter concretizado a eliminação das variáveis com informações secundárias, ao ter agregado algumas das variáveis (para sintetizar a informação), ao ter eliminado os clusters de variáveis e os grupos de variáveis menos relevantes para o marketing, e ao ter selecionado os clusters de variáveis/desenvolvido os indicadores mais relevantes para o marketing, permitiu obter 43 variáveis/indicadores com informações provenientes dos censos da Austrália e 10 variáveis com dados obtidos do *OpenStreetMap* e o *OpenCellID* que se considerou ser mais relevantes para a área de mercado da Austrália, o que resultou num total de 53 variáveis/indicadores que foram incluídos(as) no modelo de regressão. Desta forma, conseguiu-se cumprir de forma adequada a primeira etapa do modelo, dado ter sido possível sintetizar as variáveis/indicadores mais pertinentes para a área de mercado da Austrália.

Para que os dados obtidos do *OpenStreetMap* e do *OpenCellID* pudessem ser contabilizados por área estatística, tal como os dados dos censos (essencial para se poder relacionar tais dados com as informações dos censos), estes foram inseridos (através da geocodificação) e calculados por área estatística no WebSIG da Mapidea, um *software* que permite aos especialistas de marketing analisar, desenvolver e criar dados geográficos sem necessitarem de conhecimentos avançados de SIG ou de aplicar qualquer procedimento complexo. A partir do momento em que todos os dados estão representados por área estatística, foi necessário ainda normalizar as variáveis/indicadores, i.e., colocar os dados das diferentes variáveis/indicadores numa única variação de unidades, para que pudessem ser relacionados(as) na análise de regressão aplicada.

Com estes procedimentos efetuados, foi possível aplicar a análise de regressão (tendo por base os dados das variáveis/indicadores para as áreas estatísticas SA1 por permitirem uma análise de regressão mais precisa), que teve como objetivo eliminar as variáveis/indicadores colineares, dado a inclusão destes(as) na análise fatorial poder perturbar o resultado da mesma, o que deve ser evitado.

Após realizada a análise de regressão, aplicou-se a análise fatorial que permitiu identificar os fatores que explicam os dados das variáveis/indicadores, através das

correlações que existiam entre eles(as), e que serviram de base à criação dos indicadores de valor acrescentado. Esta análise também permitiu obter as correlações que existem entre as variáveis/indicadores e os fatores identificados, que representam a variação da variável/indicador que é explicado(a) por um fator, o que é essencial para determinar as características de cada fator e, por sua vez, de cada indicador de valor acrescentado (dado uma variável/indicador ao ser explicado(a) por um fator significa que partilha das mesmas características que o dado fator), procedimento essencial para identificar o que significam os indicadores e para calcular o valor único que cada novo indicador vai assumir.

Assim foi possível transformar as variáveis/indicadores em oito indicadores de valor acrescentado que representam a área de mercado da Austrália, o que permitiu cumprir a segunda etapa do modelo desenvolvido.

Tendo em conta que os consumidores de um produto estão localizados num espaço geográfico preciso e que a procura para a maior parte dos produtos varia consoante as características e comportamentos dos consumidores, ao analisar-se espacialmente o conteúdo de cada um dos indicadores de valor acrescentado obtidos, os especialistas de marketing podem cruzar a segmentação dos consumidores de um ou vários produtos previamente efetuada com os indicadores de valor acrescentado obtidos e assim poder-se determinar espacialmente os lugares com maior ou menor potencial para a oferta/promoção de um produto ou que conjunto de produtos uma empresa deve disponibilizar num determinado lugar. Desta forma conseguiu-se desenvolver um sistema espacial de apoio à promoção e seleção do conjunto de produtos que mais se adequam a cada lugar, com base nos indicadores de valor acrescentado, o que permitiu cumprir a quarta etapa do modelo.

Caso uma dada empresa não esteja localizada numa determinada área de mercado, que, com base na análise espacial dos indicadores de valor acrescentado, se verificou que poderá ter potencial para a oferta de um ou vários produtos, pode-se aplicar um ou vários modelos de localização que irão identificar os lugares dessa área de mercado com maior potencial para a oferta do produto com base nesses indicadores. No seguimento deste relatório aplicaram-se os modelos de localização-alocação p-mediana, útil para uma empresa que se queira localizar na proximidade dos consumidores dos seus produtos, maximização da cobertura, que permite a uma empresa maximizar a sua área de mercado e maximização da participação de mercado, que pode-se revelar importante para uma empresa que queira dominar o negócio de um conjunto de produtos num dado lugar, i.e., que queira aumentar a quota de mercado face à presença de empresas concorrentes. Assim, com base nos modelos de localização, foi possível identificar os lugares com maior potencial de negócio, cumprindo a terceira etapa do modelo.

Contudo, por forma a obter uma análise de mercado mais detalhada e ajustada às áreas/lugares com maior potencial para a oferta/promoção de um produto ou ao

conjunto de produtos com maior potencial de negócio em cada área/lugar, os especialistas de marketing podem aplicar técnicas de SIG que permitam representar, identificar, combinar (sobreposição), desenvolver e modificar dados espaciais específicos (como por exemplo dados de variáveis ou indicadores socioeconómicos e demográficos, do consumo de produtos, da localização dos locais de negócio existentes, de acessibilidade, da movimentação das pessoas, entre outros) para poderem determinar os lugares onde se deve realizar a oferta de produtos, delimitar as áreas de mercado (tendo em conta a dimensão espacial e o volume de negócio que se pretende atingir), caracterizar de forma adequada a potencial procura existente nas áreas de mercado e prever o lucro que uma empresa poderá obter. A maioria dessas técnicas estão disponíveis no WebSIG da Mapidea e podem ser aplicadas de forma simples e intuitiva sem ser preciso ter conhecimentos avançados em SIG.

