UNIVERSITY OF LJUBLJANA

SCHOOL OF ECONOMICS AND BUSINESS

MASTER THESIS

OPPORTUNITIES OF DATA ANALYTICS USE FOR CUSTOMER RELATIONSHIP MANAGEMENT: A CASE OF A FINANCIAL INSTITUTION

Ljubljana, May 2023 MEDINA INES GONDŽO

AUTHORSHIP STATEMENT

The undersigned Medina Ines Gondžo, a student at the University of Ljubljana, School of Economics and Business, (hereafter: SEB LU), author of this written final work of studies with the title Opportunities of data analytics use for customer relationship management: a case of a financial institution, prepared under supervision of prof. dr. Jaklič Jurij.

DECLARE

1. this written final work of studies to be based on the results of my own research;
2. the printed form of this written final work of studies to be identical to its electronic form;
3. the text of this written final work of studies to be language-edited and technically in adherence with the SEB LU's Technical Guidelines for Written Works, which means that I cited and/or quoted works and opinions of other authors in this written final work of studies in accordance with the SEB LU's Technical Guidelines for Written Works;
4. 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 of the Republic of Slovenia;
5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the SEB LU in accordance with the relevant SEB LU Rules;
6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;
7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;
8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the SEB LU Study Information System;
9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana
10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, _________________ Author's signature: _____________

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

CRM – Customer Relationship Management

NBO – Next Best Offer

CRISP - DM – Cross-Industry Standard Process for Data Mining

IT - Information Technology

GDPR - General Data Protection Regulation

EU – European Union

DPO - Data Protection Officer

AI - Artificial Intelligence

ML - Machine Learning

DWH - Data Warehouse

INTRODUCTION

Due to the big data revolution, financial institutions nowadays are not just collecting data, as they were doing decades before, but also it raises opportunities to analyze collected data in different manners (Srivastava & Gopalkrishnan, 2015). Companies had some amount of data in their systems even before, but tools for data analytics were not so advanced. This improvement prompted companies to transform into data-driven companies. They recognized the potential of data analytics, and they are aiming to use the data that their companies are generating in the best possible way.

The age of digitalization brought a lot of benefits for firms but also the consequences that now customers have more information and options about where and how they will spend their money. As a logical sequence of such customer empowerment, companies are facing bigger competition and trying to attract and keep customers.

Experts from the finance industry exemplify big data as a mechanism that permits an organization to form, control and manage big data sets in each period (Srivastava & Gopalkrishnan, 2015). Since extensive data is stored on an everyday basis, there are a lot of opportunities for the exploration of data in financial institutions. Although companies today have the data that their companies are generating, it is challenging to find the best possible ways to make revenue out of it.

Although more and more opportunities for making data-driven businesses are emerging, companies still must handle data by respecting General Data Protection Regulation (GDPR) and other data protection regulations. Of course, data privacy must be handled, also.

There are differences in how CRM is perceived and how the company will form it inside, but scientists agree that the essential building block of customer relationship management is customer, meaning that every financial institution needs to understand what their customers want, how they behave, and whether there are differences among different groups and that common goal is to optimize customer experience (Coltman, 2007).

Since today's amount of data is not a problem, in terms of firms needing data about their customers, other problems are raised. Companies are in disposal of a great amount of data and a very important issue is how to use it most effectively in CRM processes as well as choosing the best fitting data analytics technique for the enhancement of the CRM processes (Gončarovs, 2017). Consequently, the purpose of the thesis is to address opportunities and barriers of data analytics for CRM in banks and offer suggestions to make it more effective.

A good example of a financial institution that developed an approach for customer relationship management (CRM) and set a goal to amplify customers' value by using its information is Royal Bank of Nova Scotia, which is the largest bank in Canada (Khirallah, 2001). Their goal was to develop a suitable solution in the scope of CRM, by collecting information that they had from their customers from all interactions. In their effort to develop a successful CRM they analyzed in detail their customer's transaction data from 3 months and data from 18 months stored in data storage to gain knowledge of their clients and consequently to create purposeful actions (Khirallah, 2001).

There were a lot of challenges and although they were using technology and advanced analytics heavily, a lot of effort was invested into the business, cultural issues, and people in the bank. They concluded that the best approach to creating strategies is to balance people, technology that powers advanced analytics, and CRM business processes.

The purpose of this thesis was to give guidelines for financial institutions that were considering implementing analytical Customer Relationship Management (CRM). Every financial institution must first understand what they can expect from analytical CRM, then whether to adopt it, understand its opportunities, what they will need to face along this journey, and what kind of critical factors. Financial institutions would gain knowledge regarding tackling the problem of implementing analytical CRM.

Regarding case analysis, the thesis provided a better understanding after describing in detail one real-life company. The thesis gave a clear picture from the conducted interviews with the employees, what company's strengths were, where they were struggling, and what their plan was.

My thesis had several goals. The first of them was to explore the literature and give a better perception of how data is used in the banking business for CRM.

The second was to identify, describe and provide a better understanding of CRM data-driven cases that are present in the banking field and how each of them can contribute to satisfying customers' needs.

The third goal was to analyze aspirations, challenges, and opportunities for CRM for the selected company as well as the process of implementing the Next Best Offer (NBO) model. Then, to conclude what the benefits and obstacles for banks when dealing with the customer's data were. The thesis tried to address the importance of data quality and draw the line with the selected bank and investigate how they used data analytics in their company.

Next, I analyzed data that I provided from selected banks on the selected case, NBO, and stated the findings and whether the bank had benefited from its implementation.

For the first part of my thesis, I used secondary data, and it contained theoretical background regarding data analytics, CRM, and its application in the financial industry. The literature review was divided into three chapters. I systematically reviewed the literature and explored what the current state of literature regarding this topic is, the level of large amount of data that is generated every day, and in what purposes. I stated some good examples of the financial institutions that recognized the potential of analyzing the data to reach the satisfaction of customers and used a strategic framework for defining CRM. Then, I conducted interviews with employees before and after data analysis to respond to my research question, which is following: what are the opportunities and barriers to using data analytics for CRM in the financial industry?

In chapter 4, I presented the methodology regarding interviews that were conducted with employees and explained the dataset that I received from the chosen bank.

In chapter 5, for the empirical part of the thesis, I first presented the selected bank, and conducted interviews with the employees from the bank that were involved in the data analytics, to be aware of the status of the selected bank. Interviews were organized so that I could get a better insight into the current state of the bank, if they were facing any obstacles and what was their plan in the scope of data analytics for CRM. Interviews were semi-structured, and I provided a list of questions for the employees in the appendices. The roles of each employee were presented and what their part in the firm regarding CRM and data analytics was. After data analysis, I talked with employees to evaluate the findings from data analysis so that I could understand additional barriers and some opportunities that financial institutions were facing when they were implementing analytical CRM.

Chapters 6 addressed data analysis, including all the prerequisites to the analysis of data. I presented the dataset that was used for the analysis, explained in detail the data preprocessing, and presented analysis and visualization, as well as the result of the analysis which is a group of clients with personalized offers for personal loan, which brings profit to the company. In the end a model was developed, and I addressed model metrics and validation.

In chapter 7 results from data interviews were presented. In chapter 8 I gave proposals on how to make CRM more effective. This answered my research question by first analyzing the existing literature on the subject and then analyzing one financial institution and the whole process of data analysis for CRM, including opportunities and barriers.

1 CONCEPT OF CUSTOMER RELATIONSHIP MANAGEMENT

1.1 CRM background

In the next chapter I stated a review of the literature regarding the concept of CRM in general, concept of CRM in the banking industry and data analytics for CRM. I started by finding the literature for concept of CRM, reviewing its background and different definitions of CRM that were explained from different backgrounds. A literature review's objective is to give a general view of the topic's current state of knowledge and to identify gaps in existing research. This can help to shape future research and the evolution of CRM practices in organizations.

Understanding clients is the key to successful customer relationship management. Companies need to be aware of what a customer’s needs, wants and activities are and not forget to keep in mind the possibility of differences among them. Nowadays, knowing your customer is not good to have the condition but must have.

The concept of CRM has its background in sales automation and call centers management since the middle of the 90s when it was believed that connecting customer data from the sales with call centers would lead to more knowledgeable communication with customers (Osarenkhoe & Bennani, 2007). It first emerged in developed economies, increasing particularly in businesses where keeping current customers is on the list of priorities, especially in a competitive business environment (Laketa, Sanader, Laketa & Misic, 2015).

For a long time, large-scale marketing and large retail were successful since consumers were pleased with systematic products and services (Chen & Popovich, 2003). The effectiveness of mass marketing strategies began to decline as more businesses entered the market, so customer relationship marketing techniques increased their focus on individual customers (Chen & Popovich, 2003).

Over time, marketing has developed as a way for companies to approach closer to their customers and thus attract them to become their customers and show that they can be trusted. Relationship marketing at that early stage tried to collect information from databases regarding customer preferences (Osarenkhoe & Bennani, 2007).

Consequently, branding was developed to assist customers in differentiating items to counteract this notion of being like all the competitors (Peppers & Rogers, 2004). Essentially marketing addressed the needs and wants of customer groups, because differentiating customers individually was exceedingly complex (Kumar & Reinartz, 2018). The first reason why companies, following market segmentation, categorized its customers into sub-groups and advertised their products through mass marketing, is that data for the individual customers was not assessable before and the second is that looking into every individual customer was costly (Kumar & Reinartz, 2018).

CRM was created to assure and maintain specific individual relationships, as well as to construct profitable and continuing relationships with customers (Osarenkhoe & Bennani, 2007).

Long-term communication or exchange between a consumer and a bank that finally results in improved profit margins are the cornerstones of one-to-one marketing (Alam, Al Karim & Habiba, 2021). In the meanwhile, Internet-based tools including e-commerce, internet marketing personalization, and self-help were developing concurrently (Osarenkhoe & Bennani, 2007).

As time went by, relationships between customers and companies evolved and changed to a great extent. Customers became more empowered and engaged in communication with companies.

In a typical CRM project, in addition to the conventional responsibilities of marketing and sales, additional functional departments that directly or indirectly interact with customers are usually included (Shum, Bove, & Auh, 2008). New protocols for information sharing, technology, and processes must therefore be put in place at the corporate level (Shum, Bove, & Auh, 2008). CRM, when used properly, gives businesses the ability to comprehend their customers' needs, serve them differently based on those needs, and keep their loyalty (Han-Yuh, 2007).

Today, many companies, including banks, know about the implication of customer relationship management and its ability to increase client lifetime value (Abbas & Hafeez, 2017). Technology advancements, particularly in data storage capacities, data warehousing software, and data mining approaches, have popularized CRM activities (Chye and Gerry, 2002).

1.2 Definition of CRM

As a new idea in marketing that is sometimes considered to have replaced data-driven marketing, CRM, is defined as a business approach that tries to recognize, foresee, and govern the requirements of an organization's existing and prospective clients (Ogbadu & Usman, 2012).

Through the literature, numerous definitions and applications of Customer Relationship Management have been enumerated and explained. Scientists from various industries look at the concept differently and how it affects the company and its employees. There was a lot of controversy among scientists about whether CRM is just an IT solution, or it is primarily a strategy.

CRM is stated to be the main corporate approach that connects inner processes and external structures to find and deliver value to target clients (Chuang & Hu, 2014). Also, CRM should be a "cross-functional method for maintaining communication with customers through all of the points of contact, treating the most valuable customers personally, to enhance customer retention and the success of marketing campaigns," according to its ideal definition (Gončarovs, 2017).

The reason for so many different opinions is because the company’s and customer's relationship has changed and improved over the years, as well as the emergence of the Internet and information systems.

As a concept, CRM stands for the synchronization of business objectives and plans through an examination of customer wants and requests, with customers serving as the focal point of all business actions and events (Stevanović & Gavrilović, 2018). It is a process that helps organizations build, maintain, and improve relationships with attractive customers, in such a way that satisfaction among customers and retention of those customers leads to profitability and better performance (Wang & Feng, 2012).

CRM is an approach that allows banks to assess client profiles, identify their needs, and implement different measures to achieve competitive advantage and profitability (Pokharel, 2011).

When discussing technology in the field of CRM, the software includes applications that support sales, marketing, and service; an application that integrates and analyses data about customers (Wang & Feng, 2012). CRM technology may advance an organization's capacity to maintain profitable customer relationships by analyzing information about cost-effective customers (Wang & Feng, 2012). Nonetheless, many companies have discovered that when technology is the only focus of an initiative, it seldom produces the desired results.

Companies frequently attempt to "rewrite success" by learning, for instance, what technology their competitors employ (Kim, Kim & Park, 2010). Will companies that purchased and enforced the same solution of CRM machinery and software for the sake of copying their competitor's company have the same result? Imitation like that will never be successful since although tangibles are copied, what a firm cannot copy is intangibles such as people, processes, organizations, and culture. Such combinations are unique to each company and thanks to them, some companies achieve more profit and success than others.

1.3 Types of CRM systems

There are two types of CRM systems, operational and analytical. Operational CRM is designed to be utilized in daily customer interactions, such as while conducting marketing campaigns (Tsiptsis & Chorianopoulos, 2011).

On the practical side, companies interact with their consumers through a variety of channels (online, phone, email, meetings, etc.), and they want to use knowledge about previous purchases daily. The following 3 procedures are often included in operational CRM: selling processes, when trying to bring in new customers, the marketing process is used, and customer support is a process that takes place after a customer has purchased anything from us.

Some typical functionalities of Operational CRM are automation of the sales force, marketing, and services. Consumer information and contact histories are tracked. Additionally, they make sure that at every customer "touch" point (interaction points), a consistent image of the customer's connection with the company is available.

Analytical CRM is the exploration of indicative and behavioral statistics, to complement the organization's objectives for customer management (Tsiptsis & Chorianopoulos, 2011). By analyzing consumer behavior using analytical CRM, companies can more readily market to them if they are aware of their buying habits.

Customer data serves as the foundation for analytical CRM since companies can gain some insights from analytical CRM, recognize customers as distinct, addressable persons, and differentiate them based on their worth and demands. The importance of analytical CRM is growing in businesses. In analytical CRM, it is all about data, including internal data on sales, transactions, and complaints as well as external data from sources like social media. If companies want to develop strong relationships with their customers, all that data must be connected and integrated.

That leads to the conclusion that Analytical CRM is concerned with data and input on one side and on the other side information understanding gleaned from data analysis that could help in the following processes of decision-making.

Since banks usually invest in analytical CRM, most of them have excellent analytical CRM proficiency but many banks fall short when it comes to operational and collaborative CRM (Pokharel, 2011). To overcome it, they could integrate channel management to reinforce collaborative CRM such as advising their clients to use more online channels which make consumer information pursuance simpler (Pokharel, 2011).

1.4 CRM development

When seen exclusively as technology, CRM deployment is extremely difficult, and even if it is done properly, the strategy's success is not guaranteed (Kim, Kim & Park, 2010). For example, Sweeden company enforced a software system for CRM and six components that made it successful are: an accent on quality, customer satisfaction, people contribution, communication with customers, measuring performance, and technology (Osarenkhoe & Bennani, 2007).

The first stage in implementing CRM in a business is to create a strategy, and when doing so, the primary objectives and constraints for the organization's CRM operations will be specified (Payne & Frow, 2006).

Many firms are still unaware of the effects of customer relationship management strategies across organizations and are incapable to recognize gaps for additional development of a client-oriented organization, mainly for this reason customer relationship management is constantly viewed as a technical rather than a strategic growth (Panda, 2001).

Secondly, starting with the selection of the most suitable channel options for market segments, including the integration between different channels (Payne & Frow, 2006). Some firms decide to begin with developing channel management where actions are documented as customer data from the front office system, and then business-wide database brings together data from front and back-office system (Chuang & Hu, 2014). Front office system implies programs that enable office workers to communicate with clients, sales departments, and marketing, while back office implies administration processes that are needed for daily business operations, human resources, accounting, information technology department.

Thirdly, the goal of the information management process is to collect and compile customer information from all sources of contact with customers to understand their requirements and improve the quality of their experience (Payne & Frow, 2006). As the result of efficient handling of client information, businesses can build customer loyalty and grow their business (Ranjan & Bhatnagar, 2011). Reduced expenses, higher customer satisfaction, more customers, revealing ineffective operational procedures, and long-term profitability and sustainability are among the advantages that customer relationship management may offer a company (Ogbadu & Usman, 2012).

Lastly, the performance evaluation process ensures that strategic CRM goals are being achieved (Payne & Frow, 2006).

More relationship-building is not necessarily better, according to prior empirical research; rather, the secret to performance improvement is relationship development of the correct kind (Coltman, 2007). Useful findings for managers are that their focus should be on finding the right combination of organizational structure, people capabilities and then IT that are difficult for competitors to imitate (Coltman, 2007).

2 DATA ANALYTICS FOR CRM

2.1 Use of data analytics

Analytical CRM has a goal to store, analyze and employ understanding regarding the approach to clients, and it frequently uses data mining (Bihari & Murdia, 1970). Success of a firm depends on examining client interactions from a lifetime perspective (Bihari & Murdia, 1970).

One of the first companies that started using data mining intensively through CRM were financial institutions, more precisely banks, due to the need to assess the profitability of their clients and to monitor the whether the marketing programs are effective (Voican, 2020). The information about customer needs and behavior will allow banks to identify their most valuable and profitable clients, cultivate relationships with potential clients, accurately calculate the revenue generated by each client, and predict possibilities for future business endeavors with those clients (Cvijović, Kostić-Stanković & Reljić, 2017). In banking day-to-day business there are numerous opportunities to exploit data, which can be then further investigated, analyzed, and used to make distinct models, depending on distinct customer needs, such as evaluating loan applications and in target marketing (Voican, 2020).

In today's world using private or public data produces so many opportunities on the market. A massive quality of data, including details on customer accounts, transaction information, and other financial data, is produced daily by the banking sector (Patil & Dharwadkar, 2017). Many companies will try to become data-driven organizations. The perks of using data analytics can be seen in terms of profit, keeping profitable customers, sophisticated analysis, addressing specific client demands, and increasing customer happiness and loyalty (Patil, 2020).

When it comes to customer data, it is gathered by CRM at every transaction and during every customer activity (Khan, Ehsan, Mirza & Sarwar, 2012). To enhance company processes, this data is evaluated. (Khan, Ehsan, Mirza & Sarwar, 2012).

Soon, scientists predict that the whole world economy will become a data-driven economy. Businesses need to be aware of these changes and respond to them. A lot of companies are already on this path, those that are not yet will join soon or they will not make it to survive. On top of this trend, there are also Artificial Intelligence and Machine Learning raising thanks to an abundance of available data.

