Lisbon School of Economics & Management
Universidade de Lisboa

MASTERS IN MANAGEMENT (MIM)

MASTERS FINAL WORK

PROJECT

HOW DATA SCIENCE CAN SUPPORT INDUSTRY ANALYSIS: THE CASE OF SMARTPHONES

RAFFAELE FOSCHINI

MARCH – 2023

Lisbon School of Economics & Management
Universidade de Lisboa

MASTERS IN MANAGEMENT (MIM)

MASTERS FINAL WORK

PROJECT

HOW DATA SCIENCE CAN SUPPORT INDUSTRY ANALYSIS: THE CASE OF SMARTPHONES

RAFFAELE FOSCHINI

SUPERVISOR: PROF. CARLOS J. COSTA

JURY:
PRESIDENT: PROF. JOSÉ MANUEL CRISTÓVÃO VERÍSSIMO
RAPPORTEUR: PROF. ISABEL PEDROSA
SUPERVISOR: PROF. CARLOS J. COSTA

MARCH - 2023

ACKNOWLEDGMENTS

The beautiful thing about journeys is that you never finish them the way you started them.

There is an infinite list of people that made this journey great in its best moments, and bearable in its worst.

First of all as trivial and corny as it may sound, my parents. Thank you for letting me make my mistakes, they really thought me a lot.

My brother, you probably never realized how much of a point of reference you have been to me, thank you being there, and most of all thank you for helping me bear our parents

My girlfriend, Re you know me better than I know myself, thank you for making me believe that I was worth more than I was giving me credit for. Sometimes I still think the worst of myself, but when I don't it's thanks to you.

Jack and the rest of My Lisbon flat mates, thank you for bringing me home when I couldn't walk straight, and for putting up with my mess and cooking skills.

My favorite Lisbon's Austrian friends, I wouldn't have graduated if it wasn't for you

La finestra di via Foppa 27,

I think it is a good idea to mention this place so that anybody who transited from here, even if not personally mentioned, shall fill my gratitude as they surely contributed somehow, to get me where I am.

Tommaso Albano, no sorry, t_Albans, no Sorry Salmonata Boom, I honestly don't know how are we still alive, but somehow it looks like we are being pretty good at this being adults thing. Anyway thank you for everything, I would do it back all over again a hundred times

Gio, I am incredibly glad that we went from being cousins to being friends. Thanks for all the great nights in via Foppa 27. Also thank you for letting me win at FIFA every once in a while.

Gabri, GABibbo, sGABello, thank you for being the great friend that you are. If they'll find anything to improve that condition of yours, I will let you know.

Lafo, thank you for being there probably more than you realized. Just one thing, is there a form I can fill in or some other way to complain and get back all the time wasted waiting for you?

ABSTRACT

The following work presents a comprehensive study of the smartphone industry, with a focus on the development of an automated tool for the monitoring of the market and the recognition of direct competing products. The thesis begins with an overview of the smartphone market. The study then investigates the attractiveness of the market in the framework of Porter's five competitive forces model (Porter, 1979), and different management styles different companies should implement to deal with the high level of environmental Turbulence that characterizes this sector. The Ansoff model for the response to turbulent environments is used to further support this investigation. On the basis of said analysis, it is then presented the development of a Python programmed script that uses web scraping and clustering algorithms to collect the technical specifications of a large number of the smartphone available on the market and proceeds to cluster the devices in order to identify similar and directly competing products. The study will show how the industry is shaped and present the main critical points that small and big firms need to deal with in order to thrive in this industry. The results will present the study of the industry in the framework of the two models and the proof of concept of a practical tool that aims at supporting firms in dealing with the findings given by the application of said models. This thesis contributes to the understanding of the smartphone market by offering a holistic view of its current state and potential trends for the future. The application of such models ultimately points towards the crucial importance of the constant and consistent monitoring of the market as a defense towards the constant evolution and unpredictability of the Smartphone industry.

Keywords: Smartphones, Web Scraping, Clustering, Ansoff, Porter

INDEX

GRAPHICS INDEX (IF APPLICABLE)

1. INTRODUCTION

Having spent the last six month at Motorola Mobility, an historical icon of the mobile phone industry, I realized how unique and unprecedented the tech market really is. The smartphone market is a relatively young industry that has transformed personal communication, business, politics, and almost every other aspect of life. Its unique characteristics set it apart from other industries, making it an area of innovation and disruption. Despite being a nascent market, it has had an unparalleled impact on human life, introducing new functionalities that have changed the way we interact with the world.

1.1 Contextualization

The smartphone market is a relatively young industry that has transformed personal communication, business, politics, and almost every other aspect of life. Its unique characteristics set it apart from other industries, making it an area of innovation and disruption. Despite being a nascent market, it has had an unparalleled impact on human life, introducing new functionalities that have changed the way we interact with the world.

The first commercial phone was released in 1984 by Motorola, but it was not until the late 1990s that the "smart" prefix was added to the phone's functionality, giving birth to a new industry. Smartphones are now a ubiquitous part of modern life, offering a range of functionalities that go beyond their original core function of communication. They can send emails and faxes, access the internet, and run apps, among other things.

As smartphones have become a platform for building new technology on, the focus has shifted from producing the best "Phone" to identifying and satisfying end-users' needs. (Agar, 2013) This is a challenging task, as it requires constant innovation to stay ahead of the competition. Each day, there is a potential danger of the advent of a new device packing a new technology that could set a new standard in the market.

The iPhone's arrival marked a turning point in the tech industry, demonstrating that even established players were vulnerable in this fast-paced market. Blackberry and Nokia, once dominant players, were unable to adapt to the changing market demands, leading to their downfall. Blackberry's insistence on retaining its keypad and failure to update its operating system ultimately led to its exit from the smartphone industry. Nokia's inability to quickly transition to the "smart" world, paired with a poor software strategy, sealed its fate as well.

Samsung emerged as the only real competitor to Apple, introducing the Android-powered Galaxy S series and establishing itself as the king of the Android world. Meanwhile, Chinese brands like Xiaomi and Huawei rapidly rose to become major players in the industry, contributing to China's position as a smartphone manufacturing superpower. As a result, the market became less concentrated, with a growing number of new players entering the industry and increasing the competitiveness of the market. Today, more than 10 brands share the global market, with varying degrees of activity in different regions.

1.2 Motivation

Based on the premises of the previous section, it is easy to conclude that this industry is characterized by constantly evolving competition and a rapidly changing environment. In this context, the release of a single piece of disruptive technology might impact the market to the point of shaking its structure to the core and forcing all the players involved into rethinking their strategy from the ground up. Obviously, events like this remain rare. Nonetheless, without going this far, the advent of new players implementing innovative features in their devices might still pose a great threat even to the largest manufacturers in the industry. Because of this, attentively monitoring the products that are periodically released and how these relate to other devices on the market becomes of crucial importance. The issue is that the number of different devices that are released on yearly bases by all the manufacturers in the market is incredibly large, and it is growing year by year. As of today, more than one thousand three hundred different brands have released more than twenty-four thousand different Android Devices (Android.com, 2023). The large number of products that populate the market makes it very difficult to monitor the market on a regular basis and with objective parameters, and this, in turn, makes it extremely easy to miss the rise of new threats or to delay the recognition of a new trend in the offer. At the same time, the high volatility of the prices of the devices may complicate the recognition of directly competing products as these, for lack of time and resources, are often simplistically identified by associating products with the same price point. This mechanism, though, might lead to misleading conclusions, as smartphones are frequently affected by large and sudden price modifications, and even the launch price itself is more often than not fake. With this in mind, it was decided to develop the proof of concept of a tool that would support Smartphone manufacturing firms in the process of market monitoring and identification of competing products.

1.3 Objectives

Overall, the project aims to qualitatively analyze and identify the most critical issues of the industry. Development of the proof of concept of a tool that may support Smartphone manufacturing firms in the constant monitoring of the market and surrounding environment and in the identification of new directly competing products that may potentially originate a discontinuity in the economic environment.

