

UNIVERSITY OF LJUBLJANA  
FACULTY OF ECONOMICS

MASTER'S THESIS

**BUSINESS CUSTOMERS SEGMENTATION WITH THE USE OF K-MEANS  
AND SELF-ORGANIZING MAPS: AN EXPLORATORY STUDY IN THE CASE  
OF A SLOVENIAN BANK**

Ljubljana, December 2015

SLAVIŠA SIMIĆ

## **AUTHORSHIP STATEMENT**

The undersigned Slaviša Simić, a student at the University of Ljubljana, Faculty of Economics, (hereafter: FELU), declare that I am the author of the master's thesis entitled Business customers segmentation with the use of k-means and Self-Organizing Maps: an exploratory study within a Slovenian bank, written under supervision of PhD Aleš Popovič and co-supervision of PhD Roberto Henriques.

In accordance with the Copyright and Related Rights Act (Official Gazette of the Republic of Slovenia, Nr. 21/1995 with changes and amendments) I do not allow the text of my master's thesis to be published on the FELU website.

I further declare

- the text of my master's thesis to be based on the results of my own research;
- the text of my master's thesis to be language-edited and technically in adherence with the FELU's Technical Guidelines for Written Works which means that I
  - cited and/or quoted works and opinions of other authors in my master's in accordance with the FELU's Technical Guidelines for Written Works and
  - obtained (and referred to in my master's thesis) all the necessary permits to use the works of other authors which are entirely (in written or graphical form) used in my text;
- to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code (Official Gazette of the Republic of Slovenia, Nr. 55/2008 with changes and amendments);
- to be aware of the consequences a proven plagiarism charge based on the submitted master's thesis could have for my status at the FELU in accordance with the relevant FELU Rules on Master's Thesis.

Ljubljana, December 28, 2015

Author's signature: \_\_\_\_\_

# TABLE OF CONTENTS

<b>INTRODUCTION .....</b>	<b>1</b>
<b>1 THEORETICAL BACKGROUND .....</b>	<b>2</b>
1.1 CRM.....	2
1.1.1 CRM building blocks .....	3
1.1.2 The process of the CRM.....	5
1.1.3 CRM success factors .....	5
1.1.4 CRM system implementation issues .....	7
1.1.5 The benefits of the CRM .....	8
1.2 Client segmentation .....	8
1.2.1 The significance of the client segmentation field.....	10
1.2.2 The reasons for the client segmentation .....	11
1.2.3 The process of the client segmentation.....	12
1.2.4 Consumer versus industrial client segmentation .....	13
1.2.5 Segmentation bases and variables .....	14
1.2.5.1 Consumer market segmentation variables .....	15
1.2.5.2 Business-to-business market segmentation variables .....	16
1.3 Client segmentation in the banking industry .....	20
1.3.1 Banking industry characteristics.....	21
1.3.2 Products versus services .....	22
<b>2 BUSINESS CUSTOMERS SEGMENTATION WITH THE USE OF K-MEANS AND SELF-ORGANIZING MAPS: AN EXPLORATORY STUDY IN THE CASE OF A SLOVENIAN BANK.....</b>	<b>23</b>
2.1 Business case presentation.....	23
2.1.1 Introduction of Bank S .....	23
2.1.2 Bank's business needs identification.....	24
2.2 Analysis design .....	25
2.2.1 Data collection.....	25
2.2.2 Methods selection .....	31
2.2.3 Data pre-processing and variables selection.....	37
2.3 Analysis .....	41
2.3.1 Determining the number of clusters .....	41
2.3.2 Selecting the method on which to cluster.....	44

2.3.3	Clusters profiling and interpretation .....	46
2.3.3.1	Cluster 1.....	47
2.3.3.2	Cluster 2.....	48
2.3.3.3	Cluster 3.....	48
2.3.3.4	Cluster 4.....	48
2.3.3.5	Cluster 5.....	49
2.3.3.6	Cluster 6.....	49
<b>3</b>	<b>DISCUSSION .....</b>	<b>50</b>
3.1	Practical implications .....	50
3.2	Limitations of the master's thesis .....	52
3.3	Future work.....	53
	<b>CONCLUSION.....</b>	<b>53</b>
	<b>REFERENCE LIST .....</b>	<b>55</b>

## LIST OF FIGURES

Figure 1.	Building blocks of CRM.....	3
Figure 2.	Number of items published over the last 20 years.....	10
Figure 3.	Marketing strategy process .....	12
Figure 4.	Segmentation process .....	13
Figure 5.	Influences on buyer's personal characteristics .....	20
Figure 6.	Individual customer segments in % .....	24
Figure 7.	SOMs – a rough graphical representation.....	35
Figure 8.	Relationship Duration histogram .....	38
Figure 9.	Product Coverage histogram.....	38
Figure 10.	Client Value histogram .....	39
Figure 11.	Cubic Clustering Criterion.....	41
Figure 12.	Pseudo F.....	42
Figure 13.	R-squared .....	43
Figure 14.	Sum of the distances .....	43
Figure 15.	30x20 SOM U-matrix .....	44

## LIST OF TABLES

Table 1.	Company size categories by three criteria .....	18
Table 2.	Consumer versus business demographics .....	18
Table 3.	The initial customer list exportation criteria .....	26
Table 4.	Product indicator description .....	27
Table 5.	SOM algorithm – a formal description .....	36

Table 6. Correlation between variables .....	40
Table 7. k-means and SOMs comparison based on three clustering statistics .....	45
Table 8. Clusters' sizes and mean values .....	46









## INTRODUCTION

Many companies of the contemporary economy have a large number of customers, and each of these represents almost as many different sets of needs and expectations which have become more and more complex, demanding and sophisticated over time. As it is impossible to treat every customer completely individually, let alone to provide them with fully customized products and services, it is clearly evident that they should be divided into a few groups in a reasonable manner, of course.

Even though client segmentation has been present for many years, companies still struggle to use it correctly. They are trying to implement it properly as well as to integrate it into marketing strategy (Dibb & Simkin, 2009, p. 219). Instead of helping in more important, strategic areas, such as products and services innovation, pricing, and distribution channel selection, market segmentation has often been narrowly used for the needs of advertising (Yankelovich & Meer, 2006, p. 1).

While the consumer market segmentation has been a challenging task for marketers, it has been an even more difficult job for those of industrial markets, or as Kukulias (2012, p. 2) had neatly illustrated with an example; whereas consumer marketers go fishing, business-to-business marketers have to fish for sharks. The business market segmentation is known to be much less developed in comparison to the consumer segmentation. However, some techniques of the latter can be also applied to the industrial markets. Yet, unless they want to be led into the wrong direction, practitioners have to be very careful about choosing and refining the appropriate variables on which to segment (Zimmerman & Blythe, 2013, p. 121).

The purpose of this master's thesis is to examine approaches to the industrial market segmentation through a review of the scientific literature on the respective topic. The objective of the master's thesis is to find out whether the use of contemporary, non-traditional methods can help business marketers to classify their customers more completely, thoroughly, and to a greater degree. With my own findings and the concise insights from other researchers, I would like to contribute to a better understanding of the business customers segmentation, especially in reference to the banking industry. Based on the real customer data acquired from one of the major Slovenian banks, I will identify the corporate customers segments that would eventually serve as the fundamentals for appropriate addressing and customized marketing campaigns development, reflecting later on in higher profit rates of that particular bank.

The first part of the thesis consists of a comprehensive theoretical-analytical review of the academic literature, articles, studies, and research papers on the discussed topic. Here, I used the descriptive method, as well as the compilation method, that helped me unify the thoughts and beliefs of the experts. In the second part of the master's thesis, I tried to apply

theoretical observations to a real-life business case. In order for the bank to develop appropriate marketing strategies, adjusted to the characteristics of each segment, I tried to divide somewhat less than 18,000 business clients of a major Slovenian bank into a few reasonable segments by using the combination of two clustering methods: *k*-means and Self-Organizing Maps.

The outline of the thesis is as follows. The next section introduces concepts of the customer relationship management, contemporary approaches to the customer segmentation, methods of the latter that have currently been in use in the banking field, and previous studies on this subject. The second part of the thesis encompasses business case presentation, research setup, and data collection and analysis. The results are reviewed in the third section. The final part presents the conclusions based on the main findings, together with some potential future research directions.

## **1 THEORETICAL BACKGROUND**

It has been almost a century since Henry Ford's famous statement: "Any customer can have a car painted any colour that he wants so long as it is black." These words were his remark on the Model T, a car that was produced from October 1908 to May 1927 (Alizon, Shooter, & Simpson, 2009, p. 590), and they defined the beginning of the so-called "one size fits all" approach to production. Ford's objective was to build cars for the multitudes; he wanted cars to be both durable and cheap. In order to achieve that, he cut out most of the extravagances and options, including the colour of the paint (Goss, n.d.).

However, this kind of thinking about production and marketing has not been prevalent for quite some time. In an effort to satisfy the diverse desires of each and every one of them, companies nowadays attempt to provide more one-on-one-oriented services to their customers. To be able to do this, they have to understand their clients very well. A modern approach that helps them attain this is the customer relationship management (hereinafter: CRM), which is presented in more detail in the following chapter.

### **1.1 CRM**

There have been numerous attempts to provide the definition of the CRM. Some authors perceive CRM as a philosophy (Zablah, Bellenger, & Johnston, 2004, p. 478), others comprehend it as a business strategy (Karakostas, Kardaras, & Papathanassiou, 2005, p. 862), while others believe that CRM is a technology (Bose, 2002, p. 89). As a consequence of such inconsistency about the CRM perception, we have witnessed many failures of the CRM adoption and implementation, as well as success measuring (Rababah, Mohd, & Ibrahim, 2011, p. 221).

Nevertheless, I would agree with Payne and Frow's (2005, p. 168) definition of the CRM and describe it as a strategic approach that is concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments. It unites the potential of relationship marketing strategies and IT to create profitable, long-term relationships with customers and other key stakeholders. CRM provides enhanced opportunities to use data and information to both understand customers and co-create value with them. This requires a cross-functional integration of processes, people, operations, and marketing capabilities that is enabled through information, technology, and applications.

### 1.1.1 CRM building blocks

Very similar to the definition given above, Radcliffe, Kirkby and Thompson (2001, p. 2) define eight principal CRM components, touching a wide range of business aspects: information technology, its applications throughout the organization in processes, and the resulting customer experience, as well as the organization, the strategy, and the vision of the company (Peelen, Beltman, Van Montfort, & Klerkx, 2006, p. 5). Figure 1 displays the eight CRM components.

Figure 1. Building blocks of CRM



Source: Adapted from J. Radcliffe et al., *The eight building blocks of CRM*, 2001.

CRM vision is the first and also the most important component of the CRM, since all subsequent components are developed pursuant to it. It could be referred to as “personality” or “DNA of a company”, and it represents the values the company is dedicated to, taking the customer centricity into account (Kirkby, 2001, p. 2). While the

CRM vision answers the questions on “what” and “why” the company is going to do, the second component, CRM strategy, answers the question on “how” it is going to be done – it describes the way the organization wants to realize its vision (Peelen et al., 2006, p. 6).

To be able to build up a fruitful customer experience, which is the third component of the CRM, the company has to meet or even surpass the expectations of its customers, wherein the product and the service, the price, and the communication form have to be properly set (Plakoyiannaki & Tzokas, 2002, p. 230). In order to do this, the organization has to transform its model from the traditional to the client-oriented business model that will allow it to better understand its customers and, consequently, to better react to their needs. Organizational culture – the fourth building block, however, plays here a significant role since it settles the norms for appropriate behaviour of all employees (Jayachandran, Sharma, Kaufman, & Raman, 2005, p. 179).

To act as a customer-centric organization that develops products and services that fit customer needs, first, the company has to define and understand its processes, analyse them, and, if necessary, redesign its core business processes starting from the customer point of view (Chen & Popovich, 2003, p. 682). Thus, CRM processes are the fifth component of the CRM.

Customer information or customer knowledge, the sixth element of the CRM, is one of the major assets to the organization (Peelen et al., 2006, p. 7); the more the company knows about its customers, the better it can respond to their needs in terms of offer customization and consumption experience improvement, both of which can then lead to a higher customer satisfaction and customer loyalty (Mithas, Krishnan, & Fornell, 2005, p. 202). Collection, accumulation, storage, maintenance, and distribution of the customer knowledge throughout the organization are the domain of the penultimate CRM component – CRM technology (Chen & Popovich, 2003, p. 677).

The last and also the most challenging component of the CRM is performance management. As any other metrics, the CRM metrics can also be used to control, learn and collaborate; yet, it seems to be extremely difficult to identify which metrics are critical in driving CRM benefits. Instead of being dissociated from each other, the financial performance and the CRM performance have to be linked in order to pursue the enterprise-wide benefits (Kirkby, Thompson, & Buytendijk, 2001).

As we can see, CRM framework is quite a complex and sophisticated concept. Hence, since it would exceed its frames, this master’s thesis will focus only on the sixth and seventh building block, i.e. CRM information and CRM technology. But first, let us see how CRM process should take place.

### **1.1.2 The process of the CRM**

While the components of the CRM reveal what are the important aspects when talking about CRM, the process of the CRM tells us how things need to be done for the CRM system to be implemented successfully (Kim, 2004, p. 22). A number of recommendations on how CRM should take place has been published, still, Roberts, Liu, and Hazard (2005, pp. 317–322) present their six-step CRM process model which is further described below.

Since not all customers are equally profitable, hence unequally attractive, profitable customers or customers with a high customer lifetime value (CLTV) have to be identified in order to develop strategies to retain them or to make them even more profitable. Thus, a customer strategy development is the first step in the CRM process. The second step – customer objectives settlement – involves the CLTV analysis where both actual value and potential value have to be distinguished. Here, a suitable customer database is of essential value (Roberts et al., 2005, pp. 318–319).

The third step in the CRM process is organizational readiness accessing. In order to organize company around the customer needs, instead of the products, changes (e.g., process reengineering, organizational culture transformation, personnel adjustment and training, etc.) throughout the organization are inevitable; this requires close collaboration of all employees, especially the managers at all levels (Roberts et al., 2005, p. 320). An enterprise-wide program of change management, which – inter alia – includes powerful guiding and vision construction and communication (Kotter, 1995, p. 61), is a matter of the fourth step – enterprise alignment.

CRM programs execution is the next step in the CRM process, and it requires programmatic CRM activities identification. Those activities have to be well designed, carefully executed, and judiciously measured. Program effectiveness measurement, thus, represents the last step in the CRM process. It covers rigorous metrics identification as well as some new eventual metrics development (Roberts et al., 2005, p. 322).

### **1.1.3 CRM success factors**

Despite the fact that CRM has been one of the fastest growing businesses of the new millennium and a myriad of instructive researches on how to undertake it, critics still imply the high failure rate of the CRM projects (Foss, Stone, & Ekinici, 2008, p. 68). It is because companies keep forgetting some crucial points that have to be addressed when dealing with CRM embracement.

With a review of publications by researchers and practitioners, based on the degree of acceptance (percentage of the factor occurrence in the literature) and the CRM failure

causes criteria, Almotairi (2009, p. 6) offers a summarized list of ten elements for success never to be forgotten when adopting a CRM system; they are briefly described below.

Top management commitment is considered to be a key success factor for the CRM implementation and performance; it is needed in order for the CRM initiative to even get off the ground, as well as to provide sufficient and necessary resources needed. Without it, even the most brilliant CRM undertaking is doomed to failure (Kale, 2004, p. 45).

A clearly designed and stated business strategy is also important because it sets sales and profitability objectives, helps making segmentation and targeting decisions, determines the level of the desired offer customization, and so on (Dimitriadis & Stevens, 2008, p. 499).

According to Chalmers (2006, p. 1017), data management is the next essential element; it concentrates on collecting, processing, and distributing customer information of just a right amount and quality.

Culture change is by far the most delicate factor for the CRM success; conforming to Horne (2003, p. 53), who compares corporate culture changing to “steering a battleship with an oar”, people’s inability to change and adapt is the main reason for the CRM programs failure. Harding, Cheifetz, DeAngelo, and Ziegler (2004, p. 31) claim that end-users involvement at all stages tends to help in overpassing the “soft” obstacles, such as technology deployment and utilization refusal.

Redesign of business processes and organizational structure is often needed in order to align the organization to the customers. Instead of being traditionally organized around their business functions, companies have to arrange their operations around cross-functional processes. This, however, requires an engagement from the entire organization beyond just a marketing department, which is an area of yet another success factor – inter-departmental integration (Ryals & Knox, 2001, p. 540).

Information technology factor is critical since it leverages the CRM-related activities and by that helps improving the company performance. It comprises systems and processes integration, data integration, distribution channels and customer touch-points integration, and so forth (Dimitriadis & Stevens, 2008, p. 501).

Skilful, motivated and trained staff factor focuses on the availability of the high level and the adequate variety of skills among employees. It demands cooperation from both business and IT experts, and, if needed, hiring outside consulting or system development companies in order to embellish staff’s skills (Kim, 2004, p. 26). Besides training and specialized skills development, it also covers staff motivation through rewards/incentives policies, communication, re-training, and so on (Dimitriadis & Stevens, 2008, p. 501).

Motivated and empowered employees seem to collect, disseminate, and use the customer information better. They respond to customer needs faster, and customize the service to fit the expectations more easily. Usually, they interact with customers with greater hospitality and passion, thus strengthening the relationship almost at a personal level, which allows them to involve a customer to a higher degree (Plakoyiannaki, Tzokas, Dimitratos, & Saren, 2008, p. 273). All of this is the domain of the customer consultation success factor.

Last but not least, monitoring, controlling, measuring, and feedback factor is also an important element. It occupies formation and employment of the measures of the CRM implementation and the CRM impact to company performance (Almotairi, 2009, p. 8). Authors have proposed a variety of approaches to measure the CRM success; market share, increased lifetime value, cross-sell ratios, and balanced scorecard are just some of them (Sear, Hartland, Abdel-Wahab, & Miller, 2007, p. 5).

