

MESTRADO
**MÉTODOS QUANTITATIVOS PARA A DECISÃO ECONÓMICA
E EMPRESARIAL**

TRABALHO FINAL DE MESTRADO
PROJETO

**ANÁLISE METODOLÓGICA DA FORMAÇÃO DOS TRÊS GRANDES
DURANTE AS ÉPOCAS 2012-13 A 2018-19**

PEDRO MIGUEL PESEIRO CARAPAU

NOVEMBRO - 2020



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

MESTRADO
MÉTODOS QUANTITATIVOS PARA A DECISÃO
ECONÓMICA E EMPRESARIAL

TRABALHO FINAL DE MESTRADO
PROJETO

ANÁLISE METODOLÓGICA DA FORMAÇÃO DOS TRÊS
GRANDES DURANTE AS ÉPOCAS 2012-13 A 2018-19

PEDRO MIGUEL PESEIRO CARAPAU

ORIENTADOR:
PROF. CARLOS J. COSTA



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

NOVEMBRO – 2020

Glossário

AIC - Akaike Information Criterion

BIC - Bayesian Information Criterion

CIES - Centro Internacional de Estudos Desportivos

CRISP-DM - Cross Industry Standard Process for Data Mining

FCP – Futebol Clube do Porto.

KDD - Knowledge Discovery in Database

MLB - Major Baseball League

NBA - National Basketball Association

NFL - National Football League

SCP – Sporting Clube de Portugal.

SEMMA - Sampling, Exploring, Modifying, Modelling and Assessing

SLB – Sport Lisboa e Benfica.

RESUMO

O sector do futebol tem um grande impacto na sociedade atual (quer a nível social como financeiro). Este impacto é generalizado a nível mundial, mas tem uma força mais acentuada nos continentes Europeus e Sul Americano. Portugal é um dos países Europeus onde esta influência mais se faz sentir, sobretudo ao nível do impacto junto da população, sendo uma das razões para ser um dos países com mais história neste desporto. Esta influência faz ainda com que exista uma grande rivalidade entre adeptos no nosso país, sobretudo entre os três maiores clubes portugueses, vulgarmente conhecidos como Benfica (SLB), Sporting (SCP) e Porto (FCP), em qualquer tema de futebol, onde se insere também a qualidade da formação dos clubes. Este trabalho serve para proporcionar uma resposta à questão de saber qual a melhor academia de futebol entre os três, pelo menos durante o período em estudo – entre as épocas 2012-13 e 2018-19. Analisaram-se trabalhos já efetuados na área de forma a ver que variáveis afetam o valor de mercado dos jogadores, e em que medida esta é um reflexo da qualidade dos mesmos. Após se identificar as variáveis a ser estudados, foram recolhidos e tratados os dados, seguindo a metodologia CRISP-DM. Na análise dos mesmos, fez-se uma regressão linear básica, cujo único objetivo foi fazer a ponte entre os trabalhos anteriores e a análise crítica realizada no subcapítulo seguinte, verificando também alguns resultados defendidos por outros autores. Concluiu-se que as variáveis incluídas na regressão são válidas estatisticamente (à exceção de duas) na análise do Valor de mercado dos jogadores (corroborando os autores estudados). De seguida procedeu-se à análise, e comparação de resultados, entre as três academias com os dados disponíveis e chegou-se à conclusão de que a formação do SLB viveu um período de crescimento e acaba por se tornar a mais forte neste período. Já a do SCP, que começa na liderança, acabou por viver períodos de estagnação levando a que tenha sido ultrapassada pela formação do SLB. O FCP por seu lado, embora tenha tido algum crescimento nesta área, continua muito longe do nível apresentado pelos seus rivais.

ABSTRACT

The football industry has a major impact on today's society (both socially and financially). This impact is widespread worldwide, but it is more pronounced on the European and South American continents. Portugal is one of the European countries where this influence is most felt, especially in terms of its impact on the population, being one of the reasons for it to be one of the countries with most history in this sport. This influence also causes a great rivalry between supporters in our country, especially between the three biggest Portuguese clubs, commonly known as Benfica (SLB), Sporting (SCP) and Porto (FCP), in any football theme, where it also inserts the academy quality of the clubs. This work is to provide an answer to the following question: what is the best football academy among the three? The analysis was restricted to the period under study - between the seasons 2012-13 and 2018-19. Work already carried out in the area was analysed to see which variables affect the market value of the players, and to what extent this reflects their quality. After identifying the variables and objects to be studied, the data was collected and treated, following the CRISP-DM methodology. In the data analysis, a basic linear regression was made, whose sole objective was to be a link between the previous works studied and the critical analysis carried out in the following sub-chapter, and verifying some results defended by other authors. It was concluded that the variables included in the regression are statistically valid (except for two) in the analysis of the market value of the players (corroborating the previously studied authors). Then, the analysis and comparison of results between the three academies with the available data were carried out, and it was concluded that the formation of the SLB lived a period of growth and ended up becoming the strongest in this period. The SCP academy, which starts in the leadership, ends up experiencing periods of stagnation leading to it being overtaken by the SLB's academy. The FCP, although it has had some growth in this area, remains far from the level presented by its rivals.

ÍNDICE

Glossário.....	i
Resumo	ii
Abstract.....	iii
Índice	iv
Índice de Figuras	vi
Índice de Gráficos.....	vii
Índice de Tabelas	vii
Agradecimentos	viii
1. Introdução.....	1
1.1 Enquadramento.....	1
1.2 Objetivos do Trabalho	2
1.3 Abordagem de Investigação	2
1.4 Estrutura do Trabalho	3
2. Revisão Bibliográfica	4
2.1 Abordagem Estatística no Desporto	4
2.2 Abordagem Estatística no Futebol	5
2.3 Indicadores explicativos do Valor de Mercado	7
2.3.1 Trabalho de Muller, Simons e Weinmann (2017).....	8
2.3.2 Herm, Callsen-Bracker e Kreis (2014).....	10
2.4 Conclusão das variáveis explicativas do valor de mercado	13
3. Metodologia na recolha de dados	13
4. Análise da formação dos três grandes em Portugal	17
4.1 Introdução.....	17
4.2 Análise dos dados	17

4.3 Análise das Academias.....	22
4.3.1 Número de jogadores e respetivo tempo de jogo	22
4.3.2 Valor de Mercado e Transferências	25
4.3.3 Presenças na Seleção Nacional	28
4.3.4 Títulos conquistados e participação de jogadores formados.....	29
5. Conclusões e Trabalhos Futuros	32
5.1 Conclusões.....	32
5.2 Trabalhos Futuros.....	34
6. Bibliografia.....	35

ÍNDICE DE FIGURAS

Figura 1 - Evolução de subscritores dos canais SPORT TV, 1998 – 2008,	1
Figura 2 - Modelo do Transfermarket	7
Figura 3 - Fatores usados por Müller, Simons, e Weinmann (2017).....	9
Figura 4 - Fatores usados por Herm, Callsen-Bracker, e Kreis (2014)	10
Figura 5 - Modelos usados por Herm, Callsen-Bracker, e Kreis (2014).....	11
Figura 6 - Processo do CRISP-DM	13
Figura 7 - Figura explicativa das várias fases do CRISP-DM.....	14
Figura 8 - Tabela de Correlação das Variáveis.....	18
Figura 9 - Output da Regressão Linear.....	19

ÍNDICE DE GRÁFICOS

Gráfico 1 - Evolução do valor de mercado médio dos jogadores.....	26
Gráfico 2 - Percentagem de jogadores no plantel principal.....	31

ÍNDICE DE TABELAS

Tabela 1 - Quatro Modelos desenvolvidos	9
Tabela 2 - Variáveis utilizadas na regressão linear.	18
Tabela 3 - Síntese dos minutos jogados pelos jogadores.....	21
Tabela 4 - Dados das épocas praticados pelos jogadores em ligas de diferentes	22
Tabela 5 - Percentagem de Épocas jogadas nos escalões de Coeficientes.	23
Tabela 6 - Valores de Mercado dos Jogadores ao longo das Épocas.	25
Tabela 7 - Número de transferências definitivas dos clube e respetivos valores.	25
Tabela 8 - Convocatórias para a Seleção Nacional de Futebol Portuguesa Sénior ..	28
Tabela 9 - Vencedores quatro competições nacionais 2012-13/2018-19	29
Tabela 10 - Número de jogadores formados, e respetiva percentagem, no plantel. .	29

AGRADECIMENTOS

Em primeiro lugar queria agradecer ao meu professor, e orientador, Carlos Costa por toda a sua ajuda e paciência ao longo deste período e que, sem as quais, não teria sido possível finalizar este trabalho.

De seguida queria agradecer ao meu amigo, e colega de casa, Valentino Cunha por toda a ajuda e motivação dada neste período de quarentena e confinamento, essencial para conseguir conciliar o meu trabalho atual com a realização deste trabalho escrito.

Por fim queria agradecer a todos que me acompanharam nesta viagem incrível que foi o meu percurso académico, com uma menção especial à Joana Bica, pelo apoio dado, à Tuna Económicas, por todos os momentos vividos, e à minha família que nunca deixou de me apoiar durante a minha vida e sempre lutou pelo melhor para mim.

1. INTRODUÇÃO

1.1 Enquadramento

O futebol, desde há várias décadas, tem um grande peso em Portugal, com a população, desde os tempos de Estado Novo, a virar a sua atenção para o futebol como forma de distração dos problemas do quotidiano.

Esta paixão tem crescido de mãos dadas com o crescimento da indústria do futebol Europeu em si e prova disso são, entre outros, as audiências televisivas, e consequentes receitas, cada vez maiores, os novos estádios com lotação superior, muito deles criados aquando da realização do Europeu de 2004 em Portugal, e o crescente sucesso, tanto de treinadores como de jogadores, de portugueses ao mais alto nível.

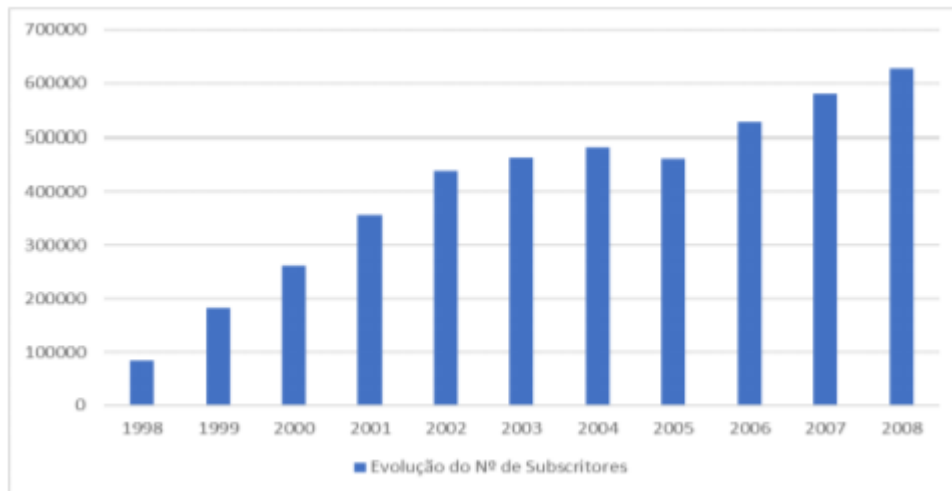


Figura 1 - Evolução do número de subscritores dos canais SPORT TV, 1998 – 2008, Dionísio (2009)

Em Portugal esta paixão pelo jogo distribui-se de maneira muito irregular. Segundo uma sondagem realizada pela Intercampus em 2019, o SLB reúne 4,5 milhões de adeptos em Portugal, seguido por FCP com 2,4 milhões e SCP com 2,3 milhões.

