

## **Financial Programming Group Project – Prepared for Prof Minh Phan**

Shabenoor Kamal

Omar Masrouji

Fabian Celi

### **Business Case**

This case is prepared for XYZ Bank, an established Eastern European bank offering services to customers across Czech Republic. The services offered by the bank include creation and management of accounts, credit card offers and loans to eligible customers.

The business problem the bank needs help with is to identify a good client (someone they can offer additional services to at an expected profit) vs a bad client (who to watch carefully in order to minimize bank losses). The bank requires help in order to understand who different segments of customers are, the common characteristics that define them, and the tailored strategy and services to target them with in order to maximize profit.

To aid this process XYZ Bank provided us with multiple data tables that contain detailed data on clients including accounts, transactions, credit cards, orders, loans and demographics. The time range of the data is from 1993 to 1998. The current year when we take on this project is assumed to be 1999. Our goals in assisting this institution were multi-fold:

- 1) Our initial aim was to clean and merge the different data tables provided by the bank in a way that it is easy for the bank to get meaningful data on each client in their database
- 2) Second, to create some additional variables to allow deeper insight into each client. This includes their demographics, past behaviors, relationship history with other accounts and banks, as well as history of credit cards and loans
- 3) Third, to segment the clients in a way that is valuable for the bank in order to understand what customers in each segment have in common, and how they differ from other segments
- 4) Lastly, to recommend some relevant strategies for each segment in the hope that this will allow the bank to maximize their profit from well-tailored offers

### **What does the client base look like?**

The initial steps of this project were to understand and define the company's problems and understand their requirements.

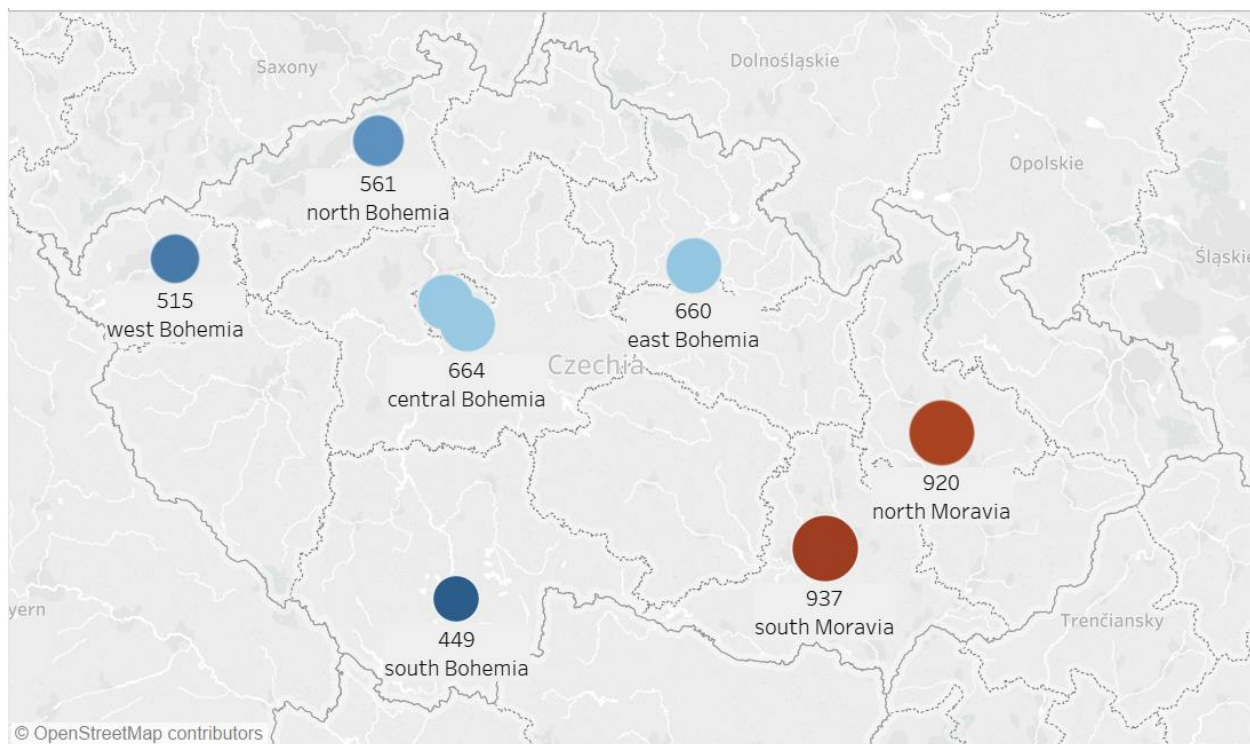
Next, we focused on understanding the data provided to us by the bank. All the data tables provided were explored in detail and cleaned. During the process some new variables were created, missing values were dealt with and data was aggregated to a client level. The updated tables were then combined into an aggregated basetable that includes one row per unique client ID. Our goal here was to be able to provide the bank with a table that would allow them to easily extract data and insights on each client, as an aggregate of all of their previous history and transactions with the bank. This table should be considered an asset for current and future projects and can be used in multiple applications, by many different departments.