1.4 Research Approach

The approach followed for the development of said tool may be considered a mixed method of quantitative and qualitative approaches. Generical, the CRISP-DM was applied (Shearer, 2020, Costa & Apacirio, 2020). As anticipated in the objective section, the study will initially focus on the qualitative study of the market. This part is mainly supported by qualitative data such as historical study, specially in a context where data science is increasing importance (Aparicio et al. 2019). Specifically, data research is quantitative as the data itself was retrieved via web scraping, that is, coding scripts and web APIs to gather and process data from web pages (Mitchel, 2015). On the other hand, the description and classification of complex devices like Smartphones cannot be reduced only to the mere set of components that makes the device. In fact, the quality of these machines is also defined by several software and user experience parameters. Thus, to define if the final Clustering of the smartphones was consistent, it was necessary to qualitatively analyze the final output of the algorithms based on the personal expertise of someone who has operated in the market and knows how the products function besides their physical characteristics. Considering that the final result is an artifact, it may also be considered a design science approach (Aparicio et al., 2023).

1.5 Structure

The thesis is divided into several sections, starting with an overview of the smartphone market. This section (Market Analysis – Porter's Five Forces Model Framework) analyses the smartphone market through the Porter Five Forces model. This model evaluates the industry's competitive environment, which is essential for understanding the market's structure and dynamics. The section discusses the five forces of competition: the bargaining power of suppliers, the bargaining power of buyers, the

threat of new entrants, the threat of substitutes, and the rivalry among existing competitors.

In addition, the thesis examines the different management styles that the management of the various players in the smartphone business may follow Market Analysis – Porter's Five Forces Model Framework (Porter, 1979, 1980).

The high level of environmental Turbulence in the smartphone market necessitates agile management styles that can adapt to changes quickly. This analysis was carried out via the Ansoff model for the analysis of the response to a turbulent environment (Spina, 2012), which is a strategic planning tool used to identify management styles to implement based on the velocity and the degree of novelty and complexity of the changes that affect the market in question.

Finally, the thesis's central topic is an empirical work that the author personally conducted (Smartphones data gathering and Clustering). The study involved developing a Python code that utilizes web scraping and a clustering algorithm to scrape the technical specifications of many devices available on the market and clusters similar phones to detect competitors and study the development of the market's technology. The section discusses the data collection process and analysis, highlighting the key findings and their implications for the smartphone market.

Overall, this thesis provides a comprehensive smartphone market analysis, covering the market's growth, structure, and dynamics. The thesis also examines the different management styles and growth strategies applicable in the smartphone market, with a focus on the technological trends shaping the industry. The empirical work using web scraping and a K-means clustering algorithm provides valuable insights into the market's development, and the findings can inform firms' strategic decision-making.

2. LITERATURE REVIEW

In this section, the reader will find the theoretical framework of the project. From the analysis of similar works that were carried out in the past and from which it was taken inspiration from to the explanation of the theories of the models with which the market is analyzed.

2.1 Strategic Answer to Environmental Turbulence

In order to navigate a market of such kind, it is crucial to define tools that may help to monitor the industry on a regular basis. In this context, being capable of identifying and even anticipating a general shift towards a specific kind of technology is the main differential between the players that succeed and those that eventually fail. Once the tools to monitor the market are in place, it is fundamental to define how the company or management will react to identifying "Change" within the market environment.

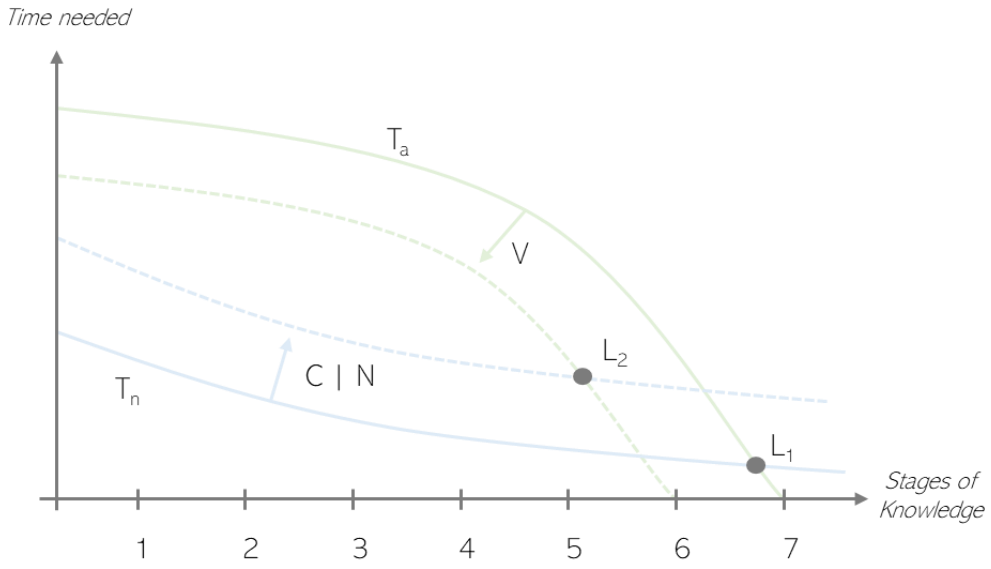

Figure 1 Description of Ansof's turbulence management model

L = Last Available Moment

T_a = Time Available

V = Velocity of the change

T_n = Time Needed

C = Complexity

N = Novelty of the Change

T_n = Time of competences development + Time of implementation

$T_n < T_d$

- *Limited information, High risk decisions*

- *First mover advantage*

$T_n > T_d$

- *Full knowledge*
- *Late decision-making, risk of not being able to catch up*

The Ansoff Model helps managers respond to changes in a turbulent market. There are seven stages in a paradigm shift: 1. Sense of general Turbulence; 2. Identification of the source of the change; 3. Identification of the impact; 4. Response definition; 5. Outcomes evaluation; 6. First Impact; 7. Full Impact.

The earlier a company recognizes a paradigm shift, the more time it has to adapt. Ta is the time between the first sense of Turbulence and the full impact of the environmental change on the market. Tn is the time necessary to adapt to the environmental change. The Latest Point is where Ta equals Tn. Environmental Turbulence affects Ta and Tn, with velocity affecting Ta, and novelty and complexity affecting Tn. Ansof argues that as time passes by a decision maker acquires more information as the novelty unravel itself through its impact on the environment. Therefore the later a manager makes a decision the more information he or she will have to make a more informed decision.

A manager's response should depend on where the Latest Point falls. If the point is between stages 6 and 7, a manager can delay making a decision until the change has had its first impact on the market. If it falls before stage 6, the manager needs to act before the innovation's first impact on the market, which is riskier.

2.1 Related Works

According to Han & Cho (2016) the evolution of the technical specifications, in terms of both hardware and software, of the smartphones released between 2012 and 2015. The way the authors perform the analysis is close to the one implemented in this project. The main difference is the focus on benchmark performances, used as a term of comparison between older and newer generations of devices. Han & Cho (2016) explore several benchmarks that test the performances of the different components of the devices. Specifically, the evolution of the devices was investigated through the score of benchmarks that tested system, memory, graphics, and web browsing performance.

Han & Cho (2016) explore the evolution of hardware components and reaches interesting conclusions regarding the future development of core components like RAM, CPU and battery as a response to the increasing computational power requirements. Han & Cho (2016) findings prove that the use of web scraping and the constant monitorization

of the competition via said techniques, does allow the detection of market trends which may be of help in the recognition of probable future developments of the market. Stoehr, et al. (2020), despite being connected to a different industry (the automotive industry), they deal with concepts like the ones that will be unraveled in the next chapters. The automotive industry is getting more dependent and connected to the technological industry's evolution. Cars are getting "smarter" and the human interaction with them is getting closer and closer to the one we have with other technologies, like smartphones and computers. This makes the automotive field much more subjected to the potential arrival of new game-changing technologies that may disrupt the current equilibrium of the market. In this perspective, the authors have found that an automated web scraping and machine learning based tool for monitoring the top three car manufacturers in the world (Toyota, Volkswagen, and Hyundai) may prove to be extremely useful in understanding and anticipating the dynamics of their strategies. The study focuses on the use of network analysis of car manufacturer web pages to reveal internal corporate positioning and innovative trends (Stoehr et al., 2020). Overall, this study is a valuable contribution to understanding digital transformation's impact on the automotive industry and highlights the importance of monitoring shifting industries. Many authors also relate the smartphones' technical characteristics with the market's characteristics (e.g. Jamalova, Constantinovits, 2019). Some of those characteristics are especially relevant to the diffusion of smartphone usage (e.g. Jamalova & Constantinovits, 2020). The role of supplier innovation performance and strategies on the smartphone supply market is also relevant, as pointed by several researchers (e.g. Varriale el al. 2022).