For example, customers can be identified in the bank through an overview of the accounts they have opened, as well as through the customer database, because the bank must have certain information about the client before opening an account, such as name, surname, date of birth, address, identification document, phone number, and so on. In their paper, where they analyzed two classification models as well, they proposed a practical and competent framework for the prediction of individual action which can be used by companies.

All companies on this path of survival will have to create Data Architecture and Data Strategies. These strategies will enable efficient usage of the data that is collected by different kinds of applications. Data Management, AI, and Advance Analytics will change the world that we know. Data, together with digitalization/mobility is shaping the world economy, but at the same time bringing some new opportunities.

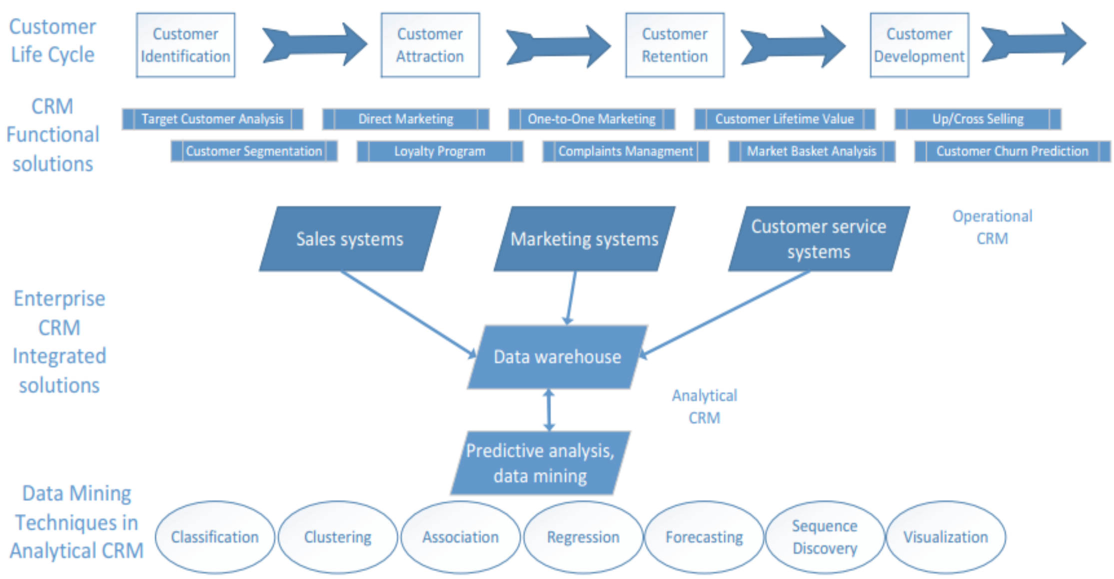
During time periods, banks realized that consumers will undoubtedly substitute one bank with the other on any occasion they realize they have better conditions elsewhere (Mahalakshmi, Saravanaraj & Umarani, 2013).

Companies will create applications, and shape products with data and security in mind. In the future will be even easier than today to bring business decisions based on data that we have available. Also, one of the trends will become data ecosystems that will enable companies to share data. This trend is very interesting for the financial industry.

In parallel with the rising importance of data, we needed improved tools for investigating the data, creating business use-cases, models etc. The rapid development of tools for analyzing data is then a natural step in our attempt to use as much data as it is possible.

In Figure 1, it can be seen how CRM solutions support customer life cycles. Within four customer life cycles, the CRM framework can be separated into operational, and analytical in terms of business angles (Gončarovs, 2017).

Figure 1: CRM endorse customer life cycle



Source: Gončarovs (2017).

2.2 Data mining techniques and algorithms

Different algorithms were investigated in this section, and in my practical section, I showed the Random Forest algorithm using the data set provided from the bank. Therefore, the Random Forest algorithm was explained in more dept within this chapter.

To detect and anticipate their customers' demands, banks can employ a variety of approaches to examine client's data and identify important information which enables banks to develop a targeted marketing approach (Pascu, 2018).

A wide range of banks are using one of the following techniques to satisfy their needs: Customer and Market Analysis (Segmentation/Profiling), Customer Profitability Analysis, Develop Sales, Retention of Customers, Risk Assessment, Fraud Detection, Reduce Operating Cost, Improving Operational Efficiency (Chuang & Hu, 2014). New customers can be attracted by their correlation with existing customers (Chuang & Hu, 2014). Various methods of data processing are together referred to as "data mining" (Aggarwal, 2015). Data mining is the process of gathering, cleansing, organizing, evaluating, and drawing insightful conclusions from data (Aggarwal, 2015).

Descriptive and predictive data mining are two broad categories. Descriptive data mining has its roots in the conventional model-building process of statistics and focuses more on explaining, categorizing, and outlining the data as opposed to predictive data mining, which examines historical data to derive trends or conclusions for future prediction (Kulkarni & Sinha, 2012).

The goal is to find similarities in the data by using probability, which can help answer certain questions (Nitsche, 2002). Finding hidden patterns, relationships, and correlations in the business information held in databases using data mining techniques can aid a company greatly in solving business challenges (Bhambri, 2012). Customer segmentation, by using large databases and dividing the market into sections, was one of the first business utilizations of data mining (Voican, 2020). Numerous ways can data mining be used in the banking sector; first, it can be used for banking services that are offered to customers, it might have exterior uses (Caron, 2019). Second, if it is integrated into a bank's operating architecture, it might have internal applications (Caron, 2019). Fraud detection" and "credit risk analysis" are two data mining utilizations that are crucial for banks to use to prevent the loss of billions in revenue each year (Soltani Delgosha, Hajiheydari, & Fahimi, 2020). These two applications are crucial for the bank's long-term stability as well as its capacity to maintain stability in the long term (Soltani Delgosha, Hajiheydari, & Fahimi, 2020).

The data mining procedure could be divided into a few phases. Firstly, business understanding where business goals and issues are identified and transformed into data mining issues, and then also understanding of collected data. Afterward, data preparation and modeling take place (Jahnavi & Katyayani, 2014). In data arrangement, data will be selected and divided into subsets, cleaned, and integrated (Pascu, 2018). Evaluation and Deployment are the last phases business objectives are used to validate models and the modeling process phases (Jahnavi & Katyayani, 2014).

The various forms of problems that can be modeled and resolved are specified by data mining algorithms that are divided into two groups: Supervised Learning and Unsupervised Learning (Chitra & Subashini, 2013). A set of independent qualities or predictors are used in directed data mining to attempt to explain the behavior of the target (Chitra & Subashini, 2013).

A supervised model must be trained, which entails having the computer examine numerous instances where the goal value is previously familiar. (Chitra & Subashini, 2013). The model "learns" how to make predictions during the training process like a model that aims to predict which customers are most seemingly to respond to a promotion, for instance, must be trained by studying the features of numerous customers who are known to have replied or not to a campaign in the past (Chitra & Subashini, 2013). With a training set of precisely identified observations, classification is one example of supervised learning (Gončarovc, 2017).

Because of a training set of data that includes perceptions whose classification is known; determining to which subset of groups a recent observation belongs is the problem of classification (Gončarovc, 2017). Common algorithms that are used are decision tree, random forest, kNN, and Naïve Bayesian classification. Decision trees are effective and widely used for classification and prediction, partly because they express rules and allow for the study of data to comprehend the links between a large number of potential input variables and a target variable (Berry & Linoff, 2004).

The banking industry has collected adequate amounts of valuable consumer information, thus the present implementations in this area are not only concentrating on the marketing side of things. There is a lot of potential and worthwhile knowledge out there just waiting to be found (Hassani, Huang, and Silva, 2018). In the banking sector, where new client acquisition is more difficult than in other sectors, customer retention is more efficient financially and requires less work (Britto & Gobinath, 2020).

As soon as new information about the banking industry became available, to retain their current clientele, banks had to establish new strategies. The increased competition in the banking sector has forced banks to find new ways to keep customers. Banks have a lot of information about their customers and would like to use it to generate more revenue.

As I have employed the random forest algorithm in the practical part of my thesis, the next section provided a more comprehensive elaboration of the algorithm.

A computer scientist at IBM Research, Tin Kam Ho has made contributions to classification, data mining, and machine learning (Ho, 2014). She was the first who designed and suggested the method of random forest, which later was extended by Leo Breiman and Adele Cutler, as well as branding it in 2006. The "split and conquer" method is used by random forest, creating randomized tree predictors on each tiny bit of data using sampled portions of the data, then pasting (aggregate) these predictors together (Biau & Scornet, 2016).

A decision tree is primarily thought of as a classification system that may build a tree and a set of rules reflecting the model of various classes from a certain dataset (Suri, Singh, Singh, Rajappan, & Florence, 2020). In simple words, a single decision tree's goal is to ask numerous questions and then provide relevant answers. Decision trees occasionally try to understand nonlinear relationships by posing a series of queries and providing a classification response (Appiahene, Missah & Najim, 2020). One of the main problems with decision trees is overfitting where the entire training set can be described if the tree is permitted to be deep or overly complex (Appiahene, Missah & Najim, 2020). Using bagging, users would sample and resample from a training set, model each sample, and then average them out.

With a decision tree alone, users might take a training set and produce one model, but bagging suggests producing a few random subsets whereas a result, the option is to sample with re-sampling, train a model on, for instance, simply 100 samples, and then repeat the process to get N models (Carbo-Valverde, Cuadros-Solas & Rodríguez-Fernández, 2020). A vote is then requested from each of the N models. If users want to choose a subset of all the features that are accessible, random forest is a good example of bagging that goes further than simple bagging (Carbo-Valverde, Cuadros-Solas & Rodríguez-Fernández, 2020).

The fundamental principle of random forests is training a decision tree using a very small subset of features, restricting its scope and depth, and asking it to learn using only a small portion of training data, repeatedly (Carbo-Valverde, Cuadros-Solas & Rodríguez-Fernández, 2020). The consequence is a forest of trees, which users then average or decide by asking for their votes. The definition of random forest is stated as "a classifier consisting of a group of tree-structured classifiers $\{h(x, \Theta k)\ k=1, 2, \ldots.\}$, where the $\{\Theta k\ \}$ are independent identically spread random vectors and each tree casts a unit vote for the most popular class at input x [11]" (Kulkarni & Sinha, 2012).

To categorize an input vector, the random forest classifier consists of multiple different kinds of tree classifiers, each of which is built using a random vector sampled independently from the input vector (Breiman, 1999). The random forest classifier employed in this study involves growing a tree by employing randomly chosen features or a mix of features at each node. For each feature/feature combination chosen bagging is used. Creating a training data set using the bagging technique involves randomly selecting N replacement instances, where N represents the initial training set's size (Breiman, 1996).

Therefore, the bootstrap sample is used when the N records are picked from the training set and at random but with substitution from the initial data (Kulkarni & Sinha, 2012). This sample will serve as the tree's training set (Kulkarni & Sinha, 2012). If there are M input variables, a number $m \ll M$ is chosen so that for each node, m variables are selected at random out of M (Kulkarni & Sinha, 2012). If there are M input variables, the node is split using the best split on these m attributes. M is kept as a constant value as the forest expands. Every tree is developed to its full potential, without prunin (Kulkarni & Sinha, 2012).

In this approach, the forest is made up of many trees, the number of which is predetermined by N tree criteria (Kulkarni & Sinha, 2012). The number of variables (m) selected at each node is sometimes named as m try or k in the literature. (Kulkarni & Sinha, 2012). The node size option, which refers to the number of instances in the leaf node, can be used to determine the depth of the tree and is typically set to 1 (Kulkarni & Sinha, 2012). Once the forest has been trained or built as previously mentioned, it goes through all the forest's cultivated trees to categorize the most recent case (Kulkarni & Sinha, 2012). Each tree classifies the new instance, and each categorization counts as one vote (Kulkarni & Sinha, 2012). After adding the votes from each tree, the class with the highest number of votes is selected as the classification for the new instance 1 (Kulkarni & Sinha, 2012).

The next section concentrated more on the concept of CRM in the financial industry after reviewing the literature that discusses CRM in general.

3 CONCEPT OF CRM IN BANKING INDUSTRY

3.1 State of CRM in financial institutions

CRM is a system that encompasses all facets of communication and interaction with customers today due to the widespread use of electronic communication technologies, meaning it decides how to interact with clients, resolve their issues, convince them to use banking goods and services, foster a feeling of loyalty, and sustain financial interrelationships with customers (Cvijović, Kostić-Stanković & Reljić, 2017).

To shorten the time and expense of processing an application for different products, as well as to ultimately improve economic performance, traditional one-on-one customer contact is being replaced by electronic points of contact in the financial services business worldwide (Moin & Ahmed, 2012).

The impact of the financial sector's restructuring intended is to boost the monetary structure's efficiency and competitiveness (Gayathry, 2016). Therefore, the financial services sector today has seen significant changes, including new rules, altered customer behavior, expanded use of information and communication technologies, and fierce rivalry (Cvijović, Kostić-Stanković & Reljić, 2017).

Due to intense competition with retail banking, it is becoming harder to keep customers and grow the customer base. The primary approach for a bank to attract or maintain a customer is to be proactive, to be aware of what clients want, and to live up to expectations (Pascu, 2018). It is challenging to offer clients something different and better, which will attract them and keep them with you in the long term. It was found out that clients manage to isolate advantages that they have from connection with banks, from actual products and services that are offered to them (Cvijović, Kostić-Stanković & Reljić, 2017).

What they perceived as benefits regarding communication with bank are: to have a sense of security, to feel that they receive a unique and personalized approach, to develop and preserve social links with a bank, and to have it all convenient and easy to follow (Cvijović, Kostić-Stanković & Reljić, 2017).

To increase their service offerings banks should develop many channels, such as branches, contact centers, ATMs, iBanking, and personal digital assistants, to improve their services, connect with clients, and enhance the convenience and security of each channel (Liu, 2007).

As customer knowledge and awareness grew, it became increasingly difficult to attract next customers as well as retain the old ones. Research suggests that in the banking business profitability can increase by 35% if there is a 5% increase in customer retention, and that the reason why banks are putting the focus on keeping their customers and enhancing their market share (Mahalakshmi, Saravanaraj & Umarani, 2013).

Considering that there is a direct connection between customer contentment and utility, banks identified the need for adopting technology to improve their services and relationships with clients (Pokharel, 2011).

Since products and prices can be easily copied, a lot of banks are focused on a strong differentiation in services, because services are more challenging to replicate than products (Pokharel, 2011). The reason for that is an experience that showed that services need customer reaction and involvement (Pokharel, 2011).

Since banks provide customized product offers through online banking, mobile banking, and ATMs, there is rivalry in the banking business to expand client reach with web-based technologies (Al-Dmour, Saad, Basheer Amin, Al-Dmour & Al-Dmour, 2021).

3.2 Opportunities in CRM implementation

Technology advancement significantly influences business models by reducing transaction and delivery costs, with a focus on gaining new clients and increasing cross-selling potential (Gupta, 2019). By producing and storing data in real-time, in various formats, and by a wide range of institutions and individuals, information and internet technology has had a significant impact on today's economic and financial activity (Wibisono, Widjanarti, Zulen, & Tissot, 2019). In 2010, CRM has been invested in by nearly half of all banks, which showed that banks were enthusiastic about using CRM as a strategic asset to achieve their organizational objectives (Perwej, 2010).

With a CRM implementation within the banking industry, businesses can enjoy a variety of benefits, including decreased entry barriers, improved service levels, and a decrease in operational expenses (Ramaj, 2015). It helps to improve the effectiveness of marketing activities, including customer segmentation, personalization, and marketing planning (Ramaj, 2015).

Next, one of the opportunities is that it analyzes business performance thoroughly, considering information on clients in marketing, sales, and services, as well as the effectiveness of the sales and communication channel (Ramaj, 2015).

In the services department implementation of CRM uses the opportunity to organize services in a way that includes capabilities for multi-channel services, resource management, client service, and professional services (Ramaj, 2015). Cooperative technology gave customers the chance to obtain information from multiple sources and to gather information, while empowerment enabled customers to learn more about goods and services (Dubey, Sharma & Sangle, 2020).

Expansion into new markets, which can lead to the creation of new working processes and improved relations with clients is an important opportunity provided by CRM implementation (Gupta, 2019). By expansion, business is evolving since it includes engagement between different channels (Nure, 2018).

To provide more and more services that are focused on the needs of the client, banks that use CRM software can gather data about their customers and use it to construct a customer profile or view (Saxena & Khandelwal, 2011). Therefore, banks are getting to know their customers better, which leads to a better awareness of their expectations and customized co-produced offerings (Dubey, Sharma & Sangle, 2020). Knowing their clients gives banking activities the chance to establish and preserve long-term relationships (Nure, 2018).

CRM has concentrated on evaluating important customer satisfaction characteristics and identifying customer groups with preferences and distinctive expectations, which helps businesses increase their profitability with their clients while also making interactions seem friendlier through individualization (Nure, 2018).

3.3 Challenges in CRM implementation

When it comes to challenges in the implementation of CRM, they can be divided into 3 categories: those that are related to data, organizational and technological challenges. Silos of information, the difficulties with privacy and security, bad data quality, are just a few of the data-related issues that exist (Soltani Delgosha, Hajiheydari, & Fahimi, 2020).

One of the main problems that banks are facing today is lack of security and regulated access since the bank must ensure that all customer data will be protected from misuse. If the bank makes a mistake and does not sufficiently secure its data and several times it happens that they have a cyberattack during which some data from clients is stolen and misused, clients will lose trust in them and therefore their relationship will be ruined. Not only will they be skeptical about doing business with them in the future, but they will also tell their acquaintances and not recommend that bank anymore.

Then, challenging is also to integrate CRM software with already present infrastructure because modern CRM systems cannot easily be integrated with existing legacy systems. It is obvious that technologies must easily interact with the company's data warehouse for a CRM to be effective (Xu, Yen, Lin & Chou, 2002). A company must merge the data kept in both front-office and back-office systems to deliver the best quality and most up-to-date information and since the structures of front-office systems are so similar, information integration is comparatively simple (Xu, Yen, Lin & Chou, 2002). Integrating data from the front office and the back office is the real challenge, because bac-office systems are frequently highly unique, and they use outdated or proprietary technology (Xu, Yen, Lin & Chou, 2002).

Therefore, many banks are struggling to implement CRM systems successfully. The most significant organizational constraints that could stop banks from fulfilling their data analytics potential are the absence of skilled and experienced personnel, the high cost of implementation, and absence of senior management support (Soltani Delgosha, Hajiheydari, & Fahimi, 2020).

