

**Czech University of Life Sciences Prague**

**Faculty of Economics and Management**

**Department of Management**



**Bachelor Thesis**

**Development of marketing strategies and plans based on  
clustering, cluster analysis and segmentation**

**Anton Barinov**

**© 2021 CZU Prague**

## BACHELOR THESIS ASSIGNMENT

Anton Barinov

Economics Policy and Administration  
Business Administration

Thesis title

**Development of marketing strategies and plans based on clustering, cluster analysis and segmentation.**

---

### Objectives of thesis

The main goal is to explain and perform the K-means clustering algorithm using python programming language, analyse the data set and clusters for further development of different marketing strategies and plans for each target segment.

The final goal of the thesis is to obtain a correctly clustered data set, analyse it and develop marketing strategies and plans for all obtained clusters, in order to strengthen market position, increase sales, brand name rate and customers involvement and loyalty.

### Methodology

The methodology of the thesis is based on analysis and study a range of current and relevant technical and scientific sources in the field of making marketing strategies techniques, marketing dataset clustering and its analysis.

Emphasis will be paid towards techniques of analysing and predicting customers behaviour, gathering them into patterns and development of marketing strategies. Based on this knowledge, the clusters will be created and analysed, making them appropriate for constructing a set of marketing strategies and plans effective for each target segment. Clusters will be formed by using few parameters with aim to obtain an appropriate number of smaller groups. Developing of marketing strategies for groups of customers will be based on knowledge gained from scientific researches in the studied area.

## **The proposed extent of the thesis**

Approx 40-50 pages

## **Keywords**

Marketing dataset clustering, Cluster analysis, Marketing strategy, Marketing plan, Consumer behaviour, Unsupervised machine learning

---

## **Recommended information sources**

Girish P., Stewart D.W., (1983), "Cluster Analysis in Marketing Research: Review and Suggestions for Application." Journal of Marketing Research, vol. 20, no. 2, pp. 134–148,  
Koch, C G., (2007), The Science of Success: How Market Based Management Built the World's Largest Private Company, John Wiley & Sons, ISBN13: 9780470139882  
Kotler P., (2016), Marketing Management, Harlow UK., Pearson Education, 832pp, ISBN-13: 978-9332557185

---

## **Expected date of thesis defence**

2020/21 SS – FEM

## **The Bachelor Thesis Supervisor**

Ing. Richard Selby, Ph.D.

## **Supervising department**

Department of Management

Electronic approval: 3. 3. 2021

**prof. Ing. Ivana Tichá, Ph.D.**

Head of department

Electronic approval: 4. 3. 2021

**Ing. Martin Pelikán, Ph.D.**

Dean

Prague on 15. 03. 2021

## **Declaration**

I declare that I have worked on my diploma thesis titled "Development of marketing strategies and plans based on clustering, cluster analysis and segmentation" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the bachelor thesis, I declare that the thesis does not break any copyrights.

In Prague on 15.03.2021

---

### **Acknowledgement**

I would like to thank my thesis supervisor Ing. Richard Selby, Ph.D. for his valuable advice and great support during my work on this thesis.

# **Development of marketing strategies and plans based on clustering, cluster analysis and segmentation**

## **Abstract**

This thesis is focused on development of marketing plans and strategies targeting specified segments of customers, based on K-means clustering algorithm and clustering results analysis. The main goal is to determine the key advantages of applying clustering methods in marketing and methodology of K-means, as well as situations in which it should be applied, and modern techniques for developing marketing strategies and plans. It studies modern approaches in marketing, specifically targeting to consumer markets, usage of customer data as segmentation variables and generational marketing. The following thesis examines the main stages of data collection, analysis and application of unsupervised machine learning clustering algorithms, segmentation, targeting and positioning, the impact of customer behavioural and demographical patterns on purchasing and brand preferences. The result of the thesis is a dataset, created for application of K-means clustering algorithm for marketing purpose, which will be executed using program code, and detailed analysis of its results. After identifying the key differences between target segments and their analysis, a marketing plan and strategy will be developed for each customer group, considering this knowledge and specifics of each segment.

**Keywords:** K-means clustering, Machine learning, Segmenting, Targeting, Positioning, Customer loyalty, Customer database, Generational marketing

# Vývoj marketingových strategií a plánů založených na shlukování, clusterové analýze a segmentaci

## Abstrakt

Tato práce se zaměřuje na vývoj marketingových plánů a strategií, zaměřených na konkrétní segmenty zákazníků, na základě shlukového algoritmu K-means a analýzy výsledků clusteringu. Hlavním cílem je určit klíčové výhody aplikace shlukových metod v marketingu a metodiku K-means, jakož i situace, ve kterých by měla být použita a moderní techniky pro vývoj marketingových strategií a plánů. Tato práce studuje moderní přístupy v marketingu, konkrétně cílení na spotřebitelské trhy, využití údajů o zákaznících jako proměnných segmentace a generační marketing. Následující práce zkoumá hlavní fáze sběru dat, analýzy a aplikace clusterových algoritmů strojového učení bez dozoru, segmentace, targetingu a positioningu, dopad chování zákazníků a demografické vzorce na nákup a preference značky. Výsledkem práce je datový soubor vytvořený pro aplikaci shlukového algoritmu K-means pro marketingové účely, který bude proveden pomocí programového kódu, a podrobná analýza jeho výsledků. Po identifikaci klíčových rozdílů mezi cílovými segmenty a jejich analýze, bude pro každou skupinu zákazníků vypracován marketingový plán a strategie s ohledem na tyto znalosti a specifiku každého segmentu.

**Klíčová slova:** K-means shlukování, Strojové učení, Segmentace, Targeting, Positioning, Věrnost zákazníků, Databáze zákazníků, Generační marketing

# Table of content

<b>1 Introduction .....</b>	<b>11</b>
<b>2 Objectives and Methodology .....</b>	<b>12</b>
2.1 Objectives.....	12
2.2 Methodology .....	12
<b>3 Literature Review.....</b>	<b>13</b>
3.1 Marketing and its purpose .....	13
3.2 Segmentation.....	14
3.2.1 Features and Benefits .....	17
3.2.2 Purpose.....	17
3.2.3 Price Sensitivity .....	18
3.2.4 Customer Loyalty.....	18
3.3 Clustering and its Purpose. Data Collection.....	19
3.3.1 Data Mining, Customer Databases and Big Data .....	20
3.3.2 The Difference between Customers and Consumers .....	23
3.3.3 Geo-clustering. Claritas PRIZM .....	24
3.3.4 The Difference between Clustering and Segmentation.....	26
3.3.5 Brief History of Clustering .....	26
3.4 Cluster Analysis .....	27
3.4.1 Generational Marketing .....	27
3.5 The Benefits of Targeting.....	30
3.5.1 Focusing and Multipurpose Targeting Strategies .....	31
3.5.2 Online Advertising and Targeting.....	32
3.6 K-means Clustering Algoritihm. Python.....	34
3.6.1 Python .....	38
3.7 Developing Marketing Strategies and Plans .....	38
3.7.1 Marketing Plan.....	38
3.7.2 Marketing Strategy.....	39
<b>4 Practical Part.....</b>	<b>42</b>
4.1 Collecting Data to Create a Dataset .....	42
4.2 Setting Segmentation Criteria .....	43
4.3 Dataset Analysis .....	44
4.4 Determining the Number of Clusters. The Elbow Method .....	47
4.5 K-means Clustering.....	49
4.6 Cluster Analysis .....	49
4.7 Developing Marketing Strategy and Plan .....	52



4.7.1	Executive Summary .....	52
4.7.2	Situation analysis .....	53
4.7.3	Strengths, Weaknesses, Opportunities, and Threat Analysis (SWOT).....	53
4.7.4	Product Offerings.....	54
4.7.5	Distribution .....	55
4.7.6	Marketing Strategy .....	55
4.7.7	Positioning .....	59
4.7.8	Strategies.....	59
4.7.9	Marketing Mix .....	60
4.7.10	Financials .....	61
4.7.11	Controls.....	61
<b>5</b>	<b>Results and Discussion.....</b>	<b>63</b>
<b>6</b>	<b>Conclusion.....</b>	<b>64</b>
<b>7</b>	<b>References .....</b>	<b>65</b>
7.1	Bibliography.....	65
7.2	Internet Resources .....	65
<b>8</b>	<b>Appendix.....</b>	<b>67</b>

## List of pictures

Figure 1.	<i>Online Advertising Markets: Revenue Growth Comparison.....</i>	33
Figure 2.	<i>Historical advertising market share .....</i>	34
Figure 3.	<i>Artistic Voronoi Diagrams in Python.....</i>	36
Figure 4.	<i>Voronoi Diagram. . . . .</i>	36
Figure 5.	<i>An example of clustering where K is chosen poorly .....</i>	37
Figure 6.	<i>First and last 5 rows of Hollywood Actors dataset in CSV format .....</i>	43
Figure 7.	<i>Step 1: Importing the libraries .....</i>	44
Figure 8.	<i>Step 2: Uploading the dataset .....</i>	44
Figure 9.	<i>Step 3: Checking the integrity of the dataset.....</i>	45
Figure 10.	<i>The distribution of net worth according to “Age” and “Gender” factors .....</i>	45
Figure 11.	<i>The correlation between “Age”, “Net Worth” and “Number of children” parameters .....</i>	46
Figure 12.	<i>The distribution of values in the dataset .....</i>	47
Figure 13.	<i>Code for Elbow Method .....</i>	48
Figure 14.	<i>Step 4: Determining the proper number of clusters using the Elbow Method ...</i>	48
Figure 15.	<i>Step 5: Plotting and visualising the K-means algorithm.....</i>	49
Figure 16.	<i>Results of K-means clustering .....</i>	49
Figure 17.	<i>Distribution of customers by clusters .....</i>	50
Figure 18.	<i>Distribution of net worth by clusters .....</i>	50
Figure 19.	<i>Division of customers by generations .....</i>	51
Figure 20.	<i>Proportion of net worth by generations .....</i>	51

## List of tables

Table 1. <i>Segmentation variables related to customers</i> .....	16
Table 2. <i>Segmentational variables related to products</i> .....	17
Table 3. <i>Customer Loyalty Matrix: priorities</i> .....	19
Table 4. <i>PRIZM clusters</i> . .....	25
Table 5. <i>Description of clusters by average age, number of children and net worth</i> .....	50
Table 6. <i>Description of dataset segmented by “Age” parameter</i> .....	52
Table 7. <i>Description of premium target segments</i> .....	56
Table 8. <i>Premium segment customer needs and corresponding benefit</i> .....	56
Table 9. <i>Description of luxury target segments</i> .....	57
Table 10. <i>Luxury segment customer needs and corresponding benefits</i> .....	57
Table 11. <i>Description of super luxury target segments</i> .....	58
Table 12. <i>Super luxury segment customer needs and corresponding benefits</i> .....	58

## List of abbreviations

CRM – Customer Relationship Management
DW – data warehouse
ERP – Enterprise Resource Planning
HH – household
MVCs – most valuable customers
NW – net worth
SCM – Supply Chain Management
SUV – sport utility vehicle
VQ – vector quantisation
WCSS – Within-Cluster-Sum-of-Squares

# 1 Introduction

With the rapid advancement of technology, the opportunities for marketers are greater than ever. Modern computing power of computers allows researchers to carry out the most complex operations and work with huge amounts of data. Programming is becoming an increasingly integral part of human life, being applicable in all industries and sciences, and marketing is no exception.

Modern approaches and techniques of clustering, segmenting and targeting change the way of marketing itself and allow marketers to identify, profile and reach desired customers in a multitude of new ways. With the growth of digitalisation and the online services sector, it has become possible to collect and analyse huge amounts of data of buyers and users from all over the world. Database marketing and data mining are implemented by companies from different countries and industries, opening up a wide range of new possibilities. Targeted online advertising is rapidly replacing mass advertising on TV, radio and newspapers, due to lower costs and higher efficiency – better customer response.

Clustering algorithms are applied in many areas, such as biology, machine learning, psychology, medicine and business. The essence of their mechanism of work is division of data into smaller groups based on identified similarities in its characteristics. There are many different algorithms that allow to perform clustering, one of them is K-means algorithm. Despite being a relatively old clustering technique, it is widely applied in marketing and data science and has proven itself to be effective among scientists.

All this leads to a change in the usual approaches to the development of marketing plans and strategies – marketing is becoming more global and individual at the same time. Analysis of customer data helps to develop a better strategy and offer for each segment, considering its specifics and characteristics. This requires an understanding of the needs and desires of different segments, as well as the factors that influence them.

This dissertation aims to explain and cover all the mentioned components, hopefully in a clear and simple way.

## **2 Objectives and Methodology**

### **2.1 Objectives**

The main goal is to explain and perform the K-means clustering algorithm using python programming language, analyse the data set and clusters for further development of different marketing strategies and plans for each target segment. The final goal of the thesis is to obtain a correctly clustered data set, analyse it and develop marketing strategies and plans for all obtained clusters, in order to strengthen market position, increase sales, brand name rate and customers involvement and loyalty.

### **2.2 Methodology**

The methodology of the thesis is based on analysis and study a range of current and relevant technical and scientific sources in the field of making marketing strategies techniques, marketing dataset clustering and its analysis.

Emphasis will be paid towards techniques of analysing and predicting customers behaviour, gathering them into patterns and development of marketing strategies. Based on this knowledge, the clusters will be created and analysed, making them appropriate for constructing a set of marketing strategies and plans effective for each target segment. Clusters will be formed by using few parameters with aim to obtain an appropriate number of smaller groups. Developing of marketing strategies for groups of customers will be based on knowledge gained from scientific researches in the studied area.

## 3 Literature Review

### 3.1 Marketing and its purpose

Definitions of marketing:

- “The activity, set of institutions, and processes for creating, communicating, delivering, and exchanging offerings that have value for customers, clients, partners, and society at large.”<sup>1</sup>
- In 1980 Kotler defined marketing as “the human activity directed at satisfying needs and wants through an exchange process” and a decade later his definition transformed to “a social and managerial process by which individuals and groups obtain what they want and need through creating, offering and exchanging products of value with others.”<sup>2</sup>
- “Marketing is an attitude of mind concerning customer satisfaction rather than a set of techniques simply to “sell” products or services.”<sup>3</sup>

There is no single generally accepted wording of what marketing is, but all definitions have somethings in common:

- Marketing is a management process.
- Marketing is about giving someone what they want.
- Marketing is identifying and anticipating customer requirements.
- Marketing fulfils customer requirements efficiently and profitably.
- Marketing offers and exchanges ideas, goods and services.<sup>2</sup>

“Good marketing is no accident. It is both an art and a science, and it results from careful planning and execution using state-of-the-art tools and techniques. In the face of a digital revolution and other major changes in the business environment, good marketing today is both increasingly vital and radically new.”<sup>4</sup>

Generally speaking, marketing is not just advertising or selling. It is applied in both public and private sectors, to both profit and non-profit making organisations, to goods, services and ideas. It includes promotion, sales, advertising and market research. Marketing should

---

<sup>1</sup> American Marketing Association definition (approved 2017)

<sup>2</sup> Kotler P. In Ward A. (2002). *Introduction to Marketing*, University of York

<sup>3</sup> McDonald M., Christopher M., Bass M. (2003). *Market segmentation. In: Marketing*, p. 41

<sup>4</sup> Kotler P., Keller K.L. (2016). *Marketing Management, 15th edition*, p. 25

not be the aim itself, it should help a company to achieve its broad, short- and long-term goals. The study of the market and surveys of customer segments and their response to goods and services can save firm a lot of time and money, provide with useful information about consumer needs and change their offers according to them. It is important to establish a strong communication with customers, employees and suppliers. A brand message should always be clear and targeted.