2.3 Comparison Between Data Clustering Algorithms

To fulfil the second purpose of the project, that is the automated identification of similar competing products, it is necessary to define which clustering algorithm to implement. The aim of the whole project is in fact to develop a tool that not only, autonomously searches for all the smartphones available on the market at a given moment, but also after storing all their data, recognizes the most similar devices based on the main relevant characteristics to return a picture of the direct competing products. To do so it is necessary to identify the appropriate algorithm. Abbas (2008) describes the distinctive characteristics of the most used clustering algorithms. The papers consider four algorithms: K-Means, Hierarchical Clustering, Self-Organization Map (SOM), and Expectation Maximization (EM).

The K-Means algorithm starts by randomly selecting K centroids, the initial clusters' centers. Then, it assigns each data point to the cluster whose centroid is closest to it. After that, the algorithm recalculates the centroids of each cluster by taking the mean of all the data points in that cluster. The process of re-assigning data points to their closest centroids and updating the centroids is repeated until the clusters no longer change or a maximum number of iterations is reached.

Hierarchical Clustering is a clustering algorithm that builds a tree-like structure by initially assigning a cluster to each observation in the dataset. The algorithm subsequently merges individual clusters based on their distance to one another. This process is iterated until all the observations belong to one cluster, containing all the data points. Here the distance between different points can be computed with different approaches: Euclidian distance, Manhattan distance, and Ward distance.

Self-Organizing Maps (SOM), also known as Kohonen maps, are a type of artificial neural network that can be used for Clustering and visualization. The algorithm involves mapping high-dimensional data onto a lower-dimensional space, typically a 2D or 3D grid. The map is initialized with random weights, and the data is iteratively presented to the map. During each iteration, the weights of the closest neurons to the input data point are adjusted to match the input. This process continues until the map converges to a stable configuration. The resulting map can be used for Clustering, as similar data points tend to be mapped to neighboring neurons.

The EM algorithm is an iterative algorithm that seeks to estimate the parameters of a statistical model based on incomplete or missing data. In Clustering, the EM algorithm is used to estimate the parameters of a mixture model, which is a model that assumes that a mixture of several probability distributions generates the data.

Overall, according to Abbas (2008), EM and K-Means show better performance with large datasets, while SOM and Hierarchical Clustering are suggested for smaller datasets. Furthermore, as K rises the performance of SOM decreases while the performances of K-Means and EM improves.

The two algorithms that potentially fit the purpose the best are K-Means and hierarchical Clustering. Between the two, as specified above, the former performs better and more efficiently when dealing with large dataset, with higher values of the super parameter K and higher complexity (i.e., with many variables considered)

The dataset is not particularly large, and the number of variables considered is also low. This suggests that the hierarchical clustering algorithm would be the best choice for clustering the data. Still the goal of the project is also to provide a scalable tool. In fact, not only, the number of devices considered will pile up over the years, but it is also desirable that the end user one day have the power of choosing which and how many variables to include in the analysis. In this perspective it seems logical to evaluate the performances of both algorithms to evaluate the highest performing of them. Still, in case the two did not show significant differences in performance, it would be advisable to proceed with the K-Means algorithm as this seems to be the preferrable one also in case of a future increase of the number of observations, of the number of clusters K or in the complexity of the algorithm.

While the detailed analysis of the output will be reported in the "Results" paragraph, we can anticipate that the results obtained with the two different algorithms actually show a lot of similarities and comparable performances. With this premise, the chosen clustering algorithm is ultimately K-Means.

3. METHODOLOGY

3.1 Market Attractiveness

The attractiveness of a market is typically described with the application of Porter's five competitive forces (Porter, 1979, 1980). In this context, a qualitative approach was adopted to analyze the market and its attractiveness in the framework of Porter's five competitive forces(Porter, 1979, 1980). In fact, the primary source of information was retrieved from historical references, News, and articles that were used to extrapolate potential insights regarding the dynamics that regulate the connections of the different players that make up the market. Identifying how the different parts of this industry interact with each other is crucial to establish whether the dynamics behind the functioning of this market are sustainable in the long run.

3.2Market Turbulence

The presence of Turbulence in the market was defined with the application of Ansoff's model for the response to turbulent environments. Once again, the research followed a qualitative approach, as the data and information gathered mainly derive from field observation, historical references, and bibliographical references, and the expertise gathered during six months of hands-on experience in the field as a market data analyst at Motorola Mobility.

3.3. Data Science Approach

The structural analysis may be related to the Data Science methodology. In fact, the relevant previous steps can be interfaced with the business understanding of data science methodologies (Shearer, 2020, Costa & Aparicio, 2020). The process followed is inspired by CRISP-DM, which stands for Cross-Industry Standard Process for Data Mining, and is a widely used methodology for guiding data mining efforts (Costa & Aparicio, 2020). It provides a standardized process model for data mining projects, consisting of six phases that are typically executed in a cyclical manner:

In the Business Understanding phase, the goals and objectives of the project are defined, and the requirements for data mining and business success are identified.

In Data Understanding, data is collected, explored, and described. The quality of the data is assessed, and any issues are addressed.

In the Data Preparation phase, the data is cleansed, integrated, transformed, and formatted into the desired input for modeling. In the data preparation phase, the code deals with these issues by removing all the devices that contained the word "Tab" in their product name and all the devices that presented even only one missing data of the variables included in the clustering algorithm. The second phase, the data exploration, turned out to be extremely valuable as a few issues arose during the analysis of the gathered dataset. Firstly, not all the devices included in the smartphone section of the website were actually smartphones. A certain number of tablets were also included in the dataset, and as their characteristics are generally a lot different than smartphones', including them in the clustering algorithm may have severely affected the algorithm's overall accuracy. Secondly, not all the devices included in the analysis presented the variables that were included in the clustering process.

In the Modeling phase, various modeling techniques are applied to the prepared data to create a model that can be used to make predictions or identify patterns.

In the Evaluation phase, the model created in the previous phase is evaluated to determine its effectiveness, and any necessary adjustments are made.

In Deployment Phase, the final model is put into production, and ongoing monitoring and maintenance is performed to ensure its continued effectiveness.

One of the strengths of CRISP-DM is its flexibility. It can be customized to fit the specific needs of a project or organization.

Including all these steps in the code ensures that when the dataset is updated (as this is the purpose of the project), the clustering algorithm will always be fed with clean and consistent data.

The first step of the CRISP-DM standard was already performed as the purpose of the project was already defined.

3.3.1 Data Gathering – Web Scraping Smartphones Data from the Web

In order for the tool to be usable, the first thing to do was to find a comprehensive and updated data set containing all the information regarding the components of a large number of smartphones available as of today. Such dataset is currently not available online, or if it is, it is not updated to the most recent releases, and it also does not fulfil the purpose of monitoring the newest releases on the market as time goes by. The only thing to do in this scenario was to gather the data autonomously, from a trusted source.

Web Scraping allows the user to perform exactly this kinds of tasks.

The selected source of information is a renowned Italian website: HDblog.it.

This website provides a detailed list of all the technical specifications of the majority of the devices available in and outside the country. The reason behind the choice of this specific website instead of other similar ones (e.g. GSM Arena) is firstly that, based on the author's personal experience. HD Blog reports more accurate information about the devices. Secondly, HD Blog gives also information about the existence of different memory variants of each phone (in contrast to GSM Arena that usually do not report this piece of information), which for our purpose is crucial to know.

Once the source was found it was time the implement the web scraping code that would autonomously navigate towards the given webpage, store the HTML code of the page that contains all the text displayed on the webpage, and gather the requested information in a local file. To do so the libraries that were used the most were: BeautifulSoup and Googlesearch.

Obviously, the different phones' specification sheets were all displayed in unique and dedicated pages, which needed to be accessed individually in order to get that data of separate devices. This meant that the first thing that needed to be done was to find an automated way to get the URL of every single web page. This was achieved by firstly scraping the names of all the devices analyzed by the website. The names of all the phones are displayed in a separate section of the website. Once all the names were store in a list, it was possible to iterate through the list in order to launch a google search, that would search with the following key words: "phone_name" and "HDblog scheda tecnica". This choice of keywords assured that the first result of every google search would have always been the HDblog specification sheet web page of the given phone.