Before delving into further details on client segments, we would like to give a short summary of what this bank's current client base looks like. The basic descriptive analysis includes gender, age groups and regions to understand the customer base.

The database contains data on 5,369 customers of which we have an equal spread of male and female clients. A minority (72 clients) are underage while more than half are between the age of 45 and 60.

| Age Class        | Gender | DISPONENT | OWNER |
|------------------|--------|-----------|-------|
| underage         | F      | 7         | 23    |
|                  | M      | 8         | 34    |
| young adult      | F      | 62        | 564   |
|                  | M      | 74        | 510   |
| middle age adult | F      | 93        | 607   |
|                  | M      | 140       | 594   |
| old adult        | F      | 111       | 659   |
|                  | M      | 111       | 622   |
| senior           | F      | 164       | 355   |
|                  | M      | 99        | 532   |

The top 4 most represented regions, accounting for about three-fifths (60%) of the bank's clients, are South Moravia, North Moravia, Central Bohemia and Prague. These regions also tend to have a higher concentration of additional bank services such as loans and credit cards.



Across all age groups (besides underage) the most accounts were opened 3 years ago (1996) followed by 6 years ago (1993). In 1997 the accounts opened dropped for every age group between 30%-40% compared to 1996.

| Years since account opening | underage | young adult | middle age adult | old adult | Senior |
|-----------------------------|----------|-------------|------------------|-----------|--------|
| 2                           | 14       | 248         | 276              | 312       | 229    |
| 3                           | 15       | 354         | 456              | 450       | 349    |
| 4                           | 13       | 189         | 210              | 205       | 151    |
| 5                           | 7        | 135         | 124              | 139       | 108    |
| 6                           | 23       | 284         | 368              | 397       | 313    |

## Details on Services

Over the past few years, a growing number of clients are availing additional bank services such as credit cards. There was an 80% increase in the number of credit cards last year compared to the one before. The

number of loans taken has also been increasing, however there was a drop (approximately 20%) in the last year compared to the one before.

| Years since card | Classic Card | Gold Card | Junior Card | All |
|------------------|--------------|-----------|-------------|-----|
| 1                | 326          | 53        | 70          | 449 |
| 2                | 190          | 26        | 26          | 242 |
| 3                | 83           | 5         | 28          | 116 |
| 4                | 42           | 4         | 17          | 63  |
| 5                | 17           | -         | 4           | 21  |
| 6                | 1            | -         | -           | 1   |
| All              | 659          | 88        | 145         | 892 |

| Loan duration (years) |     |     |     |     |     |     |
|-----------------------|-----|-----|-----|-----|-----|-----|
| loan_age              | 12  | 24  | 36  | 48  | 60  | All |
| 1                     | 28  | 28  | 32  | 33  | 37  | 158 |
| 2                     | 43  | 35  | 31  | 44  | 43  | 196 |
| 3                     | 15  | 20  | 28  | 30  | 24  | 117 |
| 4                     | 14  | 24  | 13  | 18  | 21  | 90  |
| 5                     | 26  | 26  | 21  | 12  | 16  | 101 |
| 6                     | 5   | 5   | 5   | 1   | 4   | 20  |
| All                   | 131 | 138 | 130 | 138 | 145 | 682 |

Loan durations of 12 to 24 months have the highest number of clients with a contract that is finished and the loan was not paid. Higher loan durations of 48 to 60 months have more clients that have a running contract but are currently in debt.

| Loan Status |     |                                 |                                   |                                |                                   |
|-------------|-----|---------------------------------|-----------------------------------|--------------------------------|-----------------------------------|
| Duration    |     | Contract finished - No problems | Contract finished - Loan not paid | Running Contract - OK till now | Running Contract - Client in debt |
|             | 12  | 119                             | 10                                | 35                             | 1                                 |
|             | 24  | 82                              | 11                                | 68                             | 6                                 |
|             | 36  | 42                              | 7                                 | 98                             | 8                                 |
|             | 48  | 12                              | 2                                 | 137                            | 14                                |
|             | 60  | 3                               | 1                                 | 155                            | 16                                |
|             | All | 258                             | 31                                | 493                            | 45                                |

## Variable and Basetable Creation

Based on the exploratory analysis, additional variables were created to get further insights from the data. Initial variables were created in the individual tables before merging them. Missing data and outliers were handled appropriately.