### 3.2 Segmentation

To develop marketing plans and strategies, it is highly important to know who your current customers are and who can be your customers in future. For this purpose, marketers use segmentation, a part of broader concept called *STP* (Segment, Target, Position).

Segmentation is a process of identifying and profiling customer groups with different respond toward competing proposals. Usually, in addition to the segmentation strategy, for each target segment is developed a competitive offer. To create a good segmentation strategy, marketers should conceptualise, develop and evaluate proposals for each customer group. The sharding<sup>5</sup> strategy should be assessed on following dimensions:

1. Can a competitive offer be developed and implemented to be attractive to the target segment?
2. Can it remain attractive to target customers in long-term despite despite the possible actions of competitors?
3. Is it profitable to develop this specific offer for this specific target segment?

“The success of a segmentation strategy is determined by the ability to take a dominant position within a limited market space, which competitors will not want or will not be able to attack.”<sup>6</sup>

The main difficulty, which face marketers during the segmentation process, is the variety of approaches to divide the market, literally there are hundreds of them. Usually, five or more variables are used in a segmentation analysis. To watch out for a convenient way to segment consumer markets, the marketer must consider and evaluate many variables that

---

<sup>5</sup> *Sharding* – a term from database architecture, the process of partitioning large tables into smaller pieces called “shards”. It is a horizontal breaking up of data used in search engines or databases. In this case – a synonym to segmentation.

<sup>6</sup> Aaker. D. (2004). *Strategic Market Management, 7-th edition*, p. 74

fall under one of two broad groups of variables – descriptive characteristics and behavioural considerations.

According to Kotler and Keller, descriptive and behavioural characteristics help marketers to find if they influence product responses, needs and wants of a segment.<sup>7</sup> Different descriptive patterns, such as demographic, geographic and psychographic characteristics affect customers sought benefits and wishes differently. For example, citizens of European countries prefer small cars due to lack of parking spaces and narrow medieval streets, while Americans prefer big ones. Different groups, such as “professionals” or “blue collars”, may react differently to a certain product benefit, like quality, low price or safety. Customers also differ according to behavioural considerations, such as brand preference, usage occasions or product responses. For example, conservative people are not inclined to buy “liberal” brands, like Tesla, but prefer “conservative” brands like Ford or Cadillac. People seeking for a certain benefit of the product may differ in their demographic, geographic and psychographic characteristics, marketers’ task is to define these differences and identify the segments.

“Regardless of which type of segmentation scheme we use – the key is adjusting the marketing program to recognise customer differences.”<sup>7</sup>

The most used Major Segmentation Variables for Consumer Markets are listed in Table 1. Aaker gives Toyota Scion model targeting Millennials, as an example of demographic characteristics that are researched and used by marketers. It is a brand entry model, created with the aim of introducing younger customers to the brand, as the average client of Toyota is 48 years old. Scion is a sports coupe, combining the reliability Toyota is famous for, and fresh attractive design to appeal to youth.<sup>8</sup>

“To create rumors about the “Scion” and make it attractive to the next generation of drivers, Toyota is targeting 15% of the target market for echo boomers who are considered “leaders and advisors” – that is, those who pushes his peers to transition to a new style in a variety of areas, be it music, sports or cars. If a new product is accepted in this group, then this serves as proof that other buyers will like it.”<sup>8</sup>

---

<sup>7</sup> Kotler P., Keller K.L. (2016). *Marketing Management, 15th edition*, p. 268

<sup>8</sup> Aaker. D. (2004). *Strategic Market Management, 7-th edition*, pp. 75 – 76

Geographic region	Prague, Central Bohemian, South Bohemian, Karlovy Vary, Moravian-Silesian, Vysočina, Plzeň
City or metro size	Under 5,000; 5,000–20,000; 20,000–50,000; 50,000–100,000; 100,000–250,000; 250,000–500,000; 500,000–1,000,000; 1,000,000–4,000,000; 4,000,000+
Density	Urban, suburban, rural
Climate	Northern, southern
Demographic age	Under 6, 6–11, 12–17, 18–34, 35–49, 50–64, 64+
Family size	1–2, 3–4, 5+
Family life cycle	Young, single; young, married, no children; young, married, youngest child under 6; young; married, youngest child 6 or older; older, married, with children; older, married, no children under 18; older, single; other
Gender	Male, female
Income	Under \$10,000; \$10,000–\$15,000; \$15,000–\$20,000; \$20,000–\$30,000; \$30,000–\$50,000; \$50,000–\$100,000; \$100,000+
Occupation	Professional and technical; managers, officials, and proprietors; clerical sales; craftspeople; forepersons; operatives; farmers; retired; students; homemakers; unemployed
Education	Grade school or less; some high school; high school graduate; some college; college graduate; post college
Religion	Catholic, Protestant, Orthodox, Jewish, Muslim, Hindu, other
Race	White, Black, Asian, Hispanic, Other
Generation	Silent Generation, Baby Boomers, Gen X, Millennials (Gen Y)

**Table 1.** Segmentation variables related to customers

Source: Kotler P., Keller K.L. (2016). *Marketing Management, 15th edition*, p. 269

The segmentation variables related to the product or its usage, such as consumption activity, are also important and influence the business model. A food company, for example chicken egg producer, will select different strategies when selling to supermarkets that buy eggs in large quantities and to restaurants and cafes with smaller demand. A tools



producer may develop a special line of products for a large chain like The Home Depot and sell other lines through smaller distributors using a different brand name.

The characteristics related to the product or its usage are shown in Table 2.

Type of user	Equipment buyer: builder, homeowner, repair company
Frequency of usage (Consumption activity)	Concerts: season ticket holders, casual visitors, non-users
Seeking Benefits	Dessert consumers: those who are more concerned with calorie content and those who are more interested in usability
Price sensitivity	Buyer of low-cost Volkswagen Polo and buyer of expensive Mercedes-Benz
Competitors	Consumers of products of rival firms
Intended use	Professional chain saw users and homeowners
Brand loyalty	Lovers of Heinz ketchup and lovers of cheaper brands

**Table 2.** *Segmentational variables related to products*

Source : Aaker, D. (2004). *Strategic Market Management, 7-th edition*, p. 75

### 3.2.1 Features and Benefits

If a brand has excellence in some segmentation variable or has a superior offer, it should correspond with a selected target segment. Customers of airlines, for example, can be divided into those who are most interested in the comfort during the flight, price, the amount of baggage allowed for the flight, exclusive destinations, flight speed or meals on board. Strategies for each segment are fundamentally different. It is both highly important to be the best in what is needed and to find the right target market for a superior offer.

### 3.2.2 Purpose

Segmentation according to the method of use or purpose is widely applied, especially in industrial markets. Each customer pursues his own goals when choosing and buying a car for example. Customers, who want to transport things, would rather buy a truck than sedan. For people with big families the most important factors are safety and capacity, probably they would buy a crossover. For taxi firms – the service and buying costs. For people

involved in races or other auto sports – the speed and technical characteristics, they would select among sports cars. For different people selecting a smartphone the priorities are also different. Tech savvy customers will look for the most modern and technologically perfect model, active people interested in photography would select the model with a great camera, big memory and battery capacity, while gamers would buy the model with the best processor and video card. Taking all this into consideration, a firm should select to whom they want to sell their goods and research if it corresponds with what people need and how they use the product.

### **3.2.3 Price Sensitivity**

A good decision is to provide the good or service with a combination of benefits, such as modern design, excellent technical performance and relatively low price, as Honda brand did with its Civic and Accord models in 80s, which accompanied by a fresh and successful marketing campaign led to success of the brand on the US market. A balance of high quality and low price is beneficial for premium and mass-markets, but it is not applied in luxury market, as high price serves as a filter and stands for the status of owner and uniqueness. Usually, it is a clear division of customer of those seeking for the lowest prices and those who would like to pay more for higher benefits, such as additional features or higher quality.

“Shopping centers, for example, occupy a clear intermediate position between low-cost stores and upmarket department stores.”<sup>9</sup>

A good example is the automotive industry with low-cost cars like Renault’s Dacia and Toyota’s Scion on the one pole of the market, exclusive sports cars like Bugatti or luxury cars and SUVs like Bentley on the other. The “golden mean” of the auto-mobile market are luxury cars like Acura, Infinity and Lexus. In any case, the strategy of the business is defined by the specifics of the segment.

### **3.2.4 Customer Loyalty**

It is a highly important factor for the development of new products, marketing plan and strategy, resource allocation and positioning.

---

<sup>9</sup> Aaker. D. (2004). *Strategic Market Management, 7-th edition*, p. 77

According to Aaker, the profit that can be earned through increasing of loyalty is great. It is obviously that consumers become more and more profitable in long-term, as costs of attracting new customers is higher than of maintaining current. In some industries, such as banking, insurance, car maintenance, book publishing and credit cards, an increase in 5% in customer loyalty leads to an increase of buyers lifetime profitability in two times.<sup>9</sup> The key for increasing the loyalty of customers is constant communication, improvement of product and service quality, creative and customer-oriented approach for developing offers.

Customer loyalty to the brand can be represented in the form of a matrix, as shown in Table 3.

	<b>Switching from brand to brand</b>	<b>Outgoing</b>	<b>Loyal</b>
<b>Buy a product</b>	Middle	High	The highest
<b>Do not buy a product</b>	Low - middle	High	Low

**Table 3.** *Customer Loyalty Matrix: priorities*

Source: Aaker, D. (2004). *Strategic Market Management, 7-th edition*, p. 77

Each segment of the matrix has a different strategic priority for a company, from low to the highest. The offer and market plan should be developed corresponding with the priority of a certain group.

“As the loyalty matrix shows, a high priority should be given to expectant buyers, including consumers of competing firms’ goods. In practice, using this matrix means assessing the size of all six quadrants, identifying their representatives, and developing programs that positively influence your brand preferences and loyalty levels.”<sup>9</sup>

### **3.3 Clustering and its Purpose. Data Collection**

Segmentation is used when marketers know who their target customer is. If a business is selling luxury sports cars, it is logical that they should target young and middle-aged men with high level of annual income. But if going deeply, where does target customer live? People who live in areas with mountainous or other specific terrains, cold climate or in the countryside do not buy sports cars usually. When do they buy selected product? Does it

have seasonality, or it is sold with same volume through the whole year? If business is selling ice-cream it is obvious that sales drop at winter, but what if it sells a lot of goods? Where do they buy? In large stores like Walmart or in smaller local stores? As the number of factors and data increases, it gets too hard for human's brain to analyse these hundreds or even thousands of rows of numbers trying to find relationships between each, so here marketers use clustering.

Clustering is the process of identifying the relationships between data points and grouping them into clusters based on those relationships. For analysing different types of data are used machine learning and algorithms, so then the received groups can be segmented.

When a firm already has some clients, but do not know who their primary customers are and what do they have in common or want to have more deep research and probably find some new potential ways of growth unseen before, marketers can use clustering techniques.

### 3.3.1 Data Mining, Customer Databases and Big Data

Data mining is the process of retrieving useful data about segments, individuals and trends from the overall pool. Cluster analysis, as well as other techniques like neural networking and predictive modelling, finds application in data mining.

Many companies are gathering information about their customers to make their future experience better, propose more unique and attractive deals, interact with customers.

Information can be different, as well as methods of collecting this data. Marketing managers need timely and accurate reports on current sales.

“Walmart operates a sales and inventory **data warehouse**<sup>10</sup> that captures data on every item for every customer, every store, every day and refreshes it every hour.”<sup>11</sup>

Corporations like Apple and Google know more about people than their parents, friends, or spouse. They know when user wakes up, where he lives, when he goes to work and what music he is listening to on the way, which films he watches, where does he eat and who spends free time with. All this data is collected from smartphones, tablets, PCs and laptops,

---

<sup>10</sup> DW – a system in an organisation that is responsible for data analysis and reporting (performs business intelligence activities). Main sources of data are CRM, ERP and SCM. The aim of data warehouse is to give an organisation a clear understanding of past and current indicators and help to make better predictions based on analysis of available data.

<sup>11</sup> Kotler P., Keller K.L. (2016). *Marketing Management, 15th edition*, p. 91

with invention of smartwatches it is possible to analyse the indicators of body to know what user is feeling at any specific moment. When browsing web, users often meet files called “cookies”, which track their actions on the site or even outside one specific web page.

“Companies that make good use of “cookies”, records of Web site usage stored on personal browsers, are smart users of targeted marketing.”<sup>11</sup>

Social network platforms like Facebook and Instagram collect data of customers’ in-app purchases, followed brands and people, location and what is more annual income, even though this value is not specified by the user during registration.

The fact is that most buyers do not only accept cookies, but they are also satisfied to get personalised and targeted advertisements. All the data collected is stored in the form of a huge **customer database**<sup>12</sup>, used mainly for marketing purpose, to better know who you are and what you will most probably like and buy – the possibilities for marketers in Online sector are huge. Banks and credit card companies have a lot of data about their customers, not only the history of transactions, but also more personal information: names, ages, addresses, family composition and income.

Many companies, such as supermarkets, gas stations or clothes brands, introduce club cards, applications or other forms of membership. They collect information about what, for how much and how often people buy, so then they can develop and customisable offer – send customers discount coupons or vouchers on goods that they will most probably buy. The same technology is used in online stores’ carts. That allows firms to create unique strategies, which were too expensive just few decades ago, to create a different offer for closely every consumer. All of that requires a development of databases and software to store and process all these amounts of data.

Tesco is a great example of such approach. Every Club Card member receives discount coupons based on their monthly expenses in store, what is more, receive special offers, such as discounts on goods they would most probably buy. The special algorithm allows Tesco to improve these offers, so a new member may receive more, but less relevant discounts. The more transactions with the card, the more accurate become offers that customers receive, as an algorithm is learning and predicting. This approach stimulates

---

<sup>12</sup> Customer database – an organised collection of comprehensive information about individual customers or prospects that is current, accessible, and actionable for marketing purposes. (Kotler, Keller, 2016)

customer loyalty to the brand, as proposals become more personalised, it also helps the brand to try new strategies, like stimulating customers to buy more or buy complement goods. If customers get rewards or benefits for buying in a certain shop, they are more likely to buy specifically there. It is not a secret that remaining a current customer is much cheaper than finding a new one, so creating even a competitive small added value to an offer can help a firm to remain strong relationships with clients. In a highly competitive industries like food industries, offering more is gaining more.