Assim, com as técnicas de SIG, os especialistas de marketing de uma empresa podem explorar espacialmente os resultados dos indicadores de valor acrescentado obtidos com o modelo para cada território e articular os mesmos com outros dados específicos que caracterizam as áreas de mercado, por forma a obterem uma forma de marketing que tenha em conta as características específicas de cada área de mercado. Desta forma é possível cumprir a última etapa do modelo.

Através do desenvolvimento e aplicação destas etapas foi possível elaborar um modelo que permite identificar novos indicadores mais adequados à caracterização de uma determinada área de mercado (no caso deste relatório para a área de mercado da Austrália), cumprindo assim o principal objetivo deste relatório.

Este modelo desenvolvido, ao ter uma estrutura bastante flexível, tem a vantagem de possibilitar a sua aplicação no contexto de qualquer território, tendo a pessoa que o aplica apenas de mudar os dados utilizados (seleccionando os dados essenciais para a finalidade do modelo) e de fazer alguns ajustes na metodologia do modelo (consoante os princípios de seleção dos dados/informações referidos).

O processo de elaboração do modelo teve contudo a dificuldade de não ter sido possível obter dados do consumo de produtos por segmento de consumidores para a área de mercado da Austrália, o que poderia ter ajudado a determinar um ou mais produtos a utilizar como exemplo para a aplicação do modelo e, com a inserção desses dados relativos aos produtos escolhidos no modelo fatorial, ter levado à identificação de indicadores mais dirigidos ao potencial de negócio de um ou vários produtos em cada lugar.

Em trabalhos futuros sugere-se então a recolha de dados do consumo de produtos por segmentos de clientes de uma ou várias empresas em determinados lugares de uma área de mercado e a articulação desses dados com os indicadores de valor acrescentados obtidos com a aplicação do modelo, por forma a aumentar a eficácia da escolha dos lugares para realizar a oferta de produtos.

Também seria interessante aplicar o modelo MCI, em vez dos modelos de localização-alocação, na determinação da melhor localização para os locais de negócio de uma empresa, dado este ser um modelo que, ao permitir ter em conta um maior conjunto de variáveis que caracterizam os comportamentos espaciais dos consumidores, poderia determinar com maior precisão a localização ideal para os locais de negócio.

8. Bibliografia

- Aljandali, A. (2016). Bivariate Correlation and Regression. In D. Ruppert, J. Fan, E. Renault, & E. Zivot (Eds.), *Quantitative Analysis and IBM® SPSS® Statistics* (pp. 119–132). Cham, Switzerland: Springer International Publishing. <https://doi.org/10.1007/978-3-319-45528-0>
- Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2007). *Essentials of Modern Business Statistics with Microsoft® Excel®*. (J. W. Calhoun, Ed.) (3rd ed.). Mason, Ohio: Thomson South-Western. Retrieved from https://www.academia.edu/37115472/Essentials_of_Modern_Business_Statistics_3rd_Edition
- Arsham, H., & Lovric, M. (2011). Bartlett's Test. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 87–88). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_132
- Australian Bureau of Statistics. (2016). Census DataPacks. Retrieved May 20, 2019, from <https://www.abs.gov.au/>
- Baviera-Puig, A., Buitrago-Vera, J., & Rodríguez-Barrio, J. (2013). Un modelo de geomarketing para la localización de supermercados: diseño y aplicación práctica. *Docfradis, 1*, 27. Retrieved from <https://www.catedrafundacionarecesdcuniovi.es//docfradis-un-modelo-de-geomarketing-para-la-localizacion-de--supermercados--diseno-y-apl-es.html>
- Berman, B., & Evans, J. R. (2013). Trading-Area Analysis. In S. Wall (Ed.), *Retail Management: A Strategic Approach* (12th ed., pp. 256–280). London: Pearson Education.
- Bingham, N. H., & Fry, J. M. (2010). *Regression: Linear models in statistics* (1st ed.). London: Springer-Verlag. <https://doi.org/10.1007/978-1-84882-969-5>
- Birkin, M., Clarke, G. P., & Clarke, M. (1999). GIS for business and service planning. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications* (2nd ed., Vol. 2, pp. 709–722). New York: John Wiley & Sons, Inc. Retrieved from https://www.geos.ed.ac.uk/~gisteac/gis_book_abridged/files/ch51.pdf
- Bolstad, P. (2016). *GIS Fundamentals: A First Text on Geographic Information Systems* (5th ed.). Acton, Massachusetts: XanEdu Publishing, Inc.
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. (T. D. Little, Ed.) (2nd ed.). New York: The Guilford Press.
- Cavion, R., & Philips, J. (2006). Os Fundamentos do Geomarketing: Cartografia, Geografia e Marketing. In *Congresso de Cadastro Técnico Multifinalitário e Gestão Territorial* (pp. 1–9). Florianópolis, Brasil: UFSC – Departamento de Engenharia Civil.
- Cerny, B. A., & Kaiser, H. F. (1977). A Study Of A Measure Of Sampling Adequacy For Factor-Analytic Correlation Matrices. *Multivariate Behavioral Research, 12*(1), 43–

47. https://doi.org/10.1207/s15327906mbr1201_3

Chang, K. (2019). Introduction. In *Introduction to Geographic Information Systems* (9th ed., pp. 1–16). New York: McGraw-Hill Education.

Chasco-Yrigoyen, C. (2003). El geomarketing y la distribución comercial. *Investigación Y Marketing*, 79, 6–13. Retrieved from <https://econwpa.ub.uni-muenchen.de/econwp/urb/papers/0403/0403003.pdf>

Christaller, W. (1966). *Central Places in Southern Germany [Die zentralen Orte in Süddeutschland]*. (C. W. Baskin, Ed.) (Translated). Englewood Cliffs, New Jersey: Prentic-Hall, Inc.