#### **1.1.4 CRM system implementation issues**

Langerak and Verhoef (2003, pp. 79–80) stress the following directions that can serve managers when implementing the CRM system:

- Direct CRM implementation by business strategy  
Implementation should be in line with the chosen strategy, thus the type of the CRM application has to depend on it.
- Level of embedment  
Usually, the CRM systems are successful when they are strategically installed; still, it is not always that straightforward, tactical implementation can also be a good choice.
- The nature of change  
Even though the CRM implementation considers changes in the organizational structure, this kind of changes is not always hard; it depends on the CRM approach.
- Make a cost-benefit analysis  
Unless companies want to make vast investments in CRM that eventually do not pay off, they have to make a cost-benefit analysis. It can confirm not so vague CRM application already covers their needs.
- “Develop-Buy-Outsource” choice  
If companies lack appropriate resources to develop the CRM software in-house, they should consider outsourcing the development or buying a CRM solution.
- Do not focus on software  
Although a significant part of the CRM budget is assigned to software, companies must not fall into the trap of looking at it as the key to success.
- CRM provider choice  
Companies have to be careful in choosing the CRM vendor; a low price does not always indicate a poorer quality, and vice versa.

### **1.1.5 The benefits of the CRM**

As we have seen above, CRM as a concept is quite a sophisticated process that calls for changes, improvements and transformations all over the organization, demanding a tight participation of all members, and emphasizing a top management commitment. Yet, companies still swear by it and invest a great deal of effort to incorporate it into their businesses. Why is that so? They are highly aware of the benefits that CRM can provide to them.

In the past, it had been guesstimated that the benefits of the CRM differed from industry to industry (Lemon, Rust, & Zeithaml, 2001, p. 2); newer studies, however, demonstrate exactly the opposite; benefits of the CRM are not too different across industries (Reinartz, Krafft, & Hoyer, 2004, p. 301). Therefore, Richards and Jones (2008, p. 123) state the following, commonly known core benefits of the CRM:

- improved ability to target profitable customers;
- integrated offerings across channels;
- improved sales force efficiency and effectiveness;
- individualized marketing messages;
- customized products and services; and
- improved customer service efficiency and effectiveness.

In addition, Ko, Kim, Kim, and Woo (2008, p. 66) highlight even more advantages of the CRM identified in the literature:

- increased profits;
- accurate customer information collected;
- enhanced customer loyalty;
- effect of word-of-mouth;
- reduced costs of new customer acquisitions;
- greater ease in developing new products;
- increased sales by additional purchases;
- increased brand loyalty;
- increased customer lifetime value; and so forth.

## **1.2 Client segmentation**

Thus, plenty of reasons drive organizations to become more and more interested in CRM. However, they are realizing even more that customers have a different economic value to the company, and are subsequently adapting their offerings accordingly, in order to maximize the value of the customer (Chen, Zhang, Hu, & Wang, 2006, pp. 288–289). The



base of how to maximize that value is the customer segmentation, which is further described below.

As already mentioned, customers are very different and have disparate needs and desires. However, customers that share common characteristics are prone to choose the same products or services (McKechnie, 2006, p. 119). Still, it often happens that similar people buy quite different products while dissimilar people can purchase products that are unexpectedly very much alike.

Companies have to be able to distinguish among their customers. A company that “tars all of its clients with the same brush” is addressing everybody, yet still nobody. Furthermore, client distinction has to be done with a great deal of prudence, involving some state-of-the-art strategies and approaches; namely, bear averages and statistics are not sufficient and can be misleading.

For example, an average able-to-purchase member of a Slovenian household happens to be three quarters a woman and one quarter a man, has a high-school degree plus a first year of a college completed, and earns approximately 500 euros every first day of each month. Together with their family, this person drives around in a car and a half, and has to feed one third of their cat and one third of their dog in their suburban apartment. The person prefers a heavy, saturated food which is low in calories at the same time. Their favourite painting is the one that shows the nature, or a little girl, or an abstract smudge of colour. The person has no opinion about most of the stuff, and their most common answer is “I don’t know” (Valicon, 2013).

To sum up, there is no “average” consumer neither exactly the same people. Even if the company can afford to target the whole market, it is more successful when it adjusts its products and communication to individual segments. However, a company has to be capable to group its customers together along the common variables they hold as closely as possible. The process of dividing customers into distinct, meaningful, and homogeneous subgroups based on various attributes and characteristics is – according to Tsiptisis and Chorianopoulos (2009, p. 189) – known as client segmentation, also referred to as customer or market segmentation. It is used as a differentiation marketing tool which enables organizations to understand their customers and build differentiated strategies, tailored to their characteristics.

In Rosella Software – Predictive Knowledge and Analytics Platform, this kind of segmentation is described as a process that divides customers into smaller groups called segments. Segments are to be homogeneous within and desirably heterogeneous in-between. In other words, customers of the same segments possess the same or similar set of attributes while customers of different segments have differing sets of attributes.

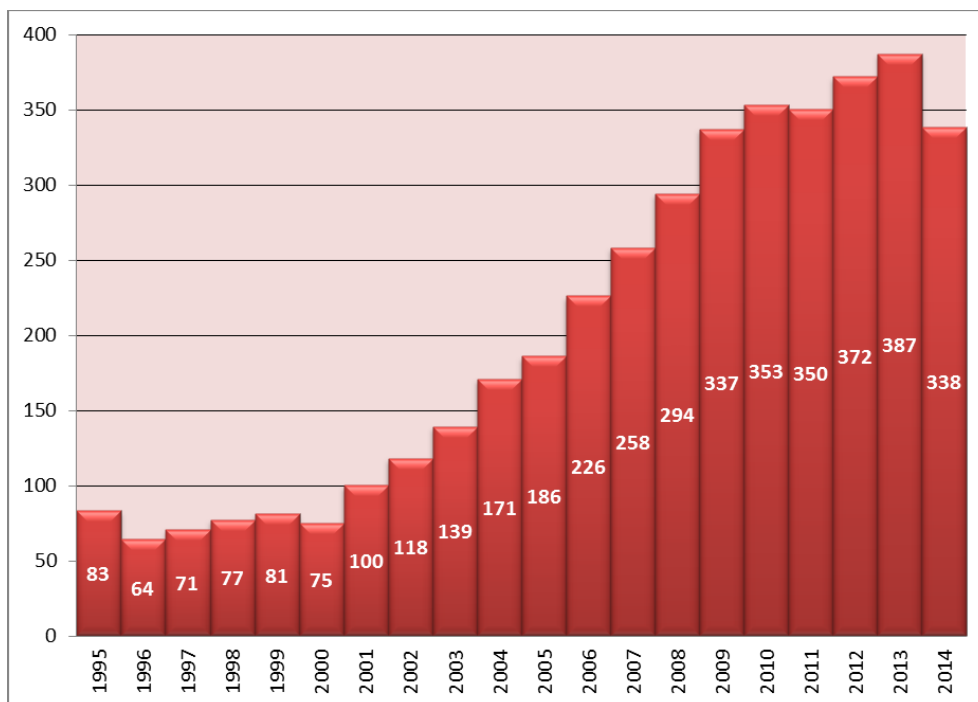
### 1.2.1 The significance of the client segmentation field

While reading the literature, I have noticed that client segmentation was a broadly attractive topic for researchers and writers. To make sure whether it is really important and how important it is, I visited Thomson Reuters' website. Thomson Reuters is the world's leading source of intelligent information for businesses and professionals that serves customers within four areas: financial and risk, legal, tax and accounting, and intellectual property and science (Thomson Reuters, n.d.).

Their citation index tool, Web of Science™, provides researchers, administrators, faculty, and students a quick, powerful access to the world's leading citation databases. An authoritative, multidisciplinary content covers over 12,000 of the highest impact journals worldwide, including Open Access journals and over 150,000 conference proceedings (Thomson Reuters, n.d.).

Thus, it made me wonder how many articles from the field of market segmentation have been published over the past years. Since Wendell Smith first proposed market segmentation as an alternative market development technique to product differentiation back in 1956 (Bailey, Baines, Wilson, and Clark, 2009, p. 229), more than four thousand and eight hundred papers have been published. As we can see in Figure 2, more and more articles were published each year. Thus, market segmentation attracts the attention of the researchers increasingly; therefore, I believe it to be quite an important issue.

Figure 2. Number of items published over the last 20 years



As a contrast, I took a look at the business market segmentation field. Similar to the market segmentation in general, the number of published articles was increasing progressively throughout the years, but to a much lesser extent, which confirms my statement in the beginning of the master's thesis that the industrial market segmentation is underdeveloped compared to the consumer market segmentation.

### **1.2.2 The reasons for the client segmentation**

The client segmentation has been considered to be a key concept in the field of marketing, which promises some great benefits that would sooner or later result in higher revenues. Therefore, it is not surprising the fact that companies are trying hard to employ it within their operations, since it enables them to understand their customers better and focus on the ones that bring them (more) profit.

Hutt and Speh (2013, pp. 91–92) listed the following benefits of the segmentation that can accrue to the firm if the necessary requirements are met:

- adjustment to the unique needs of customer segments;
- efficient and effective business marketing strategies generation;
- focus on the product-development efforts;
- profitable pricing strategies development;
- appropriate distribution channels selection;
- advertising messages development and targeting;
- sales force training and deployment; and so forth.

Even though McKechnie (2006, p. 119) claims that benefits of the client segmentation have to be weighed up against higher costs incurred in identifying and responding to the needs of specific segments (e.g., research and development, manufacturing, marketing and inventory costs), she believes segmentation offers many benefits. She argues that market segmentation provides better opportunities to:

- identify growth segments;
- create more customer value;
- increase customer satisfaction;
- charge premium prices;
- reduce competitive pressures; and so forth.

Certainly, there are so many more authors indicating lots of more benefits the customer segmentation can provide, however, their listing and discussion would probably outpace the frames of this chapter – or even the master's thesis as such.

### 1.2.3 The process of the client segmentation

In order to conquer in today's marketplace, a company has to put its customers in the centre. To do that, it has to develop such marketing strategy that will allow it to create customer value and consequently achieve profitable customer relationships (Kotler & Armstrong, 2012, p. 72). The process of the marketing strategy, which answers the questions "which customers to serve" and "how a value can be created for them", is depicted in Figure 3.

Figure 3. Marketing strategy process



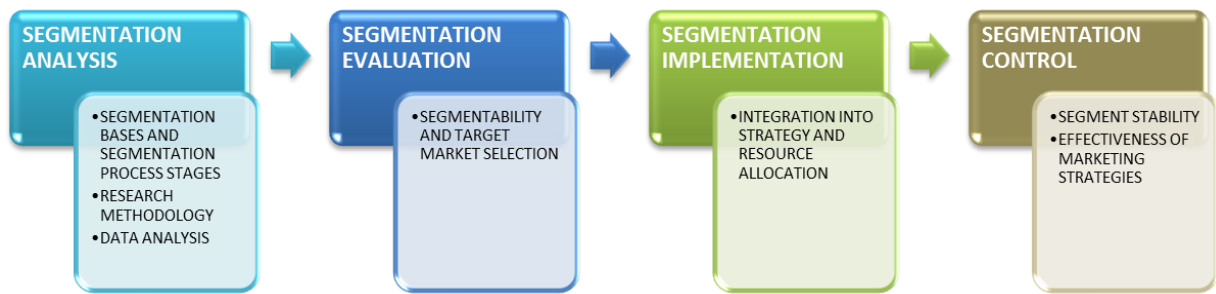
Source: P. Kotler & G. Armstrong, *Principles of Marketing*, 2012, p. 72.

The first step is the market segmentation; companies divide their markets into smaller groups of customers who respond to a given marketing effort in a similar way. This enables them to determine which segments are the most attractive, offering them the best opportunities, so they can select one or more segments to concentrate their struggles to (Kotler & Armstrong, 2012, p. 73). Finally, there come the market differentiation and positioning, which involve choosing how to present the offer to the market in a way that will take advantage of and advertise the company's competitive advantage (McKechnie, 2006, p. 119).

However, the process of the client segmentation itself is – in more detail – demonstrated in Figure 4 on the next page. As it can be seen, it is composed of four phases: segmentation analysis, segmentation evaluation, segmentation implementation, and segmentation control. Furthermore, the process of the market segmentation is described – step by step.

The segmentation analysis phase encompasses all activities involved in breaking the market into several sub-markets. It includes three elements: segmentation bases and segmentation process stages, research methodologies, and data analysis. The first element covers the appropriate models and segmentation basis selection; the task of the second element is the research methodology selection based on data requirements; while the data analysis element comprises the proper data processing tool selection (Goller, Hogg, & Kalafatis, 2002, pp. 257–261).

Figure 4. Segmentation process



Source: S. Goller et al., *A new research agenda for business segmentation*, 2002, p. 254.

The second step, segmentation evaluation, includes “segmentability” assessment and target market selection. Among all segments identified, the acceptable segments are the ones that are miscellaneous in-between and similar within each other. Based on several additional factors, such as segment size and growth, and expected market shares, the ultimate segments are then selected to be targeted (Goller et al., 2002, p. 261).

The segmentation process continues with its implementation; segmentation has to be integrated into all three levels of corporate decision making: strategic, tactical, and operational. Segments choice has to be done at a corporate level, and then translated into the tactical and operational levels through resources allocation (Goller et al., 2002, p. 263).

The last part of the segmentation process is its control. It involves the monitoring of the segment stability in terms of segments to remain homogeneous, and a potential re-segmenting of the market in case of changes in stability detection. Moreover, it relates to the effectiveness, or should I say the efficiency (i.e. segment profitability, customer retention and customer attrition rates, etc.) of the strategies implemented supervision. Those measurements enable possible gaps between the planned and the actualized strategies identification, and serve as a foundation for some corrective actions construction (Goller et al., 2002, pp. 264–265).

#### 1.2.4 Consumer versus industrial client segmentation

While business-to-business markets are analogous to consumer markets in many aspects, there are some complex forces that make them very much different. In the grand scheme of things, consumer and industrial market segmentations involve some common concepts; however, there are some basic differences that can influence segmentation strategy (Tsai, 2008, p. 39). Hague and Harrison of B2B International (n.d.) define eight main elements that indicate the difference between the two types of segmentation. These differences are further described below.

The first difference is the complexity of the decision-making process; a decision-making unit in the business market is far more complicated. Rather than to be limited to a very small family unit, as it is the case with household purchases, a purchase at an enterprise level may involve experts from assorted business functions, each of whom favouring the interests of their own function. Thus, a company may have to segment not only the organizations in which those decision-makers work, but also the decision-makers themselves (B2B International, n.d.).

Second, the difference between the two markets is also the B2B buyers to be much more rational. Instead of buying “what they want” (without placing the price in the first place), the industrial buyers typically buy “what they need”; therefore, the segmentation of business audiences should be easier, since their needs can be easier to determine than their desires (B2B International, n.d.).

Products bought by organizations are, as the decision-making process, more complex compared to consumer products. Whereas the latter are often standardized, products for companies are usually custom-made. The next difference is that the business target audiences are smaller than the consumer target audiences, but consume far more products or services (B2B International, n.d.).

Furthermore, relationships in business-to-business markets are more important. The smallness of the customer base makes it easier to give more attention to each customer, which enables a more personal interaction, resulting in relationship and trust development. It is not surprising that many companies have clients that have been loyal to them for many years (B2B International, n.d.).

The sixth difference is B2B purchases to be of a long-term essence, which is not the case with consumer buyers. Long-term purchases or purchases that are expected to be repeated over a longer extent of time are relatively rare in consumer markets. Onwards, the business-to-business markets drive innovation less than the consumer markets because, usually, the innovation is a response to a novelty that has happened further upstream (B2B International, n.d.).

Lastly, behavioural and needs-based segments are by far fewer in industrial markets, while it is not unusual for a consumer market to boast more than ten segments, an average B2B study normally produces three to four segments (B2B International, n.d.).

### **1.2.5 Segmentation bases and variables**

According to Kotler and Armstrong (2012, p. 215), there is no single way for companies to segment their customers. Instead of sticking with only one variable, they have to try different segmentation variables, alone or in combination. Thus, the aforementioned

authors outline four major market segmentation bases: geographic, demographic, psychographic, and behavioural. Each of these bases is further described below.

#### 1.2.5.1 Consumer market segmentation variables

Geographic segmentation consists of breaking the market down into smaller units, based on different geographical criteria; companies can use more sloppy classifications to divide their customers by nations, regions, states, or they can segregate their markets in a bit more profound way into counties, cities, or even neighbourhoods. It depends on an individual enterprise whether it is going to operate within just a few geographical areas, or it will cover all areas; however, local variations in needs and wants should always stick in their heads (Kotler & Keller, 2006, pp. 247–249).

Demographic segmentation is the most popular and widely used basis for segmenting the market. Demographic variables, such as age, gender, family size, income, occupation, education, religion, nationality, and so on, are very attractive groundwork for customer classification since they often reflect consumer needs, wants and expectations almost directly. Another reason for their recognition is that they are easier to measure compared to most of the other types of segmentation variables (Kotler & Armstrong, 2012, p. 215).

Another way for companies to distinguish their customer groups is by means of psychographic characteristics. According to Cambridge online dictionary (n.d.), psychographics is the science of studying and measuring customers in relation to their opinions, interests, emotions, values and lifestyles, using psychology and demographics. Therefore, knowing their customers' lifestyles, personalities and beliefs, companies can determine various customer profiles, and – later on – provide them with such products, services or brands that will match their own personality (Kotler & Armstrong, 2012, pp. 168–174).