Developing a mobile application “Clubcard Tesco” allowed brand to interact with younger groups of buyers, so instead of using plastic cards and paper coupons customers have everything in one app. In this application it is viewed how the expenses are transformed into future coupons, discounts, promotions and current campaigns, it is also possible to take part in surveys and get rewards. Switching to mobile marketing is a wise decision: “The redemption rate for mobile coupons (10 percent) far exceeds that of paper coupons (1 percent).”<sup>13</sup>

That is a great development in technologies, allowing firms to have a more personal and customisable interaction with their clients, increase their loyalty and an involvement in the life of brand. Increasing the interaction with customers is one of the main priorities in developing marketing today, as it helps consumers not to forget about your brand, without spending millions on the mass advertising. Such strategy is not for every company, but for the firms with big databases of their customers or high knowledge of their customers’ segmentation characteristics.

Peppers D. and Rogers M. were one of those developing an approach called “one-to-one” or “personalised” marketing. According to Kotler they defined 4 key steps for this strategy:

1. Business should not try to be all for all the customers – it is important to **segment** and **differentiate** customers in accordance with their current and potential value. **Profile** and **prioritise** customers, collect information on key segmentation variables to create an extensive and meaningful database.
2. “Differentiate customers in terms of (1) their needs and (2) their value to your company. Spend proportionately more effort on the **most valuable customers (MVCs)**. Apply activity-based costing and calculate customer lifetime value.

---

<sup>13</sup> Business Insider. (2013). *How Mobile Coupons Are Driving an Explosion in Mobile Commerce* [online]

Estimate net present value of all future profits from purchases, margin levels, and referrals, less customer-specific servicing costs.”<sup>14</sup>

3. Develop and perform **individual offers** for each target group. Engage and involve customers into the life of brand, provide them with full support and develop communication. Use Web pages, social media accounts or applications.
4. Offers, products, services should be **customised** in accordance with customer specifics, needs and wants.<sup>14</sup>

### 3.3.2 The Difference between Customers and Consumers

“The term “consumer” is generally interpreted to mean the final consumer, who is not necessarily the customer.”<sup>15</sup>

A brand selling goods or services to children should take into consideration needs and wants of both consumers – children, and customers – their parents. Parents act as the intermediate customers, as they do not consume the product themselves, but they are the ones who decide, so it is important to convince them of the appropriateness of the purchase.

It is highly important to understand if your customer is a final consumer of your good or service. People running a small business usually know their customers well, by sight, so if they would like to expand their business, many would prefer to have more clients similar to those they already have. Finding similarities in a current database may save firm time and money to target and reach the most valuable potential customers.

According to McDonald, it is also highly important to know where sales and profits come from not only to understand the current market positions, but also to assess potential market directions. Both in consumer and industrial markets it is highly important to develop a different offer to each market segment. Common picture for businesses is 20% of customers accounting for 80% of profits – a phenomenon called the Pareto principle (the 80/20 rule), so it is reasonable to profile and focus much more effort on such customers. The segments must be economically manageable and ‘prioritizable’. A company should spend proportionally more resources towards the most profitable segments, but it is

---

<sup>14</sup> Kotler P., Keller K.L. (2016). *Marketing Management, 15th edition*, p. 290

<sup>15</sup> McDonald M., Christopher M., Bass M. (2003). *Market segmentation. In: Marketing*, p. 42

important not to forget about rest 80% of customers, to identify other segments of the market and develop proposals for each of them.<sup>15</sup>

### 3.3.3 Geo-clustering. Claritas PRIZM

Geo-clustering is a method of combining different segmentation variables, demographical and geographical, to obtain smaller, better defined target segments.

PRIZM is a geo-clustering approach, a set of geo-demographic segments for the US developed by Claritas Inc. (from 2009 - 2016 a part of Nielsen Corporation). It classifies households in USA into one of the 68 consumer segments based on the household's buying preferences, 14 social and 11 **Lifestage groups**. Lifestage groups are divided according to the next three variables: **age**, a combination of education, income and occupation – all together called the “**socioeconomic rank**”, and the **family composition**.

PRIZM NE considers almost 40 factors from 5 broad categories:

1. Education and affluence.
2. Family life cycle.
3. Urbanisation.
4. Race and ethnicity.
5. Mobility.

PRIZM is used by marketers for better targeting, more specifically to get a portfolio of their customers and develop specific offers for them. The households are classified according to zip code. The PRIZM clusters have intuitive and descriptive labels, such as “Young & Influential”, “Executive Suites”, “Big Fish, Small Pond”, “Home Sweet Home” or “Park Bench Seniors”.

People in each cluster live quite similar lives: work on similar jobs, spend their free time in a similar way, drive similar cars and buy from similar brands.

“Geo-clustering captures the increasing diversity of the U.S. population.”<sup>16</sup>

---

<sup>16</sup> Kotler P., Keller K.L. (2016). *Marketing Management, 15th edition*, p. 268



<p><b>Money &amp; Brains</b></p>	<p>The residents of Money &amp; Brains seem to have it all - <b>high incomes, advanced degrees, and sophisticated tastes</b> to match their credentials. Many of these city dwellers are married couples with few children who live in fashionable homes on small, manicured lots with expensive cars in the driveway.</p>	<p>1.42% of US HH population  Median HH income: \$108,936  HH Age Range: 55+  Geography: Urban city  HH composition: Mostly without kids  HH Tenure: Homeowners  Education: Graduate Plus  Technology use: Average</p>
<p><b>Kids &amp; Cul-de-Sacs</b></p>	<p>Upper-middle-class, suburban, married couples with children, an enviable lifestyle of <b>large families in recently built subdivisions</b>. This segment is a refuge for college-educated, white-collar professionals with administrative jobs and upper-middle-class incomes. Their nexus of education, affluence, and children translates into large outlays for child-centred products and services.</p>	<p>1.36% of US HH population  Median HH income: \$91,063  HH Age Range: &lt; 55  Geography: Urban city  HH composition: Family Mix  HH Tenure: Most owners  Education: Graduate Plus  Technology use: Above average</p>
<p><b>Bohemian Mix (Connected Bohemians)</b></p>	<p>A collection of mobile urbanites, Connected Bohemians represent the nation's most liberal lifestyles. Its residents are a <b>progressive mix of tech savvy, young singles, couples, and families</b> ranging from students to professionals. In their funky row houses and apartments, Bohemian Mixers are the early adopters who are quick to check out the latest movie, nightclub, laptop, and microbrew.</p>	<p>1.57% of US HH population  Median HH income: \$53,504  HH Age Range: 25 – 44  Geography: Urban city  HH composition: Mostly without kids  HH Tenure: Renters  Education: College Graduate  Technology use: Highest</p>

**Table 4.** PRIZM clusters.

Source: GeoPath. *Market Segmentation* [online]

According to Kotler and Keller, geo-clustering helps marketers to profile the most valuable customers and identify which techniques and approaches will be the most efficient when marketing to these segments. Many firms target specific segments and develop their strategy based on this data, like Barnes & Nobles aiming “Money & Brains” and opening its shops in neighborhoods where they live. Another example is Hyundai, targeting districts dominated by “Pool & Patios”, “Kids & Cul-de-Sacs” and “Bohemian Mix”.<sup>16</sup>

### **3.3.4 The Difference between Clustering and Segmentation**

The main difference between clustering and segmentation is the grouping mechanism. In **segmentation** process we combine customers into groups **based on their similarities**, while in **clustering** process we are **searching for similarities** in customers to group and therefore segment them.

All customers can be divided into different clusters using different methods and approaches. The K-means clustering algorithm, which is used in this paper, is widely used among marketers nowadays. Clusters show, for example, which customer groups bring the main profit to your business and what do they have in common, so marketers, based on this information, can provide better, more specific strategies to attract profitable target groups and save advertisement budgets.

In opposite to business selling to business (B2B) or “Industrial market”, consumer market companies (B2C) sell mass consumer goods and services, for example clothes, real estate, cars or hotel rooms (hospitality). For them it is highly important to build a strong brand image through development of superior products or services, backing them with reliable performance and engaging communications. Although some marketers believe there are no huge differences in marketing for B2C and B2B, most say that approaches, perfect for one sector, should be modified or different ones used when working with another.

### **3.3.5 Brief History of Clustering**

The history of cluster analysis dates to 1932, when it was firstly used in anthropology by American scientists Harold E. Driver and Alfred L. Kroeber in their “Quantitative Expression of Cultural Relationships”. In 1938 and 1939 it was introduced to psychology by an educational psychologist J. Zubin and a behavioural psychologist R. Tryon respectively, but the world renown it got in 1943 for trait theory classification in

personality psychology – “The description of personality: Basic traits resolved into clusters” by Raymond B. Cattell.

There are many different models and approaches in clustering, K-means clustering refers to Centroid models, where each cluster is represented by a single mean vector. The term “K-means” was introduced in 1967 by James B. MacQueen in “Some Methods for classification and Analysis of Multivariate Observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics”. However, the idea was firstly mentioned by Jewish-Polish mathematician Władysław Hugo Dionizy Steinhaus in 1956 in his work “On the division of material bodies into parts”. The standard algorithm was developed by Stuart P. Lloyd of Bell Labs (nowadays owned by Nokia) in 1957 as a technique for pulse-code modulation (PCM)<sup>17</sup> but published as a journal article only a quarter of a century later – in 1982 under the name “Least square quantisation in PCM”. Basically, the same method was proposed by Edward W. Forgy in 1965 in his paper “Cluster analysis of multivariate data: efficiency versus interpretability of classifications”, that is why it is sometimes called the “Lloyd–Forgy” algorithm.

### **3.4 Cluster Analysis**

One of the most important things in clustering are good analysis and interpretation of results. A common technique used by marketers is an additional division of clusters based on profiling variables. It is used for a more detailed division of customers into groups. For example, if there are two observations in one cluster, which are similar in their income (segmentation attribute in this case), they may be totally different according to other factors, such as education, family composition, gender, age, some behavioural or geographical characteristics (profiling variables). Customers are therefore separated into smaller segments and marketers develop different plans and strategies for each of them.

#### **3.4.1 Generational Marketing**

Generational marketing is an approach of segmenting and targeting customers according to their age. People born within specified time ranges tend to be more similar in their customer behaviour and lifestyle in general.

---

<sup>17</sup> *PCM* – a method designed to digitally represent sampled analog signals, the standard form of digital audio for CDs and computers.

According to demography statistics, US population in 2021 is categorised within 6 main cohorts:

1. Generation Alpha:

- By McCrindle, the best way for marketing to this generation is through influencers.<sup>18</sup>

2. Generation Z

- *Zoomers* or *Internet Generation*, people born between 1997 and 2012, who faced the lockdowns, distant education, continuous digitalisation and robotisation.
- Majority are “digital natives”, as they grew up connected to the Web, but much less are digital literate. People of this generation tend to have worse communication skills and higher indicators of aggression.

3. Generation Y

- *Millennials* or *Echo Boomers*, people born between 1981 and 1996. They grew up with the invention of the Internet, mobile devices, and social media, became witnesses of 9/11 and the beginning of the “war on terrorism” and became adults when the Great Recession of 2007 – 2009 and COVID-19 crisis happened, scoring for historically high levels of unemployment among young people.
- Account for an annual spending power close to \$200 billion (as of 2016), but many of them have significant debts. Tend to be aware of global problems and threats, less religious than their parents, but still spiritual, highly socially conscious, less likely to get married and to have a family. In the US, young people of the 2010s were more likely to live with parents than to become a homeowner, but in 2020s they are entering their homebuying years and are the fastest-growing segment.
- According to Kotler and Keller they “want to be viewed as unique, original, and rebellious, tend to trust friends more than corporate sources of information, are more team-oriented and optimistic than Gen X. This group is receptive to cause marketing efforts.”<sup>19</sup>

4. Generation X

---

<sup>18</sup> McCrindle M., Fell A. (2020). *Understanding Generation Alpha* [online]

<sup>19</sup> Kotler P., Keller K.L. (2016). *Marketing Management, 15th edition*, p. 275

- *MTV generation*, born between 1965 and 1980, also called the *latchkey generation* due to increasing divorce rates and growing employment among women in late 60s and 70s. They grew up and became adults during the heyday of rebellious music genres as punk, heavy metal and grunge, as well as saw the end of the Cold War in 1991.
- “They were the first generation to find surpassing their parents’ standard of living a serious challenge.”<sup>20</sup> They honour the ability to overcome any difficulties and self-sufficiency, they see modern technologies as a useful tool, not as a barrier.
- The best offers for this group are those, where the benefits and value are clear, in research they are described as active, happy, and achieving a work–life balance, more pragmatic and individualistic than Gen Y. “As consumers, they are wary of hype and pitches that seem inauthentic or patronising.”<sup>20</sup>

#### 5. Baby Boomers

- People born between 1946 and 1964, who grew up and became adults during the Vietnam War era 1955 – 1975 (1964 – 1973 years of direct involvement of US military) and final separation of the world on two blocks with constructing the Berlin Wall in 1961, oil crisis in 1973, rise of hippies and times of British Invasion (the cultural phenomena when actors and singers from the UK, like the Beatles became extremely popular in the US and around the world).
- Many of them were active participants of political, social and cultural changes like “Counterculture of the 1960s”, five years ago they accounted for \$1.2 trillion in annual spending power and controlled  $\frac{3}{4}$  of the USA wealth.<sup>20</sup>
- According to Kotler and Keller, many representatives of this generation are tending to resist aging by any means and treat their 50s like new 30s. Moreover, many of them buy online, use social media and try new brands.<sup>20</sup>

#### 6. The Silent Generation

- Born between 1928 and 1945, called the “silent majority”, because people from this generation mostly do not express their opinion publicly, probably because of Cold War era experience.
- Relatively small cohort, because of Great Depression in 1920s and WWII in 1940s, when people had fewer children, many passed through the Korean War and lots became participants of the Civil Rights Movement in late 50s and 60s.

### 3.5 The Benefits of Targeting

Some of the customer characteristics do not affect the buying behaviour, while others, on the contrary, are correlated to it and each in a specific way. Target marketing studies and considers all these characteristics, as well as other available data about consumers and market. Each target segment has specific wishes and requirements. Marketers should highlight the segment, which have the best response to certain product or service, and develop an offer that will be the most beneficial specifically for this group.

“Identifying and uniquely satisfying the right market segments are often the key to marketing success.”<sup>20</sup>

Targeting is the second component of the **STP** concept, which should be applied comprehensively:

1. **Segmenting** – split market into an appropriate number of segments of customers, research their characteristics and differences in buying behaviour and preferences, calculate value and potential profit for developing a product or service for each segment.
2. **Targeting** – define the most preferable segments, there may be one or many of them. Develop a marketing strategy and plan, study customer’s response to competitive offers and probably cooperate with target customers to develop a better product.
3. **Positioning** – select the right strategy of marketing to selected segment. Design and deliver a message describing the benefits and features of the offer to the target audience, research customers’ and competitors’ response, make changes to marketing campaign according to it.