Church, R. L. (2005). Location modelling and GIS. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications* (2nd ed., pp. 293–303). New York: John Wiley & Sons, Inc. Retrieved from https://www.geos.ed.ac.uk/~gisteac/gis_book_abridged/files/ch20.pdf

Church, R. L. (2008). Location-Allocation Modeling. In K. K. Kemp (Ed.), *Encyclopedia of Geographic Information Science* (1st ed., pp. 264–267). Thousand Oaks, California: SAGE Publications, Inc. <https://doi.org/10.4135/9781412953962.n120>

Church, R., & ReVelle, C. (1974). The maximal covering location problem. *Papers of the Regional Science Association*, 32(1), 101–118. <https://doi.org/10.1007/BF01942293>

Clarke, G. P., & Hayes, S. (2006). GIS and Retail Location Models. In G. Cliquet (Ed.), *Geomarketing: Methods and Strategies in Spatial Marketing* (1st ed., pp. 165–186). London: ISTE Ltd. <https://doi.org/10.1002/9781118614020.ch7>

Cliquet, G. (2006). *Geomarketing: Methods and Strategies in Spatial Marketing* (1st ed.). London: ISTE Ltd. <https://doi.org/10.1002/9781118614020>

Cooper, L. G., & Nakanishi, M. (1983). Standardizing Variables in Multiplicative Choice Models. *Journal of Consumer Research*, 10(1), 96–108. <https://doi.org/10.1086/208948>

Cooper, L. G., & Nakanishi, M. (1988). *Market Share Analysis: Evaluating Competitive Marketing Effectiveness*. (J. Eliashberg, Ed.) (3rd ed.). Norwell, Massachusetts: Kluwer Academic Publishers. Retrieved from https://www.researchgate.net/publication/259583948_Market_Share_Analysis_Evaluating_Competitive_Marketing_Effectiveness

Davis, C. A., & Alves, L. L. (2017). Web Services, Geospatial. In S. Shekhar, H. Xiong, & X. Zhou (Eds.), *Encyclopedia of GIS* (2nd ed., pp. 2499–2503). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-17885-1_1490

Desmet, P., & Zollinger, M. (2006). Price and Geographic Information. In G. Cliquet (Ed.), *Geomarketing: Methods and Strategies in Spatial Marketing* (1st ed., pp. 217–239). London: ISTE Ltd. <https://doi.org/10.1002/9781118614020.ch9>

- Dion, D., & Cliquet, G. (2006). Consumer Spatial Behavior. In G. Cliquet (Ed.), *Geomarketing: Methods and Strategies in Spatial Marketing* (1st ed., pp. 27–56). London: ISTE Ltd. <https://doi.org/10.1002/9781118614020.ch2>
- Douard, J.-P. (2006). Geomarketing and Consumer Behavior. In G. Cliquet (Ed.), *Geomarketing: Methods and Strategies in Spatial Marketing* (1st ed., pp. 91–113). London: ISTE Ltd. <https://doi.org/10.1002/9781118614020.ch4>
- Douard, J.-P., Heitz, M., & Cliquet, G. (2015). Retail attraction revisited: From gravitation to purchase flows, a geomarketing application. *Recherche et Applications En Marketing (English Edition)*, 30(1), 110–129. <https://doi.org/10.1177/0767370114565066>
- Doyle, S. (2001). Software review: How is geography supporting marketing in today's commercial organisations? *Journal of Database Marketing & Customer Strategy Management*, 9(1), 85–89. <https://doi.org/10.1057/palgrave.jdm.3240062>
- Dyer, G. (1982). *Advertising as communication*. (J. Fiske, Ed.) (1st ed.). Abingdon-on-Thames, United Kingdom: Routledge. Retrieved from https://www.academia.edu/9183295/Advertising_as_Communication
- Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, 81(6), 358–361. <https://doi.org/10.1037/h0036316>
- Eiselt, H. A. (2011). Equilibria in Competitive Location Models. In V. Eiselt, H. A.; Marianov (Ed.), *Foundations of Location Analysis* (pp. 139–162). New York: Springer Science+Business Media. https://doi.org/10.1007/978-1-4419-7572-0_7
- Eiselt, H. A., & Marianov, V. (2011). Pioneering Developments in Location Analysis. In *Foundations of Location Analysis* (pp. 3–22). New York: Springer Science+Business Media. https://doi.org/10.1007/978-1-4419-7572-0_1
- Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory Factor Analysis*. (N. Beretvas, Ed.) (1st ed.). New York: Oxford University Press. <https://doi.org/10.1093/acprof:osobl/9780199734177.001.0001>
- Figueiredo, D. A. de O. (2016). *Modelo de Geomarketing e Estatística Espacial para Gestão das Recolhas do Instituto Português do Sangue e da Transplantação*. Relatório de estágio de mestrado, Universidade de Lisboa, Lisboa. Retrieved from <http://hdl.handle.net/10451/26029>
- Fischer, K. (2011). Central Places: The Theories of von Thünen, Christaller, and Lösch. In V. Eiselt, H. A.; Marianov (Ed.), *Foundations of Location Analysis* (pp. 471–505). New York: Springer Science+Business Media. https://doi.org/10.1007/978-1-4419-7572-0_20
- Franceschini, F., Galetto, M., & Maisano, D. (2007). *Management by Measurement: Designing Key Indicators and Performance Measurement Systems* (1st ed.). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-73212-9>
- Fu, P. (2015). *Getting to Know Web GIS* (1st ed.). Redlands, California: Esri Press.