On people-oriented challenges, banks are trying to cope with the arrangement of processes and people, a deficit of skilled people, and getting management sponsorship which can be a difficult task but can be very beneficial to the organization (Pokharel, 2011).

Customer satisfaction is a key factor in the success of CRM in banks (Pokharel, 2011). Customer service abilities are designed to establish an emotional tie with the customer which is closely related to the organizational learning philosophy and the adaptive culture of the banks (Pokharel, 2011).

In comparison with the past, today’s customers are more familiar with the services that banks offer which resulted in the expansion of customers’ expectations (Abrar, Biag, Shabbir & Hussain, 2019). In the advanced environment of this technology, the maximum challenge for commercial banks is to build sustainable relationships with our customers (Abrar, Biag, Shabbir & Hussain, 2019).

Nowadays, customers are aware that companies possess data about them so they assume that they will get customized products and have large expectations from firms since they are aware that companies have this opportunity (Gončarovs, 2017). Therefore, companies must meet these expectations to attract and retain customers. Experts for CRM know how to apply data mining techniques to gather relevant data from inputs (Gončarovs, 2017).

In organizations, data analytics supported by CRM can be used widely, starting with prognosing customer behavior up to purchasing patterns to understanding movements in particular industries (Gončarovs, 2017).

Historically speaking, the people that worked in the back offices were the first to have contact with corporate intelligence and data warehousing, then knowledge workers have started to get more involved in data analysis gaining more expertise in relation to analytic assignments (Gončarovs, 2017).

Finally, the category of technological problems includes several issues. Investment in IT is essential if banks want to survive on the market since technologies need to handle massive volumes of data (Soltani Delgosha, Hajiheydari, & Fahimi, 2020). In fact, it is true that IT is an essential and important component of CRM, but it is not sufficient in this regard; additional tools are also required (Bahrami, Ghorbani, & Arabzad, 2012).

In the next part of the thesis, I presented how the Random Forest model works using a specific example from the banking industry.

4 METHODOLOGY

4.1 CRISP-DM methodology

The CRISP-DM approach was used for my thesis. The six phases of a specific data mining project's life cycle, as defined by CRISP-DM, are adaptive, meaning that the following step in the sequence frequently depends on the results of the one before it (Laros e& Larose, 2014). CRISP-DM aims to guarantee that every stage is carried out methodically and completely, producing outcomes that are accurate and reliable (Laros e& Larose, 2014).

The first stage of the CRISP-DM process was the understanding of business, as shown in Figure 2. Declaring the project's goals and developing a plan to achieve them is crucial during this phase (Larose & Larose, 2014).

The second phase was understanding data. Data gathering was the initial step in the data understanding phase, followed by exploratory analysis and evaluation of the quality of data to become comfortable with the data and gain first insights (Larose & Larose, 2014).

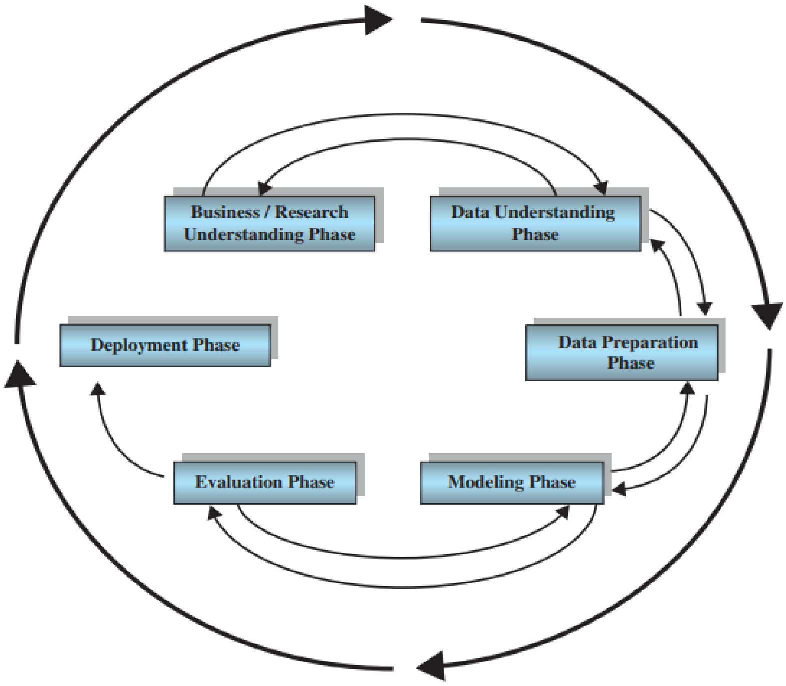
The third phase was data preparation, also known as pre-processing, which consists of cleaning the data, selecting attributes, transforming it, and so on (Bahari & Elayidom, 2015).

The previously described phase was a prerequisite for the building of models and leads to behavior prediction of customers. This time-consuming phase involves every step of creating the finally formed data set from the unclean data, which was used in the following phases (Larose & Larose, 2014).

Selection and application of modeling techniques were part of the fourth phase (Larose & Larose, 2014). To bring the form of the data into compliance with the special needs of a particular data mining approach, this phase may involve circling back to the data preparation phase (Larose & Larose, 2014).

Since one or more models were produced in the previous phase, they should be assessed for quality and efficacy in fifth phase (Larose & Larose, 2014). Looking back is helpful to determine whether the goals set for the model in the first phase were accomplished (Larose & Larose, 2014). The last phase was for model deployment, where use of the developed model is required (Larose & Larose, 2014).

Figure 2: CRISP-DM methodology



Source: Larose & Larose (2014).

4.2 Interviews with employees

I spoke with bank employees in interviews both before and after NBO was implemented. The seventh chapter contained the interview results. I prepared myself by doing research about analytical CRM in banks and prepared a list of questions to ask. At the beginning of the interview, I explained the interview's objective and how I would use it for my master's thesis in detail.

Depth and semi-structured interviews were the two types that were undertaken. To obtain a more detailed understanding of the situation inside the bank, I have opted to conduct interviews rather than questionnaires.

Instead of completing a survey, participants may feel more comfortable talking in-depth about the topic during an interview, which could result in a more relaxed environment for information gathering (Boyce & Neale, 2006). The interviews lasted from 45 minutes to 2 hours.

I have listed the job titles and responsibilities of those who work on data-related projects in the table below (see Table 1). Additionally, I received details regarding how long they have worked for this bank. A general understanding of their roles and participation in the NBO model was crucial in my opinion.

Table 1: Employee position responsibilities

Title	Position Responsibilities	Years in the company
Data Engineer	- Create systems that gather, handle, and transform unprocessed data into interpretable. - Make data available so that it can be used for analysis. - Acquire datasets that align with business needs. - Construct, evaluate and keep up database pipeline designs. - Make sure data governance and security policies are followed.	7
CRM manager	- Identifying new business opportunities through prospecting techniques. - Act as a Product owner in Data Advance Analytics teams.	5
Database Administrator	- Maintain Databases. - Handle Database Security. - Database performance tuning. - Identifying new business opportunities through prospecting techniques such as cold calling, networking, or advertising. - Responsible for supporting Data Advance Analytics teams as domain expert.	6
Head of CRM Department	- Establishing and sustaining profitable connections with important clients. - Addressing client issues promptly and successfully - Meeting with managers to plan strategically. - Act as a business owner for Data Analytics teams.	5
	- Optimizing business IT processes. - Monitoring IT efficiency, IT security and governance	

Head of IT Department	- Create, supervise, and monitor the annual budget. - Provide IT employees with advice and training. - By keeping an eye on system performance, make sure IT services are delivered and run without a problem.	10
Data Scientist	- Create algorithms that turn data into information. - Verify that data governance and security policies are being followed. - Work together with engineering and product development teams. - Evaluate a lot of information to find patterns and trends. - Create machine learning algorithms.	3

Source: Own work.

4.3 Dataset

All the data I received from the chosen bank was anonymized according to GDPR. GDPR is a directive that should help in the protection of the personal data of clients, and it is approved by European Union (EU) parliament on 27.04.2016. and since 2018 gradually must be adopted in information systems all around EU (Ferrara & Spoto, 2018). GDPR focuses on systems that are storing and manipulating personal data, and sets the rules for manipulating, storing, and deleting personal data, giving rights to clients to decide how their personal data will be handled (Ferrara & Spoto, 2018). Since the GDPR directive is very clear and strict regarding data privacy, companies, to meet these new regulations, must invest in changing their information systems, services, and applications. An associated EU program called the GDPR aims to standardize personal data protection laws among its member states and according to the "access to account" regulation, its restrictions are applicable to payment information for both payers and payees (Ferrara & Spoto, 2018).

When GDPR regulation took place, many companies hired and establish Data Protection Officer (DPO) roles in their organizations to help them to be GDPR compliant and to set foundation processes and systems to be in line with GDPR (Brodsky & Oakes, 2017). The role of DPO is to educate people on GDPR, assess and establish effective data management processes, minimize data collection etc. So, GDPR is impacting data science also in a way that data scientists must take care of data privacy and data security in line with GDPR regulations. Every data analysis process should be supported by DPO to ensure that data is handled as GDPR requested.

There are lot of techniques that can help us in handling and storing client’s data in safe manner and in line with GDPR regulations such as: Data Masking, Data anonymizing, Data Encryption, Reducing the data, etc.

In my use case all data was anonymized by a technique called „Shuffling ". In this technique data was simply mixed in the dataset. This technique was used at the end of the use case otherwise it would have impact on use case results.

Additionally, data protection was implemented on the RDBMS level, where data was stored, in a way that data encryption was implemented (Oracle TDE). Of course, the security aspect of data was on a high level since asses is controlled by Oracle database with built in advance security option.

The sample size of the data about customers and their transactions was approximately 2 years. Regarding data sources, data set was framed from General account data (all product history of clients regarding credit and non-credit products), Transactions data and Customer data (social-demographics data, clients age, gender, employment…)

Since the dataset contained a large amount of data, all data for my master thesis was stored in Oracle Database. Oracle database is intended to manipulate data in an efficient and secure way. Data was protected and available only to authorized persons.

Data preprocessing was done by SQL developer. This tool represents a native Oracle product for manipulating and managing data using SQL (Structured Query Language) Language. Analysis of prepared data was gradually maturing using Python programing language.

I did analysis and visualization of data in Jupyter notebook. Firstly, I did exploratory analysis, understood data and relations, and then preprocessed it, did the data preparation-cleaning, loading, and connection of data. Data was anonymized so that information about customers was adequately protected.

By processing a large amount of data which are generatd by a daily functioning of transactional financial systems, the organization is able to provide insight in the behavior of the clients and use that data for the purpose of providing personalized services to clients.

The end result of the analysis was a group of clients with personalized offers for personal loan. In the end a model was developed, and I addressed model metrics and validation. Results were analyzed, and I gave proposals on how to make CRM more effective.

5 CASE DESCRIPTION

5.1 Company description

Based on the material that the bank provided me; I presented a description of the organization in this chapter. The bank documents used in this research are confidential, therefore I used "XYZ" as the name of the bank.

The chosen bank is a worldwide financial and banking organization based in the EU, and until 2013, they had revenues of more than EUR 800 billion (Statut of XYZ Bank, 2019). Bank Group's headquarters is in the EU, but not all the countries that the bank is present in are EU members. The governance of the "daughter" is very centralized (Statut of XYZ bank, 2019).

Since country local regulations and laws are specific in different countries, product management is not fully centralized but technical solutions are centrally developed and managed. This way the cost of developing and maintaining technical solutions is lower but at the same time make some boundaries in sense of product development.

Also, this approach means that the time to market is somewhat slower. Since the group preferred centralized approach, knowledge is mainly based in headquarters, shared between other members of the group through collaboration tools, group projects and different trainings.

Bank offers modern products and services to over 13 million clients worldwide. They are available 24/7 - online, on mobile platforms and through the network of their branches. The bank is organized into four key regions and two sectors for corporate and individual solutions. This gives them the opportunity to be close to their customers and offer great products in all markets (Statut of XYZ bank, 2019).

The goal is to establish long-term relationships with customers from various industry sectors. Relationship managers and product specialists do this by working with clients to develop custom solutions.

Their values are sincerity, ownership, and sympathy where the first value points out the bank's honesty and transparency in their interactions internally, as well as externally-with their customers (Statut of XYZ bank, 2019).

The second value ownership states that they are responsible for their action, and the third value sympathetic means that they care for each other and their clients (Statut of XYZ bank, 2019). Around 50 employees from the Bank participated in training and education as part of the CRM concept's introduction, which began in 2019 (XYZ Bank - CRM procedures and best practices, 2020).

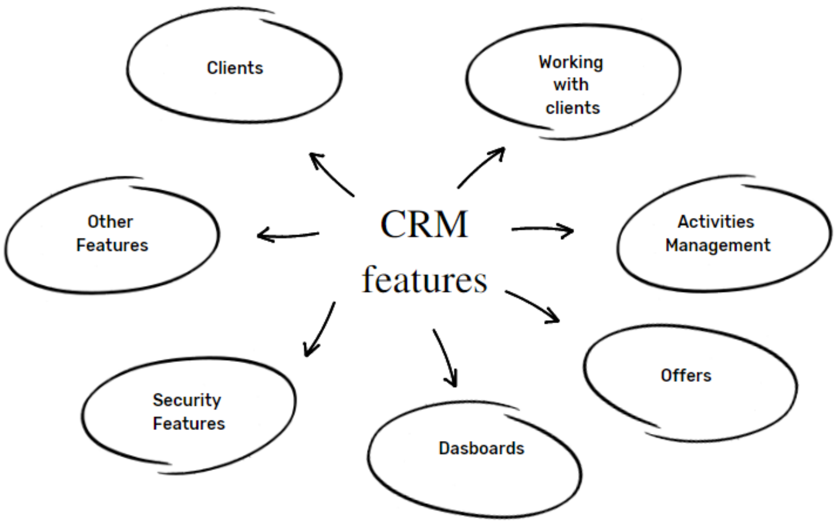
5.2 Current state of CRM

In the following chapter I presented the current situation of a bank regarding CRM and the practice of data analytics based on the material provided by the bank. Certainly, the maturity of CRM and data analytics differs across banks, based on their size and resources. Microsoft's CRM software is being used by the bank (XYZ Bank - CRM procedures and best practices, 2020). Their software belongs to the family of Banking software products, which are referred to as the Core Banking System (XYZ Bank - CRM procedures and best practices, 2018). The vendor decided to build a set of dynamic modifications for all banks because banks operate in numerous countries with diverse market conditions and environments, however, banks must adhere to bank policy because of this.

In this manner, CRM offers banks customized solutions, enabling them to customize the application's style and content. Banks must make sure that everything is established in accordance with bank group policies and strategies. CRM enables the bank to enhance client connections by utilizing a variety of processes and merging them with various banking concepts and tactics.

It offers several capabilities for keeping track of customer information and making it simple to give customer assistance. In the Figure 3, an overview of CRM product features can be seen divided into 7 parts:

Figure 3: CRM product features



Source: Documentation Department (2018).

The CRM term "activity" refers to every client encounter and aids employees in better planning and scheduling their communications. Scheduling and managing client appointments is one of the most used CRM features.

By storing information and data, users make it easier to follow appointment specifics and enhance customer satisfaction. One of the most popular features of a CRM solution is the offer management functionality. It helps users to learn from customer feedback and approach customers with the correct offer at the right time for the best outcomes. It also enables users to make wise real-time suggestions to customers and to take full advantage of the possibility to cross-sell.

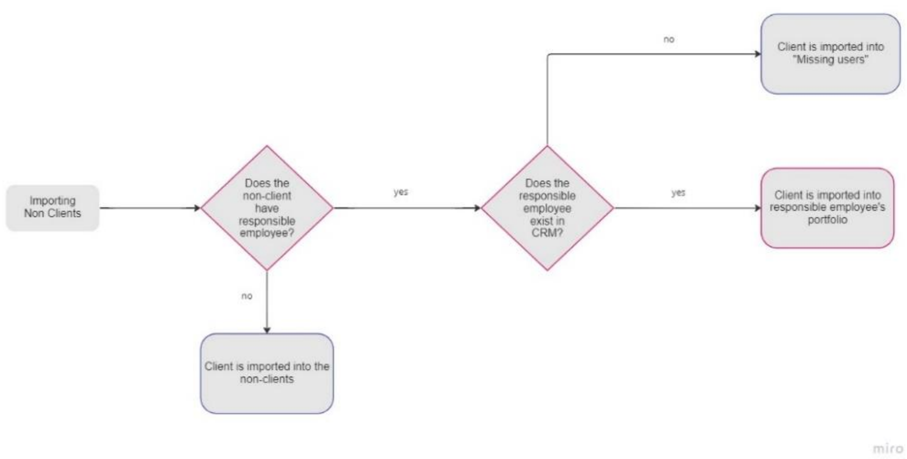
The ease of use of CRM and the availability of current data on each client in their portfolio were both mentioned by employees. They can use an ordinary web browser to access, modify, or add data to the CRM whenever they want, with a web-based application. Interaction between modules and records in the applications is straightforward and employees may simply access various records. Depending on the type of relationship, client categorization (business clients, private clients' institutional clients), responsible employee of the client, and availability of the client category, different rules are implemented for integrating clients into CRM.

In Figures 4 and 5, it is shown how the process of importing clients and non-clients into CRM system is done. For instance, business clients are imported into the responsible employee's portfolio if one has been designated for them. If there is no designated responsible employee for business clients, the import will be conducted in accordance with the client's category and branch link. CRM system makes it simpler to recognize and maintain relationships with clients because it is effective at the time of client encounter.

The same process is used to import private customers. Private clients are imported into the portfolio of the relevant employee if they have been assigned. If they do not have a designated accountable employee, they are added to the list of private clients. Also, the private clients' box imports all institutional clients. The same guidelines that apply to private clients also apply here.