For smart investment of finance and resources, developing marketing plans and strategies, it is crucial to know who your potential customer is. The market is therefore separated on well-defined groups, members of each are similar in defined characteristics. The correct division of the market requires a deep survey. To make an advertisement message more efficient and effective marketer must take into consideration the specifics of advertising to different segments.

---

<sup>20</sup> Kotler P., Keller K.L. (2016). *Marketing Management, 15th edition*, p. 267

Marketing as a process or function is not the only way of understanding this wide term, it can also be defined as a business concept or philosophy – an alternative view of marketing. According to this approach the organisation should be customer oriented, meaning that customer's needs and wants should be taken into main focus, when designing, developing and manufacturing goods or services, as it will help a firm to reach its sales and profits goals.

According to Kotler and Keller companies need to find a segment that they will be able to serve in the best way, as it is impossible to reach all customers in wide and diverse markets. To select target segment marketers should study and analyse specifics of customers in each group, understand their needs and wants.<sup>21</sup>

### 3.5.1 Focusing and Multipurpose Targeting Strategies

In marketing there are two different strategies for targeting segments: “Focusing” and “Multipurpose”.

In the first one, the point is to **focus on a specific segment** of a market, for example as a chicken eggs producer we could concentrate only on supermarkets, ignoring semi-finished products manufacturers, restaurants and mayonnaise producers.

In the beginning of its history Walmart targeted small cities in central and southern states, opening branches there, as this segment was free of big competitors. This approach led to success and expansion of the brand, which gained sustainable competitive advantages (SCAs) from the experience of working in a rural area. Key SCAs that helped Walmart to succeed are an efficient and fast warehouse supply system, a cheap and motivated workforce, relatively inexpensive retail space, and a simple, practical management style.<sup>21</sup> SCA is the time-stable value, significance created by organisation for its customers within the framework of a single market strategy, based on a unique combination of resources and abilities, which cannot be repeated by competitors in long-term. SCAs must be built on the strengths and areas of excellence of the organisation.

Other examples of focusing strategy are Coca-Cola with sugar free Diet Coke targeting people aware of sugar content in food and McDonald’s with Veggie menu developed to aim vegetarians.

---

<sup>21</sup> Aaker. D. (2004). *Strategic Market Management, 7-th edition*, pp. 78 – 79

The multipurpose strategy, in opposite to the focusing one, is tending to **serve multiple segments** at the same time. A good example of this strategy is Volkswagen AG (VAG), more known as the Volkswagen Group. Their Volkswagen, Seat and Škoda brands are targeting the price sensitive segments of the market, Audi is targeting the middle of the market, while Lamborghini and Bugatti are focusing on the less price sensitive customers. Each of Volkswagen subsidiaries has unique growth strategy and targets specific consumer segment, that gives a company further advance:

1. Increasing brand shelf and market presence and retailer dependence.
2. Attracting consumers seeking variety, who may otherwise have switched to competitor brands.
3. Increasing internal competition within the firm.
4. Yielding economies of scale in advertising, sales, merchandising and physical distribution.

“Regardless of the industry, a multipurpose service strategy is chosen by the most aggressive firms. It is expensive to develop multiple strategies at the same time, and to justify such an approach it is usually necessary that they work together to produce greater results than individually.”<sup>23</sup>

A synergy of components in multipurpose strategy is therefore highly important. The product offered to the target market should correspond with needs and have a set of benefits in most important criteria for each target segment. Campbell's implements this approach by offering different tastes of their soups in different regions of the USA, depending on the preference of the target market. For example, in the southern states the brand promotes Creole soup, while in Texas and California they sell spicy cheese and nacho soup.<sup>23</sup>

### **3.5.2 Online Advertising and Targeting**

One of the greatest breakthroughs in marketing development for last decades is online advertising and especially online targeting. Digital sector, especially mobile advertisement remains the fastest growing in marketing, leaving radio, newspapers and TV far behind – in 2018 Internet advertising passed the \$100 billion mark for the first time in history.<sup>22</sup>

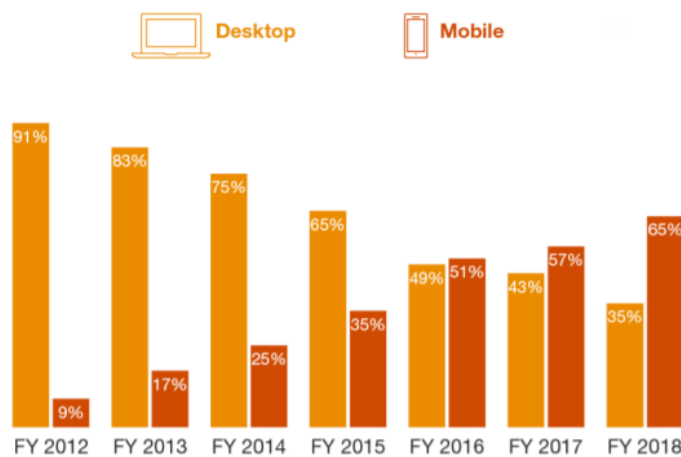
---

<sup>22</sup> IAB. (2018). *Price Waterhouse Coopers annual reports for the Interactive Advertising Bureau* [online]



Social media and other Web platforms are global marketplaces, where seller knows quite everything about the customers and can select to whom he wants to advertise. The involvement of audience is high, the costs are relatively low, the benefits are high. IT corporations such as Google or Facebook track users' actions on the Internet to collect data, based on which they can provide marketers with a higher level of response to their advertisements. For example, buyers searching for a baby chair, will see much more ads with this product. Moreover, they will also get many offers with goods that can interest people with a small child, such as children clothes, furniture or other related products. Successful online targeting requires a good knowledge of potential customers and their needs, but it can save firm a lot of money on advertising only to the right target audience. The technologies and environment, way and culture of doing business is changing, so the marketing is.

**Historical desktop vs. mobile trends, full year results**  
(\$ billions)



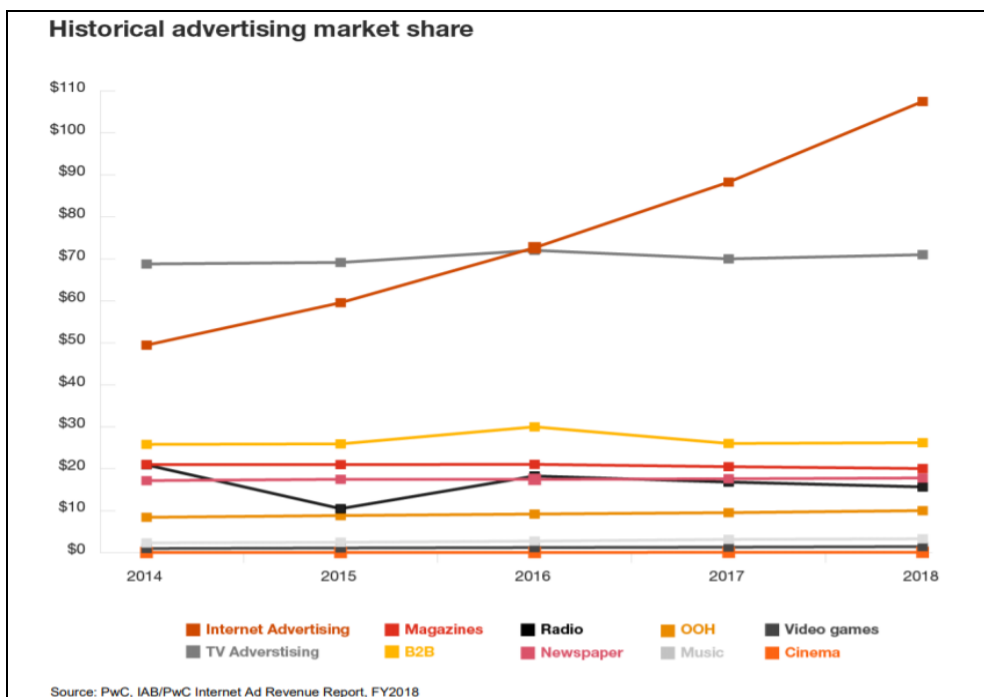
Source: IAB/PwC Internet Ad Revenue Report, FY 2018

**Figure 1. Online Advertising Markets: Revenue Growth Comparison**

Source: IAB. (2019). PWC annual reports for the Interactive Advertising Bureau [online]

Mass advertisement is too expensive and inefficient for smaller firms, which are much more interested in getting new customers, than in increasing average awareness of their brand. Online targeting is a good decision for such firms, allowing them to target only people who may most probably like and buy offered product. One of the greatest examples in online targeting is Facebook and its Instagram. With a relatively low advertising campaign budget compared to conventional advertising method and online targeting as an

amazing tool, it is logical that marketers are increasing their effort and creativity in this direction. **Facebook Ads Manager** allows anyone who makes his page a business account to select among lots of behavioral, demographical and precise geographical characteristics, tune his target audience according to their income and create multiple advertising campaigns with different targeting and budget settings. Ads Manager works with Facebook and Instagram as one system, so an advertiser decide if he wants an ad to be posted on both networks or just on one of them, as well as the place of publication – news feed, stories or an “interesting” section. It allows monitoring of key indicators of sucess of the marketing campaign, which are visualised and described in detail. All of these are just few options from the variety of actions which are possible with advertising and building brand trough social media.



**Figure 2.** *Historical advertising market share*

Source: IAB. (2019). *PWC annual reports for the Interactive Advertising Bureau* [online]

### 3.6 K-means Clustering Algorithim. Python

K-means clustering is a vector quantisation (VQ) technique, which was initially created for signal processing. Data from the data set and is being split on  $K$  – the number of clusters, the output is clusterised space – clusters, that define each data point. Data point is a unit of observation, data set consists of such point.

VQ is a method developed for signal processing, specifically for data compression, which uses the distribution of prototype vectors to create a model of probability density functions. The mechanism of work is that a set of points – vectors are split into groups, each has nearly the same number of vectors closest to it. Every group (cluster) is represented by its cluster centre. It is applied not only in K-means clustering, but also in some other algorithms.

Clustering can be based on different variables and help marketers to analyse a broad and complex variety of data. For example, commonly used strategies are **behavioural segmentation**, where clustering is performed based on customers activity, purchase history or other traits, and **demographical segmentation**, where clustering factors are age, location, income rate, ethnicity or other characteristics. It is also applied for internal use, such as inventory categorisation based on production metrics or sales rates. It is good to continue monitoring current clusters and adding new data, as the results of clustering will change with time, serving as an indicator for significant changes, transformation of the data.

The purpose of this **unsupervised machine learning algorithm**<sup>23</sup> is to **split  $n$  observations into  $K$  groups**, where each  $n$  belongs to the cluster with the nearest mean – cluster centre (cluster centroid), acting as a prototype of the cluster.

Hartigan and Wong defined the The aim of the K-means algorithm as:

“To divide  $M$  points in  $N$  dimensions into  $K$  clusters, so that the within-cluster sum of squares (WCSS) is minimised.”<sup>24</sup>

In other words, K-means clustering minimises squared Euclidean distances, as a result, the data space is separated into Voronoi diagram. WCSS is the sum of squares of the distances from each data point in all clusters to their corresponding cluster centre. The concept of WCSS is similar to nearest neighbour method, highly used in machine learning.

It is used for defining the optimal number of  $K$  for a given dataset. In case if the number of observations is equal to the number of clusters ( $n = K$ ), that will result in an ideal cluster, because the WCSS will be equal to 0, what has no sense because in this case each observation is a cluster itself. The Euclidean distance or WCSS between two observations

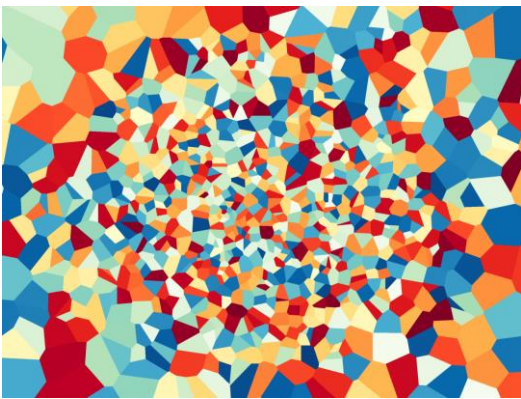
---

<sup>23</sup> *Unsupervised machine learning* – involves modeling the features of a dataset without reference to any label (in opposite to supervised machine learning, where must be some label associated with the data), and is often described as “letting the dataset speak for itself.” (VanderPlas, 2016)

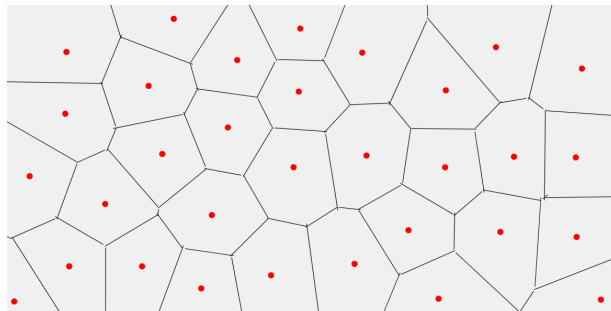
<sup>24</sup> Hartigan J., Wong M. (1979). *Algorithm AS 136: A K-Means Clustering Algorithm*

is the square root of the average of the square difference between the parameters of that observations. It is used for optimal division of data into clusters. Each data point can be visualised as a point on plot according to its parameters values (age, income or any other characteristics), the K-means algorithm calculates which of appointed cluster centers is the closest to each point using the WCSS. After performing the algorithm, new data can be easily added and assigned to the right cluster.

Voronoi diagram (sometimes called a Voronoi tessellation or partition) is a division of space into sectors, in which every section is close to each of a given set of points (objects). It was named after a Ukrainian mathematician Georgy F. Voronoy, who was the first to define this diagram in 1908, and has a wide range of applications in science, technology, and even visual arts.



**Figure 3 (on the left).** *Artistic Voronoi Diagrams in Python.* Source: Frank Ceballos. (2019). Mentor [online]



**Figure 4 (on the right).** *Voronoi Diagram.* Source : JavaLab [online]. (2019).

The K-means algorithm divides the data on cluster, but does not calculate the number of clusters –  $K$ . A commonly used approach of defining a proper number of clusters in a dataset is a heuristic called “**Elbow Method**”. The method was first proposed by an American psychometrician and educational psychologist Robert L. Thorndike in 1953. The key principle of this approach is that the goodness of fit of the model can be increasing until some limit. Any further addition of new clusters will not lead to better description of the data, but instead will dissolve the better fitting clusters as shown on Picture 1.