- Fundo Monetário Internacional. (2019). Gross domestic product per capita, current prices. Retrieved November 9, 2019, from <https://www.imf.org/external/pubs/ft/weo/2019/02/weodata/index.aspx>
- Galati, S. R. (2006). *Geographic Information Systems Demystified* (1st ed.). Norwood, Massachusetts: Artech House, Inc.
- Gallopel, K. (2006). Advertising Policy and Geographic Information. In G. Cliquet (Ed.), *Geomarketing: Methods and Strategies in Spatial Marketing* (1st ed., pp. 241–266). London: ISTE Ltd. <https://doi.org/10.1002/9781118614020.ch10>
- Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models* (1st ed.). Cambridge, United Kingdom: Cambridge University Press. <https://doi.org/10.1017/CBO9780511790942>
- Goldberg, D. W., Wilson, J. P., & Knoblock, C. A. (2007). From text to geographic coordinates: The current state of Geocoding. *URISA Journal*, 19(1), 33–46. Retrieved from <https://pdfs.semanticscholar.org/26f6/4bcacac4a64693ffe76d2e894ae6bfc88e36.pdf>
- Goodwin, P. (2018). Regression Models. In *Profit from Your Forecasting Software: A Best Practice Guide for Sales Forecasters* (pp. 109–136). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119415992.ch6>
- Greenhalgh, P., & King, H. (2013). Developing an Indicator of Property Market Resilience—Investigating the Potential of GIS to Analyse Business Occupier Displacement and Property Market Filtering: A Case Study of Tyne and Wear. *Urban Studies*, 50(2), 372–390. <https://doi.org/10.1177/0042098012453860>
- Hackley, C. (2005). Introducing Advertising and Promotion. In *Advertising and Promotion: Communicating Brands* (1st ed., pp. 1–24). London: SAGE Publications Ltd. <https://doi.org/10.4135/9781446278789.n1>
- Hill, L. L. (2006). *Georeferencing: The Geographic Associations of Information*. (W. Y. Arms, Ed.) (1st ed.). Cambridge, Massachusetts: Massachusetts Institute of Technology.
- Huff, D. L. (1964). Defining and Estimating a Trading Area. *Journal of Marketing*, 28(3), 34–38. <https://doi.org/10.1177/002224296402800307>
- Janowski, A., Nowak, A., Przyborski, M., & Szulwic, J. (2014). Mobile Indicators in GIS and GPS Positioning Accuracy in Cities. In M. Kryszkiewicz, C. Cornelis, D. Ciucci, J. Medina-Moreno, H. Motoda, & Z. W. Ras (Eds.), *Rough Sets and Intelligent Systems Paradigms. Lecture Notes in Computer Science, vol 8537* (pp. 309–318). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-08729-0_31
- Johnston, R. (2018). Geography. In *Encyclopædia Britannica*. London: Encyclopædia Britannica, inc. Retrieved from <https://www.britannica.com/science/geography>
- Joseph, L., & Kuby, M. (2011). Gravity Modeling and its Impacts on Location Analysis. In V. Eiselt, H. A.; Marianov (Ed.), *Foundations of Location Analysis* (pp. 423–443).

New York: Springer Science+Business Media. https://doi.org/10.1007/978-1-4419-7572-0_18

- Jr., J. F. H., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). Exploratory Factor Analysis. In *Multivariate Data Analysis* (7th ed., pp. 89–149). Harlow, United Kingdom: Pearson Education Limited.
- Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401–415. <https://doi.org/10.1007/BF02291817>
- Kalábová, J., Mokrý, S., & Turčínková, J. (2013). Regional differences of consumer preferences when shopping for regional products. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *61*(7), 2255–2259. <https://doi.org/10.11118/actaun201361072255>
- Keith, T. Z. (2019). *Multiple Regression and Beyond: An Introduction to Multiple Regression and Structural Equation Modeling* (3rd ed.). New York: Routledge. Retrieved from https://www.researchgate.net/publication/330698260_Multiple_Regression_and_Beyond_An_Introduction_to_Multiple_Regression_and_Structural_Equation_Modeling
- Khan, M. (2006). *Consumer behavior and advertising management* (1st ed.). New Delhi: New Age International (P) Ltd. Retrieved from https://ibave.weebly.com/uploads/1/0/7/4/10741354/consumer_behaviour_and_advertisement_mgmt.pdf
- Kotler, P. (2000). Marketing in the Twenty-First Century. In *Marketing Management, Millenium Edition* (10th ed., pp. 1–18). Upper Saddle River, New Jersey: Prentice-Hall, Inc. Retrieved from https://www.researchgate.net/publication/235362523_Marketing_Management_The_Millennium_Edition
- Kotler, P., & Armstrong, G. (2018). Marketing: Creating Customer Value and Engagement. In *Principles of Marketing* (17th ed., pp. 26–61). Harlow, United Kingdom: Pearson Education Limited.
- Kulawiak, M., Prospathopoulos, A., Perivoliotis, L., Łuba, M., Kioroglou, S., & Stepnowski, A. (2010). Interactive visualization of marine pollution monitoring and forecasting data via a Web-based GIS. *Computers & Geosciences*, *36*(8), 1069–1080. <https://doi.org/10.1016/j.cageo.2010.02.008>
- Larose, C. D., & Larose, D. T. (2019). Regression Modeling. In D. T. Larose (Ed.), *Data Science Using Python and R* (1st ed., pp. 151–166). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119526865.ch11>
- Lassarre, S., Bonnet, E., Bodin, F., Papadimitriou, E., Yannis, G., & Golias, J. (2012). A GIS-based methodology for identifying pedestrians' crossing patterns. *Computers, Environment and Urban Systems*, *36*(4), 321–330. <https://doi.org/10.1016/j.compenvurbsys.2011.12.005>
- Li, D. (2018). Geocoding and Reverse Geocoding. In B. Huang (Ed.), *Comprehensive*

