Group 8: Rémy Doyen, Harshit Paliwal, Justine Deleval

Final Report: Python Project – Financial Services

1.1 INTRODUCTON

In today's world, while insights are built on the consumer level and all the financial institutions base their strategies on customer behavior, this project is a small attempt to build insights on Financial dataset at the customer level. The entire model was built on Python and exploits the concepts of basic till advanced analytics. Various key variables were generated using the information provided in the dataset in order to build a final da.

The base-table consists of various parameters which informs the end reader with detailed insights about the clients, their spending behavior, the preferred mode of transaction, loan repayment behavior, the demographic roots and many more. Using these details, one can generate accurate ideas and strategies on how to further communicate with clients and helps the decision makers segregate the clients into Good and Bad clients.

Using various visualizations in the project, multiple trends were analyzed and new information were revealed, showing the bigger picture of the reality.

With this project, the technical complexities of financial dataset are resolved to quite an extent, thus, taking ahead the steps in Data Analytics and implementing its concepts in fields like Finance.

1.2 DATA CLEANING

1.2.1 Insights on the clients:

The data cleaning part was based on Python, and we worked on it using the Spyder environment. The data cleaning part started with finalizing the key variables to be included in the final base-table.

The main idea was to generate the insights on the 'OWNER' level, ie, the person who owns the account. The reason for keeping the owners in the table was that the owners are eligible to avail a loan or a permanent order. It allowed to avoid lots of misleading and duplicate information which we were getting on the client level (owner + user). Indeed, since an account can be shared, the owner and the user of an account would have the same information because it is impossible to determine who performed the different transactions. One way to solve this issue would be to replace most of the value with NA for the users, but these rows would become useless. This is the reason why we decided to keep only the owner in the final DataMart. We carried the analysis of the clients for only the owners of account to be able to target the one eligible to get a loan or a permanent order.

1.2.2 Explaining the variables:

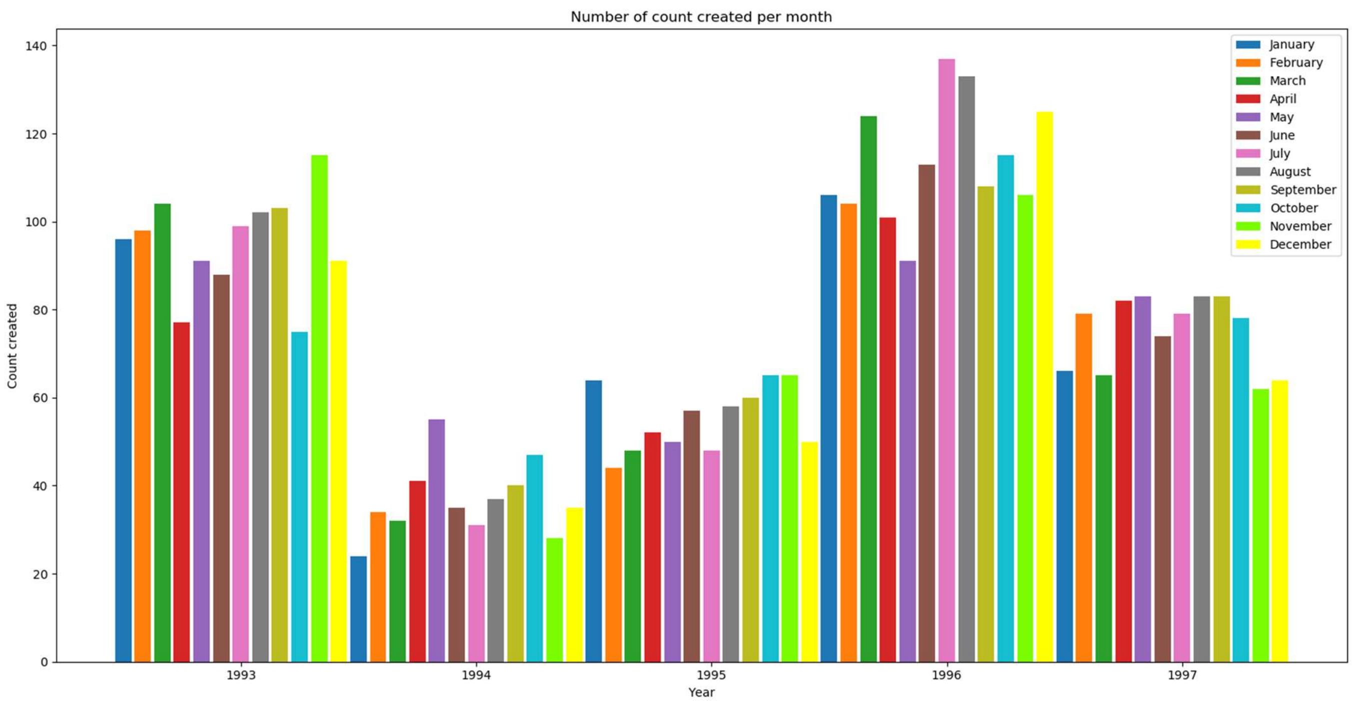
The final DataMart consists of the following main variables (every variable has been created per account id, which was then merge to the client ID level):

- ***Nbr_Credit_Trans/Nbr_Debit_Trans :*** Calculated as the total number of credit/debit transactions made from 01/01/93 to 31/12/97 from the type of transactions.
- ***Amount_Credit_Trans./Amount_Debit_Trans:*** Calculated as the total amount of credit/debit transactions made from 01/01/93 to 31/12/97 from the type of transactions.
- ***Average_Balance :*** Calculated as the average balance in the account during the total length of the period (01/01/93 to 31/12/97)
- ***Nbr_Credit_Cash/Nbr_Credit_Bank/Nbr_Credit_Other:*** Calculated as the number of credit transactions for each operation made from 01/01/93 to 31/12/97.
- ***Nbr_Debit_Cash/Nbr_Debit_Card//Nbr_Credit_Bank:*** Calculated as the number of debit transactions for each operation made from 01/01/93 to 31/12/97.
- ***Amount_Credit_Cash/Amount_Credit_Bank/Amount_Credit_Other:*** Calculated as the total amount of credit transactions for each operation made from 01/01/93 to 31/12/97.
- ***Amount_Debit_Cash/Amount_Debit_Card//Amount_Credit_Bank:*** Calculated as the total amount of debit transactions for each operation made from 01/01/93 to 31/12/97.
- ***Nbr_Statement_Trans./Nbr_Interest_Credited_Trans/Nbr_Sanction_Interest_Trans/Nbr_Pension_Trans/Nbr_Insurance_Trans/Nbr_Household_Trans/Nbr_Other_Trans /Nbr_Loan_Trans*** **:** Calculated as the number of transactions per distinct categories (k_symbol) made from 01/01/93 to 31/12/97.
- ***Amount_Statement_Trans/Amount_Interest_Credited_Trans/Amount_Sanction_Interest_Trans/Amount_Pension_Trans/Amount_Household_Trans/Amount_Other_Trans/Amount_Loan_Trans:*** Calculated as the total amount of transactions per distinct categories (k_symbol) made from 01/01/93 to 31/12/97.
- ***Nbr_PO_Leasing/Nbr_PO