**The mechanism of K-means algorithm** is quite simple.

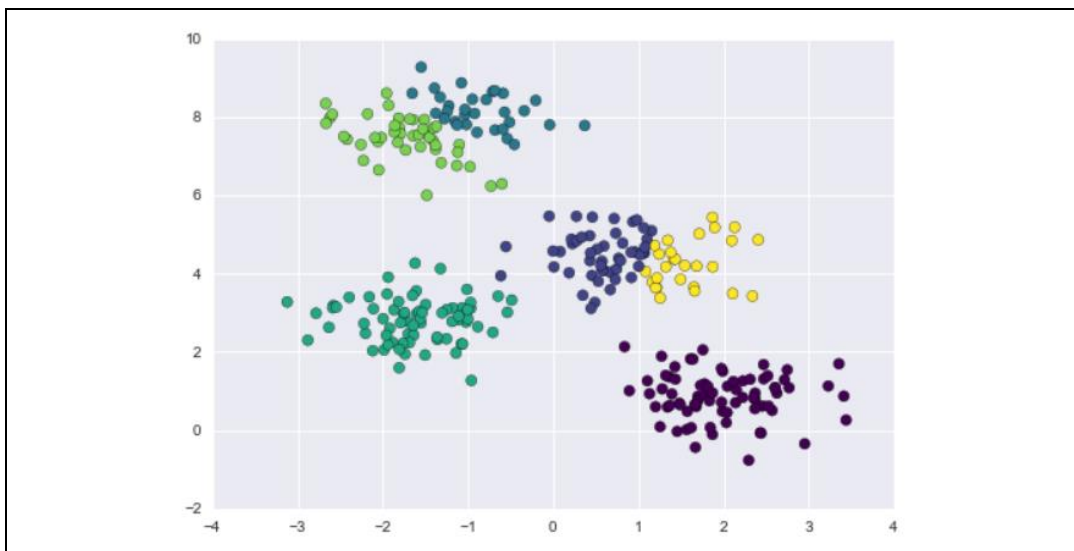
1. Firstly, the cluster centroids are estimated. They are randomly selected among the data objects in a data set or generated.
2. Secondly, the data is assigned to clusters.

According to VanderPlas, the conception of optimal clustering lies on two key hypotheses, which are the basis of K-means. The first assumption is that data is assigned to each cluster in such way, that each point is closer to its own cluster centre than to other cluster centres. The second one is that the “cluster centre” there is: “the arithmetic mean of all the points belonging to the cluster.”<sup>25</sup>

3. Finally, the cluster centroids are recalculated according to the arithmetic means.

“It is not practical to require that the solution has minimal sum of squares against all partitions, except when  $M, N$  are small and  $K = 2$ . We seek instead “local” optima, solutions such that no movement of a point from one cluster to another will reduce the within-cluster sum of squares.”<sup>25</sup>

Steps number 2 and 3 are performed until the the sum of the distances within each cluster (WCSS) is minimised and there are no chages in data appointment. That is why it is reccomended to perform the whole algorithm several times until there are no changes in clustering results, in order to improve the goodness of fit of the model.



**Figure 5.** An example of clustering where  $K$  is chosen poorly

Source: VanderPlas J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data*, p.468

Main advantages of K-means clustering method are high speed of execution, intuitiveness and easiness in learning. Main disadvantages are the need to manually determine the number of clusters and limitations, caused by separation of space into Voronoi cells.

---

<sup>25</sup> VanderPlas J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data*, pp. 463 – 470

### 3.6.1 Python

Python is a programming language created in the late 1980s and released in 1991 by Dutch programmer Guido van Rossum. According to TIOBE's index in December 2020 Python is the third of most popular programming languages, after C and Java, and is awarded as language of the year 2020.<sup>26</sup>

The main advantages are that it is:

- **Interpreted** – that means that the original code written in Python is not transformed into machine code but is executed via a special program – interpreter.
- **High-level** – in opposite to low-level C programming language Python is much easier for understanding and mastering for people who do not have deep knowledge in programming.
- **General-purpose** – it is designed for writing software in a wide variety of applications, Python is used for data analysis, developing applications, machine learning and so on.

Another advantage of using Python is that there is a variety of packages with different features and properties, that can be used in accordance to researcher needs.

A package is a set of functions, designed for specific purpose and application area, in simply words it is a tool. Some of them are written in low-level programming language to improve the script performance, which is better for low-level programming languages, allowing code written in Python to be executed faster.

## 3.7 Developing Marketing Strategies and Plans

### 3.7.1 Marketing Plan

A marketing plan is needed to provide **direction and focus for a brand, product, or company**. Any business will be better prepared to launch an innovative product or increase sales to current customers with a detailed plan. Most firms develop marketing plans that covers one year, however some of them create plans longer or shorter periods. A marketing plan is not a business plan, with its broad overview of company aims and ethics, it is a plan of applying different methods and techniques to achieve the firm's strategic objectives. An important factor is the ability of a business to implement the plans, as all departments must

---

<sup>26</sup> TIOBE. (2021). *TIOBE Index for Python* [online]

be ready to increase production, change financing to cover the expenses and add more staff to the team. Good implementation of the marketing plan requires perfect organisation of the company.

“The purpose of marketing planning is the identification and creation of competitive advantage.”<sup>27</sup>

As the starting point in developing marketing plans is customer, it is very important to **choose the target segment**. For this purpose, marketers use external research and internal databases. It helps not only to get detailed data about customers, but also to improve the results of their plan – more information about customer needs, expectations, desires and loyalty. As the company has started the realisation of their marketing plan, marketers continue to research the results of a plan, how good it is and what can be improved. Finally, marketers should focus on the research towards the general condition of the market, the tendencies, competitors, opportunities for growth and threats.

A marketing plan should also cover a spread variety of factors, as its implementation will lead to many changes in firm’s internal and external relations. It affects the relationship between marketing division and other departments inside the firm, as well as the relationships between the business and its external stakeholders – suppliers, partners, media sources, community as the whole and many others. All these relationships are highly important for a company to achieve success not only in a realisation of a certain plan, but for success in running a business in general.

To prepare a good marketing plan, marketers need time to have a research, analysis, organisational review and to coordinate other departments. Probably the most important factor in business is budget and its correct calculation. Good marketing plan usually should not exceed the budgets, and if exceeds – the achievements should be much greater than additional costs. Finally, it should be clearly defined how to measure the progress toward aims.

### **3.7.2 Marketing Strategy**

According to Kotler and Keller, following a successful concept test, the new-product manager will develop a preliminary three-part strategy plan for introducing the new product into the market. The first part describes the target market’s size, structure, and

---

<sup>27</sup> McDonald M., Christopher M., Bass M. (2003). *Market segmentation. In: Marketing*, p. 42

behavior; the planned brand positioning; and the sales, market share, and profit goals sought in the first few years.<sup>28</sup>

A strategy plan should consist of three parts:

1. The first part describes the target market and its size, the aims – expected sales, revenues and profits in first few years, the planned brand positioning.
2. The second part aims to describe the planned price of a good or service, marketing budget and distribution strategy for the first year.
3. The third part outlines the profit, sales and market share objectives, as well as **marketing-mix** strategy in the long-run.

Marketing-mix is a classification designed by E. Jerome McCarthy in his book “Basic Marketing: A Managerial Approach” in 1960. It groups different marketing activities into four classes, so called **the four Ps (4Ps)** of marketing:

- **Product** – product variety, quality, design, features, brand name, packaging, sizes, services.
- **Place** – channels, coverage, assortments, locations, inventory, transport.
- **Promotion** – sales promotion, advertising, sales force, public relations, direct marketing.
- **Price** – list price, discounts, allowances, payment period, credit terms.

Combined with modern understanding and defining of 4Ps, they create an extended marketing mix, so called “**eight Ps**” (**8Ps**), especially popular in the service marketing:

- **People** – training and recruitment of new staff, queuing systems, service errors, scripting, uniforms, waits management, resolution of complaints, managing social interactions.

It is a more broad and sophisticated view on customers, as people, not buyers and the fact that employees are the key to marketing success.

- **Processes** – service processes design, developing and measuring key performance indicators (KPIs), diagnosing failures and critical incidents, analysing resource requirements and its distribution, customisation vs standardisation solutions, service performance tracking and monitoring.

---

<sup>28</sup> Kotler P., Keller K.L. (2016). *Marketing Management, 15th edition*, p. 471



It is a way – an amount of processes by which service is delivered. Avoid templates and ad hoc planning. Bring the creativity, the state-of-the-art ideas and concepts to marketing.

- **Programs** – an importance of synergy of all marketing consumer-directed activities.
- **Performance** – the range of possible outcomes, such as profits, brand and customer equity, social responsibility, environment, legal and ethical implications.

Marketing strategy is highly connected with **marketing channel system**. There are two common strategies, but the best marketers use both. They are called a “**push**” and “**pull**” **strategies**. By Kotler and Keller, the push strategy uses the producer’s sales force or other means, so that the products are carried, promoted and sold to the final customer by intermediaries. This strategy better fits markets with goods that are bought impulsively, product benefits are clear, customers do not have preferred brands or brand loyalty is low. In the pull strategy, on the opposite, manufacturer advertises and promotes its products to increase customer awareness and force intermediaries to order and sell it. It is most benefiting for markets with goods, where products and brands are studied and selected beforehand, customers show interest and high brand loyalty.<sup>29</sup>

---

<sup>29</sup> Kotler P., Keller K.L. (2016). *Marketing Management, 15th edition*, p. 516

## 4 Practical Part

### 4.1 Collecting Data to Create a Dataset

There are many public sources where you can find data sets, Kaggle<sup>30</sup> is maybe the most famous one. It is an online community, owned by Google since 2017, that is studying machine learning and data science, and where scientists and programmers can share and receive feedback on their studies. Kaggle helps novices to master basic techniques and methods applied in different areas, providing each case study with notes, dataset and written code. Most of datasets published on Kaggle have no application in marketing, as they are artificial and made for educational purposes only, to learn how to work with datasets. Companies usually do not publish information of their customers publicly – it serves for the internal use only.

To show how the K-means clustering algorithm works, there was created a dataset of randomly selected 256 Hollywood actors and actress, 128 of both respectively. All the data used for creation this dataset was taken from open sources and does not violate General Data Protection Regulation (GDPR) or any personal rights. It was decided to choose celebrities, since there is a lot of information about them in the public domain, so there are many criteria to analyse. This data set consists of the following factors:

- ID – number of rows from 1 to 256, each represents a single actor/actress
- Age
- Gender
- Net worth, Millions \$ - the total wealth or welfare, including all financial assets and liabilities.
- Number of children
- Name

The data set was created using Microsoft Excel in CSV UTF-8 (Comma delimited) format – (\*.csv), as it is a common format used by data scientists to work with datasets.

CSV is a text file, where values are separated by commas and each line of file is a data record. In this case, each record consists of 6 fields. A CSV file typically stores tabular

---

<sup>30</sup> Kaggle. Main page [online]

data (text and numbers) in plain text, each row has the same number of fields. Each point on a graph corresponds with a specific datapoint – actor or actress.

ID	Gender	Age	Net Worth, Millions \$	Number of children	Name
1	Male	78	300	5	Harrison Ford
2	Male	47	50	2	Neil Patrick Harris
3	Male	72	16	2	Jeremy Irons
4	Male	41	10	2	Oscar Isaac
5	Male	47	7	2	Patrick Wilson
...	...	...	...	...	...
252	Female	85	30	5	Julie Andrews
253	Female	24	10	0	Hailee Steinfeld
254	Female	50	16	1	Minnie Driver
255	Female	74	50	0	Liza Minnelli
256	Female	60	50	2	Julianne Moore

256 rows × 5 columns

**Figure 6.** First and last 5 rows of Hollywood Actors dataset in CSV format

## 4.2 Setting Segmentation Criteria

The main parameters by which the clustering will be performed are “Age” and “Net Worth” – so called “Segmentation variables”. The correlation between “Age”, “Net Worth” and “Number of children”, where “Number of children” is a profiling attribute – a kind of demographic and behavioural data used for profiling the gained clusters, should be computed and analysed.

The clustering will be performed via code written in **Python** programming language, using several packages commonly used for scientific research and in particular for K-means clustering:

- **NumPy**
- **Pandas**
- **MatPlotLib**

All of them are part of **SciPy** – free and open-source Python library used for scientific and technical computing, originally created and released in 2001 by T. Oliphant, P. Peterson and E. Jones.

For attractive and informative visualisation of the results, a **Seaborn** data visualisation library, based on MatPlotLib, is used.

```
import numpy as np
from sklearn.cluster import KMeans
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

**Figure 7.** Step 1: Importing the libraries

The code will be written using **Google Colaboratory**, or “Colab” for short.

It is a platform created by Google Research, that allows anyone to write Python code on their site through the browser and execute it on Google’s servers. It is a free Jupyter notebook environment used by developers and researchers for data analysis, machine learning and education. The key advantage is that for code execution are used a single 12GB NVIDIA Tesla K80 GPU and TPU, which have extremely high computing power. They can be used up to 12 hours (1 session) and then have to be restarted, which is enough for most cases. Another plus is that it runs entirely in the cloud, notebooks are saved on Google Drive, allowing people from different cities or countries to work on same project, the only thing needed is Google account. The dataset can be uploaded into Colab to work with it, the only inconvenience is that it should be reuploaded after each session to continue working on it. The code is saved in a form of Jupyter Notebook.

```
df=pd.read_csv('/content/ActorsDataset.csv', index_col=0)
```

**Figure 8.** Step 2: Uploading the dataset

### 4.3 Dataset Analysis

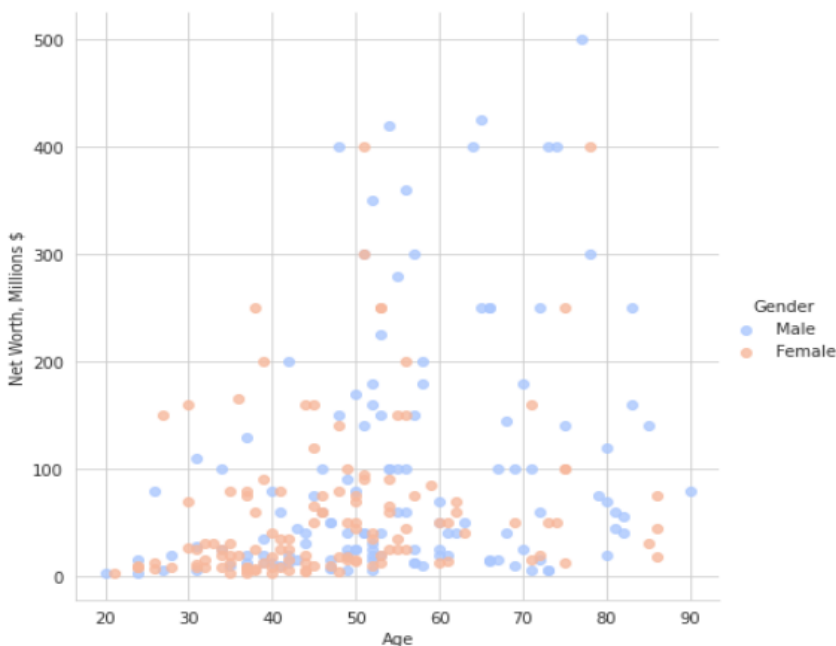
Next step is to check the dataset using the *df.info()* command as shown on Picture 6 to be sure that data is complete and there are no missing values.

From the Figure 10 it is seen that among observations with “Net Worth” value higher than \$200 million there tend to be more than twice more actors than actress. It also seen that there are more young women than man and that the youngest celebrity to pass \$200 million is a woman. Among the  $\geq$  \$100 million sector there are twice as many males as females.