Initially we had 8 separate tables. A list of the tables along with their unique keys, used to merge them, is mentioned below:

- Client – each record describes characteristics of a client.  
Keys: client\_id (used to merge with disposition table) and district\_id (used to merge with demograph table)
- Disposition table – each record relates a client to an account.  
Keys: client\_id (used to merge with client table), disp\_id (used to merge with credit card table) and account\_id (used to merge with account table)
- Credit card – each record describes a credit card issued to an account  
Key: disp\_id (used to merge with disposition table)
- Demograph – a record describes demographic characteristics of a district  
Key: district\_id (used to merge with the client and account tables)
- Account - each record describes static characteristics of an account  
Keys: account\_id (used to merge with disposition, permanent order, loan and transactions tables) and district\_id (to merge with the district table)
- Loan - each record describes a loan granted for a given account  
Key: account\_id (used to merge with the account table)
- Permanent order - each record describes characteristics of a payment order  
Key: account\_id (used to merge with the account table)
- Transactions - each record describes one transaction on an account  
Key: account\_id (used to merge with the account table)

Variables were created within the existing tables to aggregate them in a way that would allow us to get relevant information (for example number of transactions, permanent orders etc) at a unique id level (such as account\_id). These aggregated tables were then merged together to create a basetable with one row per unique client ID.

After this step, other variables were created that required information previously spread across multiple tables.

## **Segmentation and Analysis**

To help the bank understand and differentiate 'good' from 'bad' customers we first had to define who a 'good customer' is. The definition we used, after extensive discussion with the client, was a good customer is one who is profitable for the bank when targeted with the right services and offers.

The definition was essential for deciding what variables should be included in the customer segmentation. A large part of the project work consisted of creating relevant variables for the segmentation and selecting the ones that would bring most meaningful results for the analysis. A manual approach was used and our team looked at each individual variable to pick the ones that we believe could directly impact the probability that a customer will make a profit or loss for the bank (such as loan defaults, transaction behavior) or could give some other useful information to segment on (such as gender and age). After confirmation from the bank regarding the relevance of these variables, we continued with the cluster analysis. The following steps were taken:

- 1) Some columns such as those for 'amount' (eg amount of loan, transactions etc) were log transformed to fix their skewness
- 2) All numeric variables were standardized to ensure they are on the same scale. This improves the quality of the analysis, as once standardized, the different ranges of variables will not affect the distance calculations used by the clustering algorithm
- 3) The algorithm used was kmodes that is part of the "kmodes" package in python. The advantage of this algorithm is that it allows both numeric and categorical variables to be included in the cluster analysis
- 4) The analysis was run many times, with different combinations of variables and cluster sizes. Based on the results the ideal cluster size was determined to be 3 clusters
- 5) Based on this we were able to assign each customer in the basetable to a segment (from 1 to 3)

Some of the unique defining characteristics of these segments (based on the cluster centroids) are described below. All characteristics mentioned should be considered an average measure of the clusters members.

**Cluster 1 (Younger clients):** This cluster consists of younger customers who are currently in their mid-thirties. They tend to have fewer people in each account, which may indicate that they have fewer dependents.

**Services:** They have had a credit card with this bank longer than the other segments of customers. They tend not to take loans and therefore have the lowest average loan amount, payments and loan age (years since loan started) among all segments.

**Transaction behavior:** While they have the lowest average count of transactions, the average amount of each transaction is higher than the other segments. Their average and max balance is also higher. However, their average minimum balance is considerably lower, and is a negative number, which can indicate high spending. This is something the bank should watch out for when targeting them with offers. These younger clients tend to have a higher average debit and credit amount, which could mean that they make more money, but spend more as well. The total number of transactions made is lower than other segments and their most recent transaction is approximately 2 weeks ago, that indicates that they are less active customers.