Esta grande assimetria entre estes três clubes, denominados por “Três Grandes”, e os restantes clubes leva a que exista uma grande rivalidade entre os mesmos, com os adeptos de cada um dos grandes a argumentar que o seu clube é o melhor dos três.

Esta discussão acaba por se alastrar à formação, e qual a melhor academia em Portugal, cuja atenção pelo tema cresceu no início do milénio com o sucesso de jogadores portugueses no estrangeiro, primeiro com Luís Figo e depois, num segundo momento,

com Cristiano Ronaldo e Ricardo Quaresma, todos eles provenientes do SCP levando a que a academia deste clube fosse reconhecida não só como a melhor de Portugal mas também uma das melhores do mundo – 2ª segundo um artigo de 2013 da bleacherreport (um dos sites desportivos mais conceituados do mundo), Atkins (2013).

Com o sucesso recente dos jogadores provenientes do SLB, com Bernardo Silva e João Félix à cabeça, esta luta pelo título de melhor formação em Portugal tem reacendido e ocupado grande parte das conversas desportivas entre os fãs do desporto.

1.2 Objetivos do Trabalho

Propõem-se como objetivo deste trabalho contribuir, com uma base estatística e metodológica, para a definição de qual é a melhor academia, atualmente, de entre os três grandes.

Para isso foram definidos seis objetos de análise:

- Número de jogadores formados;
- Tempo de jogo a nível sénior dos jogadores;
- Sucesso na carreira dos jogadores – analisando os destinos dos mesmos;
- Valorização e ganhos financeiros com os jogadores;
- Participação dos jogadores na Seleção Nacional;
- Títulos ganhos pelos clubes, com uma participação ativa dos jogadores formados.

1.3 Abordagem de Investigação

A investigação foi feita seguindo a metodologia CRISP DM, que consiste em seis fases: “Business Understanding”, “Data Understanding”, “Data Preparation”, “Modelling”, “Evaluation” e “Deployment”.

Após a recolha e tratamento dos dados foi realizado uma regressão linear simples, com os dados disponíveis, de forma a verificar se certas variáveis em estudo são válidas, estatisticamente, para analisar o Valor de Mercado dos jogadores. Esta regressão tem um objetivo meramente demonstrativo, de forma a fazer a ligação entre os trabalhos, e estudos, analisados na revisão de literatura e os objetos em estudo nos subcapítulos seguintes. O resultado da mesma, com os dados quantitativos disponíveis reduzidos, é um R^2 também reduzido, 0,384, que é, no entanto, semelhante aos obtidos nos estudos

realizados com uma abrangência de dados semelhante (mas bastante inferior quando comparado com os modelos mais complexos).

Num segundo momento foi efetuado a análise mais crítica do trabalho, realizando constantemente comparações entre os três clubes à medida que se analisa os objetos em estudo propostos.

Esta análise, e consequente conclusão, engloba também o conhecimento pessoal adquirido no tema, não só durante a investigação para o mesmo como também ao longo de uma vida em contacto com a modalidade. Esta é a parte mais relevante do trabalho e onde se retiram as conclusões sobre o tema em estudo.

1.4 Estrutura do Trabalho

Este trabalho está dividido em seis partes principais: Introdução, onde se introduz o tema e os objetivos do trabalho; Revisão de Literatura, que se subdivide nos temas de abordagem estatística no Desporto, os trabalhos já realizados especificamente no Futebol, o que influencia o Valor de Mercado de um jogador e a análise de dois estudos realizados, de maneiras diferentes, ambos com o objetivo de explicar o Valor de Mercado; Metodologia na Recolha dos Dados, que trata a forma como os dados foram recolhidos e como vão ser estudados; Análise dos Dados, onde é estudado os dados em cada um dos objetos em estudo definidos; e Conclusões e Trabalhos Futuros, que incorpora um conjunto de pareceres sobre os resultados obtidos, que levam à apresentação do resultado (diga-se, nomear a melhor Academia no período em estudo) e onde são apresentadas algumas sugestões para trabalhos futuros sobre o tema.

2. Revisão Bibliográfica

2.1 Abordagem Estatística no Desporto

A primeira gestão desportiva documentada de um clube, orientada por diretrizes estatísticas, aconteceu nos Estados Unidos da América, na equipa de basebol Oakland Athletics.

Esta equipa, liderada pelo seu diretor Billy Bean, tendo um dos orçamentos mais baixos da liga, começou a olhar para a hipótese de usar dados, recolhidos de forma sistemática e organizada, como o principal fator de decisão para contratar novos elementos e escolher a sua posição dentro do campo.

Esta experiência, retratada no livro de Michael Lewis (*Moneyball: The Art of Winning an Unfair Game*), demonstrou o potencial deste tipo de abordagens no desporto atual ao permitir que uma equipa com dos orçamentos mais baixos da liga definisse o recorde de vitórias seguidas neste desporto (20) com mais de 100 anos de história.

De acordo com um artigo escrito por Jason Heckl, Heckl, existem cinco lições que podemos retirar deste caso na gestão de uma entidade desportiva:

- *Uma estratégia orientada e organizada demora tempo*
- *Uma vantagem competitiva adquirida não dura sem constantes desenvolvimentos da mesma*
- *É essencial pensar fora da caixa do meio em que está inserido a equipa*
- *Uma mudança organizacional não é fácil*
- *Saber distinguir as perceções tradicionais enviesadas com a vida real*

De facto, os pontos acima referidos são transversais e aplicáveis a toda a indústria do desporto, na qual o futebol também se inclui.

O futebol, como uma indústria, começou mais tarde o seu processo de atualização estatística, sobretudo de dados individuais dos jogadores, quando comparado com outros desportos, nomeadamente o basebol, futebol americano e basquetebol (representados respectivamente por MBL [Major Baseball League], NFL [National Football League] e NBA [National Basketball Association]) que beneficiaram imenso de se localizarem no país com os maiores centros de desenvolvimento tecnológicos do mundo.

Este facto é enunciado em Frick, pp. 113, (2011), onde o autor faz a seguinte referência “Contrary to the situation in most American team sports leagues with their abundance of performance figures, measurement of individual player performance in European football can be problematic”.

2.2 Abordagem Estatística no Futebol

Como podemos então avaliar o sucesso desportivo de um jogador de futebol de uma maneira organizada e sem enviesamentos tendo em conta toda a importância e especulação que lhes é dada dentro e fora dos relvados?

Como podemos fazer esta análise de maneira séria sem ser totalmente influenciada pela especulação, cada vez maior, existente neste setor na nossa sociedade e que move mundos e pessoas?

De acordo com Tunaru e Viney (2010) as primeiras abordagens estatísticas efetuadas no futebol baseavam-se muito na previsão de resultados das equipas utilizando modelos estatísticos, tendo como exemplos Crowder et al. (2002) e McHale & Scarf, (2007), entre outros, cujos trabalhos tiveram um grande impacto na indústria de apostas do Reino Unido ao contribuírem com modelos mais fiáveis e realistas.

No entanto no últimos anos, e com o aperfeiçoamento das tecnologias de recolhas de dados e respetivas empresas, tendo como maior exemplo a Opta Sports, a maior empresa de recolha e análise de dados estatísticos relacionadas com futebol e com várias parcerias globais, tanto ao nível de clubes como ligas profissionais, a abordagem começou a ser mais completa e complexa e a centrar-se mais nos fatores que desempenham um papel importante no sucesso de uma equipa e, consequentemente, nas características que influenciam o sucesso e/ou valor de um jogador individual.

A maioria das abordagens define o sucesso de um jogador através do seu valor de mercado, onde estuda um conjunto de atributos e estatísticas que influenciam positivamente ou negativamente o sucesso de um jogador, tentando assim averiguar a relação entre performance e valor de mercado dos jogadores He et al., (2015)

O valor de mercado de um jogador pode então ser definido como sendo uma estimativa do valor que um clube estaria disposto a pagar para contratar o jogador,

independentemente do valor que é realmente pago Herm et al., (2014), sendo que esse montante pago pode ser totalmente diferente do valor de mercado real de um jogador.

Durante a história do futebol o valor de mercado de um jogador tem sido definido por jornalistas e os chamados especialistas, sem uma matriz mãe por onde se possam seguir, levando a que os valores apresentados fossem altamente suscetíveis ao contexto onde se inseriam os jornalistas e jogadores, e às suas próprias visões do desporto, ao invés de uma comparação sustentada e metodológica. Um exemplo claro que dura até aos dias de hoje é o facto de o jogador inglês (nação berço do futebol europeu e onde se encontra a liga mais histórica e com mais argumentos financeiros do mundo), ou um jogador que joga em Inglaterra, ser em média mais caro que outros jogadores de outras ligas, o que é bem demonstrado no relatório do Centro Internacional de Estudos Desportivos (CIES), divulgado no passado dia 08/06/2020, onde no top 20 se encontram 6 jogadores ingleses (nacionalidade mais representada) e 11 jogadores da liga Inglesa.

No entanto, com o contínuo crescimento de interesse pelo futebol, e com o desenvolvimento tecnológico nas últimas décadas, tem vindo a surgir cada vez mais sites, blogs e outros meios com o objetivo de definirem qual o valor de mercado real dos jogadores, sendo o mais amplamente utilizado atualmente como referência o site de estimativa “crowd based” Transfermarket.

Este site funciona através de uma comunidade de utilizadores que, ao se inscreverem, ganham acesso a discussões sobre avaliações disponíveis e podem apresentar a sua própria estimativa para determinados valores de mercado dos jogadores existentes, sendo que o valor final acaba por ser o conjunto destas avaliações individuais (daí o termo “crowd based”).

De acordo com Herm et al., (2014) o valor final não resulta de uma média de todas as estimativas apresentadas mas sim através de um princípio, a que chama “judge principle”, que se refere a um determinado membro, com uma posição hierárquica superior no processo, que vai fazer uma avaliação das estimativas individuais apresentadas e faz uma depois uma ponderação superior para o resultado final das estimativas que considere mais corretas.