```
df.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 256 entries, 1 to 256
Data columns (total 5 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Gender                                256 non-null    object
1   Age                                    256 non-null    int64
2   Net Worth, Millions $                 256 non-null    int64
3   Number of children                    256 non-null    int64
4   Name                                  256 non-null    object
dtypes: int64(3), object(2)
memory usage: 12.0+ KB
```

**Figure 9.** Step 3: Checking the integrity of the dataset



**Figure 10.** The distribution of net worth according to “Age” and “Gender” factors

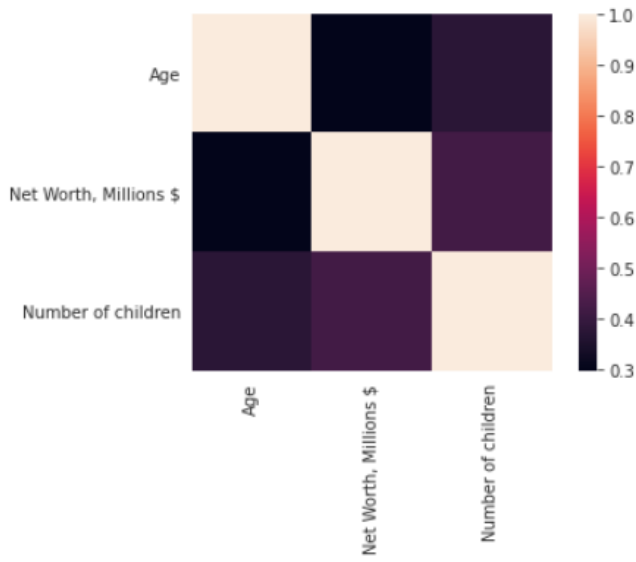
As seen from Figure 11, there are **no strong correlations** between any of the variables.

The strongest correlation is between “Number of children” and “Net Worth” factors ( $r = \sim 0.42$ , meaning that people with more children tend to be richer).

The average number of children for actors is 2.12, for actresses – 1.47. This difference is caused by the predominance of younger female and some of older actress not having children at all, most probably because of career orientation and busyness inherent to their lifestyle or personal beliefs.

An extreme value for this dataset is Mel Gibson (net worth – \$425 million) has 9 children. The richest actor – Robert De Niro (net worth – \$500 million) has 6 children.

The mean number of children in dataset is 1,79 what is close to the average value in US – 1.93.<sup>31</sup> The average Net Worth is \$79.72 million and only 25% of people in the data set have Net Worth > \$100 million.



**Figure 11.** The correlation between “Age”, “Net Worth” and “Number of children” parameters

The second one is the correlation between “Age” and “Number of children” parameters ( $r = \sim 0.38$ ).

It seems reasonable that most of young actors have no children, the point is that the average age of family building is increasing for last decades. In fact, the average age of people having their first child in developed countries is growing constantly: in USA in 1972 it was 21 and 27 for women and men respectively, while in 2016 it was 26.3 for women and 31 for men.<sup>32</sup>

The clear division is seen between urban and rural areas – people in countryside tend to have children much earlier than their peers from cities. The difference is also influenced by the education level: the average age of having first baby is 30.3 for women with college degree and 23.8 for women without, basically people with good education are having children later, as they need to finish college and build a career before entering the marriage. Marketers should take these facts in consideration when developing marketing plans and strategies nowadays, as this factor is influencing an approach of having a baby and needs of parents. Older parents are more likely to invest more money into children’s

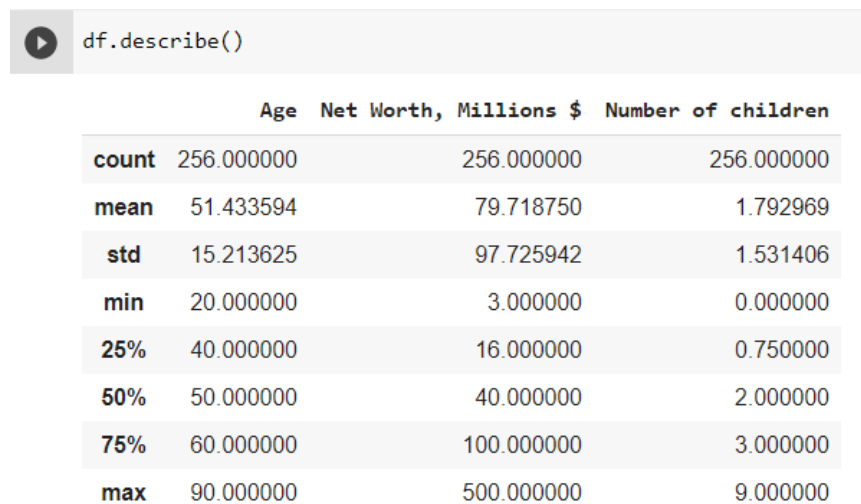
<sup>31</sup> Statista. (2020). Average number of children in US [online]

<sup>32</sup> Bui Q., Miller C. C. (2018). The Age That Women Have Babies: How a Gap Divides America [online]

education and their children in general, buying more expensive toys and clothes, paying for private school and university, etc.

The lowest correlation is between “Age” and “Net Worth” variables ( $r = \sim 0.3$ ).

It means that there is no clear dependence between richness of an actor/actress and their age. However, the average age of an actor/actress with “Net Worth”  $\geq$  \$200 million is approximately 62 years, what is more than an average age of actors in a data set – 51. The average age of the top-10 richest people in this data set is 64. That means that in general, the correlation between two variables is weak, but when having a detailed look, it is possible to find some tendencies, as that all the people richer than \$200 are older than 38, as well as all the people richer than \$300 million are not younger than 48 in this dataset.



**Figure 12.** *The distribution of values in the dataset*

Using the command *describe()* as shown on Figure 12, it is possible to visualise the standard deviation (std), mean, minimum and maximum values (min and max extremes), as well as the distribution of values according to 1<sup>st</sup> (25%), 2<sup>nd</sup> (50%) and 3<sup>rd</sup> (75%) quadrant. The values are distributed unequally, the average actor/actress in this dataset is 51.5 years old, has a Net Worth of \$79.7 million and has 1 or 2 children.

#### 4.4 Determining the Number of Clusters. The Elbow Method

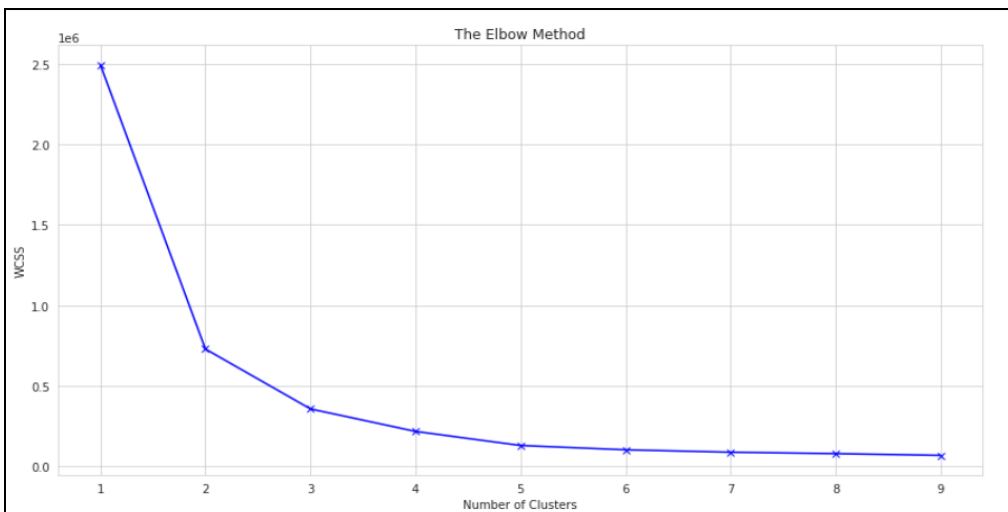
The idea of this method is using the elbow of the graph to determine the  $n$  of  $K$ , using an elbow as a cut-off point. For this, the explained variation is plotted as a function of  $n$  of  $K$ . The growing number of clusters will explain more of the variation of the data, called improvement of the fit, until some value of  $K$ , when it will become over-fitting, meaning that adding  $n$  of  $K$  is useless. This point is seen on the curve of graph, the last elbow is used

as the optimum for  $n$  of  $K$ . If the  $K$  will be greater than the optimal value, there will be an unnecessary division of actual groups of data, the split of already well-defined clusters, leading to poor results.

```
wcss = []
K = range(1,10)
for k in K:
    kmeanModel = KMeans(n_clusters=k)
    kmeanModel.fit(X)
    wcss.append(kmeanModel.inertia_)
```

```
plt.figure(figsize=(14,7))
plt.plot(K, wcss, 'bx-')
plt.xlabel('Number of Clusters')
plt.ylabel('WCSS')
plt.title('The Elbow Method')
plt.show()
```

**Figure 13.** Code for Elbow Method



**Figure 14.** Step 4: Determining the proper number of clusters using the Elbow Method

With an increase of  $n$  of clusters –  $K$  the WCSS value decreases, so we should select the  $K$  according to the level of decrease in WCSS. In Elbow Method the K-means algorithm is initialised for a range of  $K$  values, usually from 1 to 10, and plot it against the WCSS for each value of  $K$ . Basically, it means that we should select the point at which diminishing returns are no longer worth the additional cost, so the aim is to choose a  $K$  after which the decrease in WCSS is minimal – the last elbow.

To create and visualise the graph showing the “elbows” the code on Figure 13 was used. The drop in WCSS from cluster 1 to 2 is significant as seen on Figure 14. After the 3<sup>rd</sup> cluster the decrease is minimal, therefore the optimal value for  $K$  will be 3.



## 4.5 K-means Clustering

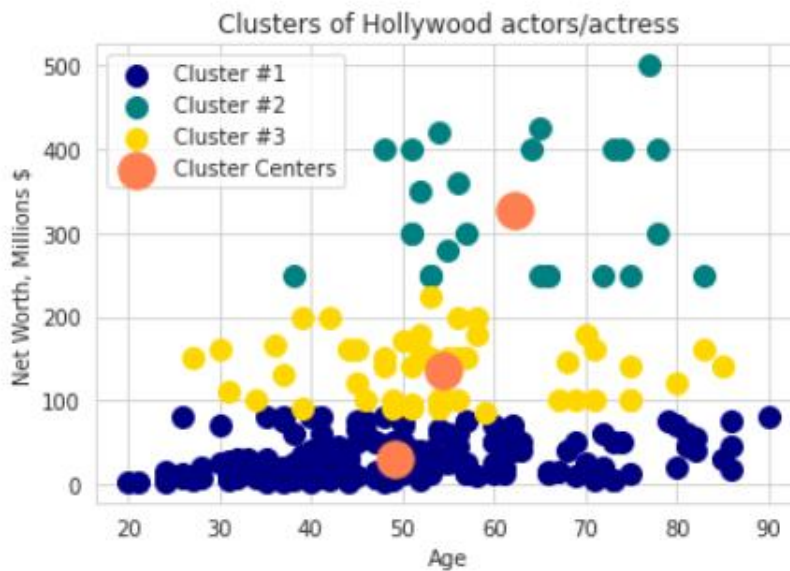
The K-means clustering algorithm was performed using the code on Figure 15.

```
kmeans = KMeans(n_clusters= 3, init='k-means++', random_state=0)
pred_y= kmeans.fit_predict(X)