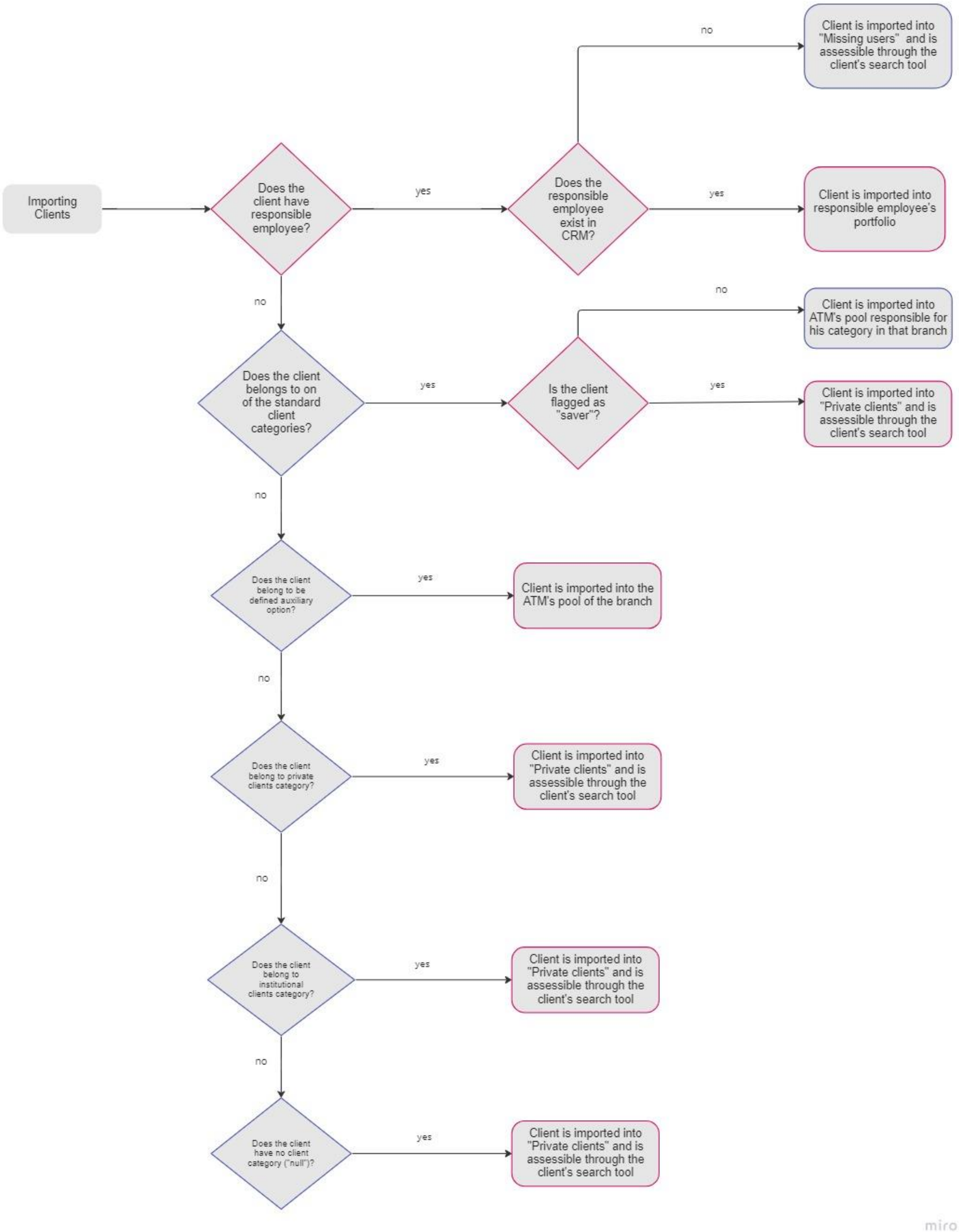
Clients who are not enrolled under one of the regular client categories will have their information imported into the box for missing users. Relationship types might alter over time, based on how the relationship with the bank develops. Clients may turn into non-clients, for instance, when they complete all their product contracts with the bank, or they may change from being non-clients to clients because of an employee's analysis.

Figure 4: Importing Non-Clients



Source: Documentation Department (2018).

Figure 5: Importing Clients



Source: Documentation Department (2018).

5.3 Current state of data analytics practices

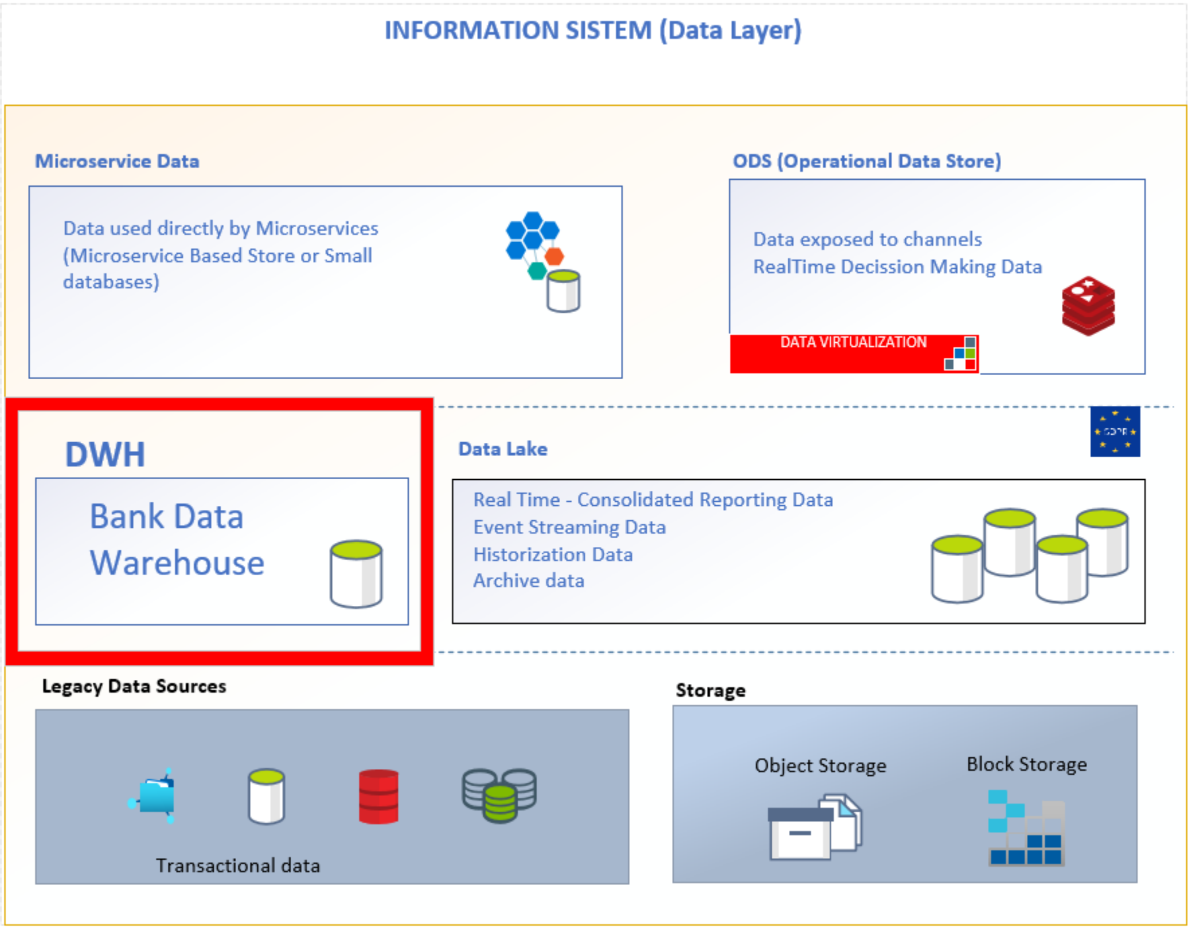
In the next section I described the bank data platform – source system in the chosen bank, according to the information provided by the bank.

Since bank has many different systems to cover all business processes, it is needed to have Data Warehouse (DWH) where all these sources will be consolidated. Daily, Extraction Transformation Load procedures are loading data to DWH. DWH is used for reporting purposes for different segments in the bank (Finance, Risk, HR, Treasury, Data analytics use analysis, reporting to regulator and management etc (IT strategy for XYZ Bank, 2018).

As it is shown in Figure 6, the bank data platform consists of several very complex data platforms that are representing a set of data stores:

- DWH
- Data Lake
- Data Analytics platform

Figure 6: Bank data platform - Data layer



Source: Documentation Department (2018).

Firstly, I stated what is NBO in general and then explained the current setting inside the chosen bank.

The "next product to offer" in the banking business chooses items to meet the needs and objectives of each customer, as opposed to traditional campaign management, which chooses qualified customers for a certain product (Lau, Wong, Ma & Liu, 2003). It is a customer-centric approach in which various client characteristics are considered when choosing the product (Lau, Wong, Ma & Liu, 2003).

The purpose of the NBO use case is to create and communicate product offers that are relevant for every individual consumer based on their prior behavior and demographic attributes.

Customers' information is used by marketers that apply the Next Best Action to make sales, keep customers, and calculate each customer's worth. Companies surpass consumer expectations, achieve personalization, and provide the best customer experience by knowing the optimal next step in the sales process for each customer.

When it comes to chosen bank. a complete set of data for the NBO use case was available in the DWH system (IT strategy for XYZ Bank, 2018). DWH system stores chronologically data that bank must keep for many years. Since reporting is depending on this data, in DWH, bank is using relational database models.

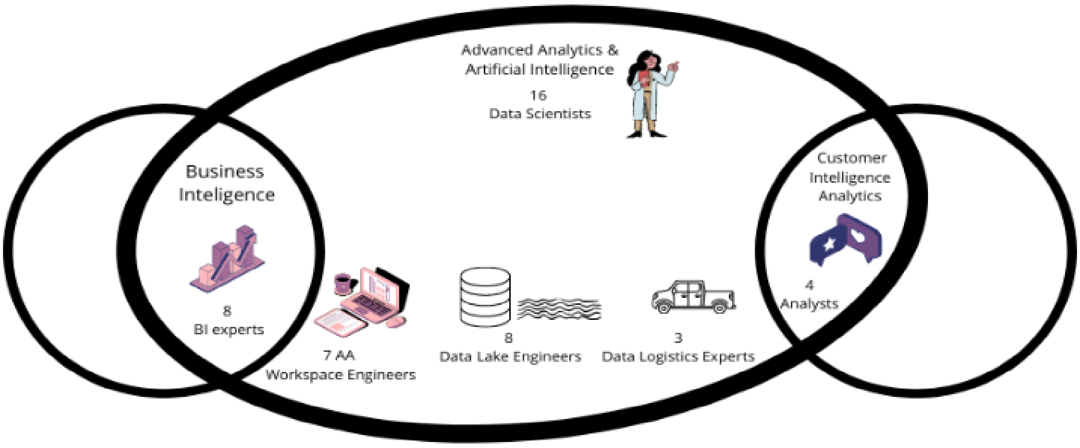
To be able to develop NBO model and integrate it to bank system, bank is also using application layer that is used for digging and analysing data in DWH (IT strategy for XYZ Bank, 2018).

Tools that bank is using are open-source solutions like the Anaconda Data analysis platform with Python programming language support, and Jupyter notebook (IT strategy for XYZ Bank, 2018).

Figure 7 showed the overview of data-related positions setup inside the bank. Over the years, the number of experts who work with data has increased, and the bank understands the necessity for people of such a profile. Many corporations claim to have strong functional knowledge in data science, data management, advanced analytics and analytics transformation and professionals for mentioned areas (Macias-Lizaso, 2018).

Employees doing these jobs have traditionally been a member of other teams, but as the demand for data analysis has grown, a team specifically focused on data has been developed. This leads to improved communication and increased production. They are aware that to offer the appropriate product and at the ideal time and price, they need to invest in deep customer insights, data science and data culture. Their top use cases are transactional scorecards, income modelling, next best offer, transaction categorization, margin optimization, customer insights and pricing optimization.

Figure 7: Overview on Advanced Analytics and Artificial Intelligence setup



Source: Documentation Department (2018).

6 DATA ANALYSIS RESULTS

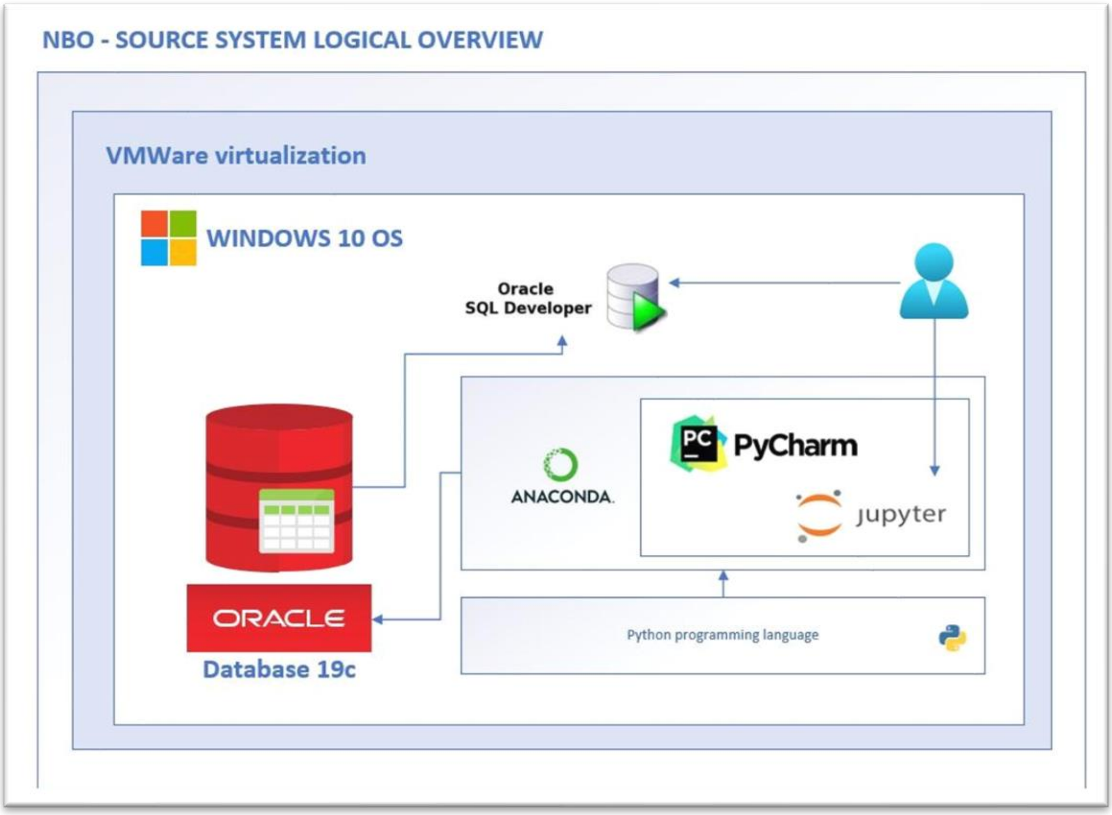
6.1 Preconditions for data delivery

The first step was the installation of the source system. Raw data that was used in the thesis represented the financial data of the banking industry clients. I processed around 300.000 clients and additionally over 1 million rows of client data including contacts, accounts, and balances. Data was extracted from the core financial system and delivered as so-called *dump exports*. Also, data that was subject to exploration was limited to 3 years. This filter of data was necessary to be able to focus on important data and shortened analysis time. Since I wanted to produce the model that is up and running in production and achieve benefits out of this analysis, I needed to have a faster time to market. Faster time to market I achieved by filtering the data in the development of the model.

Since the amount of raw data was huge, I needed to ensure that there is enough storage space to import it into the source database. The source database system hosted imported data and represented one central place for preprocessing, processing, and analyzing the data. To handle a big amount of data initially and to have sufficient performance of the database, I used Oracle Database Management System.

In Figure 8, I presented the NBO source system on the chosen bank.

Figure 8: Overview of NBO source system architecture



Source: Own work.

NBO source system consisted of:

1. Virtualization platform (VMWare)
2. Windows 10 Enterprise virtual machine (Evaluation copy of OS)
3. Oracle Database 19c and SQL Developer (SQL tool)
4. Python programing language
5. Anaconda app platform including Jupyter Notebook, PyCharm, Spider etc.

On the OS level, a windows local user was created with administrative rights to be able to install necessary programs and tools. All installation was done manually to prepare the source system for the import of data, data exploring and data analytics.

Since it was a precondition for data analysis, preparation of source system was explained in detail in Appendix 3.

6.2 Target definition

In the section above, it was explained how the import of raw data was done. All data about clients and their transactions after that were available in the Oracle database. The next step was to determine the target definition. All clients from the data set (customer table) were natural, and not legal people. They were divided into 2 groups. The groups were named Target 1 and Target 0.

Target 1 was those clients who currently have a personal loan product; and who had it in the past. The reference date was the date when the loan was approved, and it is taken from the contract table. Target 0 was a group of clients who don't have an active loan (previous loans are closed) or clients who never had a loan.

Each chosen client must be an active resident and have completed at least one credit or debit transaction in the three months preceding the chosen reference date. These transactions were also seen through POS and ATM transactions.

6.3 Exploration of data

Before focusing on the model, it was necessary to explore the data. In this context, I investigated customer data tables (data on clients), customer contracts, products that they were using and transactional data (their transaction during period of 2 years). Exploration of the data was done by the SQL Developer tool. This tool is a native Oracle SQL tool that can, in a very convenient way, by using SQL (Structured Query Language) language, manipulate the data.

6.3.1 Data set structure

The data set was containing the following information:

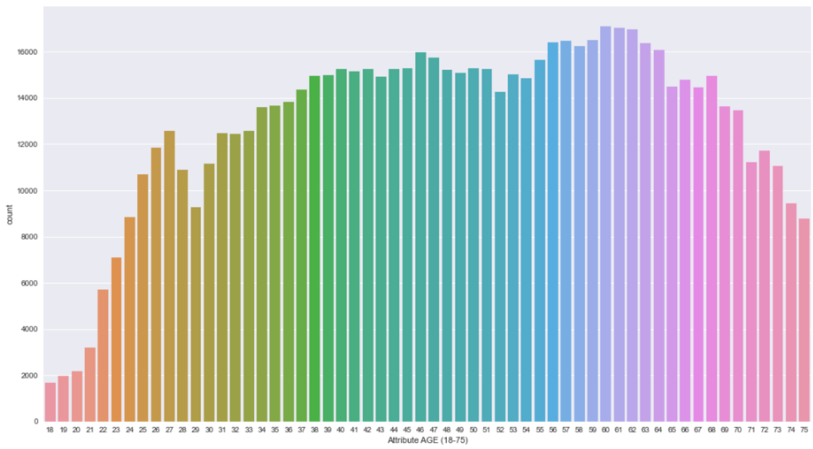
- All client product histories, including information about credit and non-credit products such as account opening dates and account types, were included in the general account data.
- Transactions data, including the history of transactions for the customers covered by the scope (relevant timestamps, quantities, debit/credit records, currency information, channel...)
- Balance information - information on all reported accounts' balances
- Customer information, including social-demographic information on clients (age, gender, industry, job, etc.). Due to data quality difficulties, campaign data, including offers and answers, were not used.