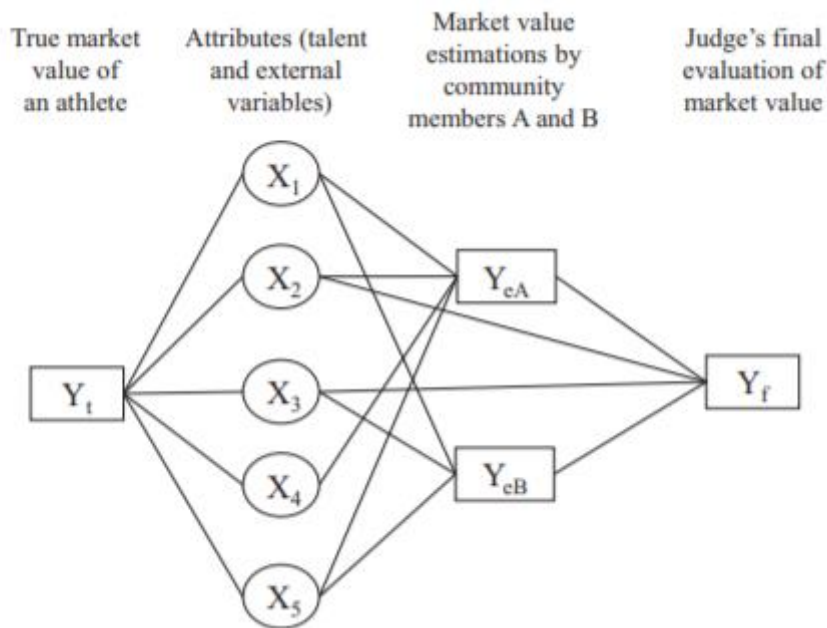


Figura 2 - Modelo do Transfermarkt segundo Herm, Callsen-Bracker e Kreis (2014)

Atualmente já assistimos no futebol a, pelo menos, uma base de comparação (maioritariamente o site Transfermarkt), o que permite obter valores mais realistas possibilitando uma melhor comparação entre jogadores de diferentes ligas, sendo até usado atualmente como base não só para negociações de transferências entre clubes e de renovações salariais entre clubes e jogadores, mas também em artigos nos jornais desportivos mais conceituados do mundo e para investigações científicas relacionadas com as remunerações dos jogadores e os fatores explicativos dos respetivos valores de mercado, como por exemplo Bryson, Frick, e Simmons (2012); Herm, Callsen-Bracker, e Kreis (2014); Franck e Nüesch (2012).

2.3 Indicadores explicativos do Valor de Mercado

Nos últimos anos têm-se multiplicado os trabalhos científicos nesta área de forma a encontrar quais são as características que mais afetam o desempenho positivo de um jogador e, por consequência, o seu valor de mercado.

Este aumento do número de trabalhos levou, conseqüentemente, a um aumento e, simultaneamente, diferenciação na justificação de quais os fatores que influenciam o valor de mercado de um jogador, e o peso de cada um. A grande maioria aborda as características físicas do jogador (por exemplo idade e altura) e estatísticas do mesmo (por exemplo minutos jogados e golos), no entanto Franck e Nüesch (2012) na sua

investigação exploraram a influência da popularidade de um jogador no seu valor de mercado através da contabilização do número de citações em jornais e revistas desportivas. Herm, Callsen-Bracker, e Kreis (2014) aprofundam esta análise ao incluírem outras estatísticas que podem refletir a popularidade de um jogador (como o número de pesquisas em motores de pesquisa).

Tendo em conta o número elevado de investigações recentes resolvi tomar como exemplo de fatores explicativos do valor de mercado de um jogador os trabalhos de Müller, Simons, e Weinmann (2017), com uma abordagem mais prática nos fatores a estudar, e Herm, Callsen-Bracker, e Kreis (2014), que explora mais os fatores extra campo, no entanto é importante referir que trabalhos como Majewski (2016), Gulbrandsen e Gulbrandsen (2011) e Tunaru e Viney (2010) também contribuíram muito para análise realizada no capítulo de análise dos dados.

2.3.1 Trabalho de Muller, Simons e Weinmann (2017)

Os autores dividem os fatores explicativos em três: Indicadores de Popularidade, Características Físicas e Desempenho.

De forma a manter a linearidade da regressão linear os autores transformaram as variáveis da Popularidade em logaritmos, pois a distribuição destas variáveis, num teste inicial, estavam enviesadas para a direita.

A variável age foi também transformada para o seu quadrado de forma a manter uma melhor relação linear com o valor de mercado.

Os dados estudados consistem em seis épocas (2009/10 até 2014/15) das 5 principais ligas Europeias (Itália, França, Alemanha, Inglaterra e Espanha), sendo que exclui os guarda-redes sobre a justificação de o seu desempenho ser classificado de maneira muito diferente das restantes posições.

Ao todo foram contabilizadas 10350 observações.

PEDRO MIGUEL PESEIRO CARAPAU
TRABALHO FINAL DE MESTRADO

Variable	Measurement	Mean	Median	St. Dev.	Min.	Max.
<i>Player valuation</i>						
Transfermarkt's market value	EUR	5588,529	3000,000	8208,470	50,000	120,000,000
<i>Player characteristics</i>						
Age	Years	26.51	26.00	4.08	17.00	40.00
Height	Centimeters	181.49	182.00	6.15	161.00	203.00
<i>Player performance</i>						
Minutes played	total p.s.	1612.39	1612.00	884.85	90.00	3420.00
Goals	total p.s.	2.39	1.00	3.85	.00	50.00
Assists	total p.s.	1.64	1.00	2.25	.00	20.00
Passes	total p.g.	29.45	28.48	13.36	1.55	110.03
Successful passes	percent p.g.	.78	.78	.07	.43	1.00
Dribbles	total p.g.	1.21	.90	1.12	.00	9.58
Successful dribbles	percent p.g.	.51	.50	.24	.00	1.00
Aerial duels	total p.g.	2.22	1.79	1.71	.00	15.50
Successful aerial duels	percent p.g.	.47	.48	.18	.00	1.00
Tackles	total p.g.	2.21	2.09	1.21	.00	9.00
Successful tackles	percent p.g.	.71	.72	.14	.00	1.00
Interceptions	total p.g.	1.35	1.25	.92	.00	7.13
Clearances	total p.g.	2.09	1.07	2.35	.00	13.44
Fouls	total p.g.	1.10	1.03	.53	.00	4.27
Yellow cards	total p.s.	3.48	3.00	2.89	.00	18.00
Red cards	total p.s.	.20	.00	.46	.00	3.00
<i>Player popularity</i>						
Wikipedia page views	total p.s.	104,509.30	23,944.00	319,022.80	.00	8786,701.00
Google Trends search index	average index p.s.	13.36	13.21	12.38	.00	91.83
Reddit posts	total p.s.	15.42	2.00	38.79	.00	789.00
YouTube videos	total p.s.	36,075.46	918.50	141,882.30	.00	1000,000.00

Notes: p.s.=per season; p.g.=per game; N = 10,350

Figura 3 - Fatores usados por Müller, Simons e Weinmann (2017)

Os autores testaram quatro modelos de regressão multinível (devido aos resíduos dos modelos não serem independentes) diferentes cuja variável dependente é o logaritmo do valor de mercado:

Modelo	Características	Desempenho	Popularidade	AIC BIC
1	N/A	N/A	N/A	17416,2 17481,4
2	X	N/A	N/A	14983,2 15070,1
3	X	X	N/A	101722 10374,9
4	X	X	X	10035,9 10267,8

Tabela 1 - Quatro Modelos desenvolvidos em Müller, Simons e Weinmann (2017)

Neste modelo 4 (o mais preciso estatisticamente) apenas uma variável do fator Popularidade não é válida 5% (valor $p > 0,05$) – Log Google Trends search index.

Do fator Características só a variável Age² é válida a 5%, já no fator Desempenho (com as variáveis mais tradicionais) existe algumas variáveis não válidas a 5%, nomeadamente: Successful dribbles, Successful aerial duels, Successful tackles, Interceptions, Clearances, Fouls e Red cards. Podemos observar que a maioria destas variáveis correspondem à área defensiva do jogo, o que pode ser uma das razões para que o valor de mercado de jogadores ofensivos seja tendencialmente superior ao de jogadores em posições defensivas.

2.3.2 Herm, Callsen-Bracker, e Kreis (2014)

Os autores dividem os fatores explicativos em dois: Variáveis Talento e Variáveis Externas (segundo os autores refletem a popularidade dos jogadores junto dos stakeholders do futebol).

Os dados utilizados dizem respeito aos jogadores que fizeram pelo menos um jogo na primeira metade da época 2011/12 da Bundesliga (liga principal alemã), correspondendo a 338 observações (número mais equiparável ao usado neste trabalho).

Player attributes and descriptives.				
Attributes (indicators for market value or income, e.g., in (source))	Description and operationalization	Talent indicator/ external variable	M (SD)	Number of posts
X ₁ : Age (Franck & Nüesch, 2012)	Age (years)/absolute deviation from age with maximum value (24,16)	Talent	25.30 (3.96)/3.32 (2.45)	>500
X ₂ : Precision (Franck & Nüesch, 2012)	% Successful passes/corrected by mean at position ^a	Talent	72.75 (23.03)/0.00 (5.63)	>500
X ₃ : Scoring (Franck & Nüesch, 2012)	Number of goals and assists/corrected by mean at position ^a	Talent	1.97 (3.03)/0.00 (2.78)	>500
X ₄ : Assertion (Franck & Nüesch, 2012)	Won duels or save to shots ratio of goalkeeper/corrected by mean at position ^a	Talent	51.48 (11.58)/0.00 (7.03)	ca. 200
X ₅ : Flexibility (Bryson et al., 2013)	Ability to play predominantly with one foot (0), or equally good with both feet (1)	Talent	0.24 (0.43)	ca. 50
X ₆ : Club management (Treadway et al., 2012)	Financial success/recruitment effectiveness operationalized by the sum of market values of all teammates (€)	External	88 960 661.76 (61 474 927.19)	>500
X ₇ : Team coach (Franck & Nüesch, 2012)	Number of league games played (Bundesliga) ^a	External	11.16 (4.87)	ca. 250
X ₈ : Sports agent (Mason & Slack, 2001)	Average market value of all other players in the portfolio of his sport agent (€)	External	1 780 392.82 (1 500 542.02)	ca. 150
X ₉ : Experts in media (Bryson et al., 2013)	Average grades (1–6) from experts of three different sports portals ^a	External	3.70 (0.52)	ca. 350
X ₁₀ : Public attention (García-del-Barrio & Pujol, 2007)	Total number of links reported by the Internet search engine Google	External	536 265.59 (953 119.50)	ca. 400
Y _i : Market value (Franck & Nüesch, 2012)	Market value (€), defined by the online community ^b	Depend. variable	4 447 205.88/(5 990 666.25)	All

Data sources: Y_i: X₁; X₅; X₆; X₇; X₉: www.transfermarkt.de; X₂; X₃; X₄: www.bundesliga.de; X₈: www.kicker.de; www.sportal.de; www.sportbild.de; X₁₀: www.google.de.

^a In winter break of season 2011/12.

^b At the end of season in July 2012.

Figura 4 - Fatores usados por Herm, Callsen-Bracker, e Kreis (2014)

Neste trabalho, os autores atribuíram pesos diferentes às variáveis talento (como os golos ou duelos ganhos) de acordo com a posição de um jogador, de forma a obter um resultado mais representativo da realidade, seguindo o exemplo desenvolvido por Tavana et al. (2013).

Os autores testaram então dois modelos:

Impact of talent and external attributes on market values.