In conformance to many marketers, behavioural segmentation can be the best cornerstone for setting up customer segments. Customers can be grouped under several behavioural characteristics, such as the time of purchase, the reasons for purchase, product usage, and so on. Below, some of the main behavioural variables, stated by Kotler and Keller (2006, pp. 255–257), are explained:

- buying occasions; which can be delimited in terms of time of day, week, month, year, or in terms of some important lifecycle events of a customer. Customers can be grouped correspondently to the occasions when they develop a need, actually make a purchase of a product, or use the bought;
- purchase benefits buyers are seeking for a certain product; it can be a powerful segmentation form;

- user status; customers can be divided by the status they hold to the product. Markets can be segmented into nonusers, ex-users, potential users, new users, or regular users, and so addressed differently, correspondingly to the status they possess;
- usage rate; a good way of grouping customers can be the rate of product usage, thus a company can identify light, medium, and heavy users;
- loyalty status; company can group its customers by the degree of their loyalty. Some buyers are absolutely loyal and buy one brand all the time, others are partially loyal and choose between two or three favourite brands, still other buyers show no loyalty to any brand.

Combining different behavioural bases tends to provide a more comprehensive and cohesive view of the market and its segments (Kotler & Keller, 2006, p. 257). However, a combination of multiple segmentation bases in general has proven to be the best decision for marketers when they want to identify smaller and better-defined target groups (Kotler & Armstrong, 2012, p. 222).

#### 1.2.5.2 Business-to-business market segmentation variables

Segmenting business-to-business markets, however, is not that different from the consumer market segmentation. Many authors, including Anderson and Narus (2004, p. 90), are certain that organizations should consider their corporate clients (companies) as individuals. According to Zimmermann and Blythe (2013, p. 121), many variables from consumer market segmentation can be used by business marketers as well. They say business buyers can be divided geographically, demographically and behaviourally, while psychographic variables are unadaptable in business markets, as they claim. Some other researchers, on the other hand, including Barry and Weinstein (2009, p. 315), argue just the opposite and advocate the usefulness of B2B psychographics, as we shall see later on. But now, let us take a look at how each consumer market segmentation basis can be placed in the context of the industrial client segmentation.

There is practically no difference in B2B over B2C segmentation in terms of geographic criteria. Here as well, geographic attributes are important and often represent a meaningful starting point for segmenting the market. Besides the fact it is rather low in costs and readily available, geographic analysis is known to be one of the easiest techniques for breaking the market down into possible target segments, as it is the case with the consumer market segmentation. Companies have to determine the scope of the market they are going to offer their products and services in. Therefore, they may go after the entire world, some selected countries, regional markets, specific cities, or particular neighbourhoods only. Nevertheless, regional differences may be considerable, so they can remarkably impact purchasing behaviour and product consumption (Weinstein, 2004, pp. 61–63).



Similar as with dividing a consumer market, where customers have been distinguished based on several variables, such as age, status, education, and the like, the industrial customers can be also categorized under a set of demographic factors, known as the “demographics of the firm” or firmographics.

As consumer desires and needs change with age, and thus serve as an excellent cornerstone for appropriate products and services offerings build-up, so can the age of a company, that is, the length of time a firm has been in business, be a useful segmentation dimension. For instance, a new company can be an attractive target for accounting or legal services providers since it does not have many relationships established with vendors (yet); marketing services agencies can focus on growing companies which are prone to some major investments; while trying to start a cooperation with mature companies that typically already have well-entrenched relationships with their sellers can be a waste of time (Weinstein, 2004, p. 72).

Customer distinction by means of financial factors is possible within industrial markets as well. In fact, the following type of segmentation is a long-standing practice for both the consumer and business market separations (Kotler & Keller, 2006, p. 251). While the qualification and the profession of an individual imply his or her earnings (the higher the education and the better the position, the greater the amount of the income), and allow marketers to distinguish among affluent and non-affluent customers, it is the amount of the profit, the share of the market, and the volume of the sales that come useful in the context of industrial segmentation by financial situation (Weinstein, 2004, pp. 72–73).

However, the volume of sales not only reveals the financial strength of a company, it also says a lot about its size; usually, the greater the sales, the bigger the company, and vice versa. According to Anderson and Narus (2004, p. 46), the size of a company can also be delimited by some other specific measures like number of employees or number of establishments, which usually come from secondary sources.

In order to support the development and success of the micro, small and medium-sized enterprises (hereinafter: SMEs), which play a central role in the European economy by providing 75 million jobs and representing 99% of all enterprises (European Commission, 2005, p. 5), the European Commission introduced “The new SME definition”. Therein, the size of a company is determined based on three factors: the number of employees, and either the annual turnover or the annual balance sheet. Table 1 on the next page encompasses the ceilings of the new definition.

Table 1. Company size categories by three criteria

<b>Enterprise category</b>	<b>Headcount: Annual Work Unit (AWU)</b>	<b>Annual turnover (in million)</b>	<b>or</b>	<b>Annual balance sheet total (in million)</b>
Medium-sized	< 250	≤ €50	or	≤ €43
Small	< 50	≤ €10	or	≤ €10
Micro	< 10	≤ €2	or	≤ €2

Source: European Commission, *The new SME definition. User guide and model declaration*, 2005, p. 14.

Another way of grouping industrial customers together is by industries in which they operate. Namely, the industry factor has been a very powerful unit of analysis for years as the industrial marketing doctrine consensually assumes that firms within a particular industry share several common characteristics (Mauri & Michaels, 1998, p. 212).

In order to help businesses, financial institutions, governments and other operators in the market by providing them with reliable and comparable statistics, countries have developed generalized business classification systems. These systems represent a framework for collecting and presenting a large range of statistical data in accordance with the economic activity. Statistical classification of economic activities, used within a European Union, is called NACE (acronym derives from the French name *Nomenclature générale des Activités économiques dans les Communautés Européennes*), and – at the highest level – consists of 21 different items or sections (Eurostat, 2008, pp. 5–44).

The comparison between consumer and business demographics is briefly summarized in Table 2.

Table 2. Consumer versus business demographics

<b>Demographic Category</b>	<b>Consumer Demographics</b>	<b>Business Demographics</b>
Age	Age distribution	Number of years firm has been in business
Financial factors	Income; occupation; education	Sales; profits; market share

Table continues

Continued

<b>Demographic Category</b>	<b>Consumer Demographics</b>	<b>Business Demographics</b>
Market size	Population; number of households; household size	Number of potential customers; number of establishments, stores, plants; number of employees
Industry structure/social class	Lower-lower to upper-upper class; lifestyle clusters	Industry position; goods versus services; economic activities classification

Source: Adapted from A. Weinstein, *Handbook of market segmentation: strategic targeting for business and technology firms*, 2004, p. 72.

However, segmenting customers based on demographics, geography or NACE codes often seems to fall short of answering the needs of industrial marketers, and can be – according to Neal (2002, p. 37) – suboptimal at best, and devastating at worse. That is why many authors, with Barry and Weinstein (2009, p. 315) at the forefront, believe that B2B psychographics can be a great alternative or complement to firmographics.

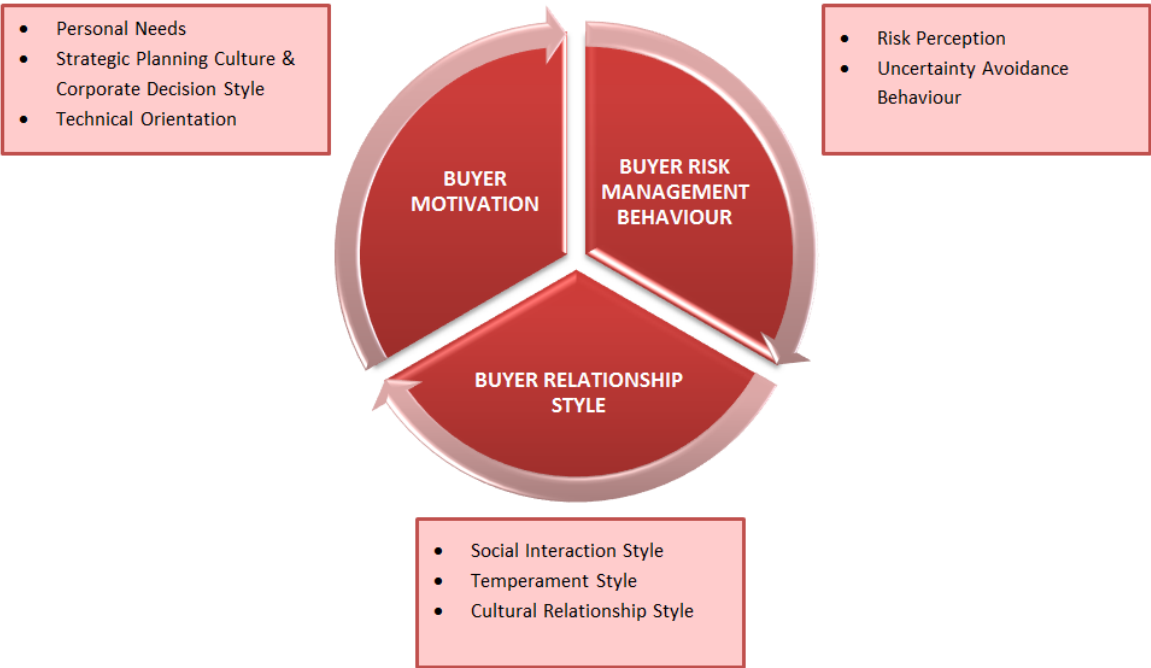
They describe organizational psychographics as the segmentation of organizational buyers into homogenous clusters of mind-sets and behaviours that are distinguished by motives, risk perceptions and social interaction styles in order to identify prospects as well as predict the predispositions of the firm’s decision-makers for the sake of adapting products, marketing messages and relational selling behaviours.

Figure 5 on the next page illustrates the above definition. As we can see, industrial customer’s buying mentality is much more complicated than it is the case of the consumer sector. By analysing purchase motives, marketers can better understand why buyers act the way they do in the marketplace; knowing a purchasing manager’s risk tendency can help in understanding and segmenting the purchasing situation; an analysis of buyer’s social interaction, temperament, and cultural relationship styles can assist in designing effective product strategies and promotional campaigns (Barry & Weinstein, 2009, pp. 320–335).

Still, the development of marketing strategies around psychographic dimensions is limited since frameworks for classifying buyers through psychographics are subjective and difficult to measure. Moreover, industrial market characteristics differ sharply from consumer markets on several dimensions; not only are many individuals involved in B2B settings, but special justifications, authorizations, and approvals often restrict the influence of personality on buying decisions (Barry & Weinstein, 2009, p. 316).

Nevertheless, the psychographic segmentation can help sales management devote the optimal amount of resources for communicating with and servicing their customers by understanding buyer’s predispositions. Knowing and understanding the motivations, risk management behaviour and relationship styles are vital for adaptive selling, account prioritization and resource allocation (Barry & Weinstein, 2009, p. 332).

Figure 5. Influences on buyer’s personal characteristics



Source: J. Barry & A. Weinstein, *Business psychographics revisited: from segmentation theory to successful marketing practice*, 2009, p. 319.

### 1.3 Client segmentation in the banking industry

Banking as well as other financial services, such as insurance or credit cards, was among the first industries where the idea of addressing different customer groups with differentiated products or services was approved (Bull, 2003, p. 594). Since Wendell Smith first proposed the market segmentation as an alternative strategy to the product differentiation back in 1956, and up until now, there have been used miscellaneous variables as a basis to segment customers within the financial services industry; from well-established segmentation bases, such as geographic and demographic, to more contemporary approaches, such as psychological or benefit segmentation (Machauer & Morgner, 2001, p. 6). However, the degree of knowledge about the corporate and the private bank customers remains in favour of the latter.

Before we take a look at the variables that are used by banks when developing appropriate segmentation and marketing strategies, let us first check, which are the banking industry particularities, and what are the main differences between products and services marketing.

### **1.3.1 Banking industry characteristics**

Like every other, the financial services industry has its own characteristics that make it unique. The banking industry itself, however, is considered to be evidently different from other industries; conformable to a large body of research, the development of a financial system is crucial for growth, development, and stability of the entire economy (Barth, Caprio, & Nolle, 2004, p. 2).

Therefore, it is important to know the aspects which beneficially affect the progress of the banking system, and consequently, the economy as a whole. Key characteristic of banking systems, pointed out by Barth, Caprio and Nolle (2004, pp. 2–39), are discussed below.

The first feature of a banking system covers the matter of the ownership of banks; indeed, we can distinguish between the private-versus-government and the foreign-versus-domestic ownership. Based on their 55-country-wide research, Barth, Caprio, and Nolle (2004, p. 8) found a negative correlation between the degree of government ownership and the amount of national income. In other words, the higher the degree to which the government is involved in the banking system, the lower the overall performance of the economy.

However, things are not that straightforward when it comes to foreign and domestic ownership comparison; the performance of foreign banks may differ among countries with different income levels, and so may their influence on the economic development. As a confirmation, Claessens, Demirgüç-Kunt, and Huizinga (2001, p. 908) found foreign banks to exceed home-country banks in developing countries, though fall behind domestic banks in already developed countries.

The second characteristic relates to the competitiveness in the banking industry, measured by concentration ratios, which are the percentages of market shares, deposits or assets, held by the largest banks in the industry. As Laeven and Claessens (2003, pp. 563-583) argue, the higher the degree of concentration, the greater is the banks' tendency to invest in gathering and processing the information about borrowers. Therefore, a high market power leads to a greater access to external finance for firms, resulting in an enhanced economic growth.

In addition, based on their research with the data from nearly 80 countries, Beck, Demirgüç-Kunt, and Levine (2006) determined the positive correlation between the concentration and the stability of a banking system; concentrated banking systems foster more stability since they are less vulnerable to banking crises.

Regulation and supervision are the next characteristic of the banking industry; since the crisis in the banking system could be felt throughout the complete economy, banks have

been regulated in every country. What is more, the banking industry may be the most regulated and supervised of all industries. However, approaches and measures vary across countries (Barth, Caprio, & Nolle, 2004, p. 17): still – according to findings of Barth, Caprio and Levine (2003, p. 234) – a more stringent control is negatively associated with bank performance.

The last, yet evenly important characteristic of a banking system encompasses the corporate governance of the banks. When the goals of the owners diverge considerably from the ones of the managers, we talk about the so-called “principal-agent” problem. Namely, managers, who were hired to “do the job” for the owners in the first place, may become egoistic, start to pursue their own interests over the goals of the owners whose motive is to maximize the value of the share. Given the special connection between banks’ performance and the overall performance of the economy as a whole, banks’ capability to overcome the aforementioned issue is of crucial importance. According to Caprio and Levine (2002, p. 19), an effective corporate governance of banks will eventually lead to a more efficient capital allocation, subsequently resulting in greater economic development.

### **1.3.2 Products versus services**

Even though many have argued that there are almost no companies that are being completely product- or service-based (Araujo & Spring, 2006, p. 802), nor are many products that are being pure goods (Johnson & Weinstein, 2004, p. 65), it is wise to know the differences between products and services, and, particularly, the marketing approaches of the both.

For many years, products and services have been distinguished based on four specific features, first ratified by Zeithaml, Parasuraman, and Barry back in 1985 (p. 43): intangibility, heterogeneity, inseparability, and perishability, also known as IHIP characteristics. The intangibility means the service cannot be seen, tasted, felt, heard or smelled before it is bought. The heterogeneity implies on the dependence of who provides it as well as when, where, and how it is provided. The inseparability refers to the fact the service is produced and consumed at the same time, thus cannot be separated from its provider. Finally, the perishability means the service cannot be stored for later use or sale (Kotler & Keller, 2006, pp. 405–408).

The four above-described characteristics have long been assumed to be making services unprecedentedly different from products, although there is an increasing number of researches suggesting a bunch of weaknesses of that almost unquestionably accepted framework, recommending reconsideration and reconceptualization. Vargo and Lusch (2004, p. 327), for instance, argue that IHIP characteristics can be misleading about the nature of marketing offerings and contradictory to customer orientation. Furthermore, Lovelock and Gummesson (2004, pp. 32–34) suggest – instead of having them separated –

services marketing and goods marketing to be reunited under a service banner. Another way they see marketing could be undertaken is based on the differences among services themselves, rather than just between products and services.

Anyway, given that the question on how to identify and take into account the differences between products and services in their marketing obviously calls for some more research, it is reasonable to check the significance of services for the economy.

## **2 BUSINESS CUSTOMERS SEGMENTATION WITH THE USE OF K-MEANS AND SELF-ORGANIZING MAPS: AN EXPLORATORY STUDY IN THE CASE OF A SLOVENIAN BANK**

Before we move to the business case examination, let us see the importance of the services sector for the Slovenian economy. Similar to the European Union, where tertiary sector holds about 75 percent of the Europe's GDP, offering more than three-quarters of all job opportunities (European Commission, 2014), in Slovenia as well, the services sector dominates the economy. In 2014, services providers contributed to 55.7 percent of the Slovenia's GDP (Statistical Office of the Republic of Slovenia, 2014a), and employed more than 62.4 percent of the whole working population (Statistical Office of the Republic of Slovenia, 2014b).

The banking industry itself, on the other hand, did not play a significant role in contributing to GDP growth in 2014; as a matter of fact, it accounted to as low as 2.3 percent of the Slovenia's GDP (Statistical Office of the Republic of Slovenia, 2014a), and was one of those industries that disseminated effects of the global economic crisis on the domestic economy.

### **2.1 Business case presentation**

The analysis, carried out in this master's thesis, is based on a real-life business case – an industrial customers dataset of one of the major Slovenian banks, which – due to data sensitivity and confidentiality – preferred its name to remain undisclosed. Therefore, in order to ensure better coherency, we will call it “Bank S”. Below, there is a brief description of the bank and its operations.