At this point it was possible to access all the URLs of the phones stored in the initial list and iterating through them in order to retrieve the information of every device. In practice the goal was achieved with the following code (note: for the sake of brevity not all the lines of the code were included in the description)

```
# Gathering all the smartphone names that were reviewd by hdblog
# and storing them in a list called phones_names

phones_list=[]

for i in range(1,17):
    url = f"https://www.hdblog.it/schede-tecniche/page/{i}/"
    page = requests.get(url)
    soup = BeautifulSoup(page.content, "html.parser")
    results=soup.find(id="cellphones_list")
    phones=results.find_all("h4")
    for i in phones:
        phones_list.append(i.text)
```

Figure 1: Smartphones' name gathering

The code in figure 1 is responsible for the extraction of all the product names of the devices that had a specification sheet page on the website.

The URL of the webpage to scrape the data from is stored in the URL variable

- page = requests.get(url) : sends an http get request to the URL specified by url, and assigns the server's response to the page variable.
- soup = BeautifulSoup(page.content, "html.parser") creates a BeautifulSoup object called soup that contains the parsed HTML content of the page.
- results=soup.find(id="cellphones_list") :searches for an HTML element with an id attribute of "cellphones_list" within the soup object and assigns it to the results variable.
- phones=results.find_all("h4"): finds all HTML elements with an h4 tag within results and assigns them to the phones variable.
- for i in phones: is a loop that iterates through each of the h4 elements in phones.
- phones_list.append(i.text): extracts the text content of each h4 element and appends it to the phones_list

All the above is part of a for loop that iterates the same code 16 times in order to scrape the data of the first 16 pages of the website, which means gathering all the smartphone names of the last two years.

The first four lines within the for loop are a standard piece of code that it is always use more or less in the same format to initiate any web scraping process with Beautiful Soup, as such they want be commented in the following instances. For the following step it was necessary to define two functions, one for the retrieval of the URL of webpage of each device, and the other to perform the iterative scraping of the devices' data itself:

```
def get_url(phone):
    query  = phone + " " + "HDblog scheda tecnica"
    for j in search(query, tld="co.in", num=1, stop=1, pause=3.5):
        url = j
    return url, phone
```

Figure 2: get_url function code

```
def specscraper1(speclist, phone):
    url, smartphone=get_url(phone)
    page = requests.get(url)
    soup = BeautifulSoup(page.content, "html.parser")
    resultsright=soup.find(id="phone_columns")
    allspecs=resultsright.find_all("li")
    resultsleft=soup.find(id="controles_titles")
    titles=resultsleft.find_all("li")
    phone_name=soup.find("h2")
    phone_name=phone_name.text

    titleslist=[]
    for i in titles:
        titleslist.append(i.text)

    specifiche=[]
    for i in speclist:
        n=titleslist.count(i)
        if n>0:
            ind=titleslist.index(i)
            if str(allspecs[ind]).count("><")>2 and "hoverred" not in str(allspecs[ind]):
                if "ok" in str(allspecs[ind]):
                    specifiche.append("yes")
                elif "wrong" in str(allspecs[ind]):
                    specifiche.append("no")
            else:
                specifiche.append(allspecs[ind].text)
        else:
            specifiche.append('-')

    specifications={f"{phone_name}": specifiche}
    return(specifications)
```

Figure 3: specscraper1 function code

3.3.2 Clustering:

Before dealing with the actual steps that were made to cluster the dataset it is relevant to mention which variables were included in the model and why. When it comes to

technical specifications, some of the most relevant ones to consider when distinguishing between low-end, mid-end, and high-end smartphones are (Agar, 2013):

Processor: The processor is the "brain" of the smartphone and determines how quickly it can perform tasks. Low-end smartphones usually have slower, less powerful processors compared to mid-end and high-end models.

RAM: Random Access Memory (RAM) is used by the smartphone to temporarily store data and run apps. Low-end smartphones typically have less RAM than mid-end and high-end models, which can impact performance and multitasking capabilities.

Camera: The quality of the camera is an important factor for many users when choosing a smartphone. High-end models often have multiple rear cameras with high megapixel counts, advanced features like optical zoom and image stabilization, and the ability to shoot 4K video.

Battery: The battery life of a smartphone can greatly impact its usefulness, and high-end models often have larger batteries with fast charging capabilities.

Connectivity: Support for faster mobile data speeds like 5G, as well as Wi-Fi 6 and Bluetooth 5, are often found in high-end smartphones.

Based on the list above the variables that were eventually included in the K-Means model were the RAM capacity, the megapixels of the front and rear facing cameras (separately) and the processor.

In general the chipset (processor) is the component that individually affects the performance of the device the most. The problem is that chipsets, are clearly not a numerical variable, and it is therefore difficult to introduce this component in the clustering model.

In order to cope with this issue I have found a reasonable and effective solution, consisting in substituting in the model the processor name with its correspondent Antutu score- Antutu is a benchmark tool that analyzes the performance of a device under stress conditions and returns a value that is comparable between different processors and allows to identify which is better performing than which. To do so, it was used a web scraping process similar to the one used for the dataset population.

In the modeling stage it was firstly necessary to assign a value to the super parameter K, at this purpose the elbow method was applied.

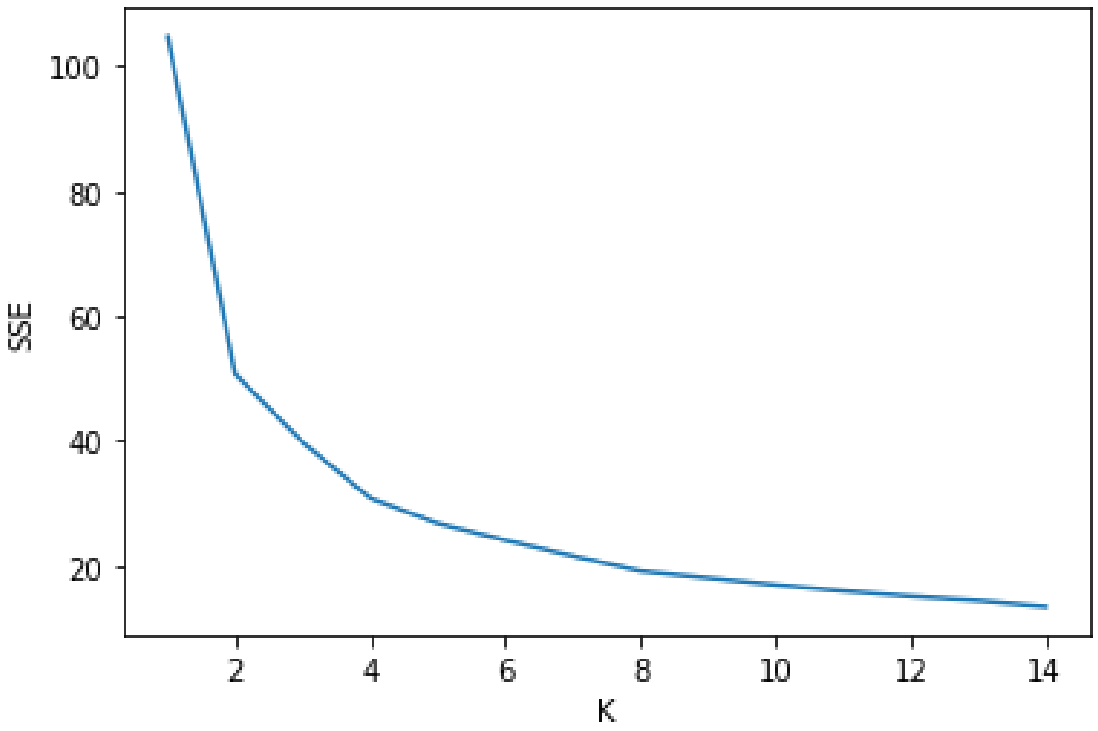

Figure 4: Elbow method representation

The results suggested the choice of K between two and four. In this scenario though, a number of cluster this low would cause the groups to have a very high numerosity, which means that the products clustered in the same group may also be significantly different. Eventually, K was set equal to nine, as the market is usually subdivided into three categories: Low, Mid, and High end, and dividing the devices into nine group is conceptually similar to subdivide each categories into three parts.