- Geographic Information Systems: GIS Methods and Techniques* (1st ed., pp. 95–109). Oxford, United Kingdom: Elsevier. <https://doi.org/10.1016/B978-0-12-409548-9.09593-2>
- Li, S., Veenendaal, B., & Dragičević, S. (2011). Advances, challenges and future directions in web-based GIS, mapping services and applications. In *Advances in Web-based GIS, Mapping Services and Applications* (1st ed., pp. 3–12). London: CRC Press/Balkema. <https://doi.org/10.1201/b15452>
- Loehlin, J. C., & Beaujean, A. A. (2017). Exploratory Factor Analysis - Basics. In *Latent Variable Models: An Introduction to Factor, Path, and Structural Equation Analysis* (5th ed., pp. 171–210). New York: Routledge.
- Longley, P. A. (2007). Some challenges to geodemographic analysis and their wider implications for the practice of GIScience. *Computers, Environment and Urban Systems*, 31, 617–622. <https://doi.org/10.1016/j.compenvurbsys.2007.10.002>
- Longley, P. A. (2008). Geographic Information Systems (GIS). In K. K. Kemp (Ed.), *Encyclopedia of Geographic Information Science* (1st ed., pp. 190–194). Thousand Oaks, California: SAGE Publications, Inc. <https://doi.org/10.4135/9781412953962.n84>
- Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic Information Science and Systems* (4th ed.). Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Lösch, A. (1954). *The Economics of Location*. (W. H. Woglom & W. F. Stolper, Eds.) (Translated). New Haven, Connecticut: Yale University Press.
- Marianov, V., & Serra, D. (2011). Median Problems in Networks. In V. Eiselt, H. A.; Marianov (Ed.), *Foundations of Location Analysis* (pp. 39–59). New York: Springer Science+Business Media. https://doi.org/10.1007/978-1-4419-7572-0_3
- Monino, J.-L., & Sedkaoui, S. (2016). Open Data: A New Challenge. In *Big Data, Open Data and Data Development* (1st ed., pp. 23–42). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119285199.ch2>
- Montejano, J. A., & Bello, G. M. C. (2018). Modelos de localización para geomarketing. *Espacialidades, Revista de Temas Contemporáneos Sobre Lugares, Política Y Cultura*, 8, 95–120. Retrieved from http://espacialidades.cua.uam.mx/vol/08/2018/01/06_Montejano_y_Cruz.php
- Murray, A. T. (2010). Advances in location modeling: GIS linkages and contributions. *Journal of Geographical Systems*, 12(3), 335–354. <https://doi.org/10.1007/s10109-009-0105-9>
- Nakanishi, M., & Cooper, L. G. (1974). Parameter Estimation for a Multiplicative Competitive Interaction Model: Least Squares Approach. *Journal of Marketing Research*, 11(3), 303–311. <https://doi.org/10.2307/3151146>
- O’Kelly, M. E. (2009). Applied Retail Location Models Using Spatial Interaction Tools. In A. S. Fotheringham & P. A. Rogerson (Eds.), *The SAGE Handbook of Spatial Analysis* (1st ed., pp. 419–442). London: SAGE Publications, Ltd.

<https://doi.org/10.4135/9780857020130.n22>

Open Knowledge Foundation. (2005). The Open Definition. Retrieved August 1, 2019, from <http://opendefinition.org/>

OpenCellID. (2019). Cell Tower data. Retrieved June 18, 2019, from <https://opencellid.org/downloads.php>

OpenStreetMap. (2019). OpenStreetMap Data Extracts - Geofabrik. Retrieved August 9, 2019, from <https://download.geofabrik.de/australia-oceania.html>

Peterson, M. P. (2008). Web GIS. In K. K. Kemp (Ed.), *Encyclopedia of Geographic Information Science* (1st ed., pp. 511–513). Thousand Oaks, California: SAGE Publications, Inc. <https://doi.org/10.4135/9781412953962.n230>

Petr, C. (2006). Direct Marketing and Geographic Information. In G. Cliquet (Ed.), *Geomarketing: Methods and Strategies in Spatial Marketing* (1st ed., pp. 267–291). London: ISTE Ltd. <https://doi.org/10.1002/9781118614020.ch11>

Quaglino, M. B., & Pagura, J. A. (2014). Factor Analysis. In J. I. Barragués, A. Morais, & J. Guisasola (Eds.), *Probability and Statistics: A Didactic Introduction* (1st ed., pp. 330–383). Boca Raton, Florida: CRC Press. <https://doi.org/10.1201/b16436>

Reese-Cassal, K. (2014). 2014/2019 Esri Diversity Index. Redlands, California: Esri. Retrieved from <https://www.esri.com/library/whitepapers/pdfs/diversity-index-methodology.pdf>

Reilly, W. J. (1931). *The Law of Retail Gravitation*. (U. of California, Ed.). New York: William J. Reilly.

ReVelle, C. S. (1986). The maximum capture or “sphere of influence” location problem: Hotelling revisited on a network. *Journal of Regional Science*, 26(2), 343–358. <https://doi.org/10.1111/j.1467-9787.1986.tb00824.x>

ReVelle, C. S., & Swain, R. W. (1970). Central Facilities Location. *Geographical Analysis*, 2(1), 30–42. <https://doi.org/10.1111/j.1538-4632.1970.tb00142.x>

Rocha, F. J. (2012). *Sistemas Complexos, Modelação e Geosimulação da Evolução de Padrões de Uso e Ocupação do Solo*. Dissertação de Doutoramento em Geografia, Universidade de Lisboa, Lisboa. Retrieved from <https://repositorio.ul.pt/handle/10451/6772>

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, 126(5), 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864>

Sendra, J. B., Jiménez, A. M., Díaz, M. F., Delgado, M. G., Medel, C. de L., Ferrero, V. O., ... García, M. J. S. (2011). Modelos de localización óptima de instalaciones y equipamientos. In *Sistemas de Información Geográfica y Localización Óptima de Instalaciones y Equipamientos* (2nd ed., pp. 73–122). Madrid: Ra-Ma Editorial y Publicaciones, S.A.