#### **2.1.1 Introduction of Bank S**

Bank S is one of the major retail banks in Slovenia that provides a broad range of various banking services. It interacts with its customers through various distribution channels; besides the high coverage with offices across the entire Slovenia (there are more than 60 of

them), it enables its customers to use their services via several contemporary banking channels such as widespread ATM network, electronic banking, telephone banking, and mobile banking. Through the branch network, together with its sales agents, it also offers all leasing products to both individuals and companies.

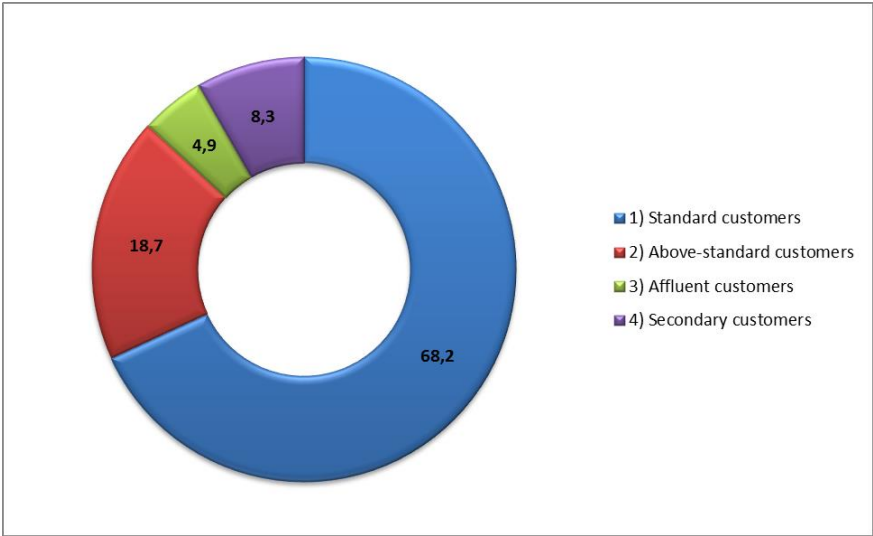
Bank S looks out to building extensive, long-term relationships with its customers. It constantly seeks to enhance the quality of its services by quick responsiveness to its clients' needs, and constant improvements of the offices in its network; it tries to satisfy the needs of its existing clients while attracting new ones. Bank S gives a special focus on individuals, sole proprietors and SMEs. For each segment, it looks for specific strategies with dedicated sales approaches and corresponding managerial support lines. At all stages of the sales process, it tries to develop a comprehensive and a standardized approach.

The range of its banking services meets all financing needs of its clients. With respect to the savings products, Bank S offers its customers a wide scope of investment opportunities as well – it is developing a wide variety of insurance and investment products in cooperation with selected insurance companies.

**2.1.2 Bank’s business needs identification**

As many other companies, Bank S is also aware of the great importance of the customer segmentation at present. Therefore, it already has in place its individual-customers segmentation; based on the data about customers and the products that those customers possess, they classify them into few smaller groups. Having done so, they identified the following segments: “Standard” customers, “Above-standard” customers, “Affluent” customers, and “Secondary” customers. The structure of individual customer segments is demonstrated in Figure 6.

Figure 6. Individual customer segments in %





However, what they still miss is the corporate customer segmentation. They want to prepare the segmentation for sole proprietors, and SMEs, and eventually, extend it across their entire business customers portfolio. As in the European Union, where SMEs play a central role in the economy by providing more than 75 million jobs and representing 99 percent of all enterprises (European Commission, 2005, p. 5), in Slovenia as well, SMEs run the economy, they represented 99.82 percent of all companies in 2014, producing together about 70 percent of the whole Slovenian income (Statistical Office of the Republic of Slovenia, 2014c).

How about the Bank S? The entire corporate customer portfolio brings about one half of their net banking income. SMEs, however, contribute to about 40 percent of the income, made by companies, which represent 20 percent of the entire net banking income.

## **2.2 Analysis design**

Operations of today's companies have been a very broad concept. Since businesses have been so expansive, the number of different transactions has become endless. Banks, however, are among the companies who operate with the greatest number of transactions every day. Thus, to be able to collect all the data and store it for later use, they need to build a data warehouse. Still, in order for different departments to work most efficiently, an enterprise-wide data warehouse needs to be further sliced into several smaller subsets, that is, data marts. The case of the Bank S is the same; all of their data has been organized into a few departmental data marts, and the source for the analysis, carried out in this master's thesis, was the data mart of the Marketing Department.

### **2.2.1 Data collection**

To collect all data needed to perform the analysis, the use of the two Business intelligence tools – Oracle Discoverer 10g and MicroStrategy 9 – was required. Therefore, I was gathering the data for almost two months since the beginning of 2015. With the help of the Marketing Department, I was able to gather some client-related data, as well as specific data on each product that those clients held, from their data mart. One part of the data, more precisely a balance sheet data, was obtained from an external data source – Agency of the Republic of Slovenia for Public Legal Records and Related Services. The entire data acquisition process is described in the following paragraphs.

First, I needed to get a list of bank's entire small and medium-sized corporate-client portfolio to which I was going to add various data afterwards. In order to do this, the Marketing Department carried out for me a query from the "Client" data mart table. Since we did not want to retrieve all of their industrial customers, they had to determine certain criteria. The criteria that were used to attain the initial list of clients are briefly described in Table 3 on the next page.

Table 3. The initial customer list exportation criteria

Attribute	Criteria	Description
Size classification	Include only micro, small, and medium-sized companies	<ul style="list-style-type: none"> <li>– Micro: up to €100,000 of revenue</li> <li>– Small: more than €100,000 and up to €2,500,000 of revenue</li> <li>– Medium-sized: more than €2,500,000 and up to €5,000,000 of revenue</li> <li>– Large: more than €5,000,000</li> </ul>
Primary/Secondary Relationship Indicator	Include only primary clients	<ul style="list-style-type: none"> <li>– Primary relationship: owner of any product or channel (service)</li> <li>– Secondary relationship: holding only insurance</li> </ul>
Activity Indicator	Include only active clients	<ul style="list-style-type: none"> <li>– Active client: primary or secondary client that holds/uses at least one active* product/service</li> <li>– Inactive client: primary or secondary client that holds/uses not one active product/service</li> </ul>

*Note.* \* The definition of an active product/service is explained in the next table.

In regard to the above criteria, the original dataset consisted of 17,855 records, arranged in a tabular form. Each of the rows represented a single customer, while the columns included different client-related data. Besides some of the key data, such as Client Number, ID Number, and Tax File Number, there were also some contact details, such as Company Name, Address, Post Code and Name, Telephone Number, and Email Address. Moreover, the initial file included several data on bank-client relationship: Relationship Starting Date, Relationship Officer Code and Name, Business Unit Code and Name, and Branch Code and Name.

In addition, information about each Client Value was provided. That is to say, Bank S has been calculating each customer value over their entire relationship lifetime. The formula has been much more complicated, but to put it simply, they computed what have been the

earnings from each customer, and how much has each customer cost them from the beginning and to the present day.

The next information I got was Double Relationship Indicator, which suggested whether there had been a connection between a corporate client and an individual customer. For the ones, where that connection had been detected, I also received the data on Individual Customer Value and Individual Customer Segment (according to the outcomes of the individual-customers segmentation).

Finally, the results of the first query included information on the status of each product the original set of clients held. As already mentioned above, Bank S offers its clients an extensive assortment of banking products – apart from basic banking products, such as transaction accounts, payment cards, and electronic banking, it also provides its customers several business financing services such as overdrafts, loans, leasing products, and guarantees. Last but not least, various insurance products as well as plenty of investment opportunities have been enabled to their customers.

For most of their products, they have system-generated indicators, which suggest the usage of each product intensity, while for some of them, they need to calculate indicators by themselves based on other departments’ data-mart information. Table 4 explains particular values of each product indicator.

*Table 4. Product indicator description*

<b>Product Indicator</b>	<b>Possible values</b>	<b>Activity specifier</b>
Account	<ul style="list-style-type: none"> <li>– A: product activity in the last 3 months</li> <li>– B: product activity in the last 3 to 6 months</li> <li>– Y: product activity in the last 6 to 12 months</li> <li>– U: product activity more than 12 months ago</li> <li>– N: client never held a product</li> </ul>	1 client-initiated debit or credit transaction (regardless if from the client or secondary account holder) in a given period
Cards		1 client-initiated transaction in a given period
“Plus” Card		contract renewed (or closed) in a given period
Insurance		contract renewed (or closed) in a given period
Loans		existence of used amount of overdraft in a given period
Overdrafts		contract renewed (or closed) in a given period
Guarantees		contract renewed (or closed) in a given period
Deposits		1 connection in a given period
Electronic Business		

After gaining the above-mentioned client-related information, I had to obtain all product-pertinent data. For that reason, the Marketing Department conducted several different inquiries, one for each product. All the variables as well as the criteria used to export each product data are described below.

All account-related data was the first to be collect; considering all active, that is, non-closed, corporate accounts, I was able to get information on the following variables for the period of three months retrospectively:

- Contract Signature Date (the date when a client opened the account);
- Amount of Inflows (amount of all client initiated credit transactions, without fees and interest);
- Number of Inflows (number of all client initiated credit transactions, without fees and interest);
- Amount of Outflows (amount of all client initiated debit transactions, without fees and interest);
- Number of Outflows (number of all client initiated debit transactions, without fees and interest);
- Amount of Credit Operations (amount of all credit transactions, both client and bank initiated);
- Number of Credit Operations (number of all credit transactions, both client and bank initiated);
- Amount of Debit Operations (amount of all debit transactions, both client and bank initiated);
- Number of Debit Operations (number of all debit transactions, both client and bank initiated);
- Average Monthly Balance (average of each day in a month balance);
- End of Month Balance (the last day in a month balance);
- Average Not Allowed Balance (average of all balances that were lower than allowed negative balance); and
- Number of Not Allowed Balance Days (number of days when client was in not-allowed negative balance).

The next set to export was the data about payment cards, more precisely about credit cards transactions. Taking into consideration all non-closed and non-blocked cards, the second file included following corporate-customer credit cards information:

- Monthly Purchase Amount (sum of all client initiated transactions);
- Number of Purchases per Month (number of all client initiated transactions);
- Amount of Withdrawals per Month (sum of all ATM cash withdrawals);
- Number of Withdrawals per Month (number of all ATM cash withdrawals);

- Fees (system-generated transaction, i.e. amount, automatically charged to every client that holds credit card – besides company analogous monthly membership fee, it also includes commissions for each credit card holder within the company); and
- Interest (system-generated transaction, i.e. proportion of all ATM cash withdrawal amounts).

Information on overdrafts was the next that the Marketing Department queried for. After setting the currencies of all accounts to convert to euro, and including only those overdraft holders whose product contract expiration date was greater than date of data exportation, they were able to provide me the following variables:

- Total Authorized Amount (amount of allowed negative balance, agreed in the contract);
- Start Date (the date when a client can start drawing an overdraft amount);
- End Date (the date when drawing of an overdraft amount is disabled);
- Average Drawn Amount (average of all allowed negative balances); and
- Debit Interest Amount (amount, charged to a client for being in allowed negative balance – it is automatically calculated based on contractually specified debit interest rate, and drawn overdraft amount).

We continued with data on loans exportation; even though short-term and long-term loans were exported separately, exactly the same parameters were used for both datasets, and so were identical variables utilized. Thus, limiting data to only active, corporate loans for the course of three months retrospectively, I was provided the following information:

- Contract Signature Date (the date from when a client can start to draw on loan);
- Contract Amount (amount of a principal to be remunerated);
- Number of all Loans (number of all active contracts);
- Total Drawn Amount (amount of funds that are drawn down by a client);
- Total Outstanding Amount (a principle balance that needs to be paid off); and
- Expiry Date (the date when a client must repay all of its liabilities, including both principal and interest).

The Marketing Department provided the data about another financing product, the guarantees. Actually, the only variable I got in this regard was Contract Amount, which represents a sum for which bank gives a warranty that a client is going to pay off its liabilities.

The last information, available from the Marketing Department's data mart, was the one on deposits. Looking at all active, non-closed, corporate-customer deposits for three months backwards, we got information on the following features:

- Contract Signature Date (the date when a client started a deposit);
- Ledger Base Balance (amount of client’s funds to be safe-kept plus interest paid out);
- Amount of Deposits (amount of all client initiated credit transactions);
- Amount of Withdrawals (amount of all client initiated debit transactions); and
- Amount of Credit Operations (amount of all credit transactions, both client and bank initiated (interest)).

In order to collect the data I was still missing, I had to go wider than the Marketing Department. Therefore, to attain information I needed, I had to contact other departments of Bank S. In fact, rather than cooperating with departments within a bank, I worked with some of their subsidiary companies. Hence, Bank S Leasing subsidiary, concerned with everything related to leasing products solely, was the one who provided me data on the latter – arising from their own data mart, they were able to deliver the following details on each customer:

- Number of all leasing products (number of all different active leasing contracts by the same client);
- Total contract amount (sum of all contract amounts); and
- Average contract duration (average length of contractual obligations).

POS-terminals related data were available to me by another subsidiary company of Bank S – Bankart. In order to reduce operational and development costs as well as unification in the field of self-service and card operations, Bank S – together with twelve other Slovenian banks – decided to found a separate company back in 1997. Since then, Bankart’s mission is to provide the reliable, safe and cost-efficient processing of transactions with different bank payment instruments and to take care of the due development, building and maintenance of an appropriate information environment to enable the continuous and quality use of their services by all their clients (Bankart, 2015). Thus, information I got from Bankart was as follows:

- Number of all POS terminals (number of all rented devices);
- Number of fixed POS terminals (number of all non-portable devices);
- Number of portable POS terminals (number of all devices, possible to move);
- Number of internet POS terminals (number of devices, used for internet payments);
- Number of PIN pads (number of all PIN-pad extensions);
- Amount of transactions (amount of all transactions in the current year, i.e. from the beginning of the year and until the date of data exportation);
- Number of transactions (number of all transactions in the current year, i.e. from the beginning of the year and until the date of data exportation);
- Fees (amount of all fees charged in the current year, i.e. from the beginning of the year and until the date of data exportation); and

- Amount of monthly POS rent (sum of all contractually agreed monthly amounts, paid by client for use of each POS terminal).

Finally, balance sheet data was the ultimate to obtain. With the help of the Marketing Department, who helped me select appropriate balance sheet variables, I exported all the annual reports needed; Agency of the Republic of Slovenia for Public Legal Records and Related Services provides data on its website free of charges. Divided in two separate files, one for companies and one for sole proprietors, I was able to collect the same balance sheet data for both groups as follows:

- Industry;
- Assets;
- Inventories;
- Short-term loans to companies within the group;
- Equity;
- Non-current liabilities;
- Current liabilities;
- Net income from sales;
- Depreciation;
- Operating profit; and
- Net profit or loss for the year.

Now that I had the data I called for, that is, all client-related data (data on accounts, cards, overdrafts, loans, guarantees, deposits, POS-terminals, leasing products), and the balance sheet data, I needed to link them all together. Using Client Number as a key to combine all the above-mentioned data from different files into a single spreadsheet, a final dataset to be analysed conducted of originally retrieved 17,855 clients, each of them equipped with 86 different characteristics describing it.

However, the resulting file had not been ready for the analysis yet; I still needed to evaluate the data and decide which variables to include. But to do this, I first needed to decide which methods to use in order to properly prepare the data. Therefore, before we move to the data pre-processing procedure description, let us check the methods used in this master's thesis. The next chapter explains both of the chosen methods, and the reasons for their selection.

### **2.2.2 Methods selection**

As already mentioned, nowadays, customer segmentation is one of the most important concepts in the marketing. Its main objective is to organize a large set of customers into a few smaller groups that are homogenous within and heterogeneous between each other.

Even when dealing with just a couple of variables, this job does not seem to be that easy, and the results may vary considerably from marketer to marketer. However, handling with a large number of variables appears to be much more difficult. Thus, in order to get as precise results as we could, we had to let the numbers themselves help us. A statistical technique, used to identify groups in data, is called cluster analysis (Kaufman & Rousseeuw, 2005, p. 1).

Nowadays, cluster analysis has been applied in a broad range of domains, including astronomy, biology and ecology, geography, marketing and economics, medicine and psychology, political sciences, artificial intelligence, and so on. Not only it can be used to identify a structure already present in the data, cluster analysis as well can impose a structure on a dataset, seeming to be homogeneous at first. In pursuit of covering the need to classify data in more than three dimensions as well as to reaching the maximum objectivity, a wide variety of methods and algorithms has been developed. However, most of them usually operate on either of two input structures; the first presents the objects by means of  $p$  attributes that can be arranged in an  $n$ -by- $p$  matrix, where the rows conform to the objects and the columns to the attributes. The second is a set of proximities, which have to be available for all pairs of objects, and create an  $n$ -by- $n$  matrix (Kaufman & Rousseeuw, 2005, pp. 3–4).

According to Rokach and Maimon (2005, pp. 330–340), clustering methods can be divided into several categories – besides some newer kinds of clustering methods, such as density-based, model-based, and grid-based methods, the most common clustering methods differentiation remains the one given by Farley and Raftery back in 1998, distinguishing them into hierarchical methods and non-hierarchical or partitioning methods. Even though partitioning clustering methods suffer from a high complexity, they come very useful when it comes to dealing with large datasets, unlike hierarchical methods where that is not the case (Andritsos, 2002, p. 8).

There are many algorithms that can be classified as non-hierarchical clustering, however, one of the simplest and most commonly used is the  $k$ -means method. As Rokach and Maimon (2005, pp. 333–334) argue, there are several reasons for its popularity. First of all, it is computationally attractive due to the linear complexity, and therefore, often comes very useful when handling with substantially large datasets. According to Dhillon and Modha (2001, p. 145),  $k$ -means also boasts with an ease of interpretation, as well as its ability to converge quickly. Moreover, it is simple to implement and can be efficiently parallelized. Its adaptability to the sparse data and exploitation of the latter is another advantage of the  $k$ -means algorithm. All those characteristics were one of the reasons to select  $k$ -means as the first method to use in the thesis.