Also in the modeling phase it was decided which clustering algorithm to implement.

Ultimately, the aim of making the project more scalable in the future, by including a larger number of devices as well as potentially increase the number of clusters in order to obtain groups of smaller numerosity, made the decision lean towards the use of the K-Means clustering algorithm. Still, the Hierarchical Clustering was also performed on the same dataset in order to verify that the accuracy of the two models was comparable also with lower values of K and relatively low complexity. It was in fact explained in the previous sections that on a theoretical level at in correspondence to lower values of K and lower complexity the hierarchical Clustering would be more advisable (Abba, 2008)

In the evaluation stage the outputs of the two algorithms were analyzed and it was confirmed that the clustering accuracy was comparable, thus in the last step of the CRISP-DM, that is the final deployment of the process it was decided to proceed with the K-Means clustering that is more efficient and accurate also with larger datasets.

In practical terms the most relevant sections of the code used to perform the Clustering of the devices was the following:

```
scaler=MinMaxScaler()
std_df=scaler.fit_transform(df1[['RAM', 'CHIP' , 'FC_MP', 'RC_MP']])
```

Figure 5: variables standardization part 1

```
df1['std_ram']=std_df[:,0]
df1['std_chip']=std_df[:,1]
df1['std_fc']=std_df[:,2]
df1['std_rc']=std_df[:,3]
```

Figure 6: variables standardization part 2

Initially the variables are standardized and centered, this is a necessary step to cope with the fact that the variables have different unit of measure, which is likely to affect the performance of the algorithms in case the range of values is particularly broad.

Successively the standardized variables were fed to the K-Means clustering algorithm with k equal to nine.

```
km = KMeans(n_clusters=9)
y_predicted=km.fit_predict(df1[['std_ram','std_chip', 'std_fc', 'std_rc']])
df1['Cluster']=y_predicted
```

Figure 7: K-Means clustering

```
df['type']=df1['Cluster']
```

Figure 8: cluster labels addition to original DF

- *km = KMeans(n_clusters=9):* This line initializes a KMeans clustering model with 9 clusters. KMeans is a clustering algorithm that groups data points into a specified number of clusters based on their similarity to each other.

- *y_predicted=km.fit_predict(df1[['std_ram','std_chip', 'std_fc', 'std_rc']]):* This line fits the KMeans model to the dataset df1, specifically the four columns with the labels 'std_ram', 'std_chip', 'std_fc', and 'std_rc'. It returns an array of predicted cluster labels for each data point in the dataset.
- *df1['Cluster']=y_predicted*: This line adds a new column 'Cluster' to the original dataset df1 and assigns the predicted cluster labels to each data point based on the KMeans clustering model. Each data point is now associated with a specific cluster label based on the similarity of its features to other.

4. RESULTS

The following sections will cover the application of the market analysis via theoretical framework of Porter's (Porter, 1979, 1980) and Ansoff's model, as well as the description of the most relevant parts of the code used to develop the automated tool for data gathering and data clustering.

4.1 Market Analysis – Porter's Five Forces Model Framework:

The Quickness with which these many companies managed to take over a market with such high entry barriers is a good representation of how complicated navigating this market can be.

The industry is, in fact, mainly characterized by highly high entry costs and investments both in production capacity and research and development. The cost of serving a global market is extremely high as well, and this is part of why Chinese companies were able to grow so much and have such a steep learning curve. Chinese brand's production was not offshored as Samsung and Apple's. This saved them the huge costs of shipping their finished goods to their home markets and allowed them to focus their investments on R&D and production site development. This mechanism filled the market with new players, all competing to develop the next huge technological improvement that will set the new smartphone standard for the future.

4.1.1 Supplier bargaining power:

Identifying suppliers in the smartphone industry involves two main categories: software and hardware. Software providers supply operating systems to smartphone manufacturers, while hardware suppliers provide physical components like chipsets, camera lenses, and batteries. The bargaining power of suppliers is generally moderate, except for a few giant companies that can heavily impact smartphone manufacturers. Google is a prime example of a powerful software supplier, as evidenced by the story of Huawei, which was banned from using Google's services and suffered significant market losses as a result. However, manufacturers are generally larger and financially stronger than their software suppliers.

On the hardware side, suppliers also hold some leverage over smartphone manufacturers, but this is less pronounced than with software suppliers. Having high-quality components from suppliers like Qualcomm and Sony can improve a

manufacturer's reputation, but some players like Samsung are attempting to reduce their dependence on giant suppliers by vertically integrating more activities in their production process and R&D. Nonetheless, this requires significant financial investment, and not every player in the market can do so. Even if every player had the financial capacity to integrate production vertically, it does not guarantee better products than off-the-shelf components.

4.1.2 Buyers bargaining power:

The market itself is not particularly concentrated, still considering the large number of similar products offering very similar functionalities, the power of Smartphone manufacturing companies is equally significantly decreased.

The high amount of similar products in the market forces manufacturers to fight a constant price war, forcing prices down and causing the progressive thinning of the margins. Furthermore, the necessity to stand out from the other players in the business generally translates into the needs to increase their product line even more, which eventually cause the number of alternative products available on the market to skyrocket.

Companies are now trying to compensate for the many alternative products available by increasing customers switching costs. In practical terms, this mainly consists of creating what was defined by many as a "device ecosystem". This is described as a system of highly compatible products that aim to improve the user experience thanks to optimizing their interaction. The idea is to convince the customer to buy all the components of said ecosystem, so switching to a different brand implicates the need to replace all the system's products, which is a lot more expensive than replacing just the one they initially wanted to replace. This strategy simultaneously achieves another goal: to increase brand loyalty and allow companies to raise prices, trusting that their customers' affection for their products will convince them not to switch to a cheaper alternative.

Overall it is easy to identify the market condition as one in which manufacturers are stripped of most of the power towards their buyers. Thus, producers are led to search for different strategies to retain their customer's loyalty.

This mechanism partially moved the battlefield from the product's quality to the brand's awareness.

4.1.3 The threat of New Entrants:

The threat of new entrants is undoubtedly shallow because of the dramatically high entry barriers. The difficulty of joining this industry consists of the huge investments

necessary to set up production and the complexity of developing such products. R&D's expenses are almost unparalleled in any other industry, and since new models are set to arrive on the market every year with new and improved features, it is undeniably rare for one company to hold enough capital to sustain such expenses.

Furthermore, as already said, we also need to consider the necessity to build a reputation, a stable image of trust, quality and reliability in the minds of the as this is one of the only ways manufacturers can balance their buyers' bargaining power. This process, even though necessary, comes at the expense of considerable investments in marketing activities and communication. Together with the already exceptionally high research and development costs, the market is becoming less and less sustainable over the years.

Nonetheless, as mentioned above, some companies (especially Chinese ones) have been able to rise from nothing and take the industry by storm. So even though all the premises suggest that the fear of potential new entrants would be unjustified, the main actors in the scene should still keep a vigilant eye on the market.

The rapidity with which all these companies managed to become so competitive, starting from scratch, embodies all the characteristics that make this market so different from any other in history.

4.1.4 Fear of Substitute products:

Even though the tech world is filled with old and new communication devices and tools, the fear of substitute products remains extremely low. This is mainly because smartphones are still unmatched in functionalities and convenience. Smartphones are, in a way, unreplaceable by any other product by itself. Because of the fact that these devices incorporate so many different features, it would be tough to find another product that could replace all the functionalities of a smartphone.

4.1.5 Competitive Rivalry:

Rivalry in the Smartphone industry is highly intense. A relatively small number of players share the market, and each is constantly fighting to gain the next bit of market share. Companies like Apple, Samsung, Xiaomi, Oppo, Realme, Oneplus, and Motorola develop new and improved products yearly. Here one single mistake In terms of hardware or software can become fatal for the company that committed it.

The high level of rivalry in this field is given by the presence of different battlegrounds where the manufacturers are fighting on.

There is the technological battleground. Every company is battling the others, trying to develop the biggest innovations. This causes R&D spending to increase more and more every year as all the participants know how big of an impact one innovative feature can have on the market.

There is the price battleground. Since the market is so full of products that, especially on the low end, offer very similar characteristics, design firms are forced to battle each other on who can offer the minimum requirements at the lower price. This generates the necessity to invest in optimizing production and transportation processes, which increases the already high overheads of this business even more.