The data set also contained data campaign data, meaning data regarding different offers to customers and their responses but it was excluded from data analysis because of problems with the data quality.

6.3.2 Investigation and representation of data

To explore data contained in a data set, in this chapter I presented visualizations of the customer table, with the aim of better understanding the data set, before further analysis. The customer table contained different information about customers such as customer id, birth date, residence, age etc. Columns that were included in this table were the following: customer_id, reference_date, birth_date, residence_flag, employee_indictor, status_indicator, targert, bank_employee, age, gender, and business_type_code. For the purposes of visualization, I imported library matplotlib, which is used to create graphs by using Python scripts. Analyzing their age structure, in Figure 9, it can be observed that most customers were between 29 and 49 years old, but also a significant number of customers were between 50 and 64 years old. Therefore, it could be concluded that the customer base for the selected period was old, which indicates that the bank had many established customers, which was positive but however bank must also work on attracting younger clients.

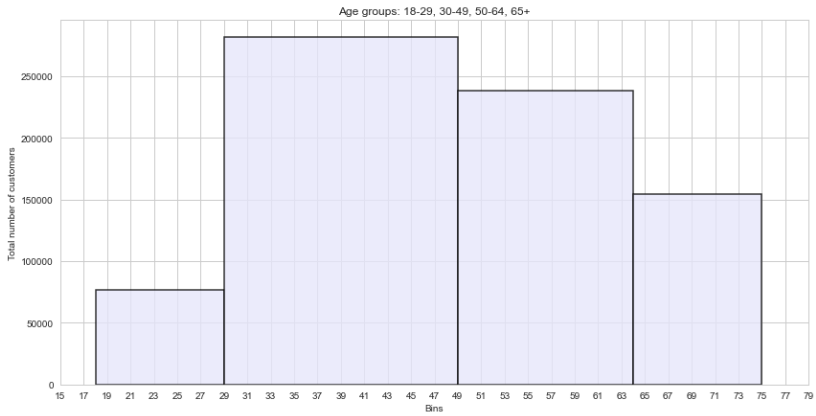
Figure 9: Customers by age



Source: Own work.

In Figure 10, the number of customers divided into age groups was presented. Many customers that were using bank services are in the range from 29 to 49 years old.

Figure 10: Age structure of customers based on categories

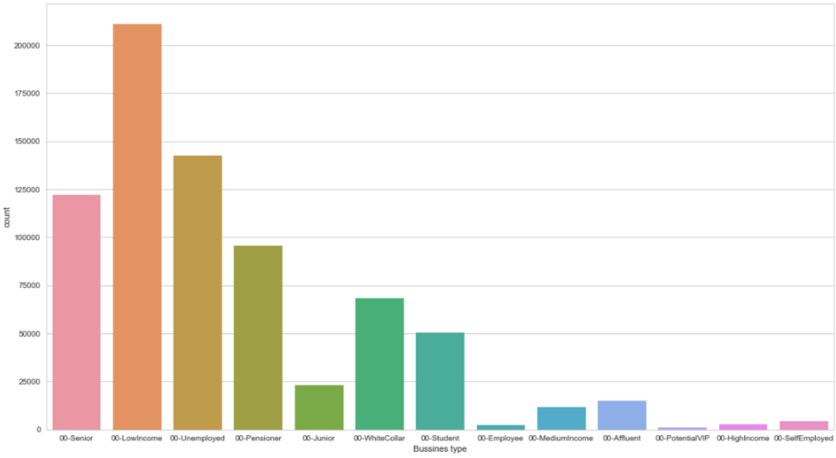


Source: Own work.

All customers were residents of the country, domestic population. It encompasses a substantial number of individuals, specifically 700,000 clients, indicating the size of the population under consideration. Each client in the dataset could be confirmed to be a resident of the country, ensuring that the data analysis and conclusions drawn from this dataset were specific to the local population. One of the preconditions was that all clients that were included in this dataset had residence flag on, meaning that clients were living in that country and receiving services.

Next, it was useful for the product personal loan to see what kind of business types were contained in the data set. After analyzing Figure 11, it can be concluded that most of them were customers with low income (below 5000 EUR of annual income).

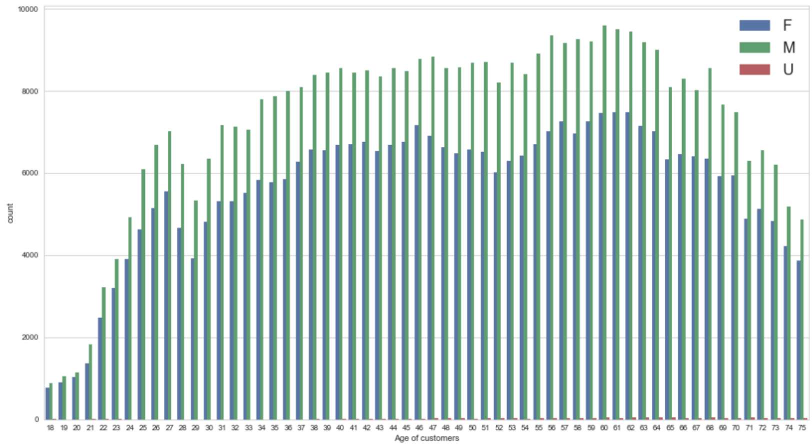
Figure 11: Business type



Source: Own work.

In Figure 12 the customer's age and gender can be seen. Gender structure of the clients was in favor of man, but there was not much difference in the segment of seniority. The reason for this was the fact that, even if there was a higher number of clients that were male, a big number was unemployed or has low income. A very low number of both genders were affluent clients.

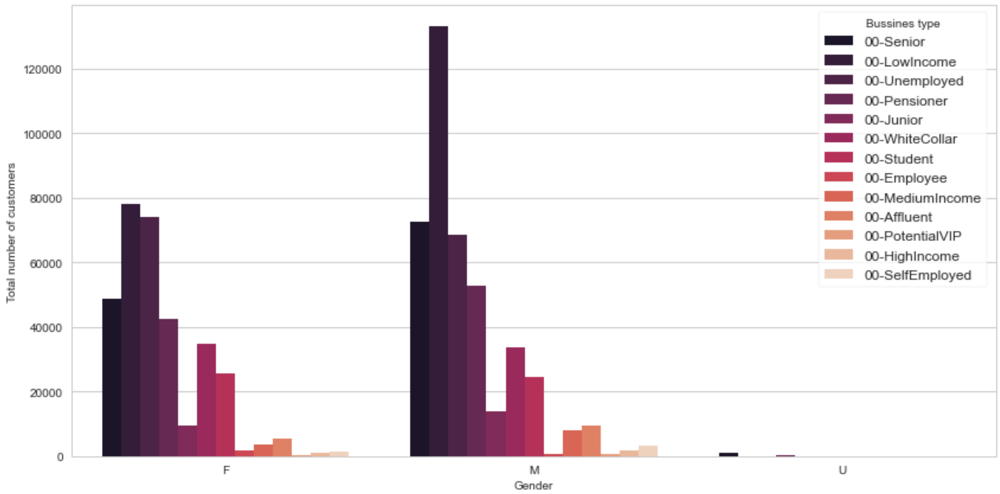
Figure 12: Age and gender of customers



Source: Own work.

Figure 13 showed gender of customers and business types.

Figure 13: Gender of customer with business types



Source: Own work.

Analyzed data from the account database was quite like product data, except it was seen from the account perspective. When investigated, date of opening accounts, it can be concluded that the date 2019-12-30 was extensive and many accounts (63.033 of them) closed on that date. Most probably, on that date, some systematic closing of inactive accounts occurred, as can be seen in Figure 14.

Figure 14: Reference date

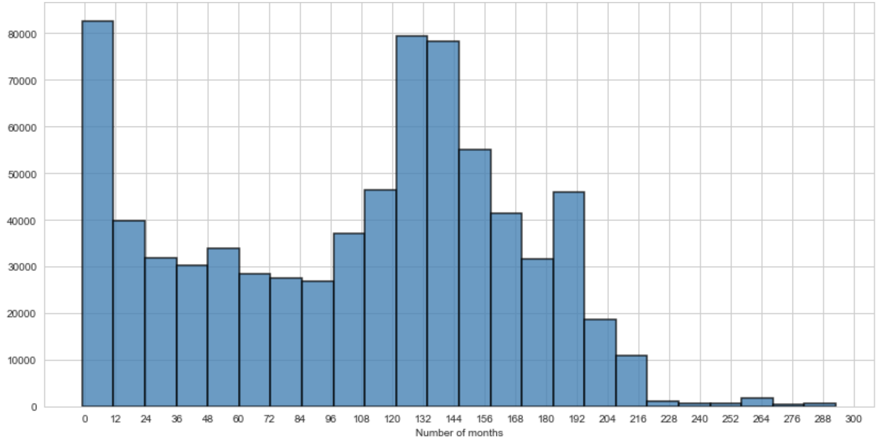
```
In [8]: accounts.groupby('REFERENCE_DATE').size()

Out[8]: REFERENCE_DATE
        2017-03-08      217
        2017-03-09      165
        2017-03-10      161
        2017-03-11      120
        2017-03-12      122
                       ...
        2019-12-25      588
        2019-12-26      475
        2019-12-29      645
        2019-12-30    63033
        2019-12-31      557
        Length: 1024, dtype: int64
```

Source: Own work.

Figure 15 showed the number of months between ref_max_date and creation_date. As mentioned before, the reference date was the date of loan approval, and it was taken from the contract table.

Figure 15: Number of months between ref_max_date and creation_date

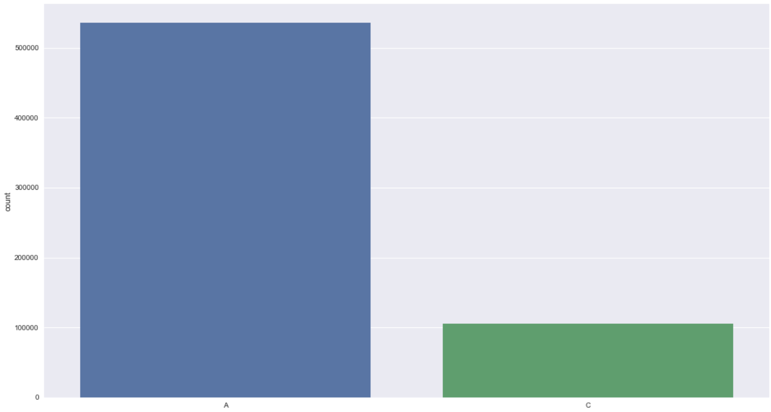


Source: Own work.

Presented in Figure 16, the relation between closed and active accounts gave a clear picture of accounts that were still active and could be targeted for models. There were 535.873 active accounts and 104.935 closed.

Understanding client retention is possible by comparing open and closed accounts. An increased attrition rate or customer discontent may be indicated by a high number of closed accounts. A higher proportion of open accounts, on the other hand, suggests a success for acquiring and keeping customers.

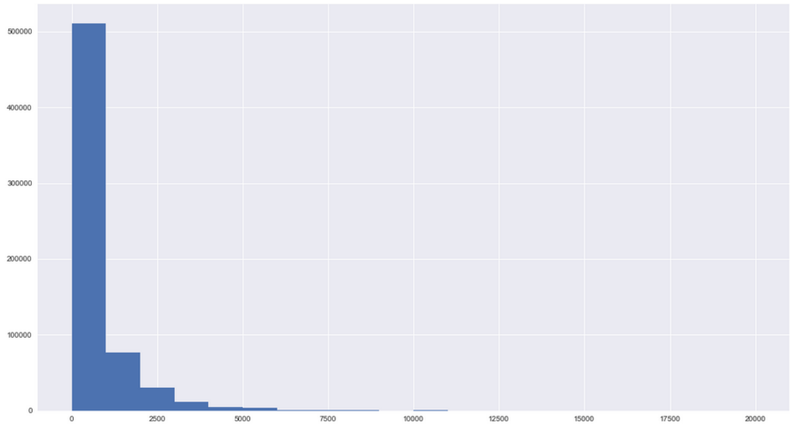
Figure 16: Status indicator



Source: Own work.

In Figure 17, I presented the amount of overdraft that employees had. An overdraft can indicate that the account holder is experiencing financial difficulties, however, depending on the data quality, it is also possible that the account holder simply committed an error in their account administration, therefore it is not always indicative of financial issues. Then, by analyzing the graph, we can evaluate the potential risk associated with the bank's exposure to overdrafts. Higher concentrations of customers with large overdraft amounts may indicate higher credit risk for the bank.

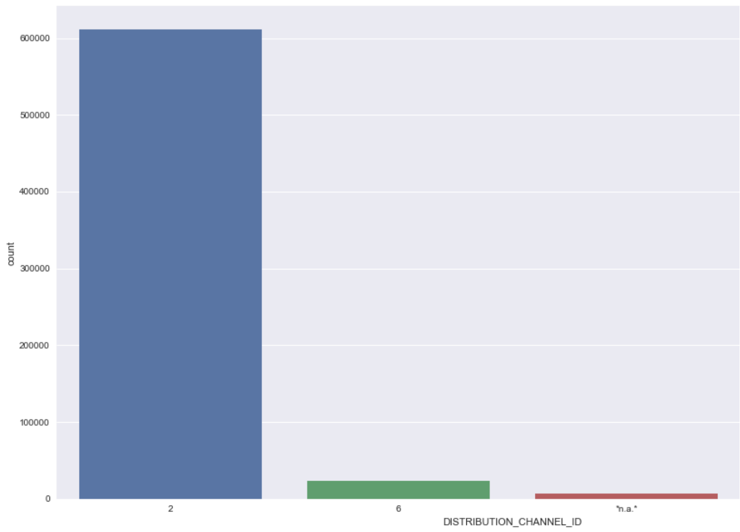
Figure 17: Overdraft amount



Source: Own work.

Regarding preferred channels for opening accounts, in Figure 18 it is presented that most clients opened accounts by visiting branches (code 2 is Branch) and a small number of them by direct sales agent (code 6 is direct sales agent).

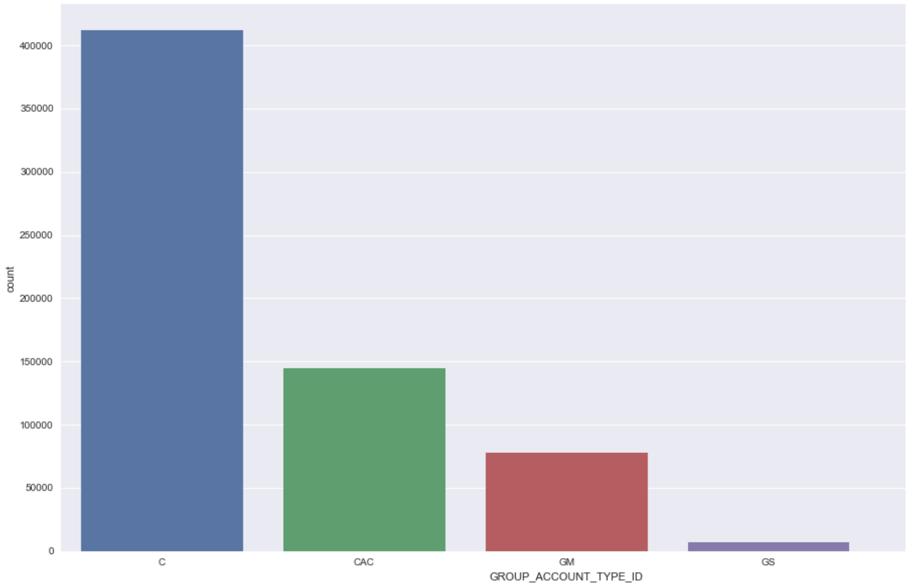
Figure 18: Distribution channel



Source: Own work.

When accounts are grouped, in Figure 19 it can be noticed that most of the accounts were current accounts (C) and current accounts for credit cards (CAC). Graph provides insights into which account types are more commonly chosen by clients and therefore indicates their preferences and banking needs. By analyzing the distribution, the bank can identify which types of accounts generate higher revenue or attract more desirable clients. The graph depicting the types of accounts that clients have opened within the bank offers insights into customer preferences, segmentation, profitability, and opportunities for product development. It can inform strategies to enhance customer satisfaction, build relationships, and optimize the bank's product and service offerings.

Figure 19: Account type



Source: Own work.

6.4 Data cleaning

Data cleaning helps to detect and remove any errors and discrepancies from data so that it is more reliable and accurate (Rahm & Do, 2000). Since poor data quality frequently produces inaccurate data analytics outputs and poor business decisions, it is one of the most significant issues in data management (Chu & Ilyas, 2016).

Data that was a source of this use-case was collected by different application modules, therefore not all data was of good quality.

Data quality should be ensured in applications that are manipulating the data, but if systems are old or poorly designed data quality will not be on a satisfactory level.

Data cleaning in data science represents a phase that will help to identify, remove, or fix data that is not relevant, wrong, or unusable in the use case. The process could be identified as data accuracy, data relevance or data correctness, but the main goal so this phase is to clean the data and make it usable for the use case. The good quality of data provides better business results and makes the use case more accurate.

If it is known that older applications as front-end systems can store data that is completely wrong, then data exploration is needed to identify data that is not relevant, missing, or bad quality. Some of the problems of low data quality that were identified in my dataset and later fixed were:

1. *Misspellings* - If data is coming from collaboration tools or some other sources typed by users without spelling corrections, for sure there will be grammar errors that can sometimes lead to wrong conclusions.
2. *Whitespaces in data* - Data where there are, due to human mistake, same values but in some cases, whitespaces are input accidentally (Rahm & Do, 2000).
3. *Blank data* - If there are blank spaces in series of data that is important for analysis that can reproduce issues in data analysis. Therefore, any blank data must be considered (Rahm & Do, 2000).
4. *Numbers represented as text* - Since, in data science use cases uses number functions, any data that is stored as text can potentially cause issues (Rahm & Do, 2000).
5. *Contradicting records* - In this use case it was important age of the clients, so if this data was not logical (e.g., 18-100) and there was data like negative numbers or >120 then that must be fixed or removed (Rahm & Do, 2000).
6. *Duplicated records* - Many times there were duplicated data in the datasets. Removing this data ensures a better quality of this case (Rahm & Do, 2000).
7. *Inconsistent formatting* - This anomaly, usually happening in cases where data is imported from different platforms, so unfirming this data is necessary (Rahm & Do, 2000).