plt.scatter(X[pred_y == 0, 0], X[pred_y == 0, 1], s = 100, c = 'navy', label = 'Cluster #1')
plt.scatter(X[pred_y == 1, 0], X[pred_y == 1, 1], s = 100, c = 'teal', label = 'Cluster #2')
plt.scatter(X[pred_y == 2, 0], X[pred_y == 2, 1], s = 100, c = 'gold', label = 'Cluster #3')
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s = 300, c = 'coral', label = 'Cluster Centers')
plt.title('Clusters of Hollywood actors/actress')
plt.xlabel('Age')
plt.ylabel('Net Worth, Millions $')
plt.legend()
plt.show()
```

**Figure 15.** Step 5: Plotting and visualising the K-means algorithm

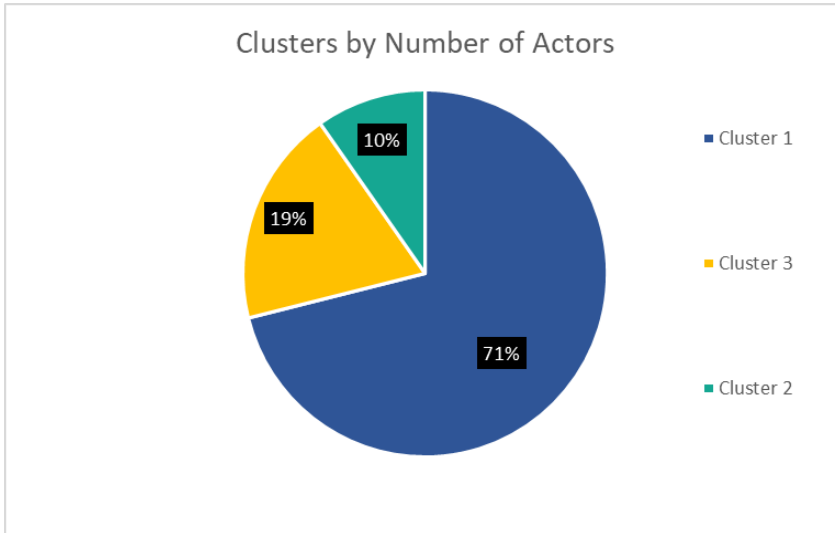
As a result, three clusters were obtained (Figure 16). The distribution varies significantly: Cluster 1 obtains 182 actors and actresses with net worth from \$3 to 80 million, Cluster 3 – 49 with net worth \$85 – 200 million, and Cluster 2 – 25 with net worth \$250 – 500 million.



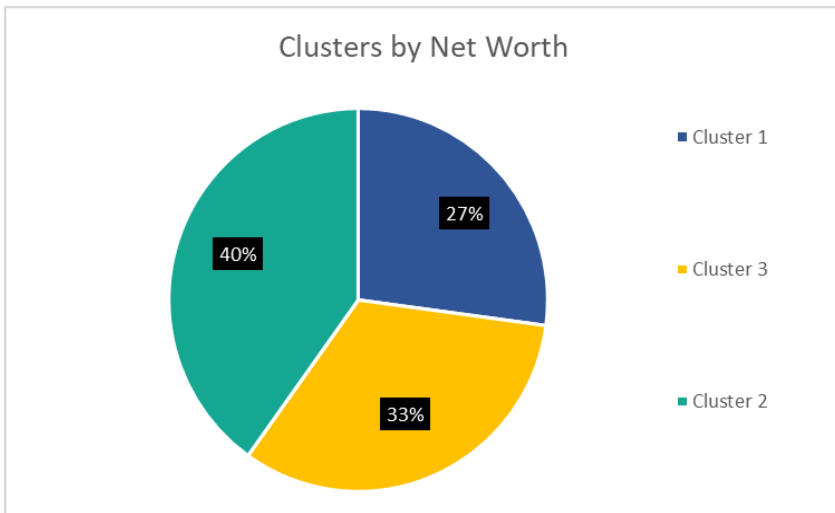
**Figure 16.** Results of K-means clustering

## 4.6 Cluster Analysis

As can be seen from clustering results, 29% of actors account for 73% of total Net Worth, what is quite close to Pareto principle or 80/20 rule. The first cluster, which accounts for the majority of actors (71%), is the smallest one according to the Net Worth (27%). The third cluster, which accounts for 19% of actors, accounts for 33% of total Net Worth. The second cluster, which is the smallest one (10% of total number of actors), accounts for the major part of total Net worth (40%), making this cluster the main priority when developing marketing plan and strategy.



**Figure 17.** Distribution of customers by clusters



**Figure 18.** Distribution of net worth by clusters

The average age and number of children also varies within the clusters (Table 5).

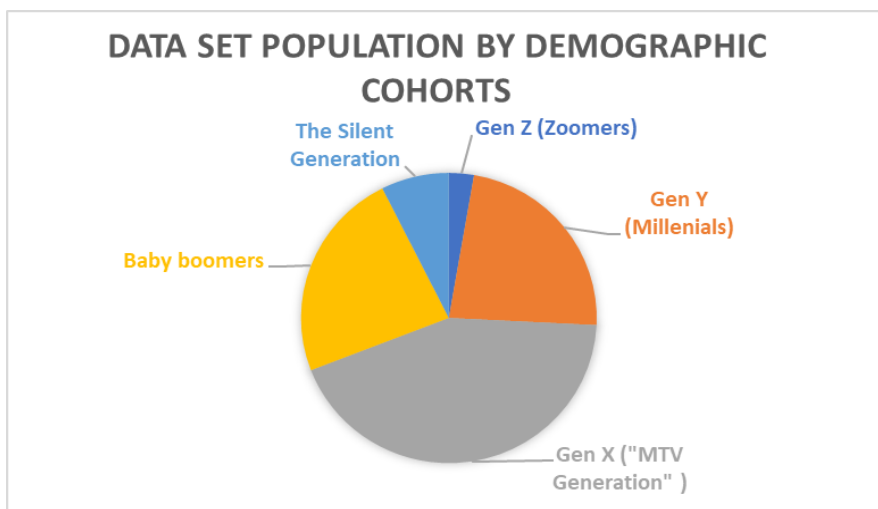
Cluster	Average age	Average number of children	Average Net Worth, millions \$
1	49	1.5	30.4
3	54	2.2	136.4
2	62	3.3	327.4

**Table 5.** Description of clusters by average age, number of children and net worth

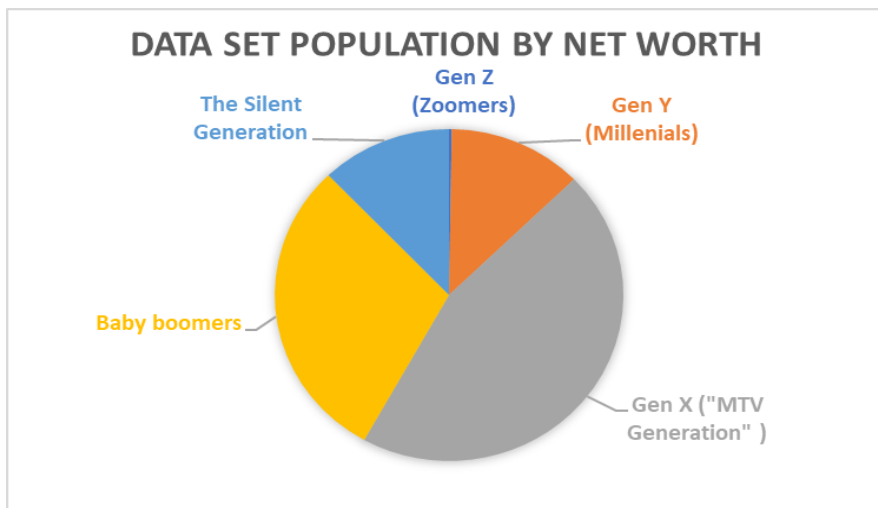
The first cluster is the youngest and the less rich one, the average number of children is only 1.5, what is less than the average value in USA – 1.93, meaning that the majority of actors in this cluster have 1-2 children. The third cluster is the mean between the first and second clusters according to all three parameters. The average number of children is 2.2, quite close to mean value in US. The second cluster consists of the richest actors in the data set, the average age is the highest, the youngest actor is 38 years old, compared to 27

in the third cluster and 20 in the first one. The average number of children is the highest, much higher than 1.93, what seems reasonable as older people tend to have more children than the younger ones. There is also a division according to gender, women account for 55.5% of total number of actors in Cluster 1, 41% in Cluster 3 and only 28% in Cluster 2. From the whole data set men account for 61.7% of Net Worth and in average have more children than women – roughly 2.1 against 1.5 respectively. The average age of female is 47.4 comparing to male 55.4.

Considering the results of clustering, it was decided to divide each of all three cluster on three smaller segments according to “Age” variable.



**Figure 19.** Division of customers by generations



**Figure 20.** Proportion of net worth by generations

A short description of the dataset population segmented according to the “Age” factor is shown in Table 6. The distribution of actors in the data set is different to overall US population according to proportion of population belonging to the respective generation.

The same is with the distribution of wealth, which does not reflect the proportion of population, as different age groups are unequally rich and are in different lifecycles. The biggest and wealthiest group is Generation X, however older generations tend to have higher average net worth per person, while younger account for more population share and lower total net worth. The number of children also varies representing the needs of target segment and the tendencies – older actors tend to have much more children and hence bigger families than younger ones.

Age Range	Generation	Share of data set population	Share of US adult population <sup>33</sup>	Total Net Worth	Share of Total Net Worth	Average Net Worth, Million \$	Average number of children
9 – 24	Gen Z (Zoomers)	3%		50	0.2%	7.1	0.1
25 – 40	Gen Y (Millennials)	23%	31.3%	2548	12.5%	43.2	0.9
41 – 56	Gen X (“MTV Generation)	43%	28%	9258	45.4%	83.4	1.9
57 – 75	Baby Boomers	23%	30.7%	6069	29.7%	101.2	2.4
76 – 94	The Silent Generation	7%	9.9%	2483	12.2%	130.7	2.7

**Table 6.** Description of dataset segmented by “Age” parameter

## 4.7 Developing Marketing Strategy and Plan

*NB: For the purposes of this thesis, it will be assumed that the author and colleague(s) have established a business in luxury real estate. It is fictitious and serves only to illustrate the technique. In the foregoing it will be seen that the marketing plan is written in the active voice (specifically the use of “We”, “Our”, etc.)*

### 4.7.1 Executive Summary

Our brand is going to start a marketing campaign with expanding market share in premium and luxury real estate markets. We already have experience in working with these segments, but as our firm is relatively young, we want to gain a reputation as a major player through sales to celebrities, to increase our future sales not only in this market, but in bigger US real estate market. By selling our services to actors and actresses of different

<sup>33</sup> Pew Research Centre. (2020). *Millennials overtake Baby Boomers as America’s largest generation* [online]

age and other demographics, we will draw potential customers' attention to our brand and improve our brand name and current customers loyalty. As a company with good organisation, high-level customer service and, most importantly, satisfied customers, we want to grow and we are targeting people in consumer market, taking advantage of high demand for housing, which has become much more than just a place of residence, and our key excellences.

Our core customer base is premium segment of the market. However, we want to move upwards the market to increase sales in luxury and its high-end segments. We are planning to achieve 10% market share in each target segment with sales of houses for total price of \$220 million, to keep expenses below \$5 million and gain profits of \$12.5 million in the first year, decreasing the expenses in second year.

#### **4.7.2 Situation analysis**

Real estate is a market with intense competition and dynamic environment, our company should use all the experience and advantages to get a sizable market share. As we are working with luxury segments the price is not the key factor, but for our brand working with premium segment, higher price must be adjusted by superior features. For the three years of running business we already have gained good positions on the premium market, our main efforts should be aimed on the smaller, but marginally more profitable segment of the real estate market.

#### **4.7.3 Strengths, Weaknesses, Opportunities, and Threat Analysis (SWOT)**

##### **Strengths**

1. Flexibility – our firm offers the variety of possibilities to choose among, orders for design and construction are possible.
2. The guarantee of quality and further services throughout the tenure – we are not one of those companies that forget about the client immediately after the purchase, our business strategy is to improve the reputation of our brand and customers loyalty, as loyal and happy customers are the best advertisement.
3. The knowledge of our customers and their needs – as a customer-oriented company we use marketing researches in order to improve our offers and understand what exactly the client wants, better than himself. Generational marketing can be very effective if used properly.

## **Weaknesses**

1. Reputation – as a relatively young company we may find it hard to quickly conquer a market, where there are many players with many years of experience. That is why our firm is thinking in the long-term and this marketing campaign is a part of a bigger plan.
2. Lack of staff – our brand faces the shortage of highly qualified personnel, that is why this year's expenses are increased to recruit and train more staff.

## **Opportunities**

1. Increasing level of technologies – allowed us to develop and construct a new generation of houses with the latest equipment and incorporated systems.
2. The size of the market – in 2018 the US real estate market was estimated \$16 trillion<sup>34</sup>. Despite the high competition the profits are high.

## **Threats**

1. Market fluctuations – the main threat for our sales and profits is the confusion and instability caused by the COVID-19, influencing the general, business and purchasing activity of people.
2. Increased competition – the constantly growing pressure on the market forces firms to seek new ways to meet customer needs.

### **4.7.4 Product Offerings**

We have developed different offers for each target segment.

**Premium segment** – price per house \$500.00 – 5.000.000 (Houses, apartments)

*Higher price for higher benefits*

We offer a great diversity of locations, architecture and interior styles, as well as special offers for couples, singles and big families, reasonable prices and incredible level of comfort.

The apartments in New York or lakeside houses in Pennsylvania, villas in Florida or loft houses in California, anyone will find the right option. Special care of details like the infrastructure and security level of the neighbourhood. Every house is provided with modern smart home and security systems.

For selling to luxury and super luxury classes, we will organise a special division, sub-brand – a common technique, frequently used in automotive industry (premium BMWs

---

<sup>34</sup> Nareit Research. (2019). *Estimating the Size of the Commercial Real Estate Market in the U.S.* [online]

and luxury Rolls-Royces) and many other markets. Since moving down the price segment can ruin customers loyalty to luxury brands, we choose branding under different names as the marketing strategy.

**Luxury segment** – price per house \$5.000.000 – 25.000.000 (detached houses, luxury apartments)

*Higher price for exclusive experience*

We offer a variety of exclusive properties in prestigious locations, such as SoHo or Sagaponack, with high level of security and many possibilities like spa, swimming pools and gyms.

Houses and apartments that are created by famous architectures and designers from all over the globe. From luxurious apartments in Manhattan to detached houses in Medina, the range of choice is huge. The built in smart-home and security systems, in addition to an unrivalled combination of beauty, comfort and wealth.

**Super Luxury segment** – price per house \$25.000.000+ (detached houses, penthouses)

*Highest price for the chosen few*

The most prestigious neighbourhoods, closed communities, all-included in your home area, special services like personal security, chiefs and service staff are included.

The best of the best architectures and designers create individual works of art, a perfect mix of luxury and technology to turn your life into a fairy tale. Modern loft bunker-style houses, classic Mediterranean-style villas, French chateaus in Beverly Hills, surrounded by a large territory with all needed infrastructure or modern penthouses in Manhattan, great choices for special people.

#### **4.7.5 Distribution**

All contacts with clients and developers are held by our company, with main office located in New York (NY) and divisions in San Francisco and Los Angeles (CA), Seattle (Washington), Miami (Florida) and Houston (Texas). Our marketing efforts will be focused on an individual approach and personalised contact with celebrities, in the broad market the mass media and Internet will be our main channels of advertisement.

#### **4.7.6 Marketing Strategy**

##### **Objectives**

- 1st year – to achieve 10% of sales to each target segment through units sales volume of 25 houses, gaining \$6.25 million profits.
- 2nd year – to increase our profits up to \$7.25 million by increasing revenues and cutting down expenses.

### Target Markets and Corresponding Benefits

Characteristic	Segment 1.1	Segment 1.2	Segment 1.3
Proportion	22.27% of dataset	29.69%	19.14%
Average NW	\$21.8 million	\$33.3 million	\$36.1 million
Family composition	no kids /1 kid mostly	1 – 2 kids	2 kids mostly
Age Range	40 ≥	41 – 56	57 ≤
Mean Age	33	48	69

**Table 7.** Description of premium target segments

Target Segment	Customer Need	Corresponding Benefit/feature
1.1	access to all the benefits of a modern metropolis property that characterises the owner housing that fits to active lifestyle diversity	modern locations in the city centre individual architecture and design locations that are the most time saving special offers for couples & singles
1.2	infrastructure for children access to all the benefits of a modern metropolis housing that fits to active lifestyle	availability of good schools and further infrastructure in the neighbourhood modern locations in the city centre locations that are the most time saving
1.3	calm and pleasant environment more than a house	villas in picturesque and safe places houses that impress not only with appearance but also with components like backyard, pool, embedded smart home systems

**Table 8.** Premium segment customer needs and corresponding benefit



Characteristic	Segment 3.1	Segment 3.2	Segment 3.3
Proportion	3.13% of dataset	9.77%	6.25%
Average NW	\$138 million	\$136.8 million	\$135 million
Family composition	1 kid mostly	2 kids mostly	2-3 kids
Age Range	40 ≥	41 – 56	57 ≤
Mean Age	34	51	70

**Table 9.** Description of luxury target segments

Target Segment	Customer Need	Corresponding Benefit/feature
3.1	infrastructure for children interesting community & environment diversity	availability of best schools and further infrastructure in the neighbourhood prestigious neighbourhoods like SoHo or Manhattan special offers for families
3.2	infrastructure for children access to all the benefits of a modern metropolis community of like-minded people all services at arm's length	availability of good schools and further infrastructure in the neighbourhood modern locations in the city centre locations that are the most time saving
3.3	calm and pleasant environment place to spend time with family community of like-minded people	detached houses in the most beautiful resorts, as well as in mountains or by the lake guesthouse on the plot prestigious districts like Medina or Sagaponack

**Table 10.** Luxury segment customer needs and corresponding benefits

Characteristic	Segment 3.1	Segment 3.2	Segment 3.3
Proportion	0.39% of dataset	3.91%	5.47%
Average NW	\$250 million	\$331 million	\$300.4 million
Family composition	2 kids	2 kids mostly	3 – 5 kids mostly
Age Range	40 ≥	41 – 56	57 ≤
Mean Age	38	52	71

**Table 11.** Description of super luxury target segments

Target Segment	Customer Need	Corresponding Benefit/feature
2.1	As this segment consists of only 1 customer, whose age is 38, it will be merged with segment 2.2	
2.2	<p>infrastructure for children</p> <p>closed community of rich, successful &amp; like-minded people</p> <p>access to all the benefits of a modern metropolis</p> <p>house as a piece of art</p> <p>all services at arm's length</p>	<p>availability of best schools and further infrastructure in the neighbourhood</p> <p>the most prestigious locations like San Roche or Silicon Valley</p> <p>penthouses in the best skyscrapers in the country</p> <p>each house as an individual work of art created by world-famous architectures and designers</p> <p>own gym, spa, pool and other services like personal chief and security</p>
2.3	<p>closed community of rich, successful &amp; like-minded people</p> <p>calm and pleasant environment</p> <p>place to spend time with family</p> <p>all services at arm's length</p>	<p>chateaus, villas, detached houses in the best neighbourhoods like San Roche</p> <p>amazing nature and pleasant climate</p> <p>guesthouse on the plot</p> <p>own gym, spa, pool and other services like personal chief and security included</p>