There is the brand battleground. Since it is incredibly easy for buyers to switch from one brand to another cheaply, companies need to build a strong connection with their customers. The only way to achieve this is by identifying the central values and drivers of the target customers and finding a way to include them in the brand communication.

4.2 Market Analysis – Ansoff Model for the Response to Turbulent Environments:

This point will make it reasonably clear that the Smartphone manufacturing industry features an extraordinarily volatile and dynamic market. Furthermore, manufacturers are dealing with highly technological devices, significantly increasing the complexity of developing new solutions to adapt to the constantly evolving market.

In other words, this industry matched all the criteria of a highly turbulent environment.

As explained in the previous paragraph, operating in such conditions means dealing with a Time Available Ta that is often very close to the Time Necessary Tn to adapt. Therefore, the L point of this industry is pulled further back in the stages of knowledge, which increases the risk behind every decision. Of course, even in this industry, not every change comes with the highest degree of novelty. However, in one way or another, more or less subtlety, the environment is constantly evolving.

Within this context, every company has its approach to dealing with the Turbulence that affects this market.

The strength of a firm both in financial and brand awareness terms plays a crucial role in the broadness of the range of possible strategies it can pursue. The more renowned and trusted a brand is, the more it is allowed to postpone its action when facing the advent of a specific innovation.

An example of such behavior can be observed in the case of Apple. The American company has succeeded in creating an amazingly loyal customer base, which enables it to be a late adopter of new technologies in their products. The average Apple consumer is not concerned about other manufacturers offering amazing features in their devices, as they believe that when Apple eventually releases its version of such features, it will likely be the best implementation available in the market.

On the opposite end, smaller companies are generally forced to constantly offer disruptive features and technologies in order to exploit the temporary absence from the market of tech giants like Apple, Google, or Samsung.

Motorola is a prime example of this phenomenon. Since its acquisition by the Chinese giant Lenovo, Motorola has been leading the way in the latest technological advancements in the smartphone industry. For instance, Motorola was the first company to offer a foldable smartphone. Since the release of the "new Motorola Razr," which was the first foldable touchscreen smartphone ever released, many other companies have followed suit with their interpretation of this technology. As a result, many companies that initially waited to join the foldable market have gained significant insights from observing the market response to Motorola's innovation.

In this picture, the concept of time to market becomes of fundamental importance. Smaller companies, as said, need to be the first on the market, and doing so generally allows them to build up their name in the minds of the consumers. On the other hand, this process often results in the release of features that the market may not be ready for or in a technology that is not finalized enough to succeed in a market with such high standards.

In this perspective, we shall frame the strategies of smaller and bigger companies into the management styles of the Ansoff model as follows:

Smaller brands like Oppo and Motorola, who are trying to build brand awareness daily, must implement creative and exploratory strategies. This means these firms must be attentive and perceptive to the market's evolution. They must constantly monitor the market to interpret even adjacent industry innovations that may affect the smartphone

manufacturing business in ways we cannot fully comprehend. The recent release of Artificial Intelligence tools such as Microsoft's CHAT GPT or Google's response to it, BARD, may represent the first sign of the next change in the smartphone market. If this is the case, these companies may be forced to be among the first to implement such technologies, for example, a new version of an AI-powered vocal assistant. However, the complexity and novelty of such innovation are incredibly high. On the one hand, being the first in the market can be a great advantage in terms of exploiting a moment of a temporary monopoly of the technology. On the other hand, firms like Apple, Samsung, and Google itself already have significant hands-on experience in this field, as they have been offering their own proprietary vocal assistant technologies in the form of Siri, and Google voice, respectively. This will allow them to wait, observe and gather a lot of information from the early release of these technologies and the consumers' response to them. This might, in turn, enable these more prominent companies to arrive on the market at the last moment with better and more complete products that will obscure the smaller companies' efforts and their version of such technologies.

Companies like Samsung or Xiaomi, thanks to the greater strength of their brand and the resources they can afford to pour into their research and development departments, are generally much more agile in reacting to possible developments in the market. In practical terms, their Tde is going to be a lot lower, which in turn lowers their overall Tn. Having a low Time Necessary to adapt, as already explained, allows them to delay their actions and make more informed decisions. They can implement Anticipative or sometimes even Reactive management styles.

Finally, at the highest end of the spectrum, we find the one company that can afford to implement a Reactive management style regularly: Apple. This company has built strong brand awareness and disposed of such high resources that it can act even after the seventh stage of knowledge (the full impact on the market.) There are several examples of such behavior from the Cupertino firm.

Companies like Samsung or Xiaomi, thanks to the strength of their brand and significant research and development resources, are generally much more agile in responding to potential market developments. This results in a lower Time to Development Tde, which lowers their overall Tn. As previously explained, having a low Time to Adapt means they can delay their actions and make more informed decisions.

These companies can implement anticipative or, in some cases, reactive management styles.

At the highest end of the spectrum, we find the company that can regularly afford to implement a reactive management style: Apple. This company has built strong brand awareness and has vast resources that it can act even after the seventh stage of knowledge (i.e., the full impact on the market). The history of Apple shows this pattern many times. Moreover, almost every instance of this behavior proves how important it is for the success of a product or a service to be released at the perfect timing.

This behavior is defined as "Strategic Followership". (Kelley, 1988)

A real life example of such behavior is Apple's strategic followership with the iPod scenario (Zoogah, 2012), where Apple's portable music player was not the first to enter the market. Analog portable music players date back to 1979 with Sony's Walkman, while the first digital MP3 players, such as the MPMAN F10 and PMP300, were released towards the end of the 90s. Despite their success at the time, these products did not stand the test of time, unlike the iPod, which owes its success to Apple's strategic followership approach.

Rather than rushing to introduce a revolutionary technology, only to be the first one to implement it, Apple adopted a patient approach and observed the market. The American company noticed that while the technology was undoubtedly revolutionary, the market was not yet ready for it. There was no established platform to get the MP3 music format. Loading times were huge, and the memory of the devices would not allow the consumers to store more than a few songs and so on. Apple gathered all the information on the market response to these products, and only after they knew they had built a device that would satisfy all the consumers' needs they released it. They knew their customers needed a secure and comprehensive platform in which to find all the MP3 music they wanted: so they released iTunes as a side service of the iPod. They knew the product would have been completely useless if it was not able to hold more than a few songs in it: they packed the iPod with a 5GB Hard Drive memory that could store thousands of songs. Finally, they packed everything in a futuristic design, and market dominance was served.

Apple has been implementing the same approach in the Smartphone industry since the launch of the first iPhone, which tells a very explicative story.

The tech industry might not be an ideal environment for first movers. Of course, the amount of a firm's disposable resources play a crucial role in defining viable strategies to implement. Still, in an environment where timing makes up a good portion of the success of a product, even smaller firms should consider observing and learning before acting. This is especially true in a world where more prominent players can always use their vast resources to redesign, improve, and introduce better versions of similar technology developed by smaller brands.

In this perspective, the first step for every company aiming to obtain a comprehensive understanding of the market and its direction is to closely monitor the industry and the evolution of the devices that constitute the business environment.

Still, monitoring the competition is not always as simple as it looks. Monitoring the competition often assumes a short term horizon and is usually narrowed to the two or three direct competitors that mainly affect a firm's sales. At Motorola, closely monitoring the competition typically involves identifying one or two competing products for each of our devices and noting the competing brands' respective promotional activities. This way of doing this may even be effective in the short-term planning of our pricing model and price adjustments, but at the same time, it prevents the company from clearly seeing the bigger picture.

Regarding smartphone manufacturers, it is important to stay on top of the competition by understanding the strengths and weaknesses of other devices in the market. To do so, it is crucial to identify direct competing products of their devices regarding technical specifications. While the price point is undoubtedly an important factor, it can sometimes be misleading when it comes to identifying true competitors. Lower-priced devices may not necessarily have the same technical specifications as higher-priced ones, making it difficult to compare them accurately.

Instead, focusing on technical specifications such as the processor speed, camera quality, storage capacity, and screen resolution can help identify direct competing products. By looking at these technical aspects, manufacturers can determine which devices are the most similar in features and capabilities.

Furthermore, understanding the technical specifications of competing products can also help manufacturers improve their own devices. By identifying areas where their

devices may fall short, they can work to make improvements and remain competitive in the market.