Data cleaning can be done by various tools but should be done before data analysis. In the data cleaning process, depending on dataset size or type, SQL tools can be used, spreadsheet tools, programming languages like Python or R.

Since raw data was stored in RDBMS, in this case it was a combination of Python and SQL language. I used Jupyter Notebook to identify this data and SQL Developer to fix the problems.

Since, for this use case AGE argument was very relevant, in Appendix 4 I described detailed process of identifying and fixing „bad“age relevant data.

6.5 Model development

To achieve the best results and develop the most accurate model I have used Python programming language together with incorporated modules for machine learning and visualization of data that was previously prepared and stored in the Oracle database.

The following were technical prerequisites for model development: pandas, NumPy, functools, sklearn, optuna, shap, matplotlib, and seaborn. To develop a model that helps to determine clients that are eligible for offering personal loans, firstly I filtered raw data through couple of extensive SQL filtering procedures. To narrow data, I have used SQL language and grouped data into 5 different tables that were used in model development: customers, products, account balances, transactions, and target client table.

The customer table kept all data on customers, the product table had all data about products, and the transaction table had transactional data connected to these customers. At the end balances were also important to analyze and data could be found in the account balances table.

Figure 20: Tables used in model development

OWNER	TABLE_NAME	STATUS	NUM_ROWS
INALJUBDS	**INA_TRX**	VALID	854580
INALJUBDS	**INA_TARGET**	VALID	854580
INALJUBDS	**INA_PRODUCTS**	VALID	856842
INALJUBDS	**INA_PL_ML_SA_CON_BALANCE**	VALID	857580
INALJUBDS	**INA_CUSTOMER**	VALID	768670
INALJUBDS	PRODUCT_CODES	VALID	3203
INALJUBDS	INA_RF_FINAL_DATA	VALID	55775
INALJUBDS	TEMP_CUSTOMER_PL	VALID	842631
INALJUBDS	TEMP_CONTRACT_PL	VALID	960829
INALJUBDS	**INA_CA_CAC_BALANCE**	VALID	855980

Source: Own work.

After the connection to the database was established, selected tables that were needed for model development were loaded. By loading this data (customer, product, transactions, accounts, and target group), a precondition to develop the model further is set. After tables were loaded separately, it was needed to merge all data into one dataset.

The next step was to calculate the growth rate and ratio of transactions, by analyzing sets of different transactions grouped over 3 and 6 months.

In the next step customers who have had 0 transactions in the past 3 months were removed, because that data was not needed for the model.

To have the possibility to use different attributes in the model, columns to be considered in modeling were defined as “Initial iteration 82 features”, “Mid iteration 51 features” and “Final iteration – 32 features”. These predefined columns were added to model calculations. An additional check of newly merged data regarding target groups now looked like this: 0 – 190600 and 1 - 202606.

Firstly, x any y data set was defined, then train and test data were split by using train_test_split, and then “Standard skaler” (standardizing features by removing the mean and scaling to unit variance) was defined. Defining the objective was very important and this step was done by defining rf_n_estimators, rf_max_depth, and rf_min_samples_leaf.

Every estimator had a score method after being trained on the data. So, when the score was called on RandomForestClassifier, the method computed the accuracy score by default.

The new study was defined using optuna_create_study and n_trials are set to 100, where optuna is optimization library in Python that could be employed for hyperparameter optimization.

Of all the trials, when algorithm went through the entire model for this metric, the results showed that the best trial was number 73.

6.6 Limitations of NBO

Firstly, only the clients for whom the company has sufficient data could be reached using the NBO model therefore it is a very targeted approach. Usually, filtering of the data is done in the process of model development, later this helps to have bigger correctness of the results.

Secondly, the significance of realizing that NBO model projections are not accurate facts and that all models have their flaws. The tendencies are that with every new version of the model, the model itself will be more accurate.

Thirdly, practice showed that despite being an NBO recommendation, the clients weren't always interested in the suggested products, but a personalized approach is giving them a sign that someone is taking care of their needs. Usually, this fact has a very positive impact on NPS (Net Promote Score).

The next limitation is the data quality, since such analysis always identifies wrong data that cannot be used in the analysis. This can improve data quality because then it can be communicated for better controls to be installed to better quality data at the entrance to the system.

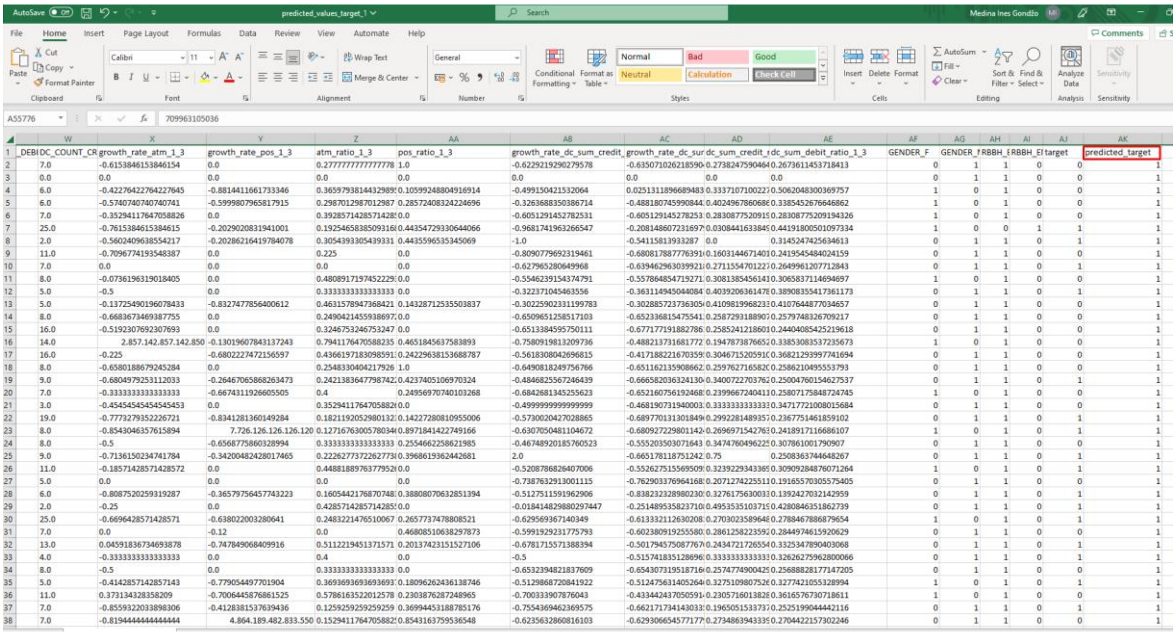
NBO data quality tackles a well-known problem of data quality. Data quality must be ensured at the very beginning of data collection, in applications. Some legacy systems are not built with enough level of data quality controls.

6.7. Final findings

For the example of NBO model development which I showed in my master's thesis, the result was an Excel table with customers to whom a bank should offer a personal loan. A personal loan is a type of loan that a person takes as an individual, and not as a company.

Otherwise, there were many more clients, and this was the result of only those to whom it was necessary to offer personal loan. Target is 1 in the predicted_target column for 55.776 clients, as shown in Figure 21.

Figure 21: Excel document - Predicted values



Source: Own work.

The most crucial component of the model is its reusability. The model built into thesis is reusable, which means that it can be applied to other sets of data as well, allowing the bank to rapidly apply the already created model to the data from a different period.

To be able to do all that, it is necessary to have in place data governance, data approach and data architecture, where data governance is the way in which data is managed, data approach is the way of exploiting that data for greater income, and data architecture refer to how the data will be stored in a technological manner. Those are different processes that are defined to pass the data through all the data pipelines and place them in usable data lakes.

After identifying a group of individuals who are eligible for a product personal loan, I investigated information about them in more detail (see in Table 2). More in-depth research could be beneficial for future studies, to see what the structure of these people is, how much they used bank products before, how old they are, what gender they are, etc.

Table 2: Findings from data analysis

Segment	**Finding**	**Explanation**	**References**
	Bank clients are too mature	Potential of the customers can be challenging in the future, because strategically	Own work – Figure 9

		it is important to have certain number of younger clients to maintain “fruit-bearing”.	
	To many clients in middle age category	Client base is concentrated in buckets between 29- and 44-year-old clients. Even do these are clients with potential, it would be good to have more clients in upper category (if we take into consideration that, up to that age they already achieve some basic goals and they are ready to invest into the future)	Own work – Figure 10
Structure of the client’s base	Business type overload with “Low income” clients	Big number of clients with low income will mean that it will be challenged to offer right product or to sell something extra on top of current account.	Own work - Figure 11
	Business type overload with “Unemployed” clients	This category is “No go” for NBO uses cases.	Own work - Figure 11
	Business type overload with “Pensioners” client’s category	On top of the upper findings on client base, there is the fact that Bank has to big number of “Pensioners”. Since NBO is focused on personal loan it will be very challenging to offer this product to this group of clients.	Own work - Figure 11

	“Senior” category of the clients is suitable for NBO	Solid number of Senior clients can be playground for offering personal loan	Own work - Figure 11
	Higher number of male clients	Even do gender structure is in favor of man, in context of seniority there is not so much difference	Own work - Figure 12
General Account data	Closing of accounts	High number of accounts closed on particular date means that Bank has automatic closing procedure in place	Own work - Figure 16
	Previous loan duration	Most of the privates’ loans are taken on period of 10-13 years	Own work - Figure 16
	Active accounts	75% of all accounts are active	Own work - Figure 16
	Account types	Most of the account are current account and current account for credit card, this indicates that bank have very low number of savings accounts in their portfolio	Own work - Figure 19
	Overdraft as a product has potential	Low number of clients have overdraft, this can be potential for NBO (product - overdraft)	Own work – Figure 17
Transactions data	Channels – transactions	Branch channel is still most preferable by clients	Own work - Figure 18

Final model results	Eligible clients selected	After model development, result is list of 55.000 clients that are eligible to be offered with personal loan product.	

Source: Own work.

After identifying a group of individuals who were eligible for a product personal loan, I investigated information about them in more detail. More in-depth research can be beneficial for future studies, to see what the structure of these people is, how much they used bank products before, how old they are, what gender they are, etc.

In Figure 22 where 1 were males and 0 were females, there were more male than female bank’s clients. It could imply that the bank is particularly good at acquiring male customers, or that the bank's services or products are more enticing to men. It can be feedback for part of the marketing in the bank, to work on attracting more women.

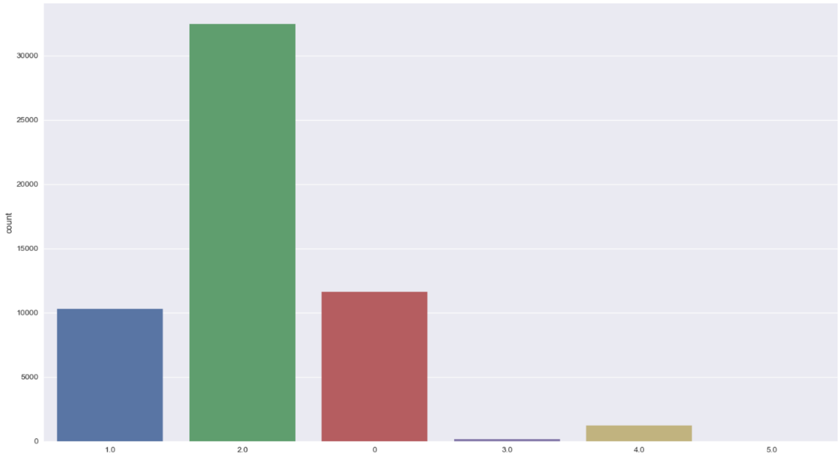
Figure 22: Gender structure of customer

Source: Own work.

The graph in Figure 23 shows how many loans these clients had in the past at the current bank. For the majority this was not their first loan, as they had already taken out some type of loan. If a client has a significant number of past loans and has struggled to repay them, that can indicate high risk for banks.

On the other side, the quantity of past loans taken out by a client can indicate their creditworthiness and chance of repaying a new loan therefore, clients who have a track record of successfully repaying loans are regarded as less risky.

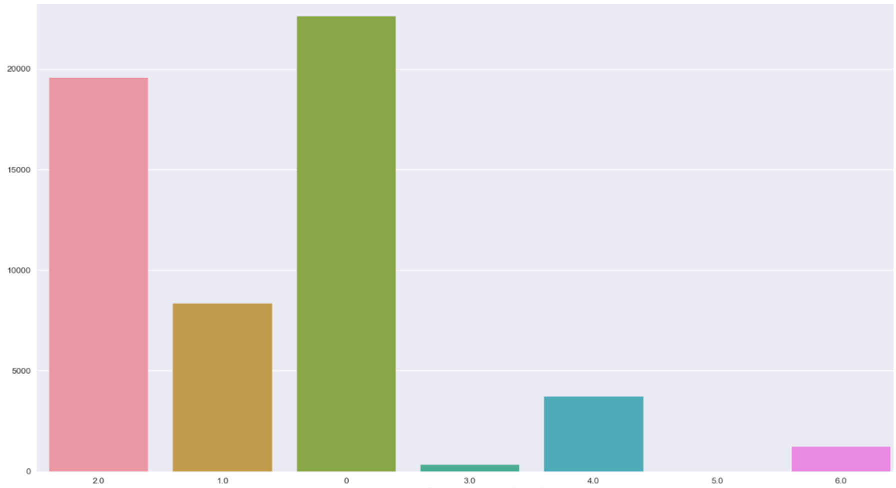
Figure 23: Number of personal loans in past



Source: Own work.

Figure 24 presented the number of current accounts for credit cards. Usually, a low number of credit card current accounts can tell that clients are less reliant on credit to manage their money and would prefer to use cash or debit cards for their transactions. A high number of credit card current accounts may indicate that the clients use credit cards frequently and are more inclined to use credit to make purchases or manage their money.

Figure 24: Number of Current Account for Credit Card



Source: Own work.

7 INTERVIEW RESULTS

In this chapter I presented results from interviews with bank employees regarding NBO model development. To increase the credibility of this study, the interviews were carefully chosen and included workers from the case organization who were most informed about the research issue. Additionally, it was beneficial talking with employees before and after NBO model development.

8.1 Interview with "Data Engineer"

Before implementation of NBO model

The Data Engineer in the bank was responsible for building systems that converted raw data into usable information. The employee was employed in the bank for 7 years, previously in the position of Database Administrator and has a lot of experience regarding data.

Main priorities for his team were uptime and accessibility of systems that were used by Data Scientists and IT Business Analyst. Additionally, the priority for Data Engineer was also building new, usable data sources for DS. Data Engineer was involved in building the NBO model as a part of an agile team that was responsible for developing NBO model.

The main goal for Data Engineer in this project was to enable enough quality data sources for DS and tight collaboration with Database Administrators. The goal was to develop a successful model that would enable a bank to offer new products to the right customers.

After implementation of NBO model

The main challenge from the Data Engineer point of view was data quality. Since Data quality is something that should be ensured at the application level (older applications do not have efficient data quality controls) hard work was required to fix this problem in later phases. Also new technologies like cloud services were a challenge. The lesson learned was that they must develop applications from the beginning with data quality and security in mind.

They overcame it successfully, but some data was excluded from the process because of bad quality. The goals were achieved when the whole team managed to produce a list of the clients and their next best offers. Performance was measured in two ways, first time of repeating whole process with different set of data, and second number of NBO accepted by clients.

8.2 Interview with "Head of IT Department"

Before implementation of NBO model

The head of IT in the bank was responsible for all IT people and their engagement in the agile teams. With 25 Years in IT and 10 years in the bank, the employee was very familiar with data architecture and data processes in the bank. His priority was to enable the IT and business agile teams to work without obstacles and to solve resource and budget problems of the team. As Manager and one of the most responsible people in the bank his goal was to get the model running on time and during that process to ensure reusability for different data sets.

After implementation of NBO model

For the Head of IT there was only one challenge related to the people. Since the IT labor market is very challenging, managing IT people's attrition was the most important task. During this period, he learned that keeping happy and engaged people in the team was the best way to achieve results. Also, that retention was always a better solution than onboarding new people on projects that are in progress. He was very satisfied with the results, especially with the reusability of project phases for other projects.

8.3 Interview with "Database Administrator"

Before implementation of NBO model

Database Administrator was the person that was responsible for maintaining databases (Oracle, SQL Server, Postgres, MySQL) in the organization. His role in NBO agile team was the role of a domain expert. He is experienced IT DBA with 6 years' experience in the bank and prior to those 10 years in other financial institutions.

His task was to help the team, especially the Data Engineer, to have access to the required data and create efficient RAW data sources. Since DBA knew data architecture and where data is, the main goal for him was to enable creation of valid data sources by helping to identify and find data.

After implementation of NBO model

The DBA challenge was to identify and find data that was needed for the model, but also to write efficient queries, since it was a large amount of data. Since Data Quality was in some cases poor, the team decided to exclude some data. All DBA goals were achieved.

Therefore, when it comes to challenges that employees from bank were facing, it must be underlined that, since data analytics and data science are relatively new to the organizations like banks or small/large corporations, and that there was a lack of:

- Knowledge of business and technical side
- Data strategies and data governance
- Processes that define how data was handled
- Operating models / Organizational structures

Knowledge of data topics was essential for a successful implementation of the CRM use cases. It can be said that technical knowledge was even easier to get, if not internally, from the market by onboarding people that were skilled enough to develop models that businesses need.

On the other side, the need for solid business knowledge that is having foundation in CRM but also "data" topic is sometimes underestimated. Organizations should continuously build know-how on data topics, focused on business segment where it will be used. This can be achieved by planning training and education of employees that will be part of teams that are building data use cases.

Also, it is important to raise awareness among the people and across organizations on data topic importance. This is especially important in the IT segment because data quality, which is essential for quality, depends on the input of data in applications. To be clear, essential for building data analytics use cases that will improve our business, is to have as much as we can controls on the first input of data into our systems. This means that awareness, among IT people, who are developing applications for capturing data should be at a high level from the beginning.

Data strategies and governance were also big challenges, since most of the traditional organizations do not have any strategies regarding use of their data or some governance bodies that will take care on whole process of managing and exploitation of data. Organizations should build, based on their business needs, governance of their data process. Even short-term strategies will help to start managing this promising source of business ideas that can boost our business results.