And how does  $k$ -means work? Before the clustering is started, user has to specify the number of clusters,  $k$ . The procedure starts by selecting  $k$  random data points to be the



seeds or centroids of each cluster. The next step is to assign all of the remaining data points to its closest cluster centre. Usually, this is done by means of Euclidean distance; each data point is assigned to a cluster whose centroid is located at a minimum distance from a data point itself. Once all of the data points are assigned to exactly one of the clusters, the next iteration of the procedure may begin. First, new cluster centroids are calculated; this time, the job is much easier since the centroids identification is simply a matter of taking the average value of each dimension for all records in the cluster. Each data point is then reassigned to the new cluster with the closest centroid. This process of assigning points to appropriate clusters and recalculating new centroids continues until clusters' structure stops changing (Berry & Linoff, 2004, pp. 354–356). A formal description of the  $k$ -means algorithm procedure is as follows (Larose, 2005, p. 153):

- Step 1: defining the number of clusters  $k$ ;
- Step 2: randomly assigning  $k$  points to be the initial cluster centre locations;
- Step 3: assigning each record to adequate cluster by finding the nearest cluster centre;
- Step 4: calculate  $k$  cluster centroids, and update the location of each cluster centre to the new value of the centroid;
- Step 5: repeat steps 3 to 5 until convergence or termination.

In spite of many advantages the  $k$ -means method has in comparison to other methods, there are also some weaknesses of the latter. According to Rokach and Maimon (2005, p. 334), its main proneness lies in the sensitivity to the selection of the initial partition. Namely, an improper selection of the initial partition may make the difference between global and local minimum. Since it requires the number of clusters in advance, it is the credibility of the results that is at stake when no prior knowledge is available. Moreover,  $k$ -means is sensitive to noisy data, and it does not perform well in presence of the outliers. In addition, it is often limited to numeric attributes only.

Therefore, I wanted to try to do some clustering using a method that is less sensitive to local optima, and is not so vulnerable to multivariate data outliers. In order to get better outcomes, using an algorithm that allows search space to be better explored, and results to be better manifested (Bação, Lobo, & Painho, 2005), I decided to choose Self-Organizing Maps as a second method to use in this master's thesis.

Self-organizing maps (hereinafter: SOMs) belong to a family of so-called artificial neural networks (hereinafter: ANNs). ANNs are a class of powerful, general-purpose tools for prediction, classification and clustering that imitate the way human brain works. Namely, the neural connections in the brain make it possible for people to make decisions and solve specific problems based on their previous experience. Thus, the ANNs simulate human ability to learn from experience by learning and generalizing from data inputs (Berry & Linoff, 2004, p. 211).

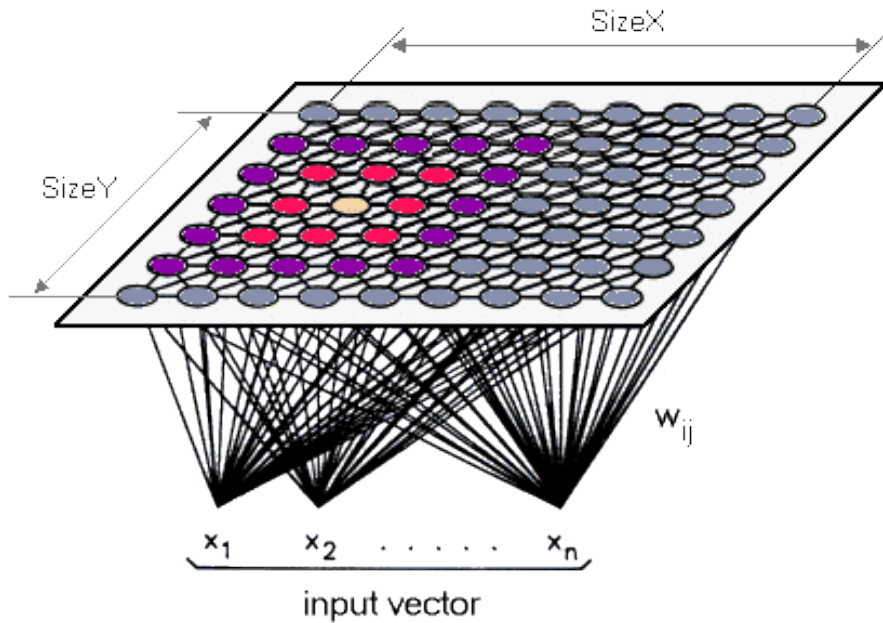
SOMs are a network that performs a non-linear projection of multidimensional input data onto a lower dimension array of units (Henriques, 2011, pp. 13–14). Invented back in 1982 by the Finnish researcher Tuevo Kohonen – this is why they are also called Kohonen networks – they were initially used for recognition of unknown patterns in images and sounds. Today, they have been used for various purposes, including undirected data mining tasks such as clusters in data discovery (Berry & Linoff, 2004, p. 249).

The concept of SOMs is to project high-dimensional data onto a one-, two-, or three-dimensional map, maintaining the initial relations between data patterns, through an unsupervised learning process. The result of this process is a low-dimensional grid – also referred to as the output space – made by the ordered array of units or neurons. On the other hand, the original dataset in which the data patterns lie is called the input space. Therefore, patterns that are close in the input space are mapped to units that are close in the output space (Henriques, 2011, p. 14).

The output space is mostly a two-dimensional grid in shape of a rectangle, however, hexagonal grid may also be an option (Kohonen, 2001). Some authors have also applied one- and three-dimensional SOMs; however, using more than three dimensions is not that common because, even though theoretically not disputed, the visualization of an output space in a familiar and intuitive manner gets challenging (Henriques, 2011, pp. 14–15).

Therefore, each unit or a node of the output space is connected to every unit of an input space, but not to each other. The number of the output space units is determined by the user – how this task is done, we shall see below – while the number of the input space units equals to a number of attributes on which the original data items are characterized. The connections between input layer units and output layer units are represented by vectors, however, the strength of those connections is measured by weights (Henriques, 2011, p. 15). This basic explanation of how the SOMs work is demonstrated in Figure 7 below.

Figure 7. SOMs – a rough graphical representation



Source: S. Joglekar, *Self-organizing Maps with Google's Tensorflow*, 2015.

Now, let take a look at the whole SOMs procedure. It starts by defining some criteria that are used to control the convergence process, and which affect the final clustering result. These contain some SOM structure related limitations (size, topology, shape, and initialization of the map used), as well as some training parameters (number of iterations, initial learning rate, and initial neighbourhood radius). Since there exist no theoretical guidelines on what are the optimal values for these initial parameters, user's experience comes to the fore (Henriques, 2011, pp. 17–20).

The size of the map depends on the SOM output space dimensionality, and it is given by the product of the number of units used in each dimension. Nevertheless, defining the number of units remains mainly an empirical process (Kohonen, 2001). Next, the topology and the shape of the SOM have to be selected. Even though the topology does not affect network's performance when training lasts long enough, user may normally choose between square and hexagonal one, and the latter tends to produce smoother maps (Jiang, Berry, & Schoenauer, 2009). Similarly, there are several shapes of SOM available for user to choose among, still, the most commonly used is the sheet.

The last parameter to set refers to the type of initialization, which can be either random or linear. If the user decides to go for random initialization, then the unit's weights are randomly drawn from the input training samples. Otherwise, the unit's weights are initialized along a linear subspace (Henriques, 2011, p. 22).

After the initial parameters are determined, the training parameters can be delimited. First, user needs to determine the duration of the training phase. This is done by setting the

number of iterations; in each iteration, one input pattern is presented to the network. Another way to do this is to define the number of the epochs. The difference between an iteration and an epoch is that one epoch consists of a group of iterations where all the input patterns were presented to the network once. Regardless to the iteration or the epoch, input patterns can be presented to the network randomly, or they can follow the order of the dataset (Henriques, 2011, p. 22).

The next parameters to set are the learning rate and the neighbourhood radius. The learning rate can have values between 0 and 1, and is initially given by the user. However, it decreases to zero during the training phase. The neighbourhood radius can have values between 0 and the maximum size of the network, and it defines the number of SOM units to be updated by each iteration (Henriques, 2011, pp. 23-24).

After all the above-described parameters are set, the learning phase may begin. For an earlier-set number of iterations, each pattern from the dataset is selected and presented to the network. Based on the Euclidean distance, the closest unit is found; this unit is called the best-matching unit, or shortly BMU. Now, the update phase can start. According to initially defined learning rate, and based on the initially delimited neighbourhood function and radius, the weights (and positions) of the BMU and its neighbours are modified in order to get closer to the data pattern each time presented. Since both the learning rate and the neighbourhood radius decrease as iterations progress, in order for SOM to reach stability, both parameters should converge to zero (Henriques, 2011, pp. 15–17). Table 5 presents a formal description of SOM algorithm.

*Table 5. SOM algorithm – a formal description*

For each input vector  $x$ , do:

---

- For each output node  $j$ , calculate the value  $D(w_j, x_n)$  of the scoring function.
- Find the winning node  $J$  that minimizes  $D(w_j, x_n)$  over all output nodes.
- Identify all output nodes  $j$  within the neighborhood of  $J$  defined by the neighborhood size  $R$ . For these nodes, do the following for all input record fields:
  - Adjust the weights:

$$w_{ij, new} = w_{ij, current} + \eta(x_{ni} - w_{ij, current})$$

- Adjust the learning rate and neighborhood size, as needed.
- 

Stop when the termination criteria are met.

Source: Adapted from D. T. Larose, *Discovering knowledge in data: an introduction to data mining*, 2005, p. 166.

### 2.2.3 Data pre-processing and variables selection

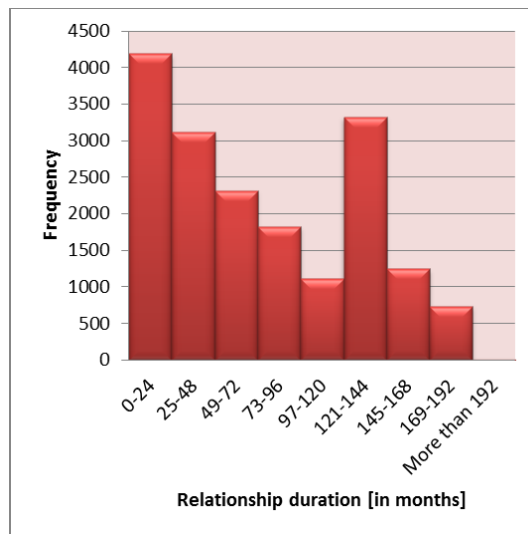
As with every other data mining technique, it is the same with clustering; in order for cluster analysis to achieve credible results, it requires some data evaluation before it can be started. In this chapter, a description of the variables selection process is provided.

Since both *k*-means and SOMs can handle only numerical data, the first thing to do after connecting all the data into a same spreadsheet was to exclude all the alphabetic, descriptive variables. Therefore, I removed all the identification data (Client Number, Tax File Number, and ID Number), company contact details, and some of the relationship-related data (Business Unit ID and Name, Branch ID and Name, Relationship Officer ID and Name, Double Relationship Indicator, Client Value – Individual, and Segment – Individual). Even though geographical data (Post code and name) could be interesting for clustering, I decided not to use them in this master's thesis. Namely, Slovenia as a country is – geographically – relatively small, thus it would be unreasonable to provide customized banking offerings to different areas. Also, most headquarters of Slovenian companies are located in one post area, but they provide their services to customers across the entire country.

Furthermore, I had to adjust all the date variables in order to transform them into numerical ones. So, I subtracted Relationship Starting Date from the date of data exportation, and rounded the result to the number of completed months. This way, I got the new variable – Relationship Duration. I did the same for all date attributes, and changed Contract Signature Date for accounts and deposits into Contract Duration. For the loans, I calculated Contract Duration attribute as the difference in subtraction between Contract Signature Date and Expiry Date.

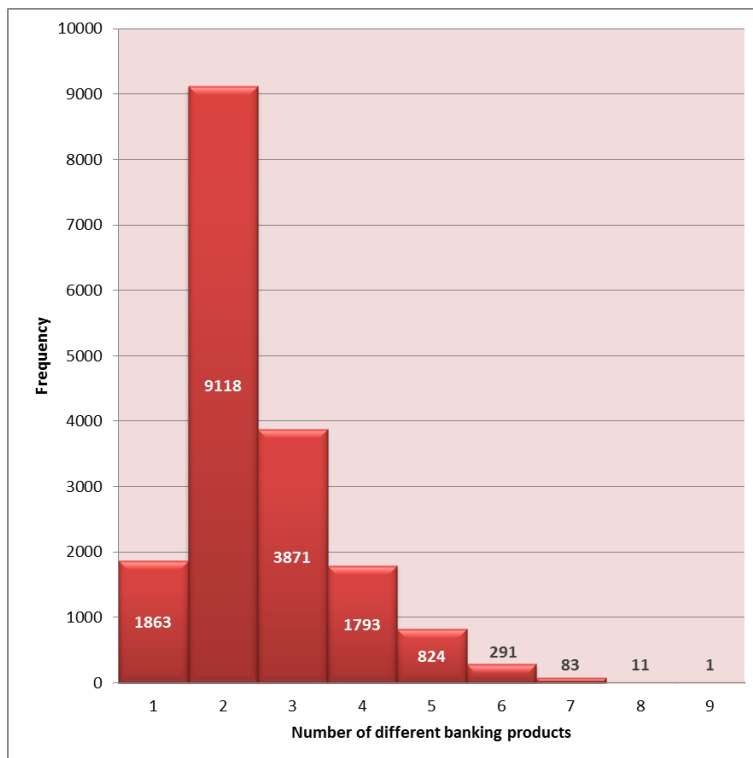
Since *k*-means is especially effective when dealing with normal distribution, I wanted to check it for the newly established Relationship Duration variable. Taking into consideration an entire industrial customer dataset, I was able to draw a histogram like the one presented in Figure 8 on the next page. As evident, having two data peaks, this was not the case with normally distributed data. In order to identify the reasons for this, I contacted the Marketing Department. They found an explanation for the first peak in Bank S becoming more and more attractive to companies with every year. However, the second peak happened to be a result of a data migration error which occurred in 2004. Therefore, even though duration variables could be interesting for the clustering, or at least clusters profiling, I decided to exclude them completely from this master's thesis.

Figure 8. Relationship Duration histogram



Since categorical variables were also less appropriate for *k*-means and SOMs, the next thing to do was to remove all the product indicators. Instead, I rather created a new variable based on those indicators – after replacing all “A”s, “B”s, and “Y”s with number 1, and all “U”s and “N”s with number 0, I could easily calculate the new variable – Product Coverage. As we can see in Figure 9, most of the clients held two different banking products (mostly a transactional account and an electronic banking), and the number was dropping each time additional product was added.

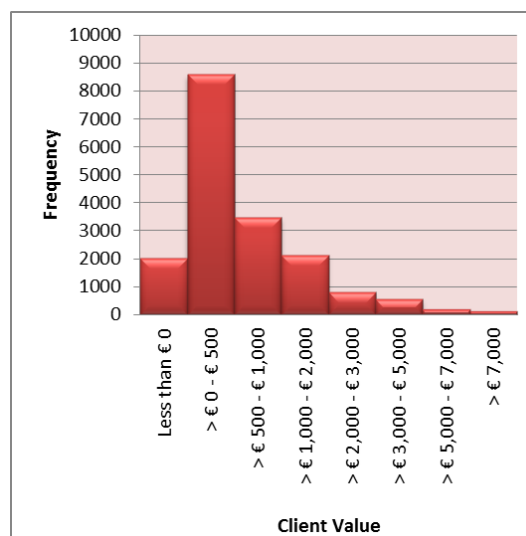
Figure 9. Product Coverage histogram



As many other method, *k*-means and SOMs also experience difficulties in digesting missing and empty values, so the next thing I had to do was to check the existence of such values. After doing some data filtering, I quickly noticed that – with the exception of the accounts – there was plenty of data missing for all products. Although this was not surprising, since not every customer had every product, I decided to remove all those variables from the analysis in order to ensure credibility of the results.

However, I needed to evaluate the variables where only a few records contained missing values. The first attribute to evaluate was the one on Client Value. By checking it, I noticed that for 1,672 customers, there was no information on value available. After some additional data drilling, I found out that all of those customers were the ones who started the relationship in the year of data exportation – they were bank customers for less than a year. Following the consultation with the Marketing Department, I realized that for those customers, Client Value calculation has not been carried out yet. Since I found this variable very interesting, I wanted to keep it for cluster analysis. Therefore, I replaced all missing values with -27.73; according to the Marketing Department, that was an average amount of the new customer acquisition cost. The histogram for Client Value attribute, shown in Figure 10, indicates the normal distribution.

Figure 10. Client Value histogram



Variables on accounts were the next to be check out. After some data filtering, I noticed that for three attributes the most frequently repeated value was 0. With 86 percent of customers having an empty value for Blocked Amount, and 71 percent of them for Not Allowed Balance and Not Allowed Balance Days, I chose to eliminate these variables from the analysis. Moreover, I found 1,383 out of 17,855 records containing missing values. Those were the customers whose Account Product Indicator did not equal to “A”. In other words, they have not been using their accounts for more than three months. Therefore, I have decided to exclude those customers from the analysis, and suggested the Bank S to

treat them separately, as they do with their individual “sleeping” customers. By doing so, the remaining file contained data on 16,472 customers.