**Types of transactions:** Since this is a younger segment pension payments are lower than those for other segments. They do not focus on saving which is why average interest credited is also lower. This group may have fewer responsibilities in life and therefore has lower payments for household, loan and insurance. However, they do have the highest average sanction interest, charged due to negative account balance. This segment has the highest credit and withdrawal in cash but the lowest remittance and collections from other banks

**Orders:** This segment has the lowest average count of orders, interacts with the least number of banks and has the lowest average amount for orders. The average amount sent to banks ST, WX and GH is higher than for other segments.

**Cluster 2 (Family oriented/ Middle aged clients):** This cluster consists of middle aged customers who are currently in their mid-forties. They tend to have higher number of people in each account, which may indicate that they have dependents.

**Services:** They have had a credit card with this bank longer than Segment 3, but for less time than Segment 1 (younger clients). However, they have had an account with this bank for the longest average duration amongst all segments. They tend to take more loans and have the highest average loan amount and payments. They generally have the longest loan age (years since loan started). The majority of these clients do not have any problems with loan payments (finished or ongoing loans).

**Transaction behavior:** The middle-aged clients have the highest average count of transactions; the average amount of each transaction is higher than Segment 3. While their average and max account balance is lower than the younger clients, their average minimum balance is considerably higher which indicates more saving and careful spending. The total number of transactions made is higher than other segments and they tend to have the most frequent (recent) transactions.

**Types of transactions:** Since this is a middle-aged segment pension payments are lower than those for Segment 3. They focus on saving which is why average interest credited is the highest for this group. The lowest sanction interest indicates they are careful spenders and their account is least likely to have a negative balance.

This may be due to the reason that this group seems to have the most responsibilities in life, indicated by the highest average household, insurance, statement and loan payments. This segment has the highest remittance to other banks, as well as the highest average amount credited to their account and debited from their account (excluding cash withdrawals) which also indicates responsibilities and possible dependents.

**Orders:** This segment has the highest average count of orders, interacts with the highest number of banks and have the highest average amount for orders. Most common orders are for household, loan and other payments. The average amount sent to banks AB, EF, YZ and IJ is higher than for other segments.

**Cluster 3 (Older clients):** This cluster consists of older customers who are currently in their fifties. They tend to have higher number of people on each account, which may indicate that they have dependents.



They also tend to reside in districts that have a slightly higher crime rate compared to those of other segments.

**Services:** They have had a credit card with this bank for the least amount of time. They tend to take less loans than Segment 2, but more than Segment 1 and might have some problems with loan payments, specifically for currently running loans.

**Transaction behavior:** These older customers tend to have the lowest average amount for transactions, as well as the lowest average bank balance. Additionally, they also have the lowest average credit and debit. One possible reason is that within this segment people may be retired and no longer have a steady income coming in.

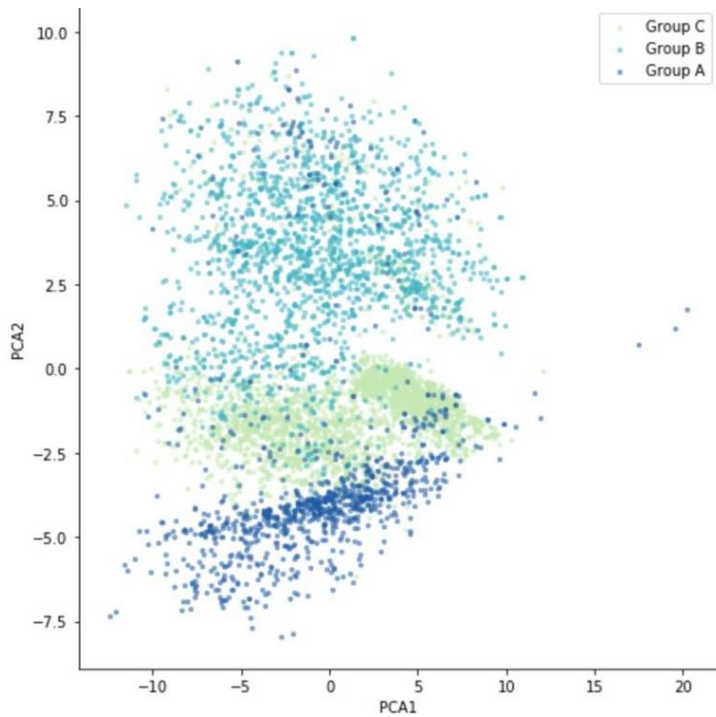
**Types of transactions:** This segment receives the highest pension payments. Due to retirement, they may be using their savings, which is why average interest credited is the lowest among all segments. They may have fewer responsibilities in life compared to Segment 3, indicated by lower household, insurance, statement and loan expenses.