In parallel, organizations should have defined processes, roles, and responsibilities in managing data topics. This process should be built and practiced in a way that covers the whole life cycle of data across information systems and at the end data models. If there is a good process in place, then it can be measured and potentially improved. Then it is easier to set up KPIs that will make the management of data better during the time of data model life cycle.

In the end, a big challenge was also the **organizational structure** that will bring a defined approach into production and keep it alive and useful for improving business. Most of the organizations that are starting with the exploration of data topics do not have teams to engage on this task. Even if they have teams usually it is not clear where these teams belong organizationally wise.

Since most tasks are technology related, companies are trying to build these teams in IT departments. This is not good practice since management wants in, in the end, to use data in the direction of improving business and customer relations, which are business topics. Today it is already clear that in organizations that are not IT related and where data science is not their core business, like banks are, it is necessary to build these structures in related business departments and use IT knowledge to bring more business value into organization.

Since these teams are dependent on data pipelines and a lot of work must be done prior developing data models, it is not easy, especially in big organizations, to make the operating model that is effective enough. In these cases, the best practice is to create small agile/adaptive teams in business departments with enough knowledge to work on data topics, but also to trough defining governance and processes, make agreements with other departments like IT, finance, risk etc.

8 DISCUSSION

Three issues can come up while discussing data analytics for CRM. The first question is whether banks have data they can utilize for CRM. Data governance, strategy, and architecture offer a solution to the problem of not all data being available since they are in out-of-date systems. As a result, the current systems require some improvement. Modern data pipelines must be established, and all required data must be brought into data lakes.

Second, do they have the procedures and expertise to put it into practice? Both the staff and the senior management are essential in this situation. Top management must be able to set objectives, decide what the business needs to do and how doing so would increase revenue and clientele, as well as to explain and communicate these ideas to staff members. To implement it, the staff must have the necessary knowledge and expertise.

Third is what are the obstacles and risks, and on the other side, what are the opportunities on that path. After the conducted interviews, bank employees stated challenges and opportunities that they experienced along their journey, which is the answer to my research question of what the opportunities and barriers are to using data analytics for CRM in the financial industry.

Lack of knowledge on the business and technical side, creation of data approach and governance and definition of the process are big challenges for financial institutions.

The cost savings, relationship quality (better relationships), and transaction quality are some of these variables (better transactions) are three areas of effectiveness for CRM adoption.

Conclusion can be drawn that a company's ability to interact, respond, and communicate with existing consumers more successfully allows company to significantly increase maintenance rates and revenue.

analyze data that I was provided by a selected bank, and present findings. Findings summarized in the exploration segment are presented in the table below. The result of this model is a list of the clients that are eligible for the product “Personal Loan”.

When it comes to practical contributions, I analyzed data that I was provided by a selected bank and presented my findings. Findings summarized in the exploration segment are presented in the table below. The result of this model is a list of the clients that are eligible for the product “Personal Loan”.

Regarding theoretical contributions, the first thing is pointing out to the existence of a large amount of data and importance of having a method on how to use existing data to increase profit, and using different opportunities to give clients better service or personalize offers. Banks are again at the beginning of this journey, after the COVID crisis. Then it became clear to companies that work will never be the same. Goods and services that clients require must go through refactoring due to new trends in the business market.

After I went through the literature review, I identified a gap regarding challenges that banks are facing from real life companies, to be described in more detail. In chapter 8, after providing results from interviews I summarized which challenges banks are facing while using data analytics for CRM.

And third, through my thesis I underlined in several parts that if cooperation between the departments was not at a good level, organizational changes are necessary to overcome this problem. It is not enough to have vision and even knowledge, but also organization culture, adjusted operating model and mindset of the people oriented towards cataloging, governance and in the exploitation of the data. Organizational changes are inevitable and adjustment of the operating model towards agile teams that are handling data topic end to end.

Advice is given for banks that are considering the implementation of analytical CRM. Other banks might use the insights gained from the analysis of the gathered data to enhance the efficiency of their banking services.

Implications and advice for banks thinking about implementing analytical CRM are:

1. Banks should devote time to strategic planning to determine their needs and difficulties. Therefore, prior to putting in place an analytical CRM system, businesses should identify clear and specific objectives.
2. Interact with both customers and staff. Moving focus from items to consumers will lead to the creation of lasting, mutually beneficial relationships with all parties involved.
3. Create a reliable infrastructure for departmental communication. Spend money on essential elements like analytical tools and data warehouses. The correct technology should be purchased by the bank to assist it with the analytical CRM system.
4. Create a unified customer view by integrating front-end systems with back-office data mining procedures. Use just the information that is pertinent to your business's needs.
5. Create a central data warehouse for storing and analyzing both new and old data. Make several data models as a solution. Use only data of the highest caliber.
6. Employ cluster analysis to unearth fresh customer insights. Predictive modeling can be used to increase client retention.
7. Create, test, and use analytical models.

8. Establish precise, quantifiable goals for each step while keeping an eye on ROI.
9. To help with future improvements, learn from past marketing failures and accomplishments.
10. Examine the client database.
11. Encourage consumer loyalty to boost profits. Identify more opportunities and client needs.
12. Deliver the same level of service across all channels and measure performance. Clearly defining KPIs and routinely monitoring and analyzing performance data are required for this. The data should be used by banks to continuously enhance the system over time.

CONCLUSION

When it comes to **meeting the goal of the thesis**, starting with the first goal of this thesis was exploring literature and giving a better perception of how data is used for Customer Relationship Management in the banking industry. The goal was met by analyzing concept of CRM in detail through the existing literature in general and then in the narrowed sense, for the banking industry.

The second goal was to identify, describe and provide a better understanding of CRM data-driven cases that are present in the banking field. When talking about opportunities for exploiting the existing data in the CRM segment it can be said that, since a large amount of raw data is stored through different channels in the bank, the usability of that data in CRM especially (because CRM is responsible to manage clients and their needs) is on very high level. There is a big opportunity for bringing more profit to the bank, by analyzing the existing data in context of client behavior. Use cases like: NBO, Customer 360 etc., can bring additional value to the company. If a bank can identify customers that are eligible to be offered some specific product based on understanding their needs and develop data analytics use cases that will be reusable, then these organizations can achieve competitive advantage. By developing these models, they can have a personal approach to every single client. CRM in essence is taking care of client needs and the best way to achieve it is to know exactly what the client needs.

The third goal was to analyze the challenges, and opportunities for CRM as well as the process of implementing the NBO model.

As demonstrated in this master thesis, there are lots of opportunities as well as challenges that come with the use of data analytics for CRM in one financial institution. The development of an analytical CRM needs to be carefully planned out and supported by the top management. It is already known that banks own enormous amounts of data but its usability in the right way is not always easy to determine. Through my master’s thesis and conducted interviews with employees that are involved in data-related work in a chosen bank and demonstrated that they have know-how for data management.

So, firstly I stated opportunities and challenges in CRM implementations and then through conversations with the employees, I stated what were opportunities and challenges of NBO model development. Considering that technology is progressively going forward, today's banks have a chance like never before to transform themselves into technology aware companies. This transformation is needed in most organizational units within the banks, especially in CRM. CRM departments can have benefits in using data for data driven decisions. Of course, banks should bring all raw data, collected in applications to a kind of data lake to be usable. This technical challenge should not be underestimated.

A lot of today's banks do have old legacy systems that were built at a time when tools were not so advanced and data science was at a very low level. This means that they should start with digital transformations of their data infrastructure. Also, integration of their new systems is facing challenges to integrate with existing legacy systems. Solving these technical challenges and smart use of these data will bring strong relations with the client, improve staff productivity, drive new business opportunities etc.

However, with the use of data in data driven organizations banks are facing a big responsibility. Questions like data security, data ethics etc. should be considered and handled with care. Not talking enough about these questions can have reputational risk for the bank. Misuse of client data or excessive exploitation of the data should not be an option for the banks. Therefore, this thesis addresses the importance of GDPR and data quality, especially because it is a bank in question and customer data must be protected and anonymized if they are used for some research, as it is the case in my thesis.

Additionally, I presented the current situation of CRM and data analytics practices inside the company. The maturity of CRMs in the banks is not equal to the opportunities that technology brings to the table. In that context it can be argued that banks must invest more into CRM departments especially in the knowledge of their employees, processes and organizational changes. The stronger the CRMs are client satisfaction is at a higher level. A satisfied customer is loyal and stays with the bank more, willing to pay more, use more services and bring more income to the bank.

When it comes to **limitations of the research**, since the case organization is distinctive and has a distinct structure, organizational culture, and a set of competencies, the research finding is not generable. The research has explained areas of opportunities for data analytics for CRM that require more investigation in different industries to acquire a more complete understanding of the subject.

Firstly, data availability is limited because data in the banking industry is frequently difficult to obtain, which might limit research. I was given data for a specific time, and it would be good to evaluate data for additional years, therefore time limits are a limitation. However, access to banking information is extremely difficult because financial firms must ensure that customer data is used fairly and lawfully. Therefore, when studying another organization, the interview questions and findings may differ. This case study research was conducted in a financial industry case organization, implying that the implementation method may differ in companies functioning in other industries.

Next, the size and structure of the organization imposes its own constraints on this study. Regulations for NBO and customer data may differ depending on whether a country operates in one or more countries.

The first area for **future research** would be determining how to calculate the return on marketing investment when the NBO model is used in combination with another targets. A future study opportunity would be to repeat this research in a different size organization. It would be good to evaluate whether the same challenges and practices apply to larger enterprises as well. It would be extremely beneficial for both firms that already use a next best offer model and those that are considering deploying the model to better understand the testing and development process. Some further research could investigate this topic more thoroughly.

Next, the area for future research could be the exploration of data analytics on customer loyalty in the banking industry. For example, the likelihood of a customer remaining if the suitable product is offered at the appropriate time.

The use of data analytics in enhancing client contentment in the banking industry. Today's clients are more and more demanding, they expect to have a service that is fast, reliable, and personalized. Organizations that succeed in offering clients what they need, fast, at reasonable prices and with a personal touch can have a competitive advantage. Therefore, understanding how data analytics could improve customer segmentation can be one of the major further research suggestions as well. By analyzing customer exposure and maturity, advanced analytics cases can be used to segment clients and prepare this data for other organizational units in the bank (for example RISK department).

REFERENCE LIST

1. Abrar, M., Biag, S. A., Shabbir, R., & Hussain, I. (2019). Role of CRM Practices in Bank's performance: Moderating role of Market Turbulence. J*ournal of Managerial Sciences*,*13*(3), 188-200.
2. Abbas, T., & Hafeez, S. (2017). Impact of CRM practices on service quality in the banking industry. *Pakistan Administrative Review*, *1*(2), 130-144.
3. Aggarwal, C. C. (2015). *Data mining - The Textbook.* New York: Springer International Publishing.

4. Alam, M. M. D., Al Karim, R., & Habiba, W. (2021). The relationship between CRM and customer loyalty: the moderating role of customer trust. *International Journal of Bank Marketing*. *39*(7), 1248-1272

5. Al-Dmour, H., Saad, N., Basheer Amin, E., Al-Dmour, R., & Al-Dmour, A. (2021). The influence of the practices of big data analytics applications on bank performance: Filed study. *VINE Journal of Information and Knowledge Management Systems. 53*(1), 119-141.

6. Appiahene, P., Missah, Y. M., & Najim, U. (2020). Predicting bank operational efficiency using machine learning algorithm: comparative study of decision tree, random forest, and neural networks. *Advances in fuzzy systems*, *2020*, 1-12.

7. Bahari, T. F., & Elayidom, M. S. (2015). An Efficient CRM- Framework for the Prediction of Customer Behaviour. *Procedia Computer Science*, 46, 725–731.

8. Bahrami, M., Ghorbani, M., & Arabzad, S. M. (2012). Information technology (IT) as an improvement tool for customer relationship management (CRM). *Procedia-Social and Behavioral Sciences*, 41, 59-64.

9. Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.

10. Bhambri, V. (2012). Data mining as a tool to predict churn behavior of customers. *India: International Journal of Computer & Organization Trends, 2*(3), 85-89

11. Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197-227.

12. Bihari, D. S. C., & Murdia, M. (1970). CRM in Banking Sector with special reference to New Age Banks. *The Journal of Internet Banking and Commerce*, *18*(1), 1-19.

13. Boyce, C., & Neale, P. (2006). Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input, *Monitoring and Evaluation,2,* 1-12.

14. Breiman, L. (1999). Random forests-Random Features. *Technical Report 567*, Statistics Department, University of California,Berkeley, 1-29.

15. Britto J., M., & Gobinath, R. (2020). REHC: Reduction through exclusive homogeneous clusters for imbalance dataset. *Materials Today: Proceedings*, S2214785320386946.

16. Brodsky, L., & Oakes, L. (2017). Data sharing and open banking. *McKinsey & Company*, 1097-1105.

17. Carbo-Valverde, S., Cuadros-Solas, P., & Rodríguez-Fernández, F. (2020). A machine learning approach to the digitalization of bank customers: Evidence from random and causal forests. *Plos one*, *15*(10), e0240362.

18. Caron, M. S. (2019). The Transformative Effect of AI on the Banking Industry. *Banking & Finance Law Review*, *34*(2), 169–214.

19. Coltman, T. (2007). Can superior CRM capabilities improve performance in banking? *Journal of Financial Services Marketing*, 12(2), 102–114.

20. Chen, I. J., & Popovich, K. (2003). Understanding customer relationship management (CRM): People, process and technology. *Business Process Management Journal*, *9*(5), 672–688.

21. Chitra, K., & Subashini, B. (2013). Data mining techniques and its applications in banking sector. *International Journal of Emerging Technology and Advanced Engineering*, *3*(8), 219-226.

22. Chye, K. H., & Gerry, C. K. L. (2002). Data mining and customer relationship marketing in the banking industry. *Singapore Management Review*, *24*(2), 1-28.

23. Chuang, C. C., & Hu, F. L. (2014). Application of CRM in banking. Actual Problems of Economics (154), 419-427.

24. Cvijović, J., Kostić-Stanković, M., & Reljić, M. (2017). Customer relationship management in banking industry: Modern approach. *Industrija*, *45*(3), 151-165.

25. Dubey, N. K., Sharma, P., & Sangle, P. (2020). Implementation and adoption of CRM and co-creation leveraging collaborative technologies: An Indian banking context. *Journal of Indian Business Research*, *12*(1), 113-132.

26. Ferrara, P., & Spoto, F. (2018, February). Static Analysis for GDPR Compliance. In *ITASEC.1*, 1-10.

27. Gayathry, S. (2016). Customer relationship management model for banks. *Journal of Internet Banking and Commerce*, *21*(S5), 1-12.

28. Gončarovs, P. (2017). Data analytics in CRM processes: A literature review. *Information Technology and Management Science*, *20*(1), 103-108.

29. Gupta, S. (2019). A Study on Customer Relationship Management in The Banking Industry. *Think India Journal*, *22*(10), 2051-2058.

30. Han-Yuh, L. (2007). Development of a Framework for Customer Relationship Management (CRM) in the Banking Industry. *International Journal of Management*, *24*(1), 15–32.

31. Hassani, H., Huang, X., & Silva, E. (2018). Digitalization and Big Data Mining in Banking. *Big Data and Cognitive Computing*, *2*(18), 1-13.

32. Ho, T.K., (2014). Knowledge from patterns: from the mail sorting facility to a world of connected devices. *The International Association for Pattern Recognition Newsletter*,*36*(1), 5-7.

33. Jahnavi, M., & Katyayani, J. (2014). DATA MINING APPLICATIONS IN BANKING AND FINANCIAL SECTORS. *CLEAR International Journal of Research in Commerce & Management*, *5*(1), 6-10.

34. Khan, A., Ehsan, N., Mirza, E., & Sarwar, S. Z. (2012). Integration between Customer Relationship Management (CRM) and Data Warehousing. Procedia Technology, 1, 239–249.

35. Khirallah, K. (2001). CRM Case Study: The Analytics that Power CRM at Royal Bank of Canada. *Research Note, TowerGroup, Needham, MA*, 1-12.

36. Kim, H.-S., Kim, Y.-G., & Park, C.-W. (2010). Integration of firm's resource and capability to implement enterprise CRM: A case study of a retail bank in Korea. *Decision Support Systems*, *48*(2), 313–322.

37. Kumar, V., & Reinartz, W. (2018). *Customer relationship management concept, strategy, and tools* (Third). Springer.

38. Kulkarni, V. Y., & Sinha, P. K. (2012). Pruning of random forest classifiers: A survey and future directions. *International Conference on Data Science & Engineering (ICDSE)*, 64-68.

39. Laketa, M., Sanader, D., Laketa, L., & Misic, Z. (2015). Customer relationship management: concept and importance for banking sector. *UTMS Journal of Economics*, *6*(2), 241-254.

40. Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons.

41. Lau, K. N., Wong, S., Ma, M., & Liu, C. (2003). 'Next product to offer'for bank marketers. *Journal of Database Marketing & Customer Strategy Management*, *10*(4), 353-368.

42. Macias-Lizaso G. (2018). Building an effective analytics organization. *McKinsey on Payments*, *11*(28), 60-67.

43. Mahalakshmi, V., Saravanaraj, M. G., & Umarani, T. S. (2013). The influence of demographic variables of bank customers towards CRM. *International Journal of Research in Business Management (IJRBM)*, *1*(2), 27-40.

44. Moin, K. I., & Ahmed, D. Q. B. (2012). Use of data mining in banking. *International Journal of Engineering Research and Applications*, *2*(2), 738-742.

45. Nitsche, M. (2002). Developing a truly customer-centric CRM system: Part one—Strategic and architectural implementation. *Interactive Marketing*, *3*(3), 207-217. Nitsche, M. (2002). Developing a truly customer-centric CRM system: Part Two — Analysis and campaign management. *Interactive Marketing*, *3*(4), 350–366.

46. Nure, G. (2018). Customer relationship management in banking system (Case of Kosovo). *Journal of Internet Banking and Commerce*, *23*(1), 1-7.

47. Ogbadu, E. E., & Usman, A. (2012). Imperatives of Customer Relationship Management in Nigeria Banking Industry. *Kuwait Chapter of the Arabian Journal of Business and Management Review*, *2*(1), 59–72.

48. Osarenkhoe, A., & Bennani, A.-E. (2007). An exploratory study of implementation of customer relationship management strategy. *Business Process Management Journal*, *13*(1), 139-164.

49. Panda, T. K. (2001). Creating customer lifetime value through effective CRM in financial services industry,2, 157-171.

50. Patil, J. S. (2020). Customer Analytics Improves Banks Performance: A Descriptive Study of Analytical CRM among Private, Cooperative and Nationalized Banks. *Test Engineering and Management*, 10464 - 10470.