Dependent variable	Market value according to transfermarkt.de					
	Model 1 (OLS, White robust s.e.)		Model 2 (OLS, White robust s.e.)		Model 2 (95% quantile reg.)	
Model and estimation	Coefficient	t	Coefficient	t	Coefficient	t
<i>N</i> = 338						
<i>Talent</i>						
Age	-539 439.66***	-5.96	-518 397.09***	-5.79	-523 858.46***	-3.94
Precision	215 718.34***	4.61	42 744.08	1.40	14 745.82	0.41
Scoring	1 242 983.01***	6.88	356 796.37**	2.25	397 328.82**	2.40
Assertion	71 288.01**	2.10	14 068.05	0.51	39 595.77	1.13
Flexibility	1 425 951.13*	1.95	637 542.95	1.41	553 709.52	0.79
<i>External attributes</i>						
Club management			0.03***	5.19	0.04***	3.65
Team coach			169 514.16***	4.03	42 528.77	0.54
Sports agent			0.17	0.79	0.08	0.20
Experts in media			-1 166 920.06***	-2.41	-2 298 657.88***	-3.16
Public attention			2.27***	3.69	3.93***	6.04
Intercept	5 897 393.31***	14.24	3 975 657.77**	1.81	11 397 816.18***	2.84
R ² /pseudo-R ²	0.42		0.70		0.65	

* *p* < 0.1
 ** *p* < 0.05.
 *** *p* < 0.01.

Figura 5 - Modelos usados por Herm, Callsen-Bracker, e Kreis (2014)

Ao observarmos os resultados do Modelo 2 podemos concluir que os as Variáveis Externas são também todas elas estatisticamente significantes a pelo menos 5% (valor $p \leq 0,05$), no entanto, e em comparação com o Modelo 1, três variáveis do fator talento deixam de ser significantes a 10% (valor $p > 0,1$). Este resultado pode ser uma consequência do número reduzido de observações, no entanto é importante salientar que o R² aumentou de 0,42 para 0,70 e que são as variáveis mais tradicionais (como a percentagem de passes) a deixarem de ser significativas e não as variáveis externas, demonstrando assim que fatores como a popularidade, merchandising e publicidade associada a um jogador são fatores a ter em conta na altura da determinação do seu valor de mercado.

É importante também salientar que quando estudado as chamadas superestrelas do jogo (Modelo 2 aplicado ao percentil 95%) embora as variáveis talento se mantenham relativamente constantes, as variáveis Externas, sobretudo “Experts in Media” e “Public

Attention” (a variável “Team Coach” deixou de ser estatisticamente significativa, que pode ser um indicador de que o valor de mercado dos melhores jogadores do mundo não é tão influenciados pelas decisões dos seus treinadores) ganham um peso extra, esta análise é também ela feita pelos autores, de notar no entanto que o R^2 diminuiu de 0,70 para 0,65.

2.4 Conclusão das variáveis explicativas do valor de mercado

Podemos concluir então que não são apenas as variáveis tradicionais que influenciam o valor de mercado de um jogador, com a crescente globalização e acesso à informação, a atenção que um jogador de futebol, e toda a polémica e/ou imagem associada, impacta cava vez mais os valores de mercado no futebol moderno, sendo que esta tendência será a de aumentar cada vez mais, como se pode ver pelo valor, cada vez maior, pago por jovens jogadores a quem é reconhecido potencial através de vídeos nas redes sociais e de vídeo jogos com impacto mundial, como a sagas FIFA e Football Manager.

3. Metodologia na recolha de dados

Com o constante crescimento que o “data science” tem tido ao longo dos últimos, veio também um aumento do número, e complexidade, de projetos que envolvam o mesmo e “machine learning” Aparicio et. al, (2019). Este facto torna cada vez mais essencial a implementação de uma metodologia de processo de trabalho (com todas as fases do mesmo associadas) que ajude a criar valor acrescentado no produto final Costa & Aparicio (2020). Com isto, analisou-se algumas abordagens existentes, tais como KDD (Knowledge discovery in database) e SEMMA (Sampling, Exploring, Modifying, Modelling and Assessing), no entanto a escolha recaiu sobre a abordagem CRISP-DM (Cross Industry Standard Process for Data Mining) por ser aquela que se encaixava melhor no problema em estudo e na visão de como resolvê-lo. Esta abordagem foi criada em 1996 por quatro líderes do crescente mercado de “data mining”, nomeadamente: Daimler-Benz, IntegralSolutions Ltd. (ISL), NCR, e OHRA, e é uma metodologia completa de recolha de dados e modelo de processos que fornece aos autores um mapa compreensivo completo para desenvolverem um projeto de recolha de dados, Costa & Aparicio (2020). Foi então seguida a metodologia CRISP-DM na recolha e tratamento de dados, cujas fases são explicadas em Wirth e Hipp (2000).

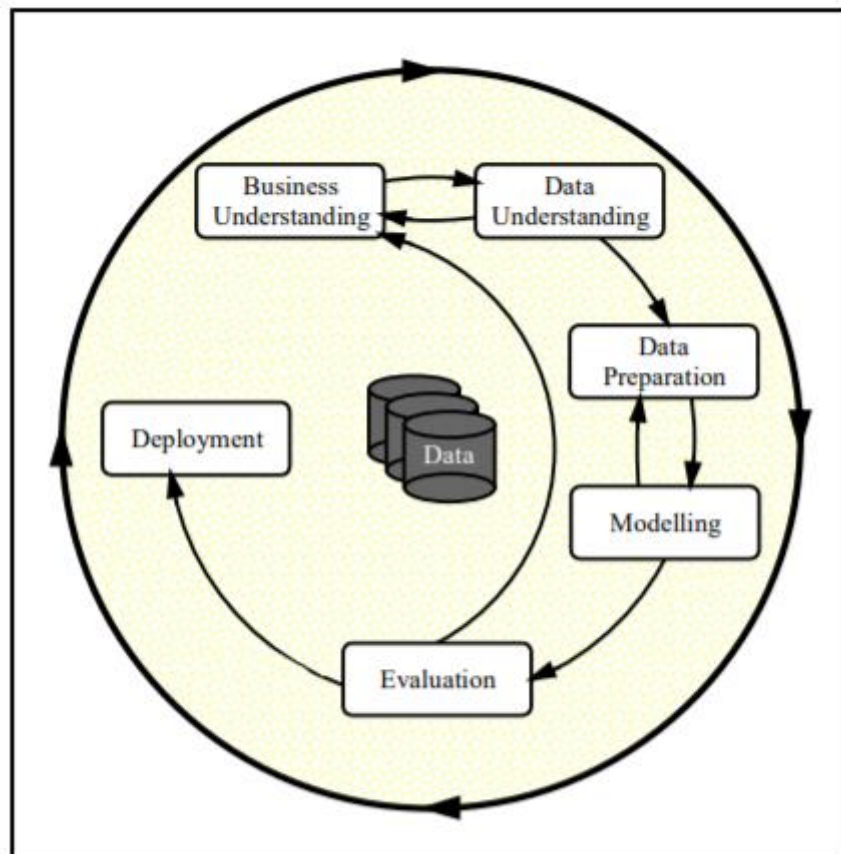


Figura 6 - Processo do CRISP-DM, Wirth e Hipp (2000).

Esta metodologia divide o processo em seis fases: “Business Understanding”, que consiste em definir o problema, os objetos em estudos e os dados a utilizar para estudá-los e chegar a conclusões; “Data Understanding”, que consiste na recolha inicial de dados, a sua análise e identificação de possíveis problemas; “Data Preparation”, que consiste na limpeza e transformação dos dados de forma a que seja possível utilizá-los na construção de modelos; “Modelling”, onde é realizado e testado os modelos; “Evaluation”, que consiste em rever os passos realizados na construção do modelo de forma a assegurar que os objetivos do estudo, do problema, são atingidos com os resultados obtidos, e que estes permitem uma análise correta do mesmo; e “Deployment”, que sendo a última fase consiste na organização dos resultados obtidos e a elaboração de um relatório final (conclusão). Estes passos estão todos definidos em Wirth e Hipp (2000).

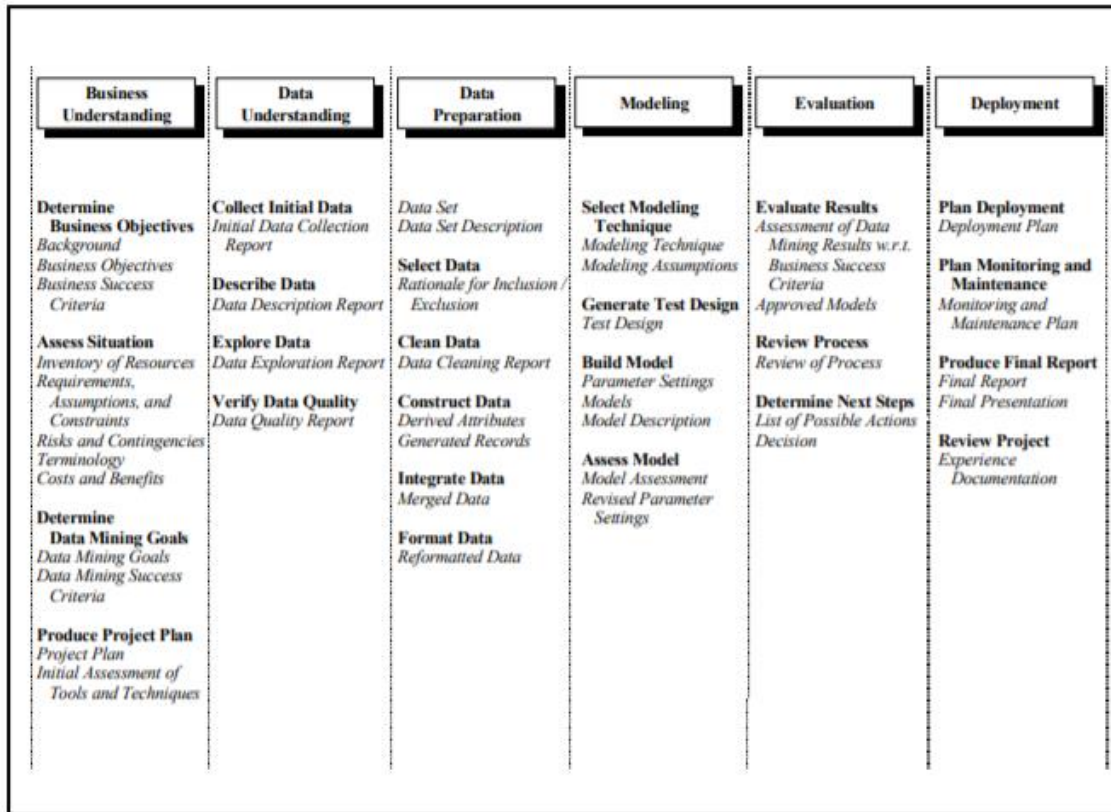


Figura 7 - Figura explicativa das várias fases do CRISP-DM, Wirth e Hipp (2000).