Overall, while the price point is an important factor to consider, it should not be the only factor when identifying direct competing products of a smartphone manufacturer's own devices. Focusing on technical specifications can provide a more accurate and comprehensive understanding of the competition and help manufacturers stay ahead in the constantly evolving smartphone market.

The issue is that the industry as a whole is made of a tremendous amount of devices. And monitoring them manually would require an entire business unit solely dedicated to this purpose.

This project aimed to provide a proof of concept of a tool that can help smartphone manufacturers identify the direct competing products and eventually study their evolution over time.

The functioning of this script will be explained in depth in the following paragraphs.

4.3 Clustering results

This segmentation shows that the k-means algorithm used to analyze the data has provided a moderately consistent and reliable subdivision of the devices.

To illustrate the consistency of the technical specification within a few of the clusters, let us consider two examples. The eighth cluster contains 15 smartphones that all possess high-end technical specifications, including rear camera systems with over 200 megapixels, top-of-the-line or second-best processors, and at least 8GB of RAM. In contrast, cluster number 4 contains entry-level devices with low benchmark scores, low RAM and ROM memories, and includes 135 devices.

The labels of the different clusters is not ordinated. Meaning that higher clusters do not correspond to higher performances of the devices in it.

The algorithm's ability to create clusters containing different number of devices is further evidence of its reliability. It is reasonable to assume that the market has more low-end devices than high-end ones, and this is reflected in the distribution of devices across different clusters.

The analysis of the output also highlights how the value attributed the k, higher than what suggested by the elbow method plays a crucial role in the definition of consistently

similar clusters. Having an objective algorithm associate similar devices guarantees the absence of the personal prejudice of a human classifier. In fact, any human based smartphone ranking, would be subjected to the relevance that one attributes to a component instead of another. A photographer would, rightfully, evaluate a phone with a better camera system as generally better, even if lacking in other departments. A gamer would be much more interested in the device's RAM or processor, and so on. Still, the goal in question is not to find the best and worst phone but rather to recognize objective similarities while balancing the differences.

Of course, the output is not immune to outliers, but these are generally limited to one of the variables used to fit the model with. An example can be found in the camera system of the two Motorola models present in this cluster: Motorola moto x30 Pro and Edge 30 Ultra. They both present a megapixel count that is way higher than the other models they share the cluster with. The reason for this lies in two distinct facts. The first is that all the other specifications are very much in line with the ones of the other model, apparently,

having one specification above average is not enough to classify a smartphone in a different cluster. Secondly, the only other category these phones might fit into, is the cluster containing gaming phones, these devices though are generally much more powerful than traditional smartphones, as they were designed from the beginning to withstand much heavier computational efforts. Thus, it is actually further proof of the algorithm's good clustering performances, as it would not make sense to cluster Motorola models with gaming phones.

Name	Brand	Chipset	Main Camera Set	Front Camera	RAM	ROM	OS	5G	Cluster
Honor Magic 4 Ultimate	Honor	Snapdragon 8 Gen1 Qualcomm SM8450	50 Mp + 64 Mp + 64 Mp + 50 Mp	12 Mp F 2.4	12 GB	512 GB	Android 12 Magic UI 6.0	yes	8
Honor Magic 3 Pro Plus	Honor	Snapdragon 888 Plus Qualcomm SM8350-AC	50 Mp + 64 Mp + 64 Mp + 64 Mp	13 Mp F 2.4	12 GB	512 GB	Android 11 Magic UI 5	yes	8
Honor Magic 3 Pro	Honor	Snapdragon 888 Plus Qualcomm SM8350-AC	50 Mp + 64 Mp + 64 Mp + 13 Mp	13 Mp F 2.4	12 GB	256 GB	Android 11 Magic UI 5	yes	8
Huawei P50 Pro (Snapdragon)	Huawei	Huawei HiSilicon Kirin 9000	50 Mp + 40 Mp + 13 Mp + 64 Mp	13 Mp F 2.4	8 GB	512 GB	HarmonyOS 2.0	-	8
Huawei P50 Pro	Huawei	Huawei HiSilicon Kirin 9000	50 Mp + 40 Mp + 13 Mp + 64 Mp	13 Mp F 2.4	8 GB	512 GB	HarmonyOS 2.0	-	8
Motorola Edge 30 Ultra	Motorola	Snapdragon 8 Plus Gen 1 Qualcomm SM8475	200 Mp + 50 Mp + 12 Mp	60 Mp F 2.2	12 GB	256 GB	Android 12 MyUX	yes	8
Motorola Moto X30 Pro	Motorola	Snapdragon 8 Plus Gen 1 Qualcomm SM8475	200 Mp + 50 Mp + 12 Mp	60 Mp F 2.2	8 GB	128 GB	Android 12 MYUI 4.0	yes	8
Redmi Note 12 Explorer	Redmi	Dimensity 1080 MediaTek	200 Mp + 8 Mp + 2 Mp	16 Mp F 2.5	8 GB	256 GB	Android 12 MIUI 13	yes	8
Redmi Note 12 Pro Plus	Redmi	Dimensity 1080 MediaTek	200 Mp + 8 Mp + 2 Mp	16 Mp F 2.5	8 GB	256 GB	Android 12 MIUI 13	yes	8
Xiaomi 12T Pro	Xiaomi	Snapdragon 8 Plus Gen 1 Qualcomm SM8475	200 Mp + 8 Mp + 2 Mp	20 Mp F 2.2	8 GB	256 GB	Android 12 MIUI 13	yes	8
Xiaomi Mi 11 Ultra	Xiaomi	Snapdragon 888 Qualcomm SM8350	50 Mp + 48 Mp + 48 Mp	20 Mp F 2.2	12 GB	256 GB	Android 11 MIUI 12.5	yes	8
ZTE Blade A41 Ultra Extreme	ZTE	Snapdragon 8 Gen1 Qualcomm SM8450	64 Mp + 64 Mp + 64 Mp	16 Mp	12 GB	512 GB	Android 12 MyOS 12	yes	8
ZTE Axon 40 Ultra	ZTE	Snapdragon 8 Gen1 Qualcomm SM8450	64 Mp + 64 Mp + 64 Mp	16 Mp F 2	8 GB	256 GB	Android 12 MyOS 12	yes	8
ZTE Blade A31 Ultra	ZTE	Snapdragon 888 Qualcomm SM8350	64 Mp + 64 Mp + 64 Mp + 8 Mp	16 Mp	12 GB	512 GB	Android 11 MyOS 11	yes	8
ZTE Axon 30 Ultra	ZTE	Snapdragon 888 Qualcomm SM8350	64 Mp + 64 Mp + 64 Mp + 8 Mp	16 Mp	8 GB	256 GB	Android 11 MyOS	yes	8

Figure 9: Output example

On the other hand, without the intervention of the human expertise, we would neglect all the value that transcends the hardware component. Smartphones might have incredibly

large batteries but if they run an extremely demanding operating system. The device can still have a below average battery life, same goes for Mega Pixel count and camera quality, or for RAM capacity and multitasking performances.

5. DISCUSSION:

Comparing the result described in the previous section, with the ones obtained by similar projects, one can appreciate the differences in the scope of work of this project and the rest. As anticipated in the literature review, the end result of Han & Cho (2016). On the one hand Han and Cho (2016) include more variables in their analysis as they also deep dive in the benchmark analysis, which is something that may in the future be considered as a potential development of the current work. On the other it is their initial purpose that diverges from the one of this project. While this work was aiming at providing a practical tool for those that operate in this industry and its complexity on a daily basis, their approach was more connected to the study from an external point of view of the evolution of the products that populate the market as a whole. Even if the scope of this project does include the possibility of studying the evolution of the technology available of the market, this is done through the study of the modification of the characteristics of the constituents of each cluster.

The second similar work that was mentioned "Mining the automotive industry, A network analysis of corporate positioning and technological trends" also aims at identifying the macro trend of the market with the use of web scraping techniques. The authors of said paper are also reminding of the importance of market monitoring in highly dynamic environments like the automotive industry is becoming. Still their approach is again significantly different as their main focus, is the use of network analysis of car manufacturer web pages to reveal internal corporate positioning and innovative trends.