However, there were still 2,984 records that were missing all balance sheet data. With the help of the Business Directory of the Republic of Slovenia, a tool for searching various types of data on all Slovenian companies and other business subjects, I soon discovered the reason for this data absence. 412 of those records were companies that shut down their businesses in the year of data exportation. However, the other 2,572 of those records were companies of specific legal forms: non-profit organizations (institutions, political parties, and trade unions), and associations. In Slovenia, those entities are not obliged to present their annual reports publicly. Thus, I have removed these 2,984 records, and my final dataset contains data on 13,488 customers.

The file was almost ready for the analysis. The only thing left to do was to analyse the relationships between the variables in order to exclude the excessive ones, and that way, contribute to minimization of the effect of the so-called curse of dimensionality. Namely, having a great number of attributes can have a significant impact on the performance of clustering algorithms and the quality of the results (Henriques, 2011, p. 114). This is why I have analysed the correlation between variables using Microsoft Excel Analysis ToolPak. Table 6 demonstrates the values of Pearson Correlation Coefficients.

Table 6. Correlation between variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1,000																					
2	0,198	1,000																				
3	0,065	0,071	1,000																			
4	0,184	0,144	0,090	1,000																		
5	0,070	0,079	0,989	0,094	1,000																	
6	0,305	0,503	0,155	0,364	0,171	1,000																
7	0,065	0,071	0,999	0,090	0,989	0,155	1,000															
8	0,184	0,144	0,090	0,999	0,094	0,364	0,090	1,000														
9	0,070	0,079	0,989	0,094	0,999	0,171	0,989	0,094	1,000													
10	0,300	0,492	0,155	0,365	0,171	0,996	0,155	0,365	0,170	1,000												
11	0,105	0,005	0,826	0,109	0,814	0,079	0,826	0,109	0,814	0,080	1,000											
12	0,217	0,016	0,109	0,249	0,106	0,156	0,109	0,248	0,106	0,157	0,481	1,000										
13	0,067	0,085	0,043	0,065	0,045	0,112	0,043	0,065	0,045	0,113	0,042	0,085	1,000									
14	0,045	0,048	0,024	0,037	0,024	0,064	0,024	0,037	0,024	0,062	0,014	0,013	0,762	1,000								
15	0,021	0,014	0,004	0,002	0,004	0,016	0,004	0,002	0,004	0,017	0,000	0,002	0,258	0,247	1,000							
16	0,068	0,070	0,035	0,035	0,035	0,091	0,035	0,034	0,035	0,092	0,053	0,119	0,886	0,705	0,249	1,000						
17	0,035	0,061	0,027	0,067	0,027	0,075	0,027	0,067	0,027	0,076	0,013	0,015	0,652	0,392	0,164	0,391	1,000					
18	0,057	0,092	0,050	0,084	0,056	0,130	0,050	0,084	0,056	0,130	0,025	0,043	0,763	0,517	0,152	0,492	0,371	1,000				
19	0,175	0,242	0,091	0,120	0,096	0,314	0,091	0,119	0,096	0,313	0,058	0,129	0,724	0,716	0,227	0,649	0,396	0,551	1,000			
20	0,072	0,193	0,058	0,071	0,061	0,204	0,058	0,071	0,061	0,205	0,034	0,066	0,746	0,591	0,237	0,676	0,547	0,474	0,727	1,000		
21	0,139	0,121	0,068	0,064	0,073	0,167	0,068	0,063	0,072	0,168	0,143	0,327	0,404	0,331	0,118	0,386	0,239	0,283	0,490	0,393	1,000	
22	0,077	0,073	0,032	0,024	0,035	0,079	0,032	0,023	0,035	0,079	0,076	0,172	0,202	0,211	0,090	0,313	-0,038	0,107	0,278	0,203	0,532	1,000

1 - Client value

2 - Product coverage

3 - Amount of inflows

4 - Number of inflows

5 - Amount of outflows

6 - Number of outflows

7 - Amount of deposits

8 - Number of deposits

9 - Amount of withdrawals

10 - Number of withdrawals

11 - Average monthly balance

12 - End of month balance

13 - Assets

14 - Inventories

15 - Short-term loans

16 - Equity

17 - Non-current liabilities

18 - Current liabilities

19 - Net income from sales

20 - Depreciation

21 - Operating profit

22 - Net profit/loss for the year

With the Pearson Correlation Coefficient of 0.90 or higher, we have five variables highly correlated to at least one other variable. Thus, I have decided to exclude the following



variables: Amount of Outflows, Amount of Deposits, Number of Deposits, Amount of Withdrawals, and Number of Withdrawals. Having done so, my final file contains 13,488 customers with 17 attributes characterizing them.

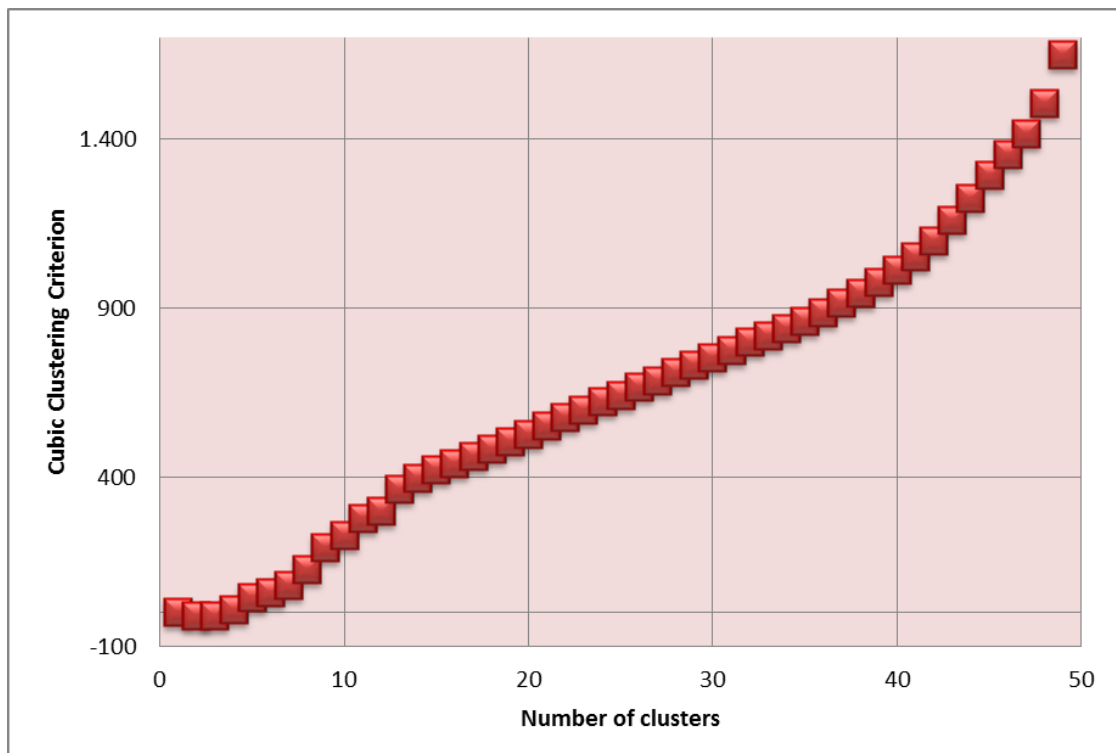
## 2.3 Analysis

### 2.3.1 Determining the number of clusters

In order to get the idea on the possible number of clusters, I have first trained  $k$ -means in SAS Enterprise Miner. Setting the specification method to automatic, and considering the range as the internal standardization criterion, and the Ward's minimum variance method as the method for clustering, I was able to compute the Cubic Clustering Criterion (hereinafter: CCC) values for a different number of clusters. Namely, the CCC is the test statistic, provided by the SAS, used to determine the optimal number of clusters; by plotting its value against the number of clusters, ranging from one cluster up to about one-tenth the number of observations, peaks on the plot can be searched for. Any peak with the CCC higher than 2 or 3 indicates good clustering.

In Figure 11, the CCC as the function of number of clusters is demonstrated. As we can see, all values – except the ones for  $k=1$ ,  $k=2$ , and  $k=3$  – are positive and higher than 3. However, we can see no peaks in the plot, and therefore, we cannot use the CCC to decide on the right number of clusters.

Figure 11. Cubic Clustering Criterion



However, there are some other statistics that may come useful in determining the optimal number of clusters. One of them is Pseudo F, the statistic that is intended to capture the tightness of clusters and is a ratio of between-cluster variance to within-cluster variance. Large values of Pseudo F usually indicate a better clustering solution (Lim, Acito, & Rusetski, 2006).

Another statistic that can be used, and which indicates the extent to which clusters differ from each other, is the  $R^2$  or R-squared. Its values lie between 0 and 1 with values close to 1 indicating high difference between clusters. The values themselves do not tell us much about the right number of clusters, however, the relative changes in the values as the number of clusters increase, can be helpful in determining the number of clusters. Namely, a substantial decrease or increase may indicate that a satisfactory number of clusters have been reached (Ritz & Skovgaard, 2005, p. 13).

Moreover, we can plot the sums of the Euclidean distances between each data point and clusters' seeds against the number of clusters and investigate the relative changes in the values as the number of clusters increase. Similar as with  $R^2$ , the considerable change indicates the stopping point.

Running  $k$ -means several times for different values of  $k$ , while keeping the Ward's method and range internal standardization, I have calculated the values of the above-mentioned statistics. Figure 12, 13, and 14 represent each of the statistics graphically. With the value of Pseudo F highest at  $k=3$ , the increase of  $R^2$  value sharpest from  $k=2$  to  $k=3$ , and the decrease of the sum of the distances between data points and cluster centroids most notable from  $k=2$  to  $k=3$ , we can conclude that  $k=3$  is the optimal solution.

Figure 12. Pseudo F

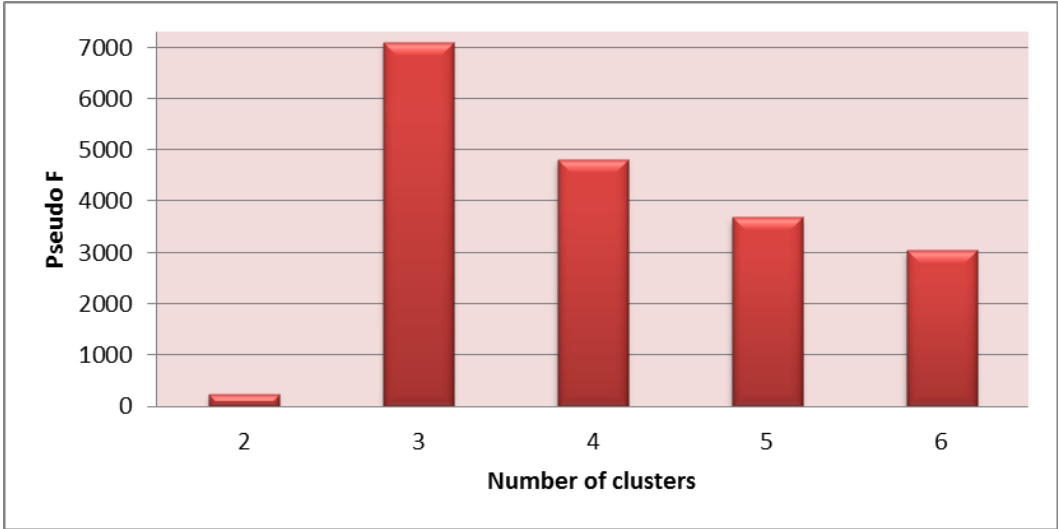


Figure 13. R-squared

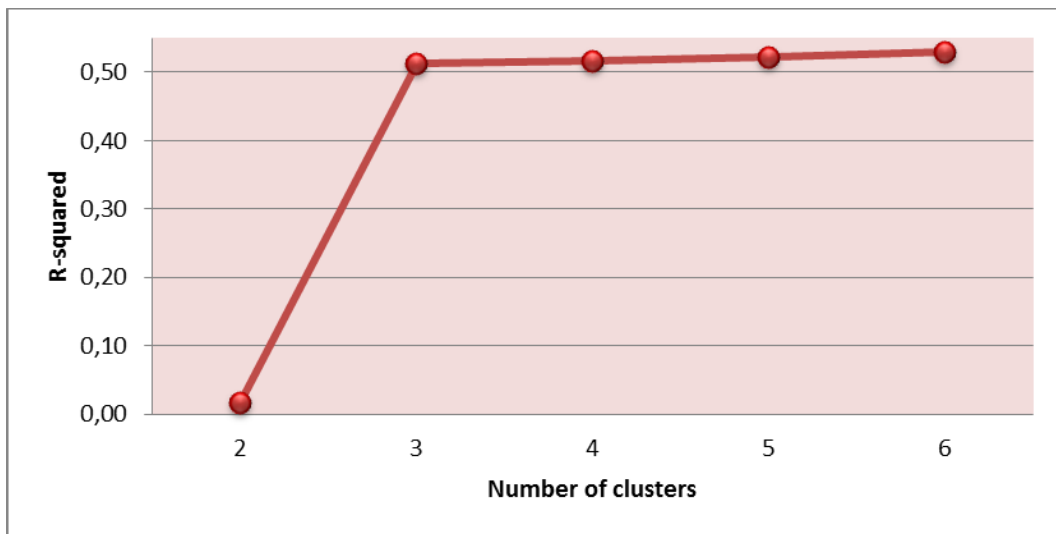
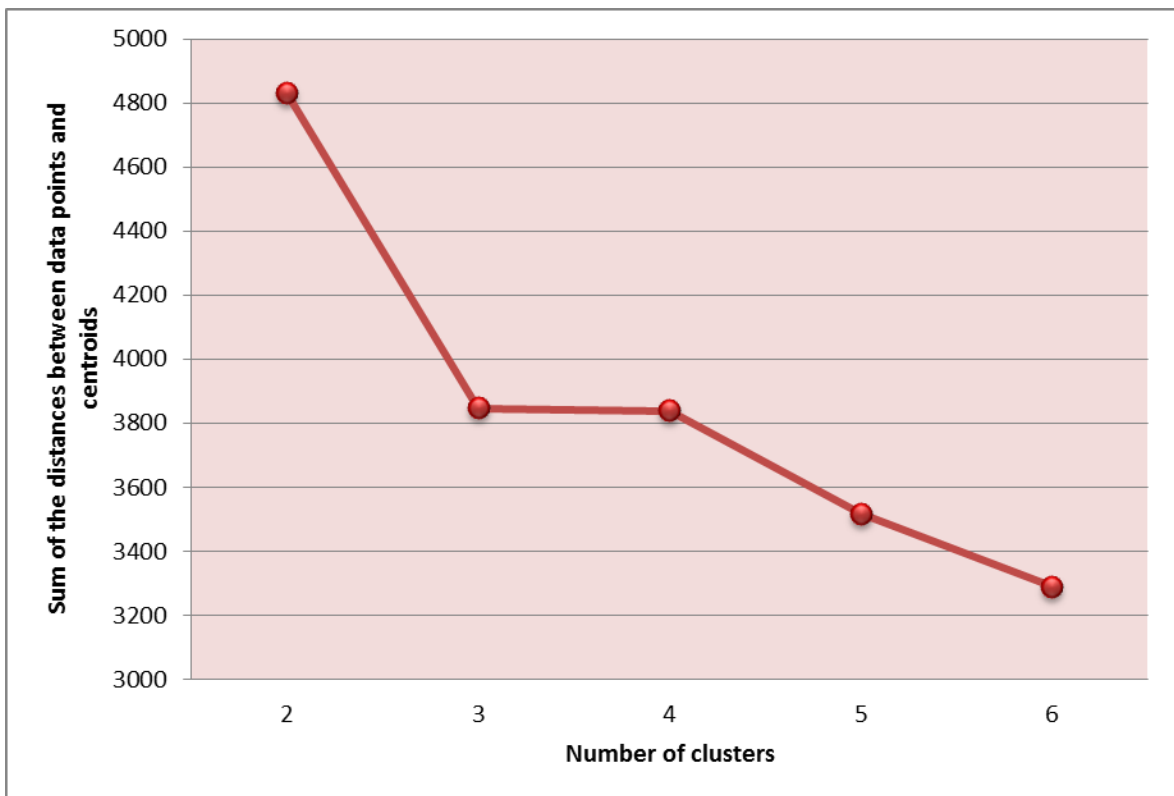


Figure 14. Sum of the distances

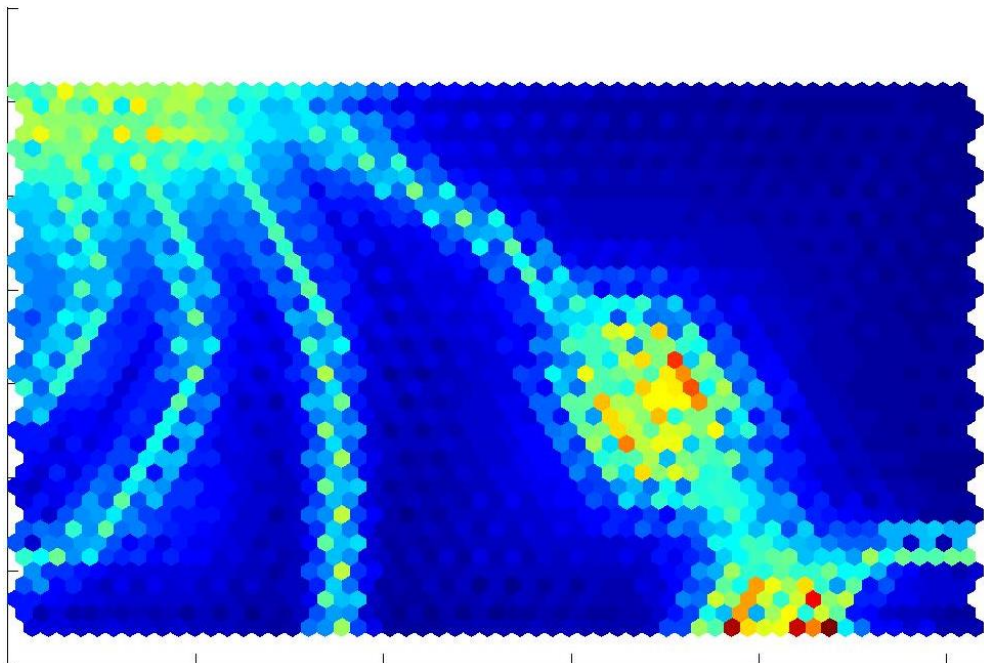


Now that the optimal number of clusters for  $k$ -means was known, I wanted to find out if the results, suggested by SOMs, would be any different. For this reason, I have trained the basic SOM Kohonen algorithm in GeoSOM Suite several times with different values of certain parameters. However, since the results were similar, I have decided to consider the following ones: I used a 30 by 20 regular SOM lattice, a random initialization of neurons,

and a range as a data normalization type. Map train has been set to sequential, consisting of 250 epochs in the rough phase with the neighbourhood radius of 10 neurons and the learning rate of 0.5, and 500 epochs in the finetune phase with the learning rate of 0.1 and the neighbourhood radius of 5 neurons. As a result, I was able to get the map, as presented in Figure 15. What we see is an example of the so-called U-matrix; according to Henriques (2011, p. 30), U-matrices are one of the most popular methods to visualize SOMs, and are used to present input space's distances in the output space.