Sendo o objetivo deste estudo analisar qual foi a melhor academia, de entre os três grandes, nas sete épocas anteriores à finita (2012-13 a 2018-19) (“Business Understanding”), foi recolhido dados referentes aos jogadores formados por cada uma das academias e que estejam inscritos no plantel sénior da respetiva equipa, na época em questão, incluindo os que já estavam no plantel na época 2012-13 que tenham estado ininterruptamente ligados contratualmente ao clube. Os dados recolhidos referem-se apenas aos jogadores que fizeram pelo menos 3 anos de formação no clube, entre os 15 e os 21 anos (critério da FIFA para um jogador ser declarado como formado localmente), e que tenham chegado ao plantel principal do clube em questão (existem vários jogadores que não conseguem dar o salto para o plantel sénior e acabam por ir para clubes de divisões inferiores onde não existe dados disponíveis para proceder a uma análise detalhada).

É importante referir que a fonte dos dados foi o site “Transfermarket”, mencionado anteriormente, e a sua recolha foi feita numa primeira fase automaticamente, recorrendo ao Web Scrapping em Python e Power BI, e numa segunda manualmente devido aos

constrangimentos que os sites (neste caso o “Transfermarkt”) impõem de forma a defenderem-se de “Scrapers” cujos objetivos são, muitas vezes, ilegais aos olhos das políticas de proteção de dados.

Foram então recolhidos dados sobre as épocas de cada jogador a partir da sua primeira na respetiva equipa sénior entre a época 2012-13 (incluindo os que já lá estavam ininterruptamente desde a sua formação [excluindo períodos de empréstimo]) e a época 2018-19, ou seja, caso um jogador que chegue à equipa principal da respetiva equipa durante a época 2012-13 teve os seus dados recolhidos até à última época em estudo, 2018-19, mesmo que tenha saído do clube em questão. Os dados recolhidos referem-se apenas a jogos nas equipas profissionais ou nas respetivas equipas B dos três grandes, pois estão inseridas num contexto de alta competição a nível sénior (2ª divisão portuguesa). Todo este processo é a fase “Data Understanding”.

Os dados, divididos por jogador e respetiva época, em questão são: nome, posição, pé principal do jogador, nacionalidade e se foi transferido/emprestado no decorrer ou final da época (dados qualitativos), juntando-se os dados quantitativos minutos jogados na época, o coeficiente da respetiva liga (0,5 a 2), idade, altura, valor de mercado e o valor recebido caso tenha ocorrido uma transferência ou empréstimo. É preciso ressaltar que não foi incluindo golos e assistências de forma a não enviesar os dados e os resultados em favor de jogadores em posições mais avançadas, já que não foi possível encontrar dados essenciais para analisar outras posições, nomeadamente percentagem de passes completados, desarmes bem-sucedidos, duelos ganhos, entre outros. Esta falta de dados é ainda maior quando observamos a segunda divisão portuguesa, que constitui uma grande percentagem das nossas observações já que é muito recorrente, desde a implementação das equipas B na Segunda Divisão Nacional portuguesa em 2012-13, os clubes recorrerem às mesmas de forma a permitir que os jogadores jovens ganhem experiência em alta competição. Este processo é designado por “Data Preparation”.

Depois da “data preparation” foi realizado o “modelling”. Nesta fase, foi realizado uma regressão linear de forma a averiguar se certas variáveis usadas na comparação são estatisticamente significativas na explicação do valor de mercado dos jogadores (irá ser usada como variável dependente pois é a valorização final de todas as características e estatísticas associadas de um jogador).

De seguida é analisado e comparado um conjunto de fatores entre as três academias de forma a podermos perceber qual a que teve mais sucesso neste período e se a tendência tem sido o melhorar, ou piorar, de resultados. Este processo inclui o que no CRIP-DM é designado por “Evaluation” e “Deployment”.

4. Análise da formação dos três grandes em Portugal

4.1 Introdução

O futebol em Portugal sempre teve uma grande importância junto da população portuguesa, e sendo um país com uma população reduzida existe uma grande assimetria no número de apoiantes dos clubes portugueses, com os denominados três grandes (Futebol Clube do Porto, Sporting Clube de Portugal e Sport Lisboa e Benfica) a recolherem a maioria dos apoiantes em Portugal. Segundo Observador (2019), numa sondagem feita pela Intercampus, 46% dos adeptos portugueses apoiam o Sport Lisboa e Benfica, 24,7% o Futebol Clube do Porto e 23,8% o Sporting Clube de Portugal, o que perfaz um total de 94,5% dos adeptos portugueses de futebol. Este forte apoio a estes clubes deve-se em grande parte ao facto de serem os clubes com mais títulos em Portugal, dado que nas últimas 84 edições da primeira liga portuguesa, apenas por duas vezes o vencedor foi um clube para além dos mencionados.

Esse histórico, associado à realidade do futebol português cuja sustentabilidade financeira depende ainda muito da venda de ativos, especialmente jogadores, levando a que sejamos um país com muita tradição em formar grandes atletas na história do futebol, faz com que os três grandes tenham um papel de preponderância na formação de novos jogadores em Portugal, com os maiores jogadores da história portuguesa a serem formados nas academias de um destes três clubes. Esta hegemonia leva a que o título informal de qual a melhor academia de futebol dos grandes seja constantemente discutida entre os adeptos portugueses.

O objetivo desta investigação prende-se então em determinar qual foi, efetivamente, a melhor academia, de entre as três em estudo.

4.2 Análise dos Dados

De forma a fazer uma ligação entre os trabalhos estudados e as variáveis utilizadas na comparação de academias, foi realizado uma regressão linear simples, cujo objetivo é

verificar se as mesmas são válidas para explicar o valor de mercado dos jogadores (corroborando os resultados nos trabalhos estudados no capítulo anterior). Os dados estudados foram todos recolhidos do site anteriormente mencionado - Transfermarket.pt.

Infelizmente os dados públicos disponíveis para os campeonatos portugueses (sobretudo a segunda divisão portuguesa) são reduzidos e de forma a não enviesar o valor dos jogadores de posições mais avançadas não foi incluído as variáveis golos e assistências (pois iria influenciar demasiado quando comparados com posições mais recuadas que com a falta de dados disponíveis [como por exemplo a percentagem de duelos ganhos pelos defesas] estariam em desvantagem na sua análise).

Foram assim apenas incluídas variáveis, na realização da regressão linear, que sejam o mais transversal possível a todas as posições. Essas variáveis estão definidas na tabela seguinte:

Número de Jogadores	Jogadores formados em determinada Época
Épocas desportivas	De forma a tentar perceber um padrão nas academias ao longo destas épocas
Posição	GK DEF MED ATA
Minutos Jogados***	Contabilizados os minutos jogados por equipas principais e equipas B em campeonatos profissionais (como a segunda divisão)
Coeficiente da Liga	De forma a ter em conta a diferente dificuldade das diferentes ligas – Segue a fórmula atual da UEFA: 2 para as 5 principais ligas europeias, 1,5 para as ligas dos países entre a 6ª e 21ª posição a 6ª e 21ª posição e 1 para as restantes ligas.
Valor de Mercado***	Irá ser a variável dependente

Ser Transferido	Se o ato de se transferir aumentou em alguma medida o valor de mercado de um jogador.
Idade ^{2***}	Seguindo o exemplo de Müller, Simons, e Weinmann (2017)
Altura*	-

Tabela 2 - Variáveis utilizadas na regressão linear.

* Presente no trabalho Müller, Simons, e Weinmann (2017)

** Presente no trabalho Herm, Callsen-Bracker, e Kreis (2014)

A variável Posição foi introduzida através de três variáveis dummies, isto é, $\delta_1 = ATAd$, $\delta_2 = DEFd$, $\delta_3 = MEDd$, em que assumem o valor 1, seguindo o exemplo de Bryson, Frick, e Simmons (2012) mas tomando como referência a posição GK.

A variável Ser Transferido é também uma variável dummy que assume valor 1 se o jogador foi transferido nessa época.

Em primeiro lugar foi realizado um teste de correlação, com a seguinte tabela como resultado:

	Época	Idade ²	Altura	Quoeficiente	Minutos	ATAd	DEFd	MEDd	Valor de Mercado	Ser Transferido
Época	1									
Idade ²	0,300	1								
Altura	0,009	-0,009	1							
Quoeficiente	0,014	0,252	-0,008	1						
Minutos	-0,127	0,081	0,064	0,2295	1					
ATAd	0,050	-0,132	-0,288	-0,0172	-0,176	1				
DEFd	-0,056	0,044	0,367	0,0910	0,105	-0,407	1			
MEDd	-0,013	0,065	-0,305	0,0495	0,051	-0,431	-0,473	1		
Valor de M	0,168	0,092	0,012	0,4491	0,421	-0,042	-0,058	0,074	1	
Ser Transfi	0,076	0,088	-0,006	0,0316	-0,078	0,014	0,032	-0,036	0,026	1

Figura 8 - Tabela de Correlação das Variáveis

De acordo com Cohen (1988), valores entre 0 e |0,29| podem ser considerados como pequenos, |0,30| a |0,49| podem ser considerados como médios (todos os valores acima destacados a vermelho encontram-se neste intervalo) e |0,5| a |1| como elevados.

Se seguirmos esta análise, ao observarmos a tabela conclui-se que não existe nenhum valor elevado na correlação das variáveis logo assume-se que as variáveis não são

correlacionadas. De seguida realizou-se uma regressão linear de forma a averiguar a relação das variáveis explicativas com a variável dependente (Valor de Mercado):

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \delta_1 x_{6i} + \delta_2 x_{7i} + \delta_3 x_{8i} + \delta_4 x_{9i} + \varepsilon_i$$

	<i>Coefficientes</i>	<i>Erro-padrão</i>	<i>Stat t</i>	<i>valor P</i>
Interceptar	-2800,249	463,205	-6,045	3,07E-09
Época	1,387	0,230	6,027	3,42E-09
Idade^2	-0,010	0,003	-2,912	0,004
Altura	-1,798	7,159	-0,251	0,802
Coefficiente	10,706	1,026	10,436	4,84E-23
Minutos	0,004	0,0004	9,669	2,88E-20
ATAd	-3,965	1,788	-2,218	0,027
DEFd	-5,807	1,585	-3,663	0,0003
MEDd	-3,309	1,729	-1,913	0,056
Ser Transferido	0,890	0,838	1,062	0,289

Figura 9 - Output da Regressão Linear

Observando-se os resultados podemos concluir que apenas as variáveis *Altura* e *Ser Transferido* não são estatisticamente significativas a 5% para explicarem o valor de mercado de um jogador.

É importante também voltar a referir que o R^2 obtido foi 0,384, sendo uma consequência do modelo simplista adotado e do pouco número de observações, mas que ainda assim apresenta um resultado semelhante ao de Herm, Callsen-Bracker, e Kreis (2014), 0,42, no seu primeiro modelo, também ele o mais simplista e com um conjunto de variáveis em estudo semelhante.

O valor positivo do coeficiente época é um sinal de que, como foi mencionado anteriormente, tem havido uma crescente valorização dos jogadores de futebol nos últimos anos (pode-se dizer que esta variável pode representar a evolução dos “bastidores” do futebol, diga-se a título de exemplo a crescente exposição mediática dos jogadores) o que se traduz diretamente no crescimento dos valores de mercado e, consequentemente, os valores de transferência envolvidos.