Snyder, L. V. (2011). Covering Problems. In V. Eiselt, H. A.; Marianov (Ed.), *Foundations of Location Analysis* (pp. 109–135). New York: Springer Science+Business Media.

https://doi.org/10.1007/978-1-4419-7572-0_6

- Socha, Y. M., Oelschlegel, S., Vaughn, C. J., & Earl, M. (2012). Improving an outreach service by analyzing the relationship of health information disparities to socioeconomic indicators using geographic information systems. *Journal of the Medical Library Association: JMLA*, 100(3), 222–225.
<https://doi.org/10.3163/1536-5050.100.3.014>
- Suárez-Vega, R., Santos-Peñate, D. R., & Dorta-González, P. (2012). Location models and GIS tools for retail site location. *Applied Geography*, 35(1–2), 12–22.
<https://doi.org/10.1016/j.apgeog.2012.04.009>
- Tsou, M.-H. (2008). Web Service. In K. K. Kemp (Ed.), *Encyclopedia of Geographic Information Science* (1st ed., pp. 513–514). Thousand Oaks, California: SAGE Publications, Inc. <https://doi.org/10.4135/9781412953962.n231>
- Vitanov, N. K. (2016). *Science Dynamics and Research Production: Indicators, Indexes, Statistical Laws and Mathematical Models*. (W. Glänzel & A. Schubert, Eds.) (1st ed.). Cham, Switzerland: Springer International Publishing.
<https://doi.org/10.1007/978-3-319-41631-1>
- von Thünen, J. H. (1966). *Von Thünen's Isolated State: An English Edition of Der Isolierte Staat*. (C. M. Wartenberg & P. Hall, Eds.). Oxford, United Kingdom: Pergamon Press.
- Wackerly, D. D., Mendenhall III, W., & Scheaffer, R. L. (2008). Multivariate Probability Distributions. In *Mathematical Statistics with Applications* (7th ed., pp. 223–295). Belmont, California: Thomson Learning, Inc.
- Widaningrum, D. L. (2015). A GIS – Based Approach for Catchment Area Analysis of Convenience Store. *Procedia Computer Science*, 72, 511–518.
<https://doi.org/10.1016/j.procs.2015.12.158>
- Wieland, T. (2017). Market Area Analysis for Retail and Service Locations with MCI. *The R Journal*, 9(1), 298–323. <https://doi.org/10.32614/RJ-2017-020>
- Yang, C., Wu, H., Huang, Q., Li, Z., Li, J., Li, W., ... Sun, M. (2011). WebGIS performance issues and solutions. In S. Li, S. Dragicevic, & B. Veenendaal (Eds.), *Advances in Web-based GIS, Mapping Services and Applications* (1st ed., pp. 121–138). London: CRC Press/Balkema. <https://doi.org/10.1201/b15452>
- Yao, X. (2009). Georeferencing, Geocoding. In R. Kitchin & N. Thrift (Eds.), *International Encyclopedia of Human Geography* (1st ed., pp. 458–465, vol. 4). Oxford, United Kingdom: Elsevier Ltd. <https://doi.org/10.1016/B978-008044910-4.00448-X>
- Zavala-Romero, O., Ahmed, A., Chassignet, E. P., Zavala-Hidalgo, J., Fernández Eguiarte, A., & Meyer-Baese, A. (2014). An open source Java web application to build self-contained web GIS sites. *Environmental Modelling & Software*, 62, 210–220.
<https://doi.org/10.1016/j.envsoft.2014.08.029>

9. Anexos

Anexo 1 - *Clusters* de variáveis fornecidos pelos censos da Austrália

<i>Clusters</i> de variáveis dos censos da Austrália
1. Sexo
2. Pessoas na noite dos censos em casa (por sexo)
3. Caracterização dos indígenas (por sexo)
4. Pessoas com cidadania australiana (por sexo)
5. Pessoas na noite dos censos consoante a propriedade de residência (por sexo)
6. Medianas e médias selecionadas
7. Lugar de residência habitual na noite do censo (por idade)
8. Idade (por sexo)
9. Estado civil registado (por idade e por sexo)
10. Estado civil perante a sociedade (por idade e por sexo)
11. Indígenas (por idade e por sexo)
12. Ascendência (por país de nascimento dos pais)
13. País de nascimento (por idade e por sexo)
14. País de nascimento (por ano de chegada à Austrália)
15. Capacidade de falar inglês e/ou outras línguas (por ano de chegada à Austrália e por idade)
16. Capacidade de falar inglês e/ou outras línguas dos pais (por idade de crianças dependentes)
17. Capacidade de falar inglês e/ou outras línguas (por língua falada em casa e por sexo)
18. Afiliação religiosa (por sexo)
19. Estudantes consoante a categoria e propriedade de instituição de ensino que frequentam (por estudantes em regime de full-time/part-time, por idade e por sexo)
20. Maior ano de escolaridade concluído (por idade e por sexo)
21. Rendimento individual total semanal (por idade e por sexo)

22. Necessidade de assistência (por idade e por sexo)
23. Trabalho voluntário para uma organização ou grupo (por idade e por sexo)
24. Trabalho doméstico não remunerado: número de horas (por idade e por sexo)
25. Assistência não remunerada a uma pessoa com deficiência (por idade e por sexo)
26. Cuidados infantis não remunerados (por idade e por sexo)
27. Relação do agregado familiar (por idade e por sexo)
28. Número de filhos (por idade dos pais)
29. Famílias e a sua composição (consoante a existência de filhos e/ou estudantes dependentes com menos de 15 anos)
30. Filhos dependentes (por composição das famílias, por ascendência dos pais e por idade)
31. Famílias combinadas
32. Rendimento familiar total semanal (por composição das famílias)
33. Rendimento do agregado familiar total semanal (por composição do agregado familiar)
34. Veículos a motor (por habitações)
35. Composição do agregado familiar (por número de pessoas geralmente residentes)
36. Habitações/residentes consoante a estrutura da habitação
37. Situação de posse da habitação e tipo de senhorio (por estrutura da habitação)
38. Prestação do empréstimo mensal - habitações (por estrutura de habitação)
39. Prestação do empréstimo mensal - famílias (por composição das famílias)
40. Prestação de renda semanal (por tipo de senhorio)
41. Ligação à internet da habitação (por estrutura de habitação)
42. Estrutura da habitação - quartos (por número de quartos)
43. Estrutura da habitação - famílias (por composição das famílias e do agregado familiar)
44. Força de trabalho selecionada, percentagem de desemprego, características de educação e migração (por sexo)
45. Lugar de residência habitual há 1 ano atrás (por sexo)