**Table 12.** Super luxury segment customer needs and corresponding benefits

#### **4.7.7 Positioning**

As a premium and luxury segments business, both of our brands are building their selling offers according to the specifics of each segment. For premium segments our main message is the reasonability of paying more for much more. For luxurious – the uniqueness and superiority of our offer, that will give invaluable impressions, emotions and experience.

#### **4.7.8 Strategies**

##### **Product**

The purchase contract specifies the services that the owner can use throughout the ownership of the property. We guarantee our customers the highest level of quality, service and support. The brand and logo will be displayed in all our commercials and marketing campaigns.

##### **Pricing**

Our prices for property for the premium brand vary between \$500,000 and \$5million, for the luxury – from \$5million and more. These prices will allow us to take share from established competitors but may be changed according to the situation on the market.

##### **Distribution**

Our channel strategy is to use individual distribution, personal offers and meetings, only with a narrow focus on the target buyer. During the first year we will hire and educate new team-members, to increase our presence in all current positions and enter new potential markets, such as Wyoming, New Jersey, Hawaii and New Mexico. During the second year we will increase our marketing costs and efforts towards gaining more share on the actors' real estate and overall market.

##### **Marketing Communications**

All our messages in all our media and advertisements should work in synergy to strengthen brand reputation and highlight key points of product differentiation. We will monitor customers response to our marketing efforts from the start of campaign till the end. The highest response is predicted to be from Generation Y (Millennials), as they are entering into the appropriate lifecycle. To generate awareness and buzz around our brand we will host events dedicated to our brand and our cause, thematical evenings, as well as

aggressive Internet advertisement and e-mails with a personal offer. We should also motivate and promote our selling staff to increase sales rates and efficiency.

#### **4.7.9 Marketing Mix**

Our offer will be presented in March, so we have described in detail our programs and tactics for the first six months of marketing campaign to achieve objectives stated by our company.

##### **March**

The first thing that we should do is to launch a sales promotion program to better prepare our team and educate newcomers, the predicted budget is \$300,000. Our brand is expanding to other states, the expected cost of opening 4 new branches is \$2 million. Our training personnel will explain the specifics of working with each segment and main advantages and benefits of our offers. We will start contacting people from the data set to promote our service and make appropriate adjustments depending on the consumers response.

##### **April**

We will start an integrated advertising campaign, using the Internet, TV, radio and print targeting customers from the data set and broad audience as well. The campaign will focus on our benefits and key excellences, the variety of our offers that corresponds to customer needs and wants, and our special, humane and customer-oriented way of doing business.

##### **May**

As the campaign continues, we will add target ads in Facebook and Instagram targeting the broad audience of premium and luxury segments. Our new branches and staff should be ready to get started, all the efforts should be aimed at sales. Constant monitoring and data analysis should be held every day to adapt to the realities of the market.

##### **June**

Until this moment we expect at least few deals to be closed, to involve celebrities into advertisement and win the doubters to our side. We will hold a trade sales contest with a significant award for the salespeople and division that achieves the highest sales during the month.

## **July**

The growing amount of sales to Hollywood actors should be used to boost sales to actors and other segments, prove our success and increase the loyalty of core customers. We will organise a festive event for our customers, invite the media and the most potentially valuable people, who can become our customers.

## **August**

Until this point, we should reach the sales mark of at least 12 houses to the target segments of actors and increase our total sales by half. It is a time for broad analysis of our campaign's results, including research of customer satisfaction level and ways how to improve our sales and marketing efforts, probably to reorganise the offers and ads for better results.

### **4.7.10 Financials**

In the US, the real estate commission is usually 5% – 6% of the home's sale price. Considering the 5% commission and the expected sales of 25 units with average price of \$9 million for all segments, we expect \$225 million in sales and our first-year revenues \$11.25 million. Our expenses, including those on marketing, are expected \$5 million, where \$3 million are the fixed costs. The assumed variable cost is \$80,000 per house and revenue is \$250,000 per house. The break-even calculation is:

$$\frac{3,000,000}{250,000 - 80,000} = 17.65 \text{ units}$$

Assuming that the break-even point is roughly 18 houses and expected sales level of 25 houses, the margin of safety is 28%.

### **4.7.11 Controls**

We have chosen the means to control the implementation and organisation of our marketing:

#### **Implementation**

Our plan is to ensure strict control measures and carefully monitor our business activities, such as monthly sales and expenses by their nature, quality of our services and customers loyalty. By doing this we will be able to respond and take all necessary measures if any problem occurs or there will be major changes in the market.

**Organisation**

The chief marketing officer (CMO) of our company bears full responsibility for our marketing activities and strategies. In order to enter the national market, we will connect our partners from the media and hire a marketing agency for the development of advertising materials and commercials. Our marketing department will continue to be responsible for drawing up proposal, ruling our marketing campaigns and monitoring our key indicators.

## 5 Results and Discussion

The result of this thesis is a performed K-means clustering algorithm, based on “Net Worth” as a segmentation variable, correlations were found between all the parameters. The dataset of 256 randomly selected actors and actress was created and divided on three clusters, then the results of clustering were analysed. After cluster analysis, each cluster was divided on three smaller segments based on profiling parameter “Age”. A marketing plan and strategy for premium and luxury market real estate business were developed considering specifics and priority of each segment. The obtained clusters and segments are of different size, due to the uneven distribution of wealth among randomly selected actors. Customers from one segment or cluster tend to be much more similar to each other in segmentation characteristics, than to customers from other segments or clusters.

For development of marketing plan and strategy, and performing clustering many different techniques and approaches were used. The K-means algorithm was performed using Python programming language and SciPy library, the number of clusters was determined using the program code for calculating WCSS applying the Elbow method, the segmentation of customers was performed according to generational marketing approach. Due to the popularity of the method, there are many useful literature sources and guidelines on K-means clustering algorithm, so there were no significant difficulties during working on it. K-means is a useful method of clustering, widely applied by researchers from different sciences and industries and being comparative easy to learn and understand. As any other technique or method, it has few limitations and specifics. In this thesis the dataset was created specifically to perform a clustering algorithm. It would be more appropriate to select method basing on a dataset, since different datasets require a different approach. K-means is better for datasets with approximately same sizes of clusters. Due to different density of the data in the dataset, the obtained clusters are different in size, this algorithm better deals with datasets where data is distributed more evenly. The right decision would be to create a dataset and select clustering method that will be more appropriate for this specific dataset.

The main limitation of this work is the number of customers in a dataset, often marketers have to work with much larger amounts of data, however, the number of participants is sufficient for performing the clustering algorithm.

## 6 Conclusion

This thesis was supposed to describe the main steps of collecting data, performing K-means clustering algorithm and developing marketing strategies and plans. The possibilities for marketers nowadays are greater than ever, digitalisation and growth of online sector grant access to customer data from all over the world, while the development of computers allows complex calculations to be performed without major investments.

The basics, features and application cases of segmentation, targeting, positioning and clustering were explained in the literature review. It also covers the modern principles of developing marketing plans and strategies, as well as some other approaches and concepts, such as customer loyalty or customer-orientation, used in marketing nowadays. The tendencies in advertising, methodology of K-means and key principles for developing plans and strategies are examined in last chapters of the literature review.

The practical part was aimed on usage of customer data in marketing. For this purpose, it was supposed to create a dataset suitable for performing the clustering algorithm and apply K-means technique. The contents of this part covers all steps from collecting data to final development of marketing plans and strategies. The additional segmentation of clusters based on profiling variable was meant to show the further possibilities of identifying and profiling customers into better defined target segments. The last chapter of practical work, dedicated to development of strategies and plans, includes SWOT analysis, analysis and definition of target markets, financial forecast, distribution and marketing mix.

The dataset with all actors “ActorsDataset.csv”, as well as Jupyter Notebook with program code for performing K-means clustering algorithm and finding correlation between variables in dataset “Hollywood\_clustering.ipynb” is attached to this thesis.

As technology advances, a huge range of opportunities opens up for marketers of the 21<sup>st</sup> century. New approaches of engagement with customers, promoting and online targeting, implementation of customer databases and data warehouses, data mining and much more change the marketing and business in general. A variety of internet sources and books aimed on studying and applying modern methods is huge, allowing anyone from anywhere to get familiar with any topic he is interested in, even without programming knowledge.



## 7 References

### 7.1 Bibliography

**Aaker D.** (2004), *Strategic Market Management, 7-th edition*. England, UK: John Wiley & Sons Ltd, 2004. ISBN: 0-471-48426-1 (English). Translated from Russian edition.

Стратегическое рыночное управление, 7-е изд.

[Strategic Market Management, 7-th edition]. St. Petersburg, Russia: LLC Peter Press, 2007. ISBN: 978-5-469-01301-3.

**Hartigan J., Wong M.** (1979), *Algorithm AS 136: A K-Means Clustering Algorithm*.

Blackwell Publishing Ltd.: Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 28, no. 1, 1979, pp. 100–108. DOI: 10.2307/2346830.

**Kotler P.** in **Ward A.** (2002), *Introduction to Marketing*, University of York.

**Kotler P., Keller K. L.** (2016), *Marketing Management, 15th edition*. England, UK: Pearson Education, Inc., 2016. ISBN: 978-1-292-092 -71-3.

**McDonald M., Christopher M., Bass M.** (2003), *Market segmentation. In: Marketing*. London, UK: Palgrave, 2003. ISBN: 978-1-4039-3741-4.

**VanderPlas J.** (2017), *Python Data Science Handbook: Essential Tools for Working with Data*. Sebastopol, CA: O'Reilly Media, Inc., 2017. ISBN: 978-1-491-91205-8.

### 7.2 Internet Resources

**American Marketing Association.** *Definitions of marketing*. 2017 [online]. Available at: [www.ama.org/the-definition-of-marketing-what-is-marketing](http://www.ama.org/the-definition-of-marketing-what-is-marketing) (Accessed 5<sup>th</sup> December 2020)

**Bui Q., Miller C. C.** *The Age That Women Have Babies: How a Gap Divides America*. 4<sup>th</sup> August 2018, New York Times [online]. Available at: [www.nytimes.com/interactive/2018/08/04/upshot/up-birth-age-gap.html](http://www.nytimes.com/interactive/2018/08/04/upshot/up-birth-age-gap.html) (Accessed 10<sup>th</sup> January 2021)

**Business Insider.** *How Mobile Coupons Are Driving an Explosion in Mobile Commerce*. 3<sup>rd</sup> June 2013 [online]. Available at: [www.businessinsider.com/mobile-coupons-driving-mobile-commerce-2013-5](http://www.businessinsider.com/mobile-coupons-driving-mobile-commerce-2013-5) (Accessed 6<sup>th</sup> January 2021)

**Frank Ceballos.** *Artistic Voronoi Diagrams in Python*. 17<sup>th</sup> December 2019, Medium [online]. Available at: <https://medium.com/i-want-to-be-the-very-best/artistic-voronoi-diagrams-in-python-928bdae85dd8> (Accessed 12<sup>th</sup> January 2021)

**GeoPath.** Market Segmentation page [online]. Available at: <https://support.geopath.io/hc/en-us/sections/360000917931-Market-Segmentation> (Accessed 18<sup>th</sup> December 2020)

**IAB.** *Price Waterhouse Coopers annual reports for the Interactive Advertising Bureau*. May 2019 [online]. Available at: [www.iab.com/wp-content/uploads/2019/05/PwC-IAB-Webinar-Presentation-Full-Year-2018.pdf](http://www.iab.com/wp-content/uploads/2019/05/PwC-IAB-Webinar-Presentation-Full-Year-2018.pdf) (Accessed 18<sup>th</sup> January 2021)

**JavaLab.** *Voronoi Diagram*. 17<sup>th</sup> September 2019 [online]. Available at: [https://javalab.org/en/voronoi\\_diagram\\_en](https://javalab.org/en/voronoi_diagram_en) (Accessed 12<sup>th</sup> January 2021)

**Kaggle.** Main page [online]. Available at: [www.kaggle.com](http://www.kaggle.com) (Accessed 24<sup>th</sup> December 2020)

**McCrindle M., Fell A.** *Understanding Generation Alpha*. July 2020 [online]. Available at: <https://generationalalpha.com> (Accessed 23<sup>rd</sup> January 2021)

**Nareit Research.** *Estimating the Size of the Commercial Real Estate Market in the U.S.* July 2019 [online]. Available at: [www.reit.com/sites/default/files/Size%20of%20CRE%20market%202019%20full.pdf](http://www.reit.com/sites/default/files/Size%20of%20CRE%20market%202019%20full.pdf) (Accessed 18<sup>th</sup> January 2021)

**Pew Research Centre.** *Millennials overtake Baby Boomers as America's largest generation*. 28<sup>th</sup> April 2020 [online]. Available at: [www.pewresearch.org/fact-tank/2020/04/28/millennials-overtake-baby-boomers-as-americas-largest-generation](http://www.pewresearch.org/fact-tank/2020/04/28/millennials-overtake-baby-boomers-as-americas-largest-generation) (Accessed 15<sup>th</sup> January 2021)

**Statista.** *Average number of children in US*. 26<sup>th</sup> January 2020 [online]. Available at: [www.statista.com/statistics/718084/average-number-of-own-children-per-family](http://www.statista.com/statistics/718084/average-number-of-own-children-per-family) (Accessed 26<sup>th</sup> January 2021)

**TIOBE.** *TIOBE Index*. 4<sup>th</sup> January 2021 [online]. Available at: [www.tiobe.com/tiobe-index](http://www.tiobe.com/tiobe-index) (Accessed 25<sup>th</sup> January 2021)

## 8 Appendix

Files attached to this thesis:

**ActorsDataset.csv** – a data set of 256 actors with following parameters: ID, Age, Gender, Net Worth, Number of children and Name.

**Hollywood\_clustering.ipynb** – Jupyter Notebook with program code for K-means clustering, including code for finding correlations between variables and executing Elbow method.