A variável *Idade*² é estatisticamente significativa a 1% (valor $p < 0,01$). Sendo uma variável quadrática com coeficiente negativo apresenta uma relação côncava, ou seja, numa primeira instância o aumento da idade leva a um aumento do valor de mercado, no entanto vai diminuindo, chegando a um ponto em que o aumento de idade leva a uma

diminuição do valor de mercado. Este resultado é o esperado pois o “normal” é os jogadores continuarem a evoluir e a melhorar até atingirem o seu pico (varia muito entre jogadores e respetivas funções em campo), sendo que a partir desse momento a sua qualidade vai diminuindo à medida que as suas capacidades físicas diminuem com a idade.

A variável altura apresenta um valor p muito alto ($>0,8$) levando à conclusão que, nesta amostra, não tem um impacto estatístico significativo no valor de mercado de um jogador. Tendo em conta que esta variável também não é válida nos modelos de (Müller et al., 2017), é um resultado esperado, no entanto é preciso ressaltar que nesta amostra os dados relativos à altura de um jogador são constantes ao longo das épocas (pois apenas está disponível o mais recente) o que pode afetar a validade deste resultado.

A variável coeficiente, como seria de esperar, é claramente significativa, a 1%, para explicar a variável dependente, sendo que é lógico que um bom desempenho numa liga de topo valorize mais um jogador do que o mesmo nível de desempenho numa liga de qualidade inferior.

A variável minutos é também ela estatisticamente significativa a 1%, o que é também lógico pois quanto mais tempo um jogador jogue mais oportunidades tem para se valorizar. É importante referir que a razão para o coeficiente ser tão reduzido se deve ao facto de a unidade de medida utilizada ser os minutos em campo, e tendo em conta que a média de minutos, ao longo de todas as épocas em estudo, dos jogadores em campo nesta amostra ser de aproximadamente 1924 minutos (por época) é um valor compreensível.

As variáveis dummies ATAd, MEDd e DEFd são todas estatisticamente significativas a, pelo menos, 10% (valores p iguais a: 0,027, 0,057 e 0,0003 respetivamente). Embora seja compreensível que a posição de um jogador afete diretamente, e de maneira diferente, o seu valor de mercado este resultado (coeficientes negativos em relação à posição de comparação, GK) não vai de encontro aos resultados apresentados, por exemplo, em Frick (2011), onde é demonstrado que os jogadores de posição atacantes (ATA) são mais valorizados, tanto no seu valor de mercado como no salário, comparativamente aos guarda-redes, seguindo-se os médios (MED) e os defesas (DEF). Esta diferença pode ser resultado da amostra reduzida estudada, $n = 473$ comparativamente ao $n = 6147$ estudado pelo autor, fazendo com que o reduzido número de Guarda-Redes aqui estudados, onde

existem alguns com valores de mercado elevado como Rui Patrício e Ederson Moraes, possam enviesar este resultado.

Em suma, os resultados obtidos corroboram os dos autores mencionados anteriormente, dando assim mais credibilidade para usar estas variáveis quantitativas, juntando-se o valor de transferências recebido, as presenças nas convocatórias da Seleção Nacional e os títulos ganhos pelos clubes, no próximo ponto - a comparação entre as três academias.

4.3 Análise das Academias

4.3.1 Número de jogadores e respetivo tempo de jogo

Em primeiro lugar vamos analisar o número de jogadores que estiveram nos plantéis principais dos clubes, provenientes das respetivas academias, ao longo das sete épocas, 2012-13 a 2018-19:

Clube	Nº Jogadores Nº Épocas	Min. Jogados nas 7 épocas	Min. por Jogador/Época	Min. Jogados pelo Clube	% Min. Jogados pelo Clube
SLB	35 153	312587	2043,1	46050	15%
SCP	49 214	396740	1853,9	107316	27%
FCP	27 106	200799	1894,3	22068	11%

Tabela 3 - Síntese dos minutos jogados pelos jogadores, dados com origem no “Transfermarket”.

Ao analisar a tabela acima podemos observar facilmente que o SCP foi o clube que formou mais jogadores durante este período, no entanto podemos também observar que os jogadores formados pelo SLB jogam, em média, mais tempo do que os jogadores dos restantes clubes, o que é um bom indicador de uma qualidade relativa por jogador superior ao dos rivais, ou de um maior sucesso de carreira dos jogadores, já que se assumem mais vezes como titulares nos respetivos clubes.

Outro ponto interessante é o facto de os jogadores do SCP jogarem consideravelmente mais tempo no próprio clube comparativamente aos rivais (apresenta uma percentagem superior em 12 pontos percentuais comparativamente ao segundo classificado). Uma possível explicação para este facto é a débil condição financeira que o clube tem vivido nos últimos anos, levando-o a ter de recorrer mais vezes aos jogadores da formação de

forma a conseguir poupar em possíveis custos de transferências. Outro possível fator é o peso histórico que a formação sempre teve no SCP, sendo algo que sempre foi reconhecido a esta clube, sobretudo no início do milénio com o sucesso de vários jogadores formados, entre eles Cristiano Ronaldo e Ricardo Quaresma, levando a uma pressão extra por parte dos seus adeptos para integrarem nos plantéis principais jogadores formados na sua academia.

É então necessário também analisar onde, e em que contexto, é que os jogadores de cada uma das academias jogaram ao longo das diferentes épocas em estudo, nomeadamente o coeficiente da respetiva liga onde esses minutos foram jogados (variável estatisticamente significativa como foi demonstrado no ponto anterior).

	SLB	SCP	FCP
Média de Coeficiente $x < 1,5^*$	74 Épocas 31 Jogadores Dif. 142098 Minutos	95 Épocas 38 Jogadores Dif. 146417 Minutos	52 Épocas 24 Jogadores Dif. 98002 Minutos
Média de Coeficiente $1,5 < x < 2^{**}$	46 Épocas 21 Jogadores Dif. 97255 Minutos	82 Épocas 36 Jogadores Dif. 171881 Minutos	46 Épocas 17 Jogadores Dif. 87103 Minutos
Média de Coeficiente $x \geq 2^{***}$	32 Épocas 12 Jogadores Dif. 73234 Minutos	37 Épocas 15 Jogadores Dif. 78442 Minutos	8 Épocas 6 Jogadores Dif. 15694 Minutos
Média Minutos/Época *	1920	1541	1885
Média Minutos/Época **	2114	2096	1894
Média Minutos /Época ***	2289	2120	1962

Tabela 4 - Dados relativos às épocas praticadas pelos jogadores em ligas de diferentes coeficientes, dados com origem no “Transfermarket”.

Ao observarmos a tabela acima, podemos concluir que volta a haver uma clara vantagem, em termos absolutos, do SCP (que é explicado pelo mesmo apresentar um maior número de jogadores formados neste período) ocupando o primeiro lugar em cada um dos escalões seguido por SLB e FCP, à exceção do segundo escalão onde tanto o FCP e o SLB apresentam 46 Épocas jogadas.

Em termos relativos (os valores das médias) o SLB volta a ganhar vantagem aos seus rivais sendo o clube onde os jogadores jogam mais tempo, em média, em cada um dos três escalões de coeficientes, seguido pelo FCP e SCP no primeiro e pelo SCP e FCP nos restantes.

Outro dado interessante é a percentagem de Épocas jogadas por cada um dos jogadores nos diferentes escalões:

Clube	Coe. < 1,5	1,5 < Coe. < 2	Coe. >= 2	Coe. = 1,5
SLB	48%	30%	22%	27%
SCP	44%	38%	17%	35%
FCP	49%	43%	8%	42%

Tabela 5 - Dados relativos à percentagem de Épocas jogadas nos respetivos escalões de Coeficientes.

Ao analisarmos a tabela acima podemos retirar que no primeiro escalão os valores são muito semelhantes, resultado de uma política transversal dos três clubes em recorrer às equipas B durante os anos iniciais das carreiras dos seus jogadores como forma de lhes dar experiência num ambiente sénior de alta competição.

Com o avançar dos escalões começamos a perceber as principais diferenças, no 2º escalão existe uma vantagem por parte do FCP, sobretudo em relação ao SLB, com 43% dos seus jogadores a realizarem épocas neste escalão, no entanto este escalão engloba todos os coeficientes entre 1,5, inclusive, e 2 exclusive (coeficiente das cinco maior ligas), ou seja engloba os jogadores que fizeram uma época completa no coeficiente de 1,5 (como a Primeira Divisão Portuguesa) e os que a meio da época se transferiram, por empréstimo ou transferência, para uma divisão de coeficiente superior. Por isto criou-se uma coluna extra que mostra a percentagem de jogadores cujo coeficiente de uma época é exatamente 1,5 (significando que ficaram toda a época no clube ou que se transferiram para uma divisão de coeficiente equiparável), analisando estes valores concluímos que, no caso do

FCP, 42 pontos percentuais, dos 43%, fizeram épocas apenas com um coeficiente de 1,5, enquanto que no SCP corresponde a 35 pontos percentuais, dos 38%, e no SLB a 27 pontos percentuais, dos 30%. Quer isto dizer que embora o FCP tenha uma percentagem superior neste escalão essa superioridade não corresponde necessariamente a um resultado totalmente positivo, por um lado é sinal que existe uma maior retenção de jogadores da academia no plantel principal, por outro é também um indicativo que os jogadores do FCP têm mais dificuldades em dar o salto competitivo para ligas de qualidade superior (neste caso representadas pelo coeficiente 2), sendo este ponto reforçado analisando os valores percentuais das épocas realizadas pelos jogadores no terceiro escalão (que engloba apenas épocas completas em ligas com coeficiente 2). Neste escalão, que pode ser considerado como o destino de excelência na carreira de um jogador, o SLB está em situação de vantagem com 22% das épocas dos seus jogadores, comparativamente aos 17% no SCP e 8% no FCP.

Mais que enaltecer a vantagem do SLB neste ponto é necessário realçar o baixo valor apresentado pelo FCP cujos jogadores tiveram neste período uma clara dificuldade em chegar às ligas mais competitivas do mundo quando comparado com os rivais. Este dado ganha ainda uma maior importância se tivermos em conta a política desportiva do clube, muito baseada no desenvolvimento de jogadores e posterior venda por valores avultados, que sempre caracterizou o FCP neste milénio, significando que embora tenha existido alguma facilidade em conseguir vendas por grandes valores para os principais clubes Europeus a esmagadora maioria das mesmas não foram com jogadores formados no clube.

4.3.2 Valor de Mercado e Transferências

O sucesso de uma academia está ligado invariavelmente ao valor de mercado dos seus jogadores (facto demonstrado ao longo deste trabalho) e aos respetivos valores de transferências que movimentaram.

Foi então analisado o valor de mercado total ao longo de todas as épocas em estudo, assim como o valor médio que representa o valor médio de mercado de uma época de um jogador de cada um dos clubes. Analisou-se também o número de transferências definitivas (excluiu-se os empréstimos) que os jogadores tiveram neste período (incluindo

as transferências ocorridas após terem saído do respetivo clube) e qual o número de transferências, e respetivo valor, que partiu diretamente de cada um dos clubes.

Estes números estão retratados nas tabelas em baixo:

Clube	Valor Total de Mercado	Valor Médio de Mercado
SLB	1341,35	8,82
SCP	1163,45	5,44
FCP	375,15	3,54

Tabela 6 - Valores de Mercado dos Jogadores ao longo das Épocas, em Euros, dados com origem no “Transfermarket”.