46. Lugar de residência habitual há 5 anos atrás (por sexo)
47. Situação da força de trabalho (por idade e por sexo)
48. Situação da força de trabalho (por sexo dos pais e por idade das crianças dependentes para famílias monoparentais (composição das famílias))
49. Situação da força de trabalho (por sexo dos pais e por idade das crianças dependentes para famílias de casais (composição das famílias))
50. Escolaridade (não estudantes por idade e por sexo)
51. Área de qualificação (não estudantes por idade e por sexo)
52. Área de qualificação (não estudantes por ocupação dos trabalhadores e por sexo)
53. Escolaridade (não estudantes por ocupação dos trabalhadores e por sexo)
54. Escolaridade (não estudantes por atividade dos trabalhadores empregados e por sexo)
55. Atividade dos trabalhadores empregados (por idade e por sexo)
56. Atividade dos trabalhadores empregados (por horas de trabalho e por sexo)
57. Atividade dos trabalhadores empregados (por ocupação dos trabalhadores)
58. Rendimento familiar total semanal para famílias de casais sem filhos (por situação de força de trabalho dos parceiros)
59. Rendimento familiar total semanal para famílias de casais com filhos (por situação de força de trabalho dos pais)
60. Rendimento familiar total semanal para famílias monoparentais (por situação de força de trabalho do parente)
61. Ocupação dos trabalhadores (por idade e por sexo)
62. Ocupação dos trabalhadores (por horas de trabalho e por sexo)
63. Método de viagem para o trabalho (por sexo)

Anexo 2 - Variáveis finais consoante os *clusters* de variáveis com variáveis agregadas.

Clusters de variáveis com variáveis agregadas	Variáveis finais
Idade	<ul style="list-style-type: none"> • Pessoas com 0 a 14 • Pessoas com 15 a 24 anos • Pessoas com 25 a 34 anos • Pessoas com 35 a 44 anos • Pessoas com 45 a 54 anos • Pessoas com 55 a 64 anos • Pessoas com 65 a 74 anos • Pessoas com 75 a 84 anos • Pessoas com mais de 85 anos <p>* Manteve-se os grupos acima consoante o sexo.</p>
	<ul style="list-style-type: none"> • Jovens entre os 0 e os 17 anos • Pessoas entre os 18 e os 64 anos • Idosos com 65 anos ou mais
Ascendência	<ul style="list-style-type: none"> • Pessoas com pais Australianos • Pessoas com pais Aboriginal • Pessoas com pais nascidos na China • Pessoas com pais nascidos nos Países Baixos • Pessoas com pais nascidos em Inglaterra • Pessoas com pais nascidos na Alemanha • Pessoas com pais nascidos na Grécia • Pessoas com pais nascidos na Índia • Pessoas com pais nascidos na Irlanda • Pessoas com pais nascidos em Itália • Pessoas com pais nascidos na Escócia • Pessoas com pais nascidos em outros países • Total de pessoas em que ambos os pais nasceram noutros países • Total de pessoas em que o pai nasceu noutro país • Total de pessoas em que a mãe nasceu noutro país

	<ul style="list-style-type: none"> • Total de pessoas em que ambos os pais nasceram na Austrália
País de nascimento	<ul style="list-style-type: none"> • Pessoas que nasceram na China • Pessoas que nasceram na Índia • Pessoas que nasceram em Itália • Pessoas que nasceram na Malásia • Pessoas que nasceram na Nova Zelândia • Pessoas que nasceram nas Filipinas • Pessoas que nasceram na África do Sul • Pessoas que nasceram no Sri Lanka • Pessoas que nasceram no Reino Unido • Pessoas que nasceram no Vietname • Pessoas que nasceram em outros lugares • Total de pessoas que nasceram fora da Austrália
Capacidade de falar inglês e/ou outras línguas	<ul style="list-style-type: none"> • Pessoas que falam apenas Inglês • Pessoas que falam Árabe • Pessoas que falam línguas indígenas Australianas • Pessoas que falam Cantonês • Pessoas que falam Mandarim • Pessoas que falam Grego • Pessoas que falam Italiano • Pessoas que falam Filipino • Pessoas que falam Hindi • Pessoas que falam Espanhol • Pessoas que falam Tailandês • Pessoas que falam Vietnamita • Pessoas que falam outras línguas • Total de pessoas que falam outras línguas • Pessoas que falam outras línguas e falam Inglês bem ou muito bem (total) • Pessoas que falam outras línguas e falam Inglês muito mal ou não falam (total) • Total de pessoas que falam outras línguas e falam Inglês
Rendimento individual total semanal	<ul style="list-style-type: none"> • Pessoas com rendimento semanal negativo • Pessoas com rendimento semanal de 1 a 449 dólares australianos • Pessoas com rendimento semanal entre 500 e 999 dólares australianos

	<ul style="list-style-type: none"> • Pessoas com rendimento semanal entre 1000 e 1499 dólares australianos • Pessoas com rendimento semanal entre 1500 e 1999 dólares australianos • Pessoas com rendimento semanal entre 2000 e 2999 dólares australianos • Pessoas com rendimento semanal acima de 3000 dólares australianos <p>* Manteve-se os grupos acima consoante o sexo e o grupo etário (mas não ambos na mesma variável).</p>
<p>Necessidade de assistência</p>	<ul style="list-style-type: none"> • Jovens que necessitam de assistência por deficiência • Jovens que não necessitam de assistência por deficiência • Pessoas em idade ativa que necessitam de assistência por deficiência • Pessoas em idade ativa que não necessitam de assistência por deficiência • Idosos que necessitam de assistência por deficiência • Idosos que não necessitam de assistência por deficiência • Total de pessoas que necessitam de assistência por deficiência • Total de pessoas que não necessitam de assistência por deficiência
<p>Rendimento familiar total semanal</p>	<ul style="list-style-type: none"> • Famílias com rendimento semanal negativo • Famílias com rendimento semanal entre 1 e 499 dólares australianos • Famílias com rendimento semanal entre 500 e 999 dólares australianos • Famílias com rendimento semanal entre 1000 e 1449 dólares australianos • Famílias com rendimento semanal entre 1500 e 1999 dólares australianos • Famílias com rendimento semanal entre 2000 e 2999 dólares australianos • Famílias com rendimento semanal superior a 3000 dólares australianos