Looking at the U-matrix retrieved, we can notice several differently coloured areas; dark blue areas can be thought as clusters, while light blue to green and yellow coloured parts may be regarded as cluster separators. Orange and red coloured nodes, however, represent the outliers who need to be excluded. With all this in mind, we can therefore identify six different clusters (which are more evident, if we squint a bit).

*Figure 15. 30x20 SOM U-matrix*



As we can see, the optimal number of clusters, proposed by each method, is different, with  $k$ -means suggesting three clusters as the right choice, and SOMs recommending customers to be divided into six groups. Therefore, in order to decide which solution to use, we have to compare both of them. The method selection step is explained in the following chapter.

### **2.3.2 Selecting the method on which to cluster**

Here, we will compare both of the suggested solutions through the same three statistics we have used for determining the optimal number of clusters in the case of the  $k$ -means algorithm. Having done so, we will end up with having a final decision about which

method to use as the basis for customer segmentation. In Table 7 below, the values of the sum of the distances, the  $R^2$ , and the Pseudo-F for results suggested by both methods, are presented.

Table 7. *k*-means and SOMs comparison based on three clustering statistics

Statistic	<i>k</i> -means (k=3)	SOMs (6 clusters)
Sum of the distances between data points and centroids	3,847.42	3,387.97
$R^2$	0.512879	0.568181
Pseudo-F	7,099.03	3,731.30

The solution suggested by the SOM algorithm outperforms the one proposed by *k*-means in terms of sum of the distances between data points and their centroids; as we can see, the sum happens to be lower with SOMs than with *k*-means, meaning that members of each cluster are closer to their centroid. In other words, clusters are more homogeneous within.

If we look at the values of the  $R^2$ , we can see that the number is higher with SOMs than it is the case of *k*-means. Since we know that  $R^2$  measures the difference between clusters, and the closer the value is to 1, the more different the clusters are to one another, we can conclude that the result, suggested by SOMs, presents better solution, which means clusters are more heterogeneous between each other.

With the value of the Pseudo-F higher for *k*-means, we could say *k*-means to be better choice than SOMs. However, it may not be that straightforward. Since we know Pseudo-F values dropping when increasing the number of clusters, it is not surprising that the value of Pseudo-F will be lower for SOMs, since they suggest separating the customers into three groups more. Therefore, if we check on how the results would be if we used the same number of clusters for both methods, maybe we could make an easier decision. Thus, I have compared the values of Pseudo-F in the case of six-group clustering; with the Pseudo-F value of 3,045.36 for *k*-means, and 3,731.30 for SOMs, we can conclude SOMs to present a better solution in terms of Pseudo-F as well.

Therefore, since all analysed statistics indicate SOMs to be a better solution, we will use it as a basis for our customer segmentation. In the next chapter, the clusters profiling and interpretation is presented.

### 2.3.3 Clusters profiling and interpretation

In order to describe the obtained clusters, I have examined each of their representatives – the cluster centroids. This included the analysis of the clusters’ sizes, as well as the values of each of the variables. Not only I went through the seventeen attributes I have used for clustering, I have also examined the variables that had been excluded from the clustering task in order to identify any additional characteristics of the clusters.

Table 8 contains the sizes and the mean values of the six clusters, suggested by SOM algorithm. As we can see, the clusters are of different sizes, but if we take a closer look at the total number of the customers, we can notice that it is lower than it was before the analysis; this is due to removal of the 137 outliers, earlier identified by the SOM algorithm. We can notice that the values of the variables increase by each cluster, and the more products customers have, the greater is their value. In general, we can distinguish between customers that only need some daily banking products, and customers that need advanced financing products for their growth. Anyway, the description of the clusters is given below, cluster by cluster.

*Table 8. Clusters’ sizes and mean values*

<b>Cluster number</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
Frequency	636.00	6,598.00	3,289.00	1,650.00	795.00	383.00
Client value	292.91	497.67	784.86	1,431.78	1,813.15	2,912.66
Product coverage	1.00	2.00	2.99	4.00	5.00	6.28
Amount of inflows	1,667.26	8,066.67	15,098.56	24,472.71	33,366.93	51,610.73
Number of inflows	2.50	6.92	17.84	31.59	31.23	60.29
Number of outflows	7.91	16.70	28.79	40.10	50.31	74.73
Average month balance	2,699.51	6,830.59	8,125.36	8,106.01	5,511.16	7,296.30

Table continues

Continued

<b>Cluster number</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
End of month balance	2,631.15	5,494.44	7,024.61	7,372.38	5,126.31	6,588.80
Assets	40,932.74	92,039.87	184,151.58	236,857.78	292,670.96	361,254.32
Inventories	2,823.01	9,653.36	19,715.58	28,571.47	31,822.71	48,134.13
Short-term loans	61.49	390.15	294.87	512.81	419.05	534.36
Equity	18,691.26	37,732.79	77,384.90	102,449.93	112,759.94	141,321.97
Non-current liabilities	4,732.41	14,818.47	34,239.90	42,126.05	64,120.19	70,435.91
Current liabilities	16,135.67	37,717.29	68,496.73	88,589.86	110,754.75	142,519.86
Net income from sales	34,246.30	86,631.88	200,277.33	299,288.26	396,619.93	591,224.99
Depreciation	1,873.92	3,949.92	9,696.91	13,335.24	19,252.31	25,655.74
Operating profit	529.14	3,142.72	7,057.10	8,965.41	11,819.85	15,115.06
Net profit/loss for the year	672.47	3,005.16	5,768.28	8,396.48	12,107.95	14,678.96

### 2.3.3.1 Cluster 1

The first cluster comprises 636 clients, which makes it our second smallest cluster. It represents the least valuable customers, not just in terms of Client value variable, but also in general. These clients hold only a transactional account, and have very small number of transactions, both influx and efflux. The amount of their monthly transactions is very low, and so are the values of all of their balance sheet items. Their net income from sales is the lowest of all clusters, and they make the least profit out of it (less than 2 percent). As individuals, they have no relationship with the bank. They are doing their businesses in industries such as construction, buildings cleaning, motor vehicles maintenance and repair, hairdressing, and similar. Thus, we can assume we are dealing with technically and vocationally educated and trained individuals, who started their own businesses in the

fields where they have a wealth of expertise; however, they do not have much of business knowledge.

#### 2.3.3.2 Cluster 2

With 6,598 customers – which is almost a half of all clients – Cluster 2 represents our biggest cluster. It is the second least valuable cluster whose members are worth a little less than €500 in average. Besides the transactional account, these customers use electronic banking services as well. The number of their transactions is relatively small, but the value of each incoming transaction is the highest of all clusters. The amounts of their balance sheet items are greater than the ones of Cluster 1, however, they are still lower compared to other four clusters. The share of their profits from their net sales is the highest of all clusters. Nearly all clients have a relationship with the bank as individuals as well, and their average value is €8.11. Considering the industries they work in, we can see they are engaged in various range of activities. Still, about one third of them do their business in professional industries such as accounting, computer programming, technical consultancy, architectural planning, and others.

#### 2.3.3.3 Cluster 3

Cluster 3 is the second biggest cluster consisting of 3,289 clients and with an average value of €784.86. It is the cluster of customers who use 3 banking products; all of them have a transactional account and an electronic banking, yet, the third product differs slightly. One third of the customers use credit cards, the other third have a loan (whether long-term or short-term), 400 customers utilize POS services, and the rest use different banking products as their third product. It is the cluster with the highest average monthly balance, which makes it the most suitable customers to be offered any savings products to. The values of their balance sheet items are satisfactory, but not too high. The majority of the clients have chosen Bank S as their bank for their private needs as well, and their average Client value as individuals is (with €15.18) almost twice as of the second cluster. The industries of their businesses are very similar to the ones of Cluster 2, however, the proportion of stores, bars, cafes, and other retail service providers is quite higher. With 12.19 percent is almost the same as the proportion of the customers that are using POS services (11.95 percent).

#### 2.3.3.4 Cluster 4

Cluster 4 is our third cluster in terms of size (1,650) and customer value (€1,431.78). Its members hold 4 products; besides the transactional account and e-banking, they also use credit cards (60 percent of the customers, where nearly all of them also use cards insurance), and loans (almost 50 percent of them). The rest are using other banking products. The number of their transactions is relatively high, but the value of each



transaction is the second lowest of all clusters. Right after Cluster 3, they have the greatest average balance per month. Their short-term loans are the second biggest of all clusters. All of the customers trust Bank S to be the right bank for their personal finances as well, and their lifetime value to the bank as individuals equals €22.54 in average. They carry out their businesses in diversified range of industries, but the most numerous for Cluster 4 are the cargo carriers, business consultants, technical consultants, and sales agents.

#### 2.3.3.5 Cluster 5

Our fifth cluster contains 795 clients, and it is the third smallest cluster. Its customers are the second most valuable clients whose Client value equals €1,813.15. In combination with transactional account and electronic banking, they mostly have loans, credit cards, and the insurance of the latter. They have the second highest amount of monthly inflows. Besides Cluster 1, they have the lowest average monthly balance. Their balance sheet items values are the second best. They have the second highest income from sales, and their share of profit from their net sales is the second biggest. As individuals, they are worth €29.08 in average. They are registered to do business in 202 different industries, and no industry stands out significantly.

#### 2.3.3.6 Cluster 6

The smallest, yet the most precious cluster of all is our last cluster. It contains only 383 clients, but they are the most valuable ones, in terms of Client value attribute (€2,912.66), and in general. With more than 6 products, they are the most equipped customers. Nearly all of them have a transactional account, an e-banking, a credit card, and a loan. The rest 2 products differ; most of the customers have cards insurance, some of them use an overdraft, a leasing, and POS services, about one fifth of them have a deposit. This is why it is not surprising their average monthly balance is not the highest. However, the amount of their transactions is the greatest of all clusters, and so is the number of transactions. Their balance sheet data values are the richest, and they make the highest net profit. All of them are bank's customers in the private life as well, and with €35.56, their value as individuals is the highest. They do their businesses in several industries, thus we can find many wholesalers, restaurants and inns owners, automobile repair shops owners, computer programmers, and others. However, the strongest industry of all in terms of number of clients who perform their business in is the cargo-carrying business (almost 13 percent of our most valuable customers).

## **3 DISCUSSION**

### **3.1 Practical implications**

Bank S is known as a modern and innovative bank that regularly drives the development of the Slovenian banking system. Its mission is to provide up-to-date technologies to its clients, empower the creation of the new business opportunities, and enable the accessibility of financing resources. It invests a great deal of effort in building solid relationships with its clients through a mutual trust and understanding. Its vision is to build a solid, safe, and healthy bank that would be visible as a bank with the highest standards. They are determined to become a leading financial group in Slovenia.

Despite a demanding economic situation and an intense competition within the Slovenian banking system, they strive to accomplish their strategic goals. Higher client satisfaction is certainly one of them. For this reason, they are constantly improving their products and services, as well as their working processes and decision-making procedures. Their goal is also their market share growth, both in retail banking and on the market of sole proprietors and SMEs. Since they promote creativity, they placed a special focus on the innovative and export-oriented SMEs, whose progress they monitor carefully.

Bank S follows the principles of sustainable development and social responsibility, thus it keeps developing quality offers, adjusted to their clients' needs, at the competitive prices. It gives a special attention to reducing negative influences on the environment, so it constantly invests into energy saving equipment. Due to the use of state-of-the-art communication technologies, they have been gradually decreasing the number of business trips as well.

Having all those strategic goals in mind, I tried to make several proposals that could eventually come useful to Bank S in developing their marketing and sales strategies. In the paragraphs that follow, the recommendations considering each of the clusters obtained are provided.

Ending a relationship with a customer often sounds ridiculous, and is for many companies an oxymoron. However, sometimes this is the best thing to do. As Rains (2010) claims, bad customers can cost a company more than they realize. Usually, the less profitable customers are the most difficult to serve; they are often very demanding and never satisfied, the service they get is never good enough and most likely, they find these services too expensive. Bank S could consider this as one of the options in dealing with its less valuable customers – members of Cluster 1.

There could be many reasons for these clients to be the least profitable; if they have no relationship with the bank as individuals, it does not mean they do not have a fruitful

relationship with some other bank, and they chose Bank S for their business needs only. Another reason could be they were good customers earlier, but they decided to switch the bank, and are now in the process of leaving Bank S. One of the possibilities is that these are newly established companies at the very beginning of their road but with a growth potential in the future. Maybe opening their own business is their only chance to ensure employment, or they are just simply bad entrepreneurs. Nevertheless, since no company's resources are boundless – and Bank S is not an exception – holding onto bad customers can prevent companies from serving the good ones.

However, since their market share growth is one of the strategies of Bank S, a better solution than getting rid of those customers would probably be to keep allowing them access its products, while using no effort in dealing with them, that is, redirecting these customers to remote channels, such as electronic or mobile banking. Eventually, in case their business would improve, it could always change its strategy towards them. In the meantime, Bank S could involve those customers in customer acquisition campaigns; since it is obvious Bank S is not their primary bank for their private needs, they could offer them special benefits in case of handing over their entire business (both private and professional) to Bank S.

Moreover, Bank S could sponsor special workshops for the new coming entrepreneurs. In Slovenia, there is a special portal named e-VEM that comprises all the information about how to start a business in Slovenia. Every once in a while, in cooperation with some specialized companies, it organizes free courses on how to do business (e.g., how to prepare a business plan, how to understand financial statements, etc.). To promote itself, Bank S could be a godparent of such initiatives, which could result in attracting many of the newly established companies to choose it as their primary bank. At least, this could contribute to strengthening their corporate social image.

Similar to Cluster 1, members of Cluster 2 use only daily banking products: transactional account and electronic banking. By looking at the amount of their inflows, we could say they are starting to do some serious business, however, they have no special need for financing products so far. Therefore, instead of focusing on these products, Bank S should adjust its business policy to these clients, and offer them lower prices of their daily banking products (e.g., free usage of an e-banking, lower account managing costs, etc.).

Even though a majority of them has a relationship with the bank also as individuals, their value as such is relatively low. Thus, Bank S should keep on remunerating clients that have double relationship in terms of special benefits, and this way, work on their loyalty increase, and the increase of their total value. Based on the simplicity of their business needs, it should address them through remote channels or within their outlets, while not offering them the specialized bank account administrators.

Since they have different needs that exceed just regular daily banking products, clients of both Cluster 3 and Cluster 4 are the potential customers to get more equipped. For this reason, Bank S should focus on treating them more actively. Certainly, the first thing they need to do in order to make them feel treated on a more personal level is to provide them specially assigned bank account supervisors. Next, they need to perform several campaigns in order to equip the existing customers. Therefore, they should develop special packages for those customers in terms of special offers when implementing new products, special offers for investments and assets financing, and so on.

The most interesting customers in terms of retaining are clients of Cluster 5 and Cluster 6, and Bank S can consider them as “the premiums”. By being almost fully equipped, it is obvious they are satisfied with the bank and its products, and have a nice picture about it. Therefore, Bank S should reward them for their loyalty by including them in special loyalty programs. Since only loyal and satisfied customers can see the real value of the bank, they could use them as the ambassadors of the bank for acquiring new customers, and also reward them for every new client they bring.

In order to cover as many needs as possible, Bank S could provide them some of the non-banking products as well, though in cooperation with some of its selected allies. For instance, if we look at the values of the assets and inventories of customers from Cluster 5 and Cluster 6, we can agree they are relatively high. Thus, there is a great possibility that these clients would want to insure their assets, and Bank S could play a major role in enabling them access to different insurance products at competitive prices.

Moreover, since large number of clients of Cluster 6 comes from cargo-carrying industry, Bank S could develop special loans and leasing offers for them with the purpose of funding the purchase of new, environmental-friendly vehicle fleet. This way, they would yet strengthen the image of their environmental consciousness, as well as their general corporate image.

### **3.2 Limitations of the master’s thesis**

Although it was carefully prepared and provides several practical implications, this master’s thesis has its limitations. First of all, it bases on a case of only one Slovenian bank and it might not represent the situation of the entire Slovenian banking system, let alone the banking system in general. Furthermore, the dataset comprises only sole proprietors, and small- and medium-sized enterprises; however, it would be better if it included the entire industrial customer portfolio of Bank S.

The product coverage of the analysed customers is quite low, which turned out to be quite a challenging issue already at the data pre-processing stage. Plenty of values were missing; therefore, a great deal of (important) variables had to be excluded. Moreover, some of the

variables (e.g., relationship duration) that could be very interesting to have within the analysis or at least at the results interpretation had to be removed due to the data migration error, which occurred back in 2004.