Clube	Nº de Transf. ao longo das épocas	Valor Movimentado	Nº de Transf. do Clube	Valor recebido pelo Clube
SLB	48	653,81	26	357,75
SCP	59	317,45	34	223,55
FCP	26	85,26	19	84,96

Tabela 7 - Número de transferências definitivas de cada clube e respetivos valores, em Euros.

Ao analisarmos as tabelas concluímos que, mais uma vez, o SLB volta a assumir a liderança no valor de mercado total e médio (com uma diferença de 177,9 e 3,39 milhões de euros para o segundo classificado [SCP]) e nos valores de transferências movimentados pelos seus jogadores, tanto ao longo das suas carreiras neste período (diferença de 336,36 milhões de euros para o SCP) como diretamente para o clube (diferença de 134,2 milhões para o SCP). Esta diferença ganha maiores contornos quando comparamos com os valores do FCP, que volta a estar atrás dos rivais.

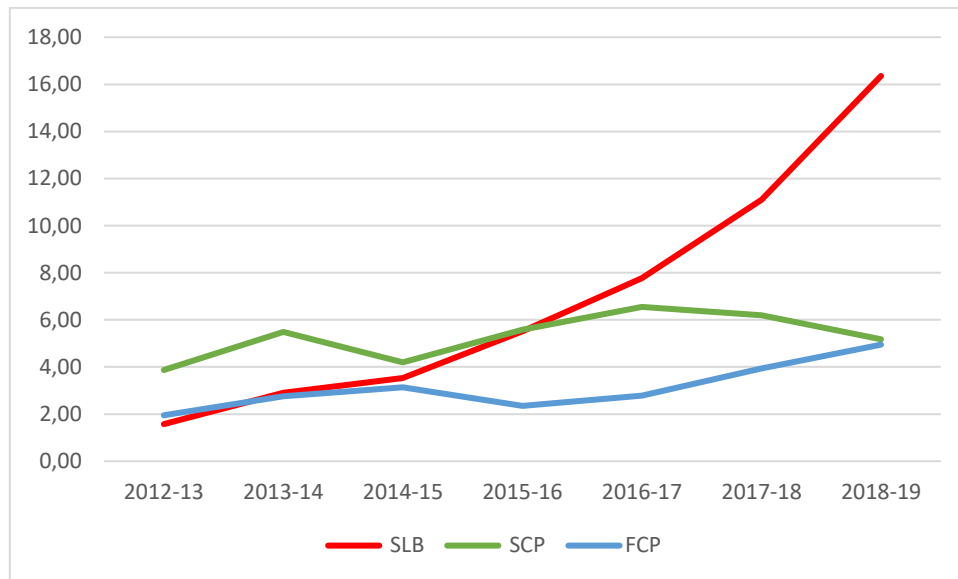


Gráfico 1 - Evolução do valor de mercado médio dos jogadores, de cada um dos três clubes, ao longo das sete épocas.

Analisando também o gráfico acima, podemos ver que o valor de mercado médio dos jogadores do SLB tem tido um crescimento constante ao longo destas épocas, passando de 1,58 milhões em 2012-13 para 16,36 milhões em 2018-19. O SCP e o FCP apresentam valores inferiores, e com uma diferença cada vez maior, sobretudo a partir da época 2015-16.

O SCP, embora tenha começado com o valor mais alto de entre os três, não tem conseguido manter um crescimento constante neste período, variando sempre entre os quatro e os seis milhões sensivelmente.

Quanto ao FCP, apresenta o valor mais baixo dos três em todo o período em estudo, no entanto, e ao contrário do SCP, tem apresentado uma tendência de crescimento nas últimas três épocas que, a continuar, deverá permitir ultrapassar o SCP na época seguinte.

Esta vantagem significativa por parte do SLB no valor de mercado médio dos seus jogadores pode ser um sinal claro, dentre os pontos analisados neste trabalho, da superioridade de rendimento que os seus jogadores tiveram, comparativamente aos rivais, pois engloba todo um conjunto de fatores resultantes das suas performances e postura dentro e fora do campo (analisados no capítulo anterior) que, ao que este gráfico indica, tem afetado positivamente as suas carreiras e, conseqüentemente, valorização.

Para além destes dados, se considerarmos a diferença entre o valor recebido por um jogador e o seu valor de mercado como “lucro” para os clubes observamos que:

- SLB apresenta um lucro de 150,45 milhões de euros nas 26 transferências realizadas,
- SCP apresenta um lucro de 20,425 milhões de euros nas 34 transferências realizadas,
- FCP apresenta um lucro de 13,06 milhões de euros nas 19 transferências realizadas.

Estes dados comprovam, ainda mais, a grande superioridade que o SLB teve, em relação aos rivais, no domínio do mercado e aproveitamento financeiro dos jogadores da sua academia.

4.3.3 Presenças na Seleção Nacional

Uma consequência positiva de uma boa academia é os seus jogadores chegarem ao nível sénior das respetivas seleções nacionais.

A tabela abaixo indica a percentagem de jogadores formados por SLB, SCP e FCP nas convocatórias da Seleção Nacional Portuguesa (pela esmagadora maioria dos jogadores estudados terem nacionalidade portuguesa) para grandes competições (Campeonatos da Europa e Mundiais de Seleções), durante o período em causa, tendo apenas em conta jogadores presentes neste estudo (formados a partir de 2012-13).

É preciso ter em conta que chegar à Seleção Nacional de Futebol Portuguesa é algo que, não tendo um critério geral, pode demorar várias épocas, variando muito de jogador para jogador.

Não foi tido em conta jogadores que, embora tenham sido formados por algum dos clubes, não encaixam no período em estudo, no entanto é preciso realçar a grande vantagem que o SCP tem para os restantes neste ponto, colaborando assim o ponto referido anteriormente relativamente ao peso histórico que a formação sempre teve no SCP.

Foi acrescentado também informação relativa à última convocatória, realizada em outubro de 2020, de forma a ter alguma informação sobre a evolução neste ponto desde o final da época 2018-19:

Clube	Europeu 2012	Europeu 2016	Mundial 2014	Mundial 2018	Convocatória Outubro 20
SLB	1	2	0	3	6
SCP	1	5	2	6	4
FCP	0	0	0	1	3

Tabela 8 - Convocatórias realizadas pela Seleção Nacional de Futebol Portuguesa Sénior, dados da Federação Portuguesa de Futebol (FPF).

Em primeiro lugar é possível observar a excelente evolução do número de jogadores convocados nos últimos anos, passando de sete no Europeu 2016 (quatro anos após o primeiro ano em estudo) para treze na última convocatória.

Quanto à comparação direta entre clubes, o clube que apresentou mais convocados na última grande competição foi o SCP com seis jogadores, no entanto na última convocatória foi o SLB que apresentou o maior número de convocados com também seis jogadores.

Neste ponto o FCP volta a estar abaixo dos rivais, tendência verificada ao longo deste trabalho. Quanto aos restantes clubes, o SCP ganha vantagem por ter, constantemente, mais jogadores formados nas convocatórias finais para grandes competições, assim como, num aspeto mais geral e referido anteriormente, ter mais jogadores fora deste período em estudo a representá-lo.

4.3.4 *Títulos Conquistados e participação de jogadores formados*

Se perguntarmos a qualquer adepto de futebol o que é mais importante para si, a esmagadora maioria vai responder “ganhar títulos”.

Assim vamos, neste último ponto, analisar os títulos conquistados pelos três clubes neste período e ver a percentagem de jogadores formados no plantel em cada uma das épocas e que tenha feito pelo menos um minuto pelo plantel principal:

Época	Supertaça	Primeira Liga	Taça de Portugal	Taça da Liga
2012-13	FCP	FCP	-	-

2013-14	SLB	SLB	SLB	SLB
2014-15	SCP	SLB	SCP	SLB
2015-16	SLB	SLB	-	SLB
2016-17	SLB	SLB	SLB	-
2017-18	FCP	FCP	-	SCP
2018-19	SLB	SLB	SCP	SCP

Tabela 9 - Vencedores das principais quatro competições nacionais desde a época 2012-13.

Época	SLB	SCP	FCP
	Nº jogadores %	Nº jogadores %	Nº jogadores %
2012-13	1 3%	12 30%	2 6%
2013-14	5 13%	10 33%	3 9%
2014-15	3 8%	11 25%	2 6%
2015-16	6 16%	12 33%	5 12%
2016-17	7 16%	12 31%	5 16%
2017-18	4 13%	9 20%	3 8%
2018-19	8 24%	8 20%	4 12%

Tabela 10 - Número de jogadores formados, e respetiva percentagem, no plantel.

Analisando as tabelas acima podemos concluir, em primeiro lugar, a grande superioridade que o SLB apresentou, ao nível de títulos, neste período com catorze títulos em vinte e oito possíveis. No entanto esse não é um critério direto para o sucesso da sua academia, é necessário averiguar a participação de jogadores formados nestas conquistas.

Mais uma vez, em termos de volume, o SCP volta a assumir uma posição de destaque estando em primeiro lugar, termos absolutos, em todas as épocas, sendo que na última está em pé de igualdade com o SLB.

Olhando para a percentagem no plantel principal o SCP assume também o primeiro lugar em todas as épocas (com grande distância para o segundo lugar), à exceção da última, em que o SLB assume o primeiro lugar.

Uma análise conjunta diz-nos que embora o SLB se tenha destacado dos rivais, no que a títulos diz respeito, essa superioridade nem sempre se consolidou tendo como base jogadores formados no clube a participarem ativamente (minutos jogados em competições oficiais). No entanto, a tendência tem sido o de aumentar a percentagem de jogadores ao longo deste período (na última época regista um valor recorde para o clube, neste período, de 24% o que se traduz em oito jogadores).

O caso do SCP é exatamente o oposto, isto é, embora tenha tido constantemente as percentagens mais elevadas de jogadores formados a participarem ativamente nas competições, esse modelo apenas rendeu cinco títulos neste período, sendo nenhum destes o de Campeão Nacional (título nacional mais prestigiado). Talvez por esta escassez de títulos, o seu modelo (diga-se percentagem de jogadores formados a serem membros ativos no plantel principal) tem vindo a mudar ao longo destas épocas, sofrendo uma quebra de onze pontos percentuais (que se traduz em uma diferença de três jogadores) na época 2017-18, comparativamente à anterior. Embora tenha tido esta descida, o seu valor continua elevado (20% na última época) quando comparado com os valores dos rivais.

O FCP, embora tenha ganho dois campeonatos nacionais, contabiliza menos títulos ganhos neste período do que o SCP (quatro para cinco, respetivamente). Aliada a esta escassez de troféus está também uma fraca utilização de jogadores formados, sendo o pior dos rivais em todas as épocas, à exceção da primeira e quinta (onde iguala o valor do SLB), e com um valor máximo de 16% (que se traduz em cinco jogadores). Este valor aumentou nas últimas duas épocas, no entanto mantém-se abaixo dos rivais (quatro jogadores comparativamente aos oito de cada um dos rivais, na última época), e no período em estudo nunca ultrapassou a barreira dos 20% (passando a barreira dos 10% apenas por três ocasiões) e esteve último lugar, em termos de número absoluto de jogadores, nas últimas seis épocas.