6. CONCLUSION

The project, since its beginning started as a proof of concept for the development of a practical tool to be used by any managerial figure in the smartphone industry. Specifically, in an environment such as Motorola that is characterized by a very small work force that manages rapidly increasing volumes, the high level analysis of the market is often neglected due to the massive amount of tasks that everyone is carrying out. In this perspective the development of a an automated tool that carries on itself the burden of monitoring the ingress of new products, classify them, and cluster them in pools of similar devices might be of great help not to lose the vision of the bigger picture. The final output, being in an excel format, is accessible by any kind of end user, which is the fundamental purpose of the project itself.

Overall, the report provides a reliable and fully automated analysis of the hardware components of smartphones, which can offer valuable insights into the current competition among smartphones in the market, solely based on hardware components and devoid of price components.

6.1 Theoretical Contribution

This work underlines the relevance of the use of data science methods to deepen the understanding of an industry. The qualitative analysis of a market remains crucial to understand the most critical aspect of a particular market. Still, the findings given by a thorough analysis of the data that defines a market are usually just as, if not more, valuable as it provides more objective explanations and solutions to the most critical aspect of managing a company in any given field.

Clustering methodologies, in particular, offer valuable insights into the products that constitute a market. Industries characterized by a vast range of diverse products can present significant challenges when it comes to identifying and prioritizing key elements. By applying clustering techniques, companies can categorize and group similar products together based on various parameters such as customer preferences, product features, or market demand. This process allows for a more comprehensive and structured understanding of the market landscape, enabling companies to allocate resources efficiently and tailor their strategies to specific product segments.

Furthermore, machine learning algorithms play a pivotal role in analyzing vast amounts of data to identify patterns, trends, and correlations that might not be immediately apparent through manual analysis alone. These algorithms can uncover hidden insights and provide valuable predictive capabilities. By training machine learning models on historical data, companies can make informed decisions based on data-driven forecasts, enabling them to anticipate market trends, customer behavior, and potential risks.

By incorporating clustering methodologies and machine learning into the analysis and decision-making processes, companies gain a more holistic understanding of their industry. The objective explanations and solutions offered by these data-driven approaches enhance management practices by providing actionable insights and reducing reliance on subjective interpretations. Moreover, the adoption of such methodologies fosters a culture of evidence-based decision-making, enabling companies to respond promptly to changing market dynamics and gain a competitive edge.

In conclusion, the utilization of clustering methodologies and machine learning techniques is instrumental in defining the key elements necessary to optimize company management across diverse fields. These advanced analytical approaches enable companies to navigate complex market landscapes, identify crucial factors for success, and make data-driven decisions that drive efficiency, growth, and profitability. Embracing data science methodologies becomes a strategic imperative for companies aspiring to thrive in today's rapidly evolving business environment.

6.2 Managerial Contributions

In practical terms, the clustering process's end result will significantly benefit companies like Motorola in various aspects of their daily operations. Firstly, utilizing automated tools allows for the efficient reallocation of resources to more operational tasks. Manually monitoring the emergence of new products and companies not only consumes a considerable amount of time but also increases the likelihood of human errors and mistakes. When dealing with large volumes of data and information, the probability of overlooking one, ten, or even more products becomes progressively higher as the dataset expands.

Secondly, by relying on automated processes, companies can eliminate subjective perspectives from the equation. Humans are naturally prone to biases, which can

compromise the objectivity of their analyses. Such behavior may result in the neglect of potential threats posed by a significant number of products simply because they are not held in high regard by the subject conducting the competitive analysis.

6.3 Limitations

The current version of the tool has several limitations. At the moment the price factor of the products is not considered in the clustering process, even though it clearly plays a crucial factor in the research for directly competing products.

Retrieving information about the current price of a smartphone is never easy as it firstly requires to define which price to consider. The minimum selling price of the week? The average price amongst the main retailers offline and online? A combination of these two values? And so on. Secondly, implementing price considerations in the clustering process implies that the web scraping script should be adapted to retrieve data from several different sources, each one with its own html structure. This process, besides being time consuming and complex to implement would also extend the running time of the program.

In fact, the script used to download the information of all the models in existence still takes more than 2 hours to complete its tasks. While this does not represent an issue by itself as the computer running the program is still usable for different activities, shortening the running time would still represent a positive achievement.

6.4 Future Work

There are many aspect of these project that should be carried on, and that were not covered more in depth because they did not entirely fit the practical scope of the project.

First of all the development of a proper graphical user interface that would allow anyone to launch the script and even specify which brand's phones data to scrape and focus on. At the moment even though the output is given in a very familiar format, it is still necessary to have Python installed on the terminal in order to run the program.

Secondly a proper statistical analysis of the different clusters evolution over time would enable to understand the technological trend that the market is following, as well as to identify outliers, brands that are building a product portfolio that is going against the current of the market and that may be about to disrupt the industry. As in the smartphone market the next great revolution is always around the corner.

REFERENCES

Abbas, O. A. (2008). Comparisons between data clustering algorithms. International Arab Journal of Information Technology (IAJIT), 5(3).

Agar, J. (2013). Constant touch: A global history of the mobile phone. Icon Books Ltd.

Android.com. (2023, March). Retrieved from https://www.android.com/everyone/#history-tab-panel-1

Aparicio, J.T., Aparicio, M., Costa, C.J. (2023). Design Science in Information Systems and Computing. In: Anwar, S., Ullah, A., Rocha, Á., Sousa, M.J. (eds) Proceedings of International Conference on Information Technology and Applications. Lecture Notes in Networks and Systems, vol 614. Springer, Singapore. https://doi.org/10.1007/978-981-19-9331-2_35

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.

Costa, C. J., & Aparicio, J. T. (2020). POST-DS: A methodology to boost data science. In 2020 15th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-6). IEEE. doi: 10.23919/CISTI49556.2020.9140932.

Costa, C. J., Silva, J., & Aparício, M. (2007). Evaluating web usability using small display devices. In Proceedings of the 25th annual ACM international conference on Design of Communication (pp. 263-268).

Dewalt, B. (2004). Jon Agar, Constant Touch: A Global History of the Mobile Phone. Material Culture Review 60 (1). https://journals.lib.unb.ca/index.php/MCR/article/view/18013.

Han, Q., & Cho, D. (2016). Characterizing the technological evolution of smartphones: insights from performance benchmarks. Proceedings of the 18th Annual International Conference on Electronic Commerce: e-Commerce in Smart connected World.

Jamalova, M., & Constantinovits, M. (2019). The comparative study of the relationship between smartphone choice and socio-economic indicators. International Journal of Marketing Studies, 11(3), 11.

Jamalova, M. &. Constantinovits, M (2020) "Smart for development: Income level as the element of smartphone diffusion." Management Science Letters 10, no. 5 1141-1150.

Kelley, R. (1988). In Praise of Followers. Harvard Business Review, 66, 142-148.

Mitchell, R. (2015). Web Scraping with Python: Collecting Data from the Modern Web (1st ed.). O'Reilly

Media, Inc.

Porter, M. E. (1979). How Competitive Forces Shape Strategy. Harvard Business Review, 57(2), 137-145.

Porter, M.E. (1980) Competitive Strategy, Free Press, New York, 1980.

Rego, L., Brady, M., Leone, R., Roberts, J., Srivastava, C., & Srivastava, R. (2022). Brand response to environmental turbulence: A framework and propositions for resistance, recovery, and reinvention. International Journal of Research in Marketing, 39(2), 583-602.

Shearer, C. (2020) The CRISP-DM Model: the New Blueprint for Data Mining, Journal of Data Warehousing, Volume 5, Number 4, page. 13-22.

Spina, G. (2012). La gestione dell'impresa. Etas.

Stoehr, N., Braesemann, F., Frommelt, M., & Zhou, S. (2020). Mining the automotive industry: A network analysis of corporate positioning and technological trends. In Complex Networks XI: Proceedings of the 11th Conference on Complex Networks CompleNet 2020 (pp. 297-308). Springer International Publishing.

Varriale, V., Cammarano, A., Michelino, F., & Caputo, M. (2022). The role of supplier innovation performance and strategies on the smartphone supply market. European Management Journal, 40(4), 490-502.

Zoogah, D. B., & Thomas, D. C. (2012). Apple's strategic followership: The iPod and the digital music revolution. The Journal of Business Strategy, 33(3), 13-19.