<p>Habitacões/residentes consoante a estrutura da habitacão</p>	<ul style="list-style-type: none"> • Habitacões de tipo vivenda • Residentes em vivendas • Habitacões de tipo moradia • Residentes em moradias • Habitacões de tipo apartamento • Residentes em apartamentos • Habitacões de outro tipo • Residentes em habitacões de outro tipo • Total de pessoas em habitacões ocupadas • Total de habitacões não ocupadas
<p>Prestacão do empréstimo mensal</p>	<ul style="list-style-type: none"> • Famílias com prestacão mensal de empréstimo entre 0 e 300 dólares australianos • Famílias com prestacão mensal de empréstimo entre 300 e 449 dólares australianos • Famílias com prestacão mensal de empréstimo entre 450 e 599 dólares australianos • Famílias com prestacão mensal de empréstimo entre 600 e 799 dólares australianos • Famílias com prestacão mensal de empréstimo entre 800 e 999 dólares australianos • Famílias com prestacão mensal de empréstimo entre 1000 e 1399 dólares australianos • Famílias com prestacão mensal de empréstimo entre 1400 e 1799 dólares australianos • Famílias com prestacão mensal de empréstimo entre 1800 e 2399 dólares australianos • Famílias com prestacão mensal de empréstimo entre 2400 e 2999 dólares australianos • Famílias com prestacão mensal de empréstimo entre 3000 e 3999 dólares australianos • Famílias com prestacão mensal de empréstimo igual ou superior a 4000 dólares australianos • Total de famílias com prestacão mensal de empréstimo
<p>Situacão da força de trabalho</p>	<ul style="list-style-type: none"> • Pessoas empregadas num emprego em full-time • Pessoas empregadas num emprego em part-time • Total de pessoas empregadas • Pessoas desempregadas entre os 15 e os 19 anos • Pessoas desempregadas entre os 20 e os 24 anos • Pessoas desempregadas entre os 25 e os 34 anos

	<ul style="list-style-type: none"> • Pessoas desempregadas entre os 35 e os 44 anos • Pessoas desempregadas entre os 45 e os 54 anos • Pessoas desempregadas entre os 55 e os 64 anos • Pessoas desempregadas com mais de 65 anos • Total de pessoas desempregadas • Total de pessoas de mão de obra trabalhadora • Total de pessoas não inseridas na mão de obra trabalhadora
Método de viagem para o trabalho	<ul style="list-style-type: none"> • Pessoas que se deslocam para o trabalho apenas de comboio • Pessoas que se deslocam para o trabalho apenas de autocarro • Pessoas que se deslocam para o trabalho apenas de barco • Pessoas que se deslocam para o trabalho apenas de metro ou transporte ligeiro • Pessoas que se deslocam para o trabalho apenas de táxi • Pessoas que se deslocam para o trabalho apenas de automóvel • Pessoas que se deslocam para o trabalho apenas de moto • Pessoas que se deslocam para o trabalho apenas de bicicleta • Pessoas que se deslocam para o trabalho com outro meio de transporte • Pessoas que se deslocam para o trabalho apenas a pé

Anexo 3 - Variáveis/indicadores considerados relevantes

Código	Nome da variável/indicador
V001	Homens
V002	Mulheres
V003	Pessoas
V025	Mediana de idade das pessoas
V027	Mediana do rendimento semanal por pessoa
V028	Mediana da renda semanal da habitação

V029	Mediana do rendimento semanal por família
V031	Média do tamanho do agregado familiar
V038	Pessoas entre os 0 e os 17 anos
V039	Pessoas entre os 18 e os 64 anos
V040	Pessoas com 65 anos ou mais
V076	Índice de diversidade
V108	Estudantes
V215	Famílias que têm filhos com menos de 15 anos
V216	Famílias sem filhos com menos de 15 anos
V219	Famílias
V223	Crianças
V224	Famílias com rendimento semanal negativo
V225	Famílias com rendimento semanal entre 1 e 499 dólares australianos
V226	Famílias com rendimento semanal entre 500 e 999 dólares australianos
V227	Famílias com rendimento semanal entre 1000 e 1449 dólares australianos
V228	Famílias com rendimento semanal entre 1500 e 1999 dólares australianos
V229	Famílias com rendimento semanal entre 2000 e 2999 dólares australianos
V230	Famílias com rendimento semanal superior a 3000 dólares australianos
V234	Habitações ocupadas com nenhum veículo a motor
V239	Habitações ocupadas com veículos a motor
V240	Habitações ocupadas
V251	Residentes em vivendas ou moradias
V253	Residentes em apartamentos
V255	Residentes em habitações de outro tipo
V257	Habitações não ocupadas
V284	Famílias com prestação mensal de empréstimo
V299	Habitações arrendadas
V311	Pessoas empregadas
V319	Pessoas desempregadas
V322	Pessoas com pós-graduação
V323	Pessoas com o diploma de graduação e certificado de graduação
V324	Pessoas com bacharelato
V325	Pessoas com diploma avançado
V328	Pessoas com certificado
V386	Pessoas que se deslocam para o trabalho apenas de transportes públicos

V387	Pessoas que se deslocam para o trabalho apenas de automóvel ou mota
V388	Pessoas que se deslocam para o trabalho apenas de bicicleta ou a pé ou outros
V390	Centros comerciais
V391	Bancos
V392	ATM
V393	Antenas
V394	Transportes
V395	Outros supermercados
V396	Bombas de gasolina
V397	Coles Express
V398	Coles
V399	Woolworths