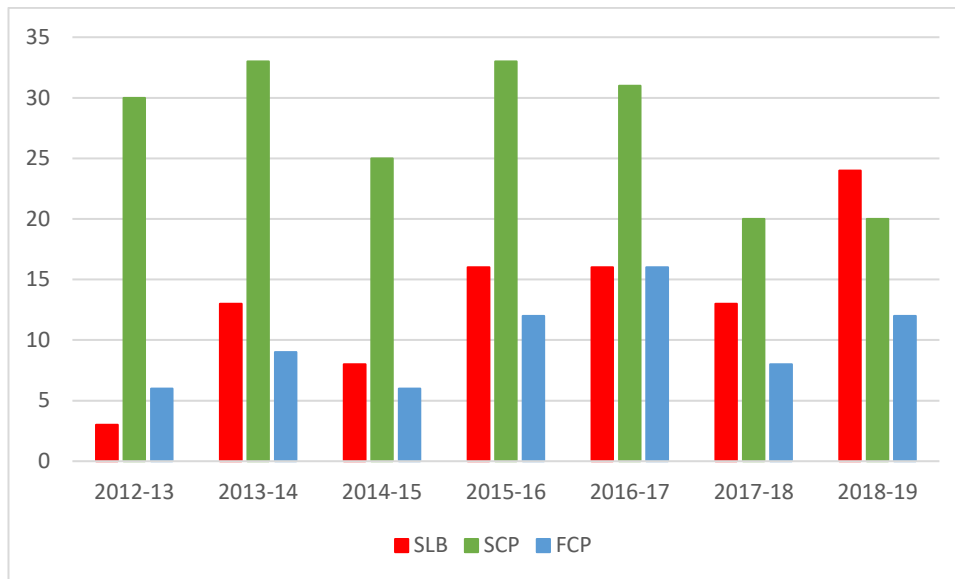


Gráfico 2 - Percentagem de jogadores no plantel principal, formados nos respetivos clubes, a fazerem pelo menos um minuto nas competições oficiais, por época.

5. Conclusões e Trabalhos Futuros

5.1 Conclusões

Analisando todos os objetivos propostos, a primeira conclusão que podemos retirar é a fraca prestação do FCP em comparação com os rivais. Se virmos cada um dos objetivos o FCP está em último lugar em todas as categorias. A principal razão que podemos apontar a esta realidade é o número reduzido de jogadores formados que apresenta o que, naturalmente, condiciona todos os possíveis ganhos, tanto desportivos como financeiros, com os jogadores da academia, levando a que caso queira ombrear com os números dos rivais os seus jogadores teriam de ter uma qualidade relativa muito superior aos demais, algo que não se verifica. Outro fator essencial é o facto de tanto o SCP como o SLB têm academias próprias e personalizadas (SCP finalizou a sua em 2002 e o SLB em 2005), o que trouxe claramente proveito a estes clubes e é essencial para o FCP ter o seu próprio complexo personalizado, caso este queira atingir os mesmos resultados que os seus rivais.

Comparados os valores entre SLB e SCP, deparamo-nos com uma grande semelhança de resultados (algo que não seria expectável tal tem sido o sucesso, e reconhecimento, internacional que os jogadores provenientes do SLB têm tido nas últimas épocas). O SCP apresenta um grande número de jogadores, algo expectável já que a formação sempre foi um dos pilares do clube, o que o beneficia na altura de recolher proveitos financeiros e

desportivos (quanto maior o número de jogadores maior a probabilidade de se ter um grande talento nas mãos). Podemos afirmar que o SCP começa muito à frente dos rivais, resultado de uma política de investimento na academia (aquando da sua construção, em 2002, era a maior de Portugal e uma das mais bem equipadas da Europa) iniciada muito antes que os seus rivais. Os benefícios deste investimento antecipado estão bem visíveis nos valores das primeiras duas épocas estudadas, assim como se olharmos para as convocatórias da Seleção Nacional que apresenta sempre mais jogadores do SCP nas grandes competições (vantagem que ganha ainda maiores contornos se tivermos em conta todos os jogadores do plantel e não apenas os que se enquadram nas épocas em estudo) o que é um claro sinal de superioridade dos jogadores provenientes do SCP.

No entanto, esta vantagem foi sendo reduzida ao longo das épocas pelo SLB, que após a ter concluído a construção do seu próprio complexo em 2005 tem vindo a implementar uma política de desenvolvimento e aproveitamento dos seus jogadores formados. Como seria de esperar, os resultados deste tipo de investimento são sempre a longo prazo e muito dependentes de um contexto favorável.

Neste caso, foi sobretudo a partir da época 2014-15 (visível no gráfico 1), onde o valor médio de mercado dos jogadores do SLB começou a ter uma subida acentuada e nunca mais parou (até à última época estudada entenda-se).

De facto, e contabilizados os objetos estudados, o SLB neste período de sete épocas ganha em quatro das seis categorias, nomeadamente:

- Tempo jogado: uma média de minutos por jogador época superior às dos rivais;
- Mais sucesso na carreira dos jogadores: apresentam uma média de minutos jogados por jogador/época superior em todos os escalões de coeficientes;
- Maior valorização, neste período, dos seus jogadores: uma diferença de 177,9 e 3,39 milhões no valor total e médio de mercado face ao SCP – e nos valores movimentados pelos mesmos, tanto no total das suas carreiras como para o clube – diferença de 336,36 e 134,2 milhões, respetivamente, face ao SCP;
- Mais sucesso desportivo que os rivais, aliado a uma participação cada vez maior e mais ativa dos seus jogadores formados no sucesso desportivo da equipa.

O SCP assume o lugar de destaque nas restantes categorias, nomeadamente:

- Volume de jogadores produzidos: quarenta e nove, mais catorze que o SLB.
- Presença na Seleção Nacional: nas últimas quatro edições das grandes competições internacionais teve sempre tantos ou mais jogadores convocados do que os rivais.

Concluimos então que durante este período o SLB conseguiu ocupar o lugar do SCP como a melhor academia em Portugal, com um crescimento constante ao longo deste período contrastando com o rival que atravessou um período de estagnação. Este facto poderá ser justificado pela crise financeira e várias crises internas (por exemplo o episódio de violência na academia em 2018) que assolaram o SCP neste período e impediram que existisse um contexto favorável para a evolução dos seus jogadores.

5.2 Trabalhos Futuros

Para trabalhos futuros sugiro tentar diversificar as fontes de dados de maneira a obter dados mais especializados (mencionados ao longo deste trabalho) de forma a ser possível realizar modelos explicativos mais fidedignos e que permita outro tipo de análise (como previsão de crescimento do valor de mercado).

Esta maior abundância de dados, sobretudo ao nível da segunda divisão nacional, iria permitir também uma análise mais detalhada nos aspetos técnicos dos jogadores o que, por sua vez, iria permitir identificar padrões de forças e fraquezas dos jogadores de cada uma das academias.

6. Bibliografia

- Aparicio, S., Aparicio, J. T., & Costa, C. J. (2019). Data Science and AI: trends analysis. In 2019 *14th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1-6). IEEE. <https://doi.org/10.23919/CISTI.2019.8760820>
- Atkins, C. (sem data). *Ranking the Best Youth Academies in World Football*. Bleacher Report. Obtido 12 de Novembro de 2020, de <https://bleacherreport.com/articles/1884751-ranking-the-best-youth-academies-in-world-football>
- Bryson, A., Frick, B., & Simmons, R. (2012). The Returns to Scarce Talent: Footedness and Player Remuneration in European Soccer. *Journal of Sports Economics*. <https://doi.org/10.1177/1527002511435118>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Edition). Routledge.
- Costa, C. J., & Aparicio, J. T. (2020). POST-DS: A Methodology to Boost Data Science. *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–6. <https://doi.org/10.23919/CISTI49556.2020.9140932>
- Crowder, M., Dixon, M., Ledford, A., & Robinson, M. (2002). Dynamic modelling and prediction of English Football League matches for betting. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51(2), 157–168. <https://doi.org/10.1111/1467-9884.00308>
- Dionísio, P. (2009). *Casos de Sucesso em Marketing Desportivo*. <https://www.wook.pt/livro/casos-de-sucesso-em-marketing-desportivo-pedro-dionisio/2751681>

- Franck, E., & Nüesch, S. (2012). Talent and/or Popularity: What Does It Take to Be a Superstar? *Economic Inquiry*, 50(1), 202–216. <https://doi.org/10.1111/j.1465-7295.2010.00360.x>
- Frick, B. (2011). *Performance, Salaries and Contract Length: Empirical Evidence from German Soccer*. ResearchGate. https://www.researchgate.net/publication/227450873_Performance_Salaries_and_Contract_Length_Empirical_Evidence_from_German_Soccer
- Gulbrandsen, A. M., & Gulbrandsen, C. M. (2011). *Valuation of football players: A complete pricing framework*. <https://openaccess.nhh.no/nhh-xmlui/handle/11250/169019>
- He, M., Cachucho, R., & Knobbe, A. (2015). *Football Player's Performance and Market Value*. 10.
- Heckl, J. (sem data). *5 Strategic Planning Lessons From Moneyball*. Obtido 19 de Setembro de 2020, de <https://www.smstrategy.net/blog/5-strategic-planning-lessons-from-moneyball>
- Herm, S., Callsen-Bracker, H.-M., & Kreis, H. (2014). When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. *Sport Management Review*, 17(4), 484–492. <https://doi.org/10.1016/j.smr.2013.12.006>
- Lewis, M. (2004). *Moneyball: The Art of Winning an Unfair Game*. W. W. Norton & Company.

- Majewski, S. (2016). Identification of Factors Determining Market Value of the Most Valuable Football Players. *Central European Management Journal*, 24(3), 91–104. <https://doi.org/10.7206/jmba.ce.2450-7814.177>
- McHale, I., & Scarf, P. (2007). Modelling soccer matches using bivariate discrete distributions with general dependence structure. *Statistica Neerlandica*, 61(4), 432–445. <https://doi.org/10.1111/j.1467-9574.2007.00368.x>
- Müller, O., Simons, A., & Weinmann, M. (2017). Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research*, 263(2), 611–624. <https://doi.org/10.1016/j.ejor.2017.05.005>
- Observador. (2019, Novembro 2). *Quase metade dos portugueses é do Benfica. Porto e Sporting em segundo e terceiro lugar.* Observador. <https://observador.pt/2019/11/02/quase-metade-dos-portugueses-sao-do-benfica-porto-e-sporting-quase-empatados-em-segundo-lugar/>
- Tavana, M., Azizi, F., Azizi, F., & Behzadian, M. (2013). A fuzzy inference system with application to player selection and team formation in multi-player sports. *Sport Management Review*, 16(1), 97–110. <https://doi.org/10.1016/j.smr.2012.06.002>
- Tunaru, R. S., & Viney, H. P. (2010). Valuations of Soccer Players from Statistical Performance Data. *Journal of Quantitative Analysis in Sports*, 6(2). <https://doi.org/10.2202/1559-0410.1238>
- Wirth, R., & Hipp, J. (sem data). *CRISP-DM: Towards a Standard Process Model for Data Mining*. 11.