51. Patil, P. S., & Dharwadkar, N. V. (2017). Analysis of banking data using machine learning. *International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 876–881.

52. Pascu, A. I. (2018). Data Mining. Concepts and Applications in Banking Sector. *Annals of Constantin Brancusi'University of Targu-Jiu, Economy Series*,*1*, 159-166.

53. Payne, A., & Frow, P. (2006). Customer relationship management: from strategy to implementation. *Journal of marketing management*, *22*(1-2), 135-168.

54. Pokharel, B. (2011). Customer relationship management: Related theories, challenges, and application in banking sector. *Banking Journal*, *1*(1), 19-28.

55. Peppers, D., & Rogers, M. (2004). *Managing customer relationships: A strategic framework*. John Wiley & Sons.

56. Perwej, A. (2010). Effective Management of Customer Relationship Management (CRM) In Banking Industry. *YOJNA The Management Journal of KITE Group*, *2*(1). 1-5.

57. Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull*, *23*(4), 3-13.

58. Ramaj, A. (2015). Customer relationship management (CRM) in banking sector in Albania. E*uropean Journal of Economics and Business Studies Articles*, Revistia Research and Publishing, *1*(2), 136-143.

59. Ranjan, J., & Bhatnagar, V. (2011). Role of knowledge management and analytical CRM in business: Data mining-based framework. The Learning Organization, 18(2), 131–148.

60. Saxena, R., & Khandelwal, P. (2011). Exploring customer perception and behavior towards CRM practices in banking sector: An empirical analysis, *5*(9), 375-391.

61. Soltani Delgosha, M., Hajiheydari, N., & Fahimi, S. M. (2020). Elucidation of big data analytics in banking: A four-stage Delphi study. *Journal of Enterprise Information Management*. *34*(6), 1577-1596.

62. Statut of XYZ bank. (2019).

63. Stevanović, S., & Gavrilović, Z. (2018). Effects of the Application of the CRM Concept to Profitability of the Banks. *Economic Themes*, *56*(3), 283-299.

64. Shum, P., Bove, L., & Auh, S. (2008). Employees' affective commitment to change: The key to successful CRM implementation. *European Journal of Marketing*, *42*(11/12), 1346–1371.

65. Srivastava, U., & Gopalkrishnan, S. (2015). Impact of big data analytics on banking sector: Learning for Indian banks. *Procedia Computer Science*, *50*, 643-652.

66. Suri, T., Singh, S., Singh, T., Rajappan, A. K., & Florence, Dr. S. M. (2020). An Effective Method to Understand Bank Customer Retention System. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(6), 4279–4283.

67. Tsiptsis, K. K., & Chorianopoulos, A. (2011). *Data mining techniques in CRM: inside customer segmentation*. John Wiley & Sons, 1-351.

68. Voican, O. (2020). Using Data Mining methods to solve classification problems in financial-banking institutions. *Economic Computation & Economic Cybernetics Studies & Research*, *54*(1), 159-176.

69. Wang, Y., & Feng, H. (2012). Customer relationship management capabilities: Measurement, antecedents, and consequences. *Management Decision*, *50*(1), 115-129.

70. Wibisono, O., Ari, H. D., Widjanarti, A., Zulen, A. A., & Tissot, B. (2019). The use of big data analytics and artificial intelligence in central banking–An overview. *IFC Bulletins, Bank for International Settlements*.

71. Xu, Y., Yen, D. C., Lin, B., & Chou, D. C. (2002). Adopting customer relationship management technology. *Industrial Management & Data Systems*, *102*(8), 442–452.

72. XYZ Bank - CRM procedures and best practices. (2020).

APPENDICES

Appendix 1: Povzetek (Summary in Slovene language)

Zaradi velike podatkovne revolucije finančne institucije dandanes ne le zbirajo podatke, kot so to počele desetletja prej, ampak se jim tudi odpirajo možnosti za analizo zbranih podatkov na različne načine (Srivastava & Gopalkrishnan, 2015). Podjetja so imela velike količine podatkov v svojih sistemih že prej, vendar orodja za podatkovno analitiko niso bila tako napredna. Ta izboljšava je spodbudila podjetja, da so se preoblikovala tako, da temeljijo svoje poslovanje na podatkih. Prepoznali so potencial podatkovne analitike in želijo podatke, ki jih ustvarjajo njihova podjetja, uporabiti na najboljši možen način.

Ker danes količina podatkov ni problem, v smislu, da podjetja ne bi imela podatkov o svojih strankah, se pojavljajo drugi problemi. Ključno vprašanje je, kako velike količine podatkov najučinkoviteje uporabiti v procesih managementa odnosov z odjemalci (CRM) ter izbrati najprimernejšo tehniko podatkovne analitike za izboljšanje procesov CRM (Gončarovs, 2017). Posledično to magistrsko delo obravnava različne priložnosti in ovire podatkovne analitike za CRM v bankah ter podati predloge za njeno večjo učinkovitost.

Razumevanje strank je ključ do uspešnega managementa odnosov z odjemalci. Podjetja se morajo zavedati, kakšne so potrebe, želje in dejavnosti strank ter ne pozabiti na razlike med njimi.

Namen magistrskega dela je dati smernice finančnim institucijam, ki razmišljajo o implementaciji analitičnega managementa odnosov z odjemalci. Vsaka finančna institucija mora najprej razumeti, kaj lahko pričakuje od analitičnega CRM, nato pa se vprašati, ali naj ga privzame, za kar pa mora razumeti njegove priložnosti in s čim se bo morala soočiti na tej poti ter kateri so kritični dejavniki uspeha.

Z analizo realnega primera, je magistrsko delo omogočilo boljše razumevanje omenjene problematike.

Magistrsko delo je imelo več ciljev. Prvi izmed njih je bil raziskati literaturo in bolje razumeti, kako se podatki uporabljajo za CRM v bančnem sektorju. Drugi je bil identificirati, opisati in prispevati k boljšemu razumevanju primerov CRM, ki temeljijo na analitiki, in so prisotni na področju bančništva in kako lahko prispevajo k zadovoljevanju potreb strank. Tretji cilj je bil analizirati želje, izzive in priložnosti za CRM za izbrano podjetje ter proces implementacije modela naslednje najboljše ponudbe. Nato pa na podlagi vseh rezultatov raziskave sklepati, kaj so koristi in kaj ovire za banke, ko imajo opravka s podatki komitenta. V delu je posebej izpostavljen pomen kakovosti podatkov. Nato analiziram podatke, ki sem jih pridobila od izbrane banke oblikovanje priporočil naslednje najboljše ponudbe in navedem končne ugotovitve ter ali bo banka imela koristi od uvedbe analitičnega CRM.

Za prvi del naloge sem proučila teoretično ozadje podatkovne analitike, CRM in njegove uporabe v finančni dejavnosti.

Potem sem opravila intervjuje z zaposlenimi pred in po analizi podatkov, da bi odgovorila na svoje raziskovalno vprašanje, ki glasi: kakšne so priložnosti in ovire za uporabo podatkovne analitike za CRM v finančni dejavnosti?

V empiričnem delu sem najprej predstavila izbrano banko in opravila intervjuje z zaposlenimi v banki, ki se ukvarjajo z analizo podatkov, da se seznanim s trenutnim stanjem v izbrani banki. Intervjuji so organizirani, da bi bolje spoznala trenutno stanje v banki, ali se soočajo s kakšnimi ovirami in kakšni so načrti na področju podatkovne analitike za CRM. Intervjuji so delno strukturirani, seznam vprašanj za zaposlene pa je naveden v prilogah. Predstavljene so tudi vloge vsakega zaposlenega in kakšna je njegova vloga v podjetju glede CRM in podatkovne analitike.

Razložila sem predpogoje za dostavo podatkov, opredelila cilj analize in nato opravila preiskovalno analizo nabora podatkov. Potem, področje rezultatov intervjujev je predstavljeno in ugotovitve analize podatkov. Končni rezultat iz pridobljenih podatkov je Excel tabela s strankami, ki naj bi jim banka ponudila osebno posojilo. Na koncu sem pojasnila izzive, s katerimi se sooča izbrana banka, in na koncu podala nekaj nasvetov za banke, ki razmišljajo o implementaciji analitičnega CRM.

Appendix 2: Interview

Before implementation of NBO model

1. What is your position in a bank?
2. How long have you been employed in the bank?
3. What are your primary duties inside the company?
4. What are the main priorities for you and your team?
5. How are you involved in implementation of the NBO model?
6. What are your goals regarding implementation of NBO model?

After implementation of NBO model

7. Did you face any challenges during implementation of model and if yes, which one?
8. What are the lessons you learned from it?
9. What has been successful and what not in the entire process?
10. Did you achieve your goals?
11. How is the performance of the NBO model measured?

Thank you for your time!

Appendix 3: Source system description

In Table 3, I listed technical specification of the source/processing system; components, its characteristics, and links from where I downloaded all.

Table 3: Technical specification of the source/processing system

Component	Characteristics
VMWare Virtual Machine	VMware-player-full-16.2.3-19376536
Windows 10 Operating system	MSEdge.Win10
Oracle Database 19c	Requires registration on Oracle portal
Pip installation and package management system for Python	
Python	Ver 3.7
Anaconda	
SQL Developer	Requires registration on Oracle portal

Source: Own work

Installation/setup of operating system

There is no need to install Operating system from the scratch. Instead, it is possible to use one of many virtual machine images that can run in one of the virtualization technologies like VMware, Oracle Virtual Box, etc. These images are preinstalled virtual machines that usually come with software that is needed. If there is a need anything that is not already installed, it can be installed on regular “bare metal” machines. For my use case, to run Oracle Database, SQL Developer, Python, and other tools for processing of data, I downloaded and use the official Microsoft virtual machine.

Microsoft virtual machine comes with Microsoft Windows 10 EE 64bit version of Operating system. Since Virtual Machine is legally downloaded from the official Microsoft repository, I was able to use it legally, for up to 90 Days. Once a Virtual image is downloaded, I could easily run it in “VMware Workstation 16 Player” which is also free for use. The machine is renamed to “INALJUBDS” and 8GB of RAM are allocated to the machine to be able to handle the large amount of data that it will process. Also, since newer laptops or desktop workstations are equipped with powerful CPUs, I allocated as many virtual CPUs as I can to run the machine without performance overhead.

When it comes to storage space, SSD disks are useful, so it will be allocated 278GB of space for importing the data. I created two Virtual Disks to store OS Data and tools on Disk 1 and Disk 2 I kept Oracle database datafiles. This is the best practice to protect data and make processing flow without performance problems. As soon as the virtual machine starts up, the Win10 operating system will be available for use.

Installation of source system database

The next step was to install the Oracle database. When I ran the installation procedure, I needed to choose what kind of installation I wanted. I chose the installation of Oracle software but also the option to automatically create one database. This option is sufficient for learning and practice purposes.

The rest of the installation steps I did as default installation without any additional setup besides the Oracle default installation procedure. As mentioned previously, I installed Oracle software on C: partition. After it was checked that all parameters are in line with the needs, I proceeded with the installation. After installation was finished, I was able to connect to the Oracle database with applications or tools like SQL Developer. The next step in the building of environment was to install the “SQL Developer” tool for data exploring, preprocessing, and manipulating. SQL Developer is a free and native Oracle tool that can be downloaded from Oracle's official support. To run it, I required Java that can be downloaded with the installation of SQL Developer.

Installation of data analytics tools and programming language

Data analysis will be done using Python language. Then, during data analysis I will need Python packages to be installed. To make this process easy and convenient (Linux-like) I have installed a tool called “PiP”. PIP comes automatically installed with Python 3.4+. PIP is a package management system that I needed to install and manage software packages/libraries written in Python. There is a large “online repository” termed Python Package Index (PyPI).

Anaconda Navigator is an Open-Source Python distribution platform. During installation, Anaconda will download all needed packages for Python and other tools for Data

Analysis. Anaconda's online repository keeps a lot of data science and machine learning packages. Anaconda is the central place for all tools that are needed when building ML or DS use cases and models.

When Anaconda is installed, then installation of additional tools can be done directly from the Anaconda environment. For my use case, I installed JupyterLab, Jupyter Notebook, and PyCharm Community.

Source Database Setup and data import

After successfully installing and configuring the Operating system and Data Science tools, the next step was to create/configure the source database. During the installation of Oracle software, the parent database is already created, and then I created a "pluggable database" named "INAPDB" that will host my data. The database will set up basic tablespaces/datafiles during the creation process, but I had to add USER and Tablespaces/Datafiles that will own/host data. The rest of the database's initial parameters will be configured by default, and there will not be any need to change any of these.

User that will own the data needed for my use-case will be created on" INAPDB" Pluggable database with DBA role. When I create users in the Oracle database I need also to grant "Create Session" system privileges. This will enable the user to connect to the database. The default tablespace will be "USERS" and "TEMP1" default temporary tablespace. Additionally, the password needs to be configured.

After that I was ready to start the database. In the figure below it can be seen that database is up and running and in "READ-WRITE" mode. This means now I can create some additional tablespaces/datafiles and import the data. To keep a large amount of data, I allocated needed space by creating additional Tablespaces/Datafiles in the database. In total, 250GB of storage space was allocated to the owner of the data. All data files are placed on virtual disk E. This is best practice for example case, in case of an eventual disaster I will be able to backup only this disk and recover in case needed. Import of data was done by Oracle *impdp* tool. This tool will import all data from Oracle dump files (format of Oracle export) into the Oracle database, and data will be owned by "INALJUBDS" user.

Appendix 4: Removing/Fixing illogical data

First, I had to connect to database (source of data). I imported two python libraries Oracle_cx and pandas.

```
import pandas as pd

import cx_Oracle

pd.set_option('display.max_columns', None)

conn = cx_Oracle.connect('INALJUBDS/oracle@inapdb')
```

After that it was important to specify the data set that I am working with. In this case I defined it as simple SQL query to fetch „AGE“ customer data from database with couple other attributes like: GENDER, CUSTOMER_ID etc.

```
query = """select CUSTOMER_ID, BIRTH_DATE, AGE, GENDER,
BUSINESS_TYPE_CODE from temp_customer_pl"""

#print (query)
```

If connecting and selecting the data is successful, then data will be read and placed into dataset (df_ora), that I would work with and rename it to customer.

```
df_ora = pd.read_sql(query, con=conn)

customers = df_ora

#print (df_ora)
```

To get more info about the dataset structure (columns, datatypes etc.) I used command info ().

```
customers.info ()
```

```
<class 'pandas.core.frame.DataFrame'>

RangeIndex: 842631 entries, 0 to 842630

Data columns (total 5 columns):

 #   Column                    Non-Null    Count       Dtype

---  -----------                 ---------     -------      ------

 0   CUSTOMER_ID           842631      non-null    object

 1   BIRTH_DATE             842631      non-null    object
```

2	*AGE*	*842631*	*non-null*	*int64*
3	*GENDER*	*842631*	*non-null*	*object*
4	*BUSINESS_TYPE_CODE*	*842631*	*non-null*	*object*

dtypes: int64(1), object (4)

memory usage: 32.1+ MB

And to find basic structure of AGE data I executed command describe ():

customers.describe()

	AGE
count	*842631.000000*
mean	*32.881646*
std	*342.835233*
min	*-7982.000000*
25%	*34.000000*
50%	*47.000000*
75%	*61.000000*
max	*119.000000*

Since I wanted to check format of data that is stored in the dataset, I used command head ():

customers.head()

	CUSTOMER_ID	*BIRTH_DATE*	*AGE*	*GENDER*	*BUSINESS_CODE*
0	*2502966188890*	*1966-02-25 00:00:00*	*51*	*F*	*00-Senior*
1	*0612955198617*	*1955-12-06 00:00:00*	*62*	*F*	*00-LowIncome*
2	*2812979190001*	*1979-12-28 00:00:00*	*38*	*M*	*00-Senior*
3	*0612960108022*	*1960-12-06 00:00:00*	*57*	*F*	*00-Senior*
4	*0612988105029*	*1988-12-06 00:00:00*	*29*	*F*	*00Unemployed*

Finally, I could proceed with grouping of data that gave me a first look of data quality.

customers.groupby('AGE').size()

```
AGE
-7982   1139
-7981    174
-7980    222
-36        1
-31        1
       ...
 111       4
 112       3
 116       1
 118       1
 119       1
Length: 122, dtype: int64
```

Already on this step it can be noticed that some data that should be relevant for use-case is not of sufficient quality. The lowest and the biggest values show that there are customers with a negative age value, but also people over 100 years old.

The rest of the process I did in SQL developer, because I used direct connection from Jupyter to Oracle database, and fixing data directly in database will impact dataset automatically.

Select all customers that have AGE = '-7982';

```
select
--CUSTOMER_ID,
BIRTH_DATE, AGE, GENDER, BUSINESS_TYPE_CODE
from temp_customer_pl
where AGE = '-7982';
```

1,139 rows selected.

BIRTH_DAT	AGE	GENDER	BUSINESS_TYPE_CODE
01-OCT-99	*-7982*	*M*	*00-Student*
01-OCT-99	*-7982*	*U*	*00-Student*
01-OCT-99	*-7982*	*U*	*00-Student*
01-OCT-99	*-7982*	*U*	*00-Student*
01-OCT-99	*-7982*	*U*	*00-Student*
01-OCT-99	*-7982*	*U*	*00-Student*
01-OCT-99	*-7982*	*U*	*00-Student*
01-OCT-99	*-7982*	*U*	*00-Student*

Since it can be noticed that all these clients are students with BIRTH_DATE = 01-OCT-99, it can be assumed that this data is systematically (machine/code driven) input and cannot be used for this part of analysis. It needs to be fixed it, not by deleting the data but by unfirming the data. The reason for this is the fact that this is very specific data about AGE, but in other cases the rest of the data will be useful. The obvious way to fix this is to update the ‘AGE' attribute for these clients with some uniform value that will be used across the whole use case, like "N/A" (Not applicable) or simply by putting 0 in case of numeric data.

Update the data (AGE attribute)

update temp_customer_pl set AGE = 0

where AGE = '-7982';

1,139 rows updated.

Commit changes in database

commit;

Commit complete.

Also, I could use SQL 'group by' function to identify illogical data and fix it (in this case uniform it), for example:

```
select count(*), AGE

from temp_customer_pl

group by AGE;
```

This showed all counts on AGE attribute with values, then not relevant data can be seen and fix it by update statement. After checking the results, it should return 0 rows, and the process can continue until all data is cleaned up.