### **3.3 Future work**

Considering the above-mentioned limitations, I provide several examples and ideas for future research. First, I would suggest Bank S to perform the same analysis as the one I did, but include their entire business customer array. Second, since almost all of their industrial clients also have relationship with the bank as individuals, I would recommend them to combine their both individual and corporate customer datasets, and see whether the results would provide any more explicit insights.

Furthermore, I suggest the analysis carried out in this master's thesis to be extended across the entire Slovenian banking industry, including more or even all Slovenian banks, when possible. Moreover, it might be interesting to compare SOMs clustering results between industrial clients of banks from the developed countries, and corporate customers of the third world's banks.

## **CONCLUSION**

Even though customer segmentation has been present for many years, it is still far from being an easy task for the companies. If the segmentation of the individuals has been a complex job, the segmentation of the industrial clients has been an even more challenging project; as Hague and Harrison from B2B International (2013) argue, companies behave much more rational than individuals since they buy what they need, instead of buying what they want. Moreover, the decision-making process in organizations is much more sophisticated, and involves far more people and even departments.

Customer segmentation in the banking industry is yet more important since it can contribute to a faster development of the bank. However, the development of the banking system is of a paramount importance for growth and stability of the entire economy (Barth, Caprio, & Nolle, 2004, p. 2). Bank S, whose case is the basis for this master's thesis, is also aware of the significance of customer segmentation. For this reason, they have already implemented segmentation in their individual-customers array. Still, they would like to do it with all of their industrial clients as well, starting with small- and medium-sized companies. Namely, an entire corporate customer portfolio contributes to one half of their net banking income.

In order to help them, I decided to try to segment their entire SMEs customer portfolio by using two clustering methods. The first one was the  $k$ -means algorithm, which is one of the simplest and most popular methods for data clustering. The second, more contemporary

one was the Self-Organizing Maps method, which requires no prior knowledge about the nature of the data to cluster, and finds patterns and relations between the data that are not apparent at first sight. After running both of the methods – one in SAS Enterprise Miner, and the other in GeoSOM Suite – I got different results; *k*-means suggested three clusters as the right choice, while SOMs proposed customers to be divided into six groups. For this reason, I compared both of the results with the help of the three clustering statistics, which in the end indicated SOMs as the better solution. Based on the interpretation and profiling of the obtained six clusters, I made several suggestions that can help Bank S develop and determine their sales and marketing strategies and approaches.

## REFERENCE LIST

1. Alizon, F., Shooter, S. B., & Simpson, T. W. (2009). Henry Ford and the Model T: lessons for product platforming and mass customization. *Design studies*, 30(5), 588–605.
2. Almotairi, M. (2009). A framework for successful CRM implementation. *6<sup>th</sup> European and Mediterranean Conference on Information Systems 2009* (pp. 1–14). Izmir: Brunel University and Dokuz Eylul University.
3. Anderson, J. C., & Narus, J. A. (2004). *Business Market Management: Understanding, Creating, and Delivering Value* (2<sup>nd</sup> ed.). New Jersey: Pearson.
4. Andritsos, P. (2002). *Data clustering techniques*. Toronto: University of Toronto, Department of Computer Science.
5. Araujo, L., & Spring, M. (2006). Services, products, and the institutional structure of production. *Industrial Marketing Management*, 35(7), 797–805.
6. B2B International. (n.d.). Market Segmentation in B2B Markets. Retrieved November 5, 2013, from <http://www.b2binternational.com/publications/b2b-segmentation-research/>
7. Bação, F., Lobo, V., & Painho, M. (2005). Self-organizing maps as substitutes for *k*-means clustering. In *Computational Science–ICCS 2005* (pp. 476–483). Berlin: Springer.
8. Bailey, C., Baines, P. R., Wilson, H., & Clark, M. (2009). Segmentation and customer insight in contemporary services marketing practice: why grouping customers is no longer enough. *Journal of Marketing Management*, 25(3/4), 227–252.
9. Barry, J., & Weinstein, A. (2009). Business psychographics revisited: from segmentation theory to successful marketing practice. *Journal of marketing management*, 25(3/4), 315–340.
10. Barth, J. R., Caprio, G., & Levine, R. (2003). Bank regulation and supervision: Lessons from a new database. *Macroeconomic stability, financial markets, and economic development*. Mexico City: Banco de Mexico.
11. Barth, J. R., Caprio, G., & Nolle, D. E. (2004). *Comparative international characteristics of banking*. Washington: Comptroller of the Currency.
12. Beck, T., Demirgüç-Kunt, A., & Levine, R. (2006). Bank concentration, competition, and crises: First results. *Journal of Banking & Finance*, 30(5), 1581–1603.
13. Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. Indianapolis: John Wiley & Sons.
14. Bose, R. (2002). Customer relationship management: key components for IT success. *Industrial Management & Data Systems*, 102(2), 89–97.
15. Bull, C. (2003). Strategic issues in customer relationship management (CRM) implementation. *Business Process Management Journal*, 9(5), 592–602.
16. Caprio, G., & Levine, R. (2002). Corporate governance in finance: Concepts and international observations. In R. E. Litan, M. Pomerleano & V. Sundararajan (Eds.),

- Financial Sector Governance: The Roles of the Public and Private Sectors*, (pp. 17–50). Washington: Brookings Institution Press.
17. Chalmers, R. (2006). Methodology for customer relationship management. *The Journal of Systems and Software*, 79(7), 1015–1024.
  18. Chen, I. J., & Popovich, K. (2003). Understanding customer relationship management (CRM): People, process and technology. *Business Process Management Journal*, 9(5), 672–688.
  19. Chen, Y., Zhang, G., Hu, D., & Wang, S. (2006). Customer segmentation in customer relationship management based on data mining. In *Knowledge Enterprise: Intelligent Strategies in Product Design, Manufacturing, and Management* (pp. 288–293). New York: Springer.
  20. Claessens, S., Demirgüç-Kunt, A., & Huizinga, H. (2001). How does foreign entry affect domestic banking markets?. *Journal of Banking & Finance*, 25(5), 891–911.
  21. Customer segmentation. (n.d.). Rosella, Predictive Knowledge & Data Mining. Retrieved November 1, 2013, from <http://www.roselladb.com/customer-segmentation.htm>
  22. Dhillon, I. S., & Modha, D. S. (2001). Concept decomposition for large sparse text data using clustering. *Machine learning*, 42(1/2), 143–175.
  23. Dibb, S., & Simkin, L. (2009). Bridging the segmentation theory/practice divide. *Journal of Marketing Management*, 25(3/4), 219–225.
  24. Dimitriadis, S., & Stevens, E. (2008). Integrated customer relationship management for service activities. *Managing Service Quality*, 18(5), 496–511.
  25. European Commission. (2005). *The new SME definition. User guide and model declaration*. Brussels: European Commission.
  26. European Commission. (2014). Goods and services. Retrieved August 21, 2014, from <http://ec.europa.eu/trade/policy/accessing-markets/goods-and-services/>
  27. Eurostat. (2008). *NACE Rev. 2*. Luxembourg: Office for Official Publications of the European Communities.
  28. Foss, B., Stone, M., & Ekinici, Y. (2008). What makes for CRM system success – or failure? *Database marketing & Customer Strategy Management*, 15(2), 68–78.
  29. Goller, S., Hogg, A., & Kalafatis, S. P. (2002). A new research agenda for business segmentation. *European Journal of Marketing*, 36(1/2), 252–271.
  30. Goss, J. L. (n.d.). Henry Ford and the Assembly Line. *About.com*. Retrieved October 4, 2013, from <http://history1900s.about.com/od/1910s/a/Ford--Assembly-Line.htm>
  31. Hague, P., & Harrison, M. (n.d.). Market Segmentation in B2B Markets. B2BInternational. Retrieved September 21, 2014, from <https://www.b2binternational.com/publications/b2b-segmentation-research/>
  32. Harding, D., Cheifetz, D., DeAngelo S., & Zielger, E. (2004). CRM's Silver Lining. *Marketing Management*, 13(2), 27–32.
  33. Henriques, R. A. P. (2011). *Artificial Intelligence in geospatial analysis: applications of self-organizing maps in the context of geographic information science*. Lisbon: ISEGI.



34. Horne, S. (2003). Needed A Cultural Change. *Target Marketing*, 26(8), 53–55.
35. Hutt, M. D., & Speh, T. W. (2013). *Business marketing management: B2B* (11<sup>th</sup> ed.). Canada: Cengage Learning.
36. Jayachandran, S., Sharma, S., Kaufman, P., & Raman, P. (2005). The Role of Relational Information Processes and Technology Use in Customer Relationship Management. *Journal of Marketing*, 69(4), 177–192.
37. Jiang, F., Berry, H., & Schoenauer, M. (2009, June). The impact of network topology on self-organizing maps. In *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation* (pp. 247–254). Nex York: ACM.
38. Joglekar, S. (2015). Self-Organizing Maps with Google’s Tensorflows. Retrieved November 30, 2015, from <https://codesachin.wordpress.com/2015/11/28/self-organizing-maps-with-googles-tensorflow/>
39. Johnson, W. C., & Weinstein, A. (2004). *Superior customer value in the new economy: concepts and cases*. Florida: CRC.
40. Kale, S. H. (2004). CRM Failure and the Seven Deadly Sins. *Marketing Management*, 13(5), 42–46.
41. Karakostas, B., Kardaras, D., & Papathanassiou, E. (2005). The state of CRM adoption by the financial services in the UK: an empirical investigation. *Information & Management*, 42(6), 853–863.
42. Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: an introduction to cluster analysis*. Hoboken: John Wiley & Sons.
43. Kim, H. W. (2004). A Process Model for Successful CRM System Development. *Software, IEEE*, 21(4), 22–28.
44. Kirkby, J. (2001). *Creating a CRM vision*. Stamford: Gartner, Inc.
45. Kirkby, J., Thompson, E., & Buytendijk, F. (2001). *Getting the Best Out of CRM Performance Metrics*. Stamford: Gartner, Inc.
46. Ko, E., Kim, S. H., Kim, M., & Woo, J. Y. (2008). Organizational characteristics and the CRM adoption process. *Journal of Business Research*, 61(1), 65–74.
47. Kohonen, T. (2001). *Self-organizing maps*. Berlin: Springer.
48. Kotler, P., & Keller, K. L. (2006). *Marketing Management* (12<sup>th</sup> ed.). Harlow: Pearson.
49. Kotler, P., & Armstrong, G. (2012). *Principles of Marketing* (14<sup>th</sup> ed.). Harlow: Pearson.
50. Kotter, J. P. (1995). Leading Change: Why Transformation Efforts Fail. *Harvard Business Review*, 73(2), 59–67.
51. Kukulas, N. (2012). *Modern trends in B2B marketing and advertising in Finland* (bachelor’s thesis). Finland, Lahti: Lahti University of Applied Sciences, Faculty of Business Studies.
52. Laeven, L., & Claessens, S. (2003). What drives bank competition? Some international evidence. *Journal of Money, Credit and Banking*, 36(3), 563-583.
53. Langerak, F., & Verhoef, P. C. (2003). Strategically embedding CRM. *Business Strategy Review*, 14(4), 73–80.

54. Larose, D. T. (2005). *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons.
55. Lemon, K. N., Rust, R. T., & Zeithaml, V. A. (2001). What Drives Customer Equity. *Marketing Management*, 10(1), 20–25.
56. Lim, L. K., Acito, F., & Rusetski, A. (2006). Development of archetypes of international marketing strategy. *Journal of International business studies*, 37(4), 499–524.
57. Lovelock, C., & Gummesson, E. (2004). Whither services marketing? In search of a new paradigm and fresh perspectives. *Journal of service research*, 7(1), 20–41.
58. Machauer, A., & Morgner, S. (2001). Segmentation of bank customers by expected benefits and attitudes. *International Journal of Bank Marketing*, 19(1), 6–18.
59. Mauri, A. J., & Michaels, M. P. (1998). Firm and industry effects within strategic management: an empirical examination. *Strategic Management Journal*, 19(3), 211–219.
60. McKechnie, S. (2006). Integrating Intelligent Systems into Marketing to Support Market Segmentation Decisions. *Intelligent Systems in Accounting, Finance and Management*, 14(3), 117–127.
61. Mithas, S., Krishnan, M. S., & Fornell, C. (2005). Why Do Customer Relationship Management Applications Affect Customer Satisfaction? *Journal of Marketing*, 69(4), 201–209.
62. Neal, W. D. (2002). Shortcomings Plague the Industry. *Marketing News*, 36(19), 37.
63. Payne, A., & Frow, P. (2005). A Strategic Framework for Customer Relationship Management. *Journal of Marketing*, 69(4), 167–176.
64. Peelen, E., Beltman, R., Van Montfort, K., & Klerkx, A. (2006). *A Study into the Foundations of CRM Success*. Breukelen: Nyenrode Research Group.
65. Plakoyiannaki, E., & Tzokas, N. (2002). Customer relationship management: A capabilities portfolio perspective. *Journal of Database Marketing*, 9(3), 228–237.
66. Plakoyiannaki, E., Tzokas, N., Dimitratos, P., & Saren, M. (2008). How Critical is Employee Orientation for Customer Relationship Management? Insights from a Case Study. *Journal of Management Studies*, 45(2), 268–293.
67. Psychographics. (n.d.). In *Cambridge Dictionaries Online*. Retrieved November 23, 2013, from <http://dictionary.cambridge.org/dictionary/business-english/psychographics?q=psychographics>
68. Rababah, K., Mohd, H., & Ibrahim, H. (2011). A unified definition of CRM towards the successful adoption and implementation. *Academic Research International*, 1(1), 220–228.
69. Radcliffe, J., Kirkby, J., & Thompson, E. (2001). *The Eight Building Blocks of CRM*. Stamford: Gartner, Inc.
70. Rains, J. (2010, February 17). Get Rid of Bad Customers. *Americanexpress.com*. Retrieved November 4, 2015, from <https://www.americanexpress.com/us/small-business/openforum/articles/get-rid-of-bad-customers-1/>

71. Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). The Customer Relationship Management Process: Its Measurement and Impact on Performance. *Journal of Marketing Research*, 41(3), 293–305.
72. Richards, K. A., & Jones, E. (2008). Customer relationship management: Finding value drivers. *Industrial Marketing Management*, 37(2), 120–130.
73. Ritz, C., & Skovgaard, I. (2005). Module 2: Cluster Analysis. Master of Applied Statistics. Retrieved August 4, 2015, from <http://www.imm.dtu.dk/~perbb/MAS/ST116/module02/module.pdf>
74. Roberts, M. L., Liu, R. R., & Hazard, K. (2005). Strategy, technology and organizational alignment: Key components of CRM success. *Database Marketing & Customer Strategy Management*, 12(4), 315–326.
75. Rokach, L., & Maimon, O. (2005). Clustering methods. In *Data mining and knowledge discovery handbook* (pp. 321–352). Springer US.
76. Ryals, L., & Knox, S. (2001). Cross-functional issues in the implementation of relationship marketing through customer relationship management. *European Management Journal*, 19(5), 534–542.
77. Sear, E. A., Hartland, T. G., Abdel-Wahab, M. S., & Miller, C. G. (2007). *Customer relationship management implementation: a case study in the construction sector*. Pontypridd: University of Glamorgan.
78. Statistical Office of the Republic of Slovenia. (2014a). *Proizvodna struktura BDP po dejavnosti 2014*. [2014 GDP Production structure by activities]. Ljubljana: Statistical Office of the Republic of Slovenia.
79. Statistical Office of the Republic of Slovenia. (2014b). *Zaposlenost po dejavnosti 2014*. [2014 Employment by activities]. Ljubljana: Statistical Office of the Republic of Slovenia.
80. Statistical Office of the Republic of Slovenia. (2014c). *Podjetja po dejavnosti in velikosti 2014*. [2014 Enterprises by activities and size class]. Ljubljana: Statistical Office of the Republic of Slovenia.
81. About us. (n.d.). Thomson Reuters. Retrieved October 31, 2013, from <http://thomsonreuters.com/about-us/>
82. Web of Science. (n.d.). Thomson Reuters. Retrieved October 31, 2013, from <http://thomsonreuters.com/web-of-science/>
83. Tsai, J. (2008). The smallest slice. *Customer Relationship Management*, 12(2), 37–41.
84. Tsiptsis, K., & Chorianopoulos, A. (2009). *Data Mining Techniques in CRM: Inside Customer Segmentation*. Chichester: Wiley.
85. Valicon, d.o.o. (n.d.). Segmentacija: Kako vedeti, kateri pristopi so najbolj učinkoviti za katere ciljne skupine? [Segmentation: How to know which approaches are the most efficient for which target groups?]. Retrieved October 30, 2013, from <http://www.valicon.net/sl/valicon/resitve/segmentacija/>
86. Vargo, S. L., & Lusch, R. F. (2004). The four service marketing myths remnants of goods-based, manufacturing model. *Journal of service research*, 6(4), 324–335.

87. Weinstein, A. (2004). *Handbook of market segmentation: strategic targeting for business and technology firms* (3<sup>rd</sup> ed.). New York: Haworth Press.
88. Yankelovich, D., & Meer, D. (2006). Rediscovering Market Segmentation. *Harvard Business Review*, 84(2), 122–131.
89. Zablah, A. R., Bellenger, D. N., & Johnston, W. J. (2004). An evaluation of divergent perspectives on customer relationship management: Towards a common understanding of an emerging phenomenon. *Industrial Marketing Management*, 33(6), 475–489.
90. Zeithaml, V. A., Parasuraman, A., & Berry, L. L. (1985). Problems and strategies in services marketing. *The Journal of Marketing*, 49(2) 33–46.
91. Zimmerman, A., & Blythe, J. (2013). *Business to Business Marketing Management: A Global Perspective* (2<sup>nd</sup> ed.). Abingdon: Routledge.