This segment has low credits, withdrawals and remittance to other banks.

**Orders:** This segment has an average amount of orders, mostly for loan payments.

### **Visualization of the segments:**

PCA (principle components analysis) was used at the end of the segmentation analysis to reduce the dimensionality effectively and to aid in visualizing the clusters in two dimensions. The graph below is a visual representation where Group A is Cluster 1 (younger customers), Group B is Cluster 2 (middle aged customers) and Group C is Cluster 3 (older customers).



**Segmentation note:**

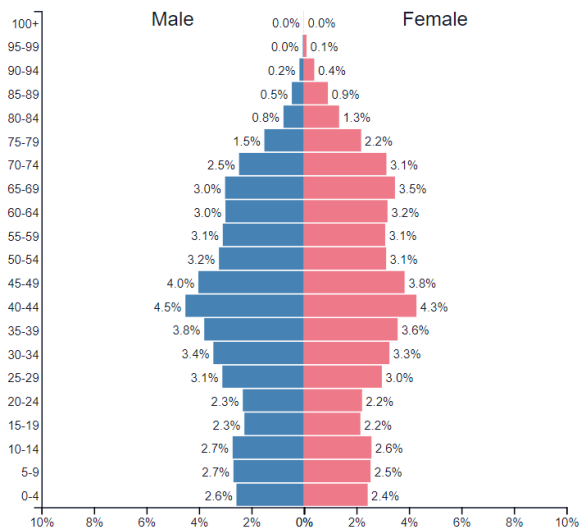
Across the three clusters, the central age we have ranges from mid-thirties to fifties. We looked into this some more to make sure the numbers we have make sense. As the chart below shows Czech Republic has an aging population with a large proportion in the 35-50 age range.

Due to this, we concluded that the age ranges for our clusters actually make sense and are representative of reality.

## Czech Republic ▼

2019

Population: 10,569,725



Source: <https://www.populationpyramid.net/czech-republic/2019/>

## Strategy

Based on the results of our segmentation we would like to recommend some strategies to the bank for each segment. Each of the segments identified has some benefits (which could allow them to make a profit for the bank) as well as some drawbacks (which could lead to losses).

### 1. Cluster 1 (Young customers):

Recommendations:

- This group is more likely to spend large amounts of money and is more likely than other segments to have an average minimum balance in the negative numbers. The bank should therefore be careful with these customers by setting a limit for negative balance, if that limit is crossed the bank can temporarily block the account. This will ensure the bank does not have to suffer large losses in case the customer is not able to pay back
- Since this cluster has had a credit card with the bank for the longest time, and is often making transactions of large amounts of money, the bank should target them with good credit card offers. Offers could include cashback offers on the money spent via credit card, or other benefits such as airline miles and no charges on international transactions.

- This group is least likely to have a loan and due to their tendency to overspend (signified by their negative average minimum balance) it would be a good idea to not offer them loans as they may have a higher risk of default

## 2. Cluster 2 (Middle aged customers):

- This segment consists of responsible, middle aged customers, who are more likely to have dependents. Due to this, it would be a good idea to offer them 'family deals' such as opening free joint accounts for their partners and children. This would also ensure that the bank is able to capture a younger customer base. Since the young population of Czech Republic is lower in proportion compared to older groups, this age range will become very important to target in the future. This is a good strategy for the bank to start thinking of, and can help them reap large future benefits and avoid losing out to competitors
- Since this segment is most likely to take out a loan, more loan related offers and services should be targeted towards this segment. Special loan offers such as those for houses or cars can be crafted, along with more flexible payments and loan durations. This segment is more responsible and has low likelihood to default so they are lower risk for the bank
- Members of this segment have a larger number of responsibilities as can be seen through their transaction behavior (large number of transactions for household, loan, insurance and other payments) along with a larger number of orders to send money to other accounts. This segment should be targeted with services that could ease the process of making transactions such as an app that allows quick transfers, automatic recurring payments and 24/7 customer service. Such a service will ensure that these customers can be retained and remain loyal to the bank

## 3. Cluster 3 (Older customers):

- Segment 3 mainly consists of older customers who are likely to be retired. This assumption is based on their average bank balance, average credit amount and average debit amount which is less than the other segments. They also have a large number of pension transactions. Due to these factors, offers tailored to this segment should be focused on savings, retirement and pension plans to allow this senior group to prepare for their retirement and make the best use of their pension payments
- Since they have a higher probability of having problems with loans, to minimize the banks losses this segment should not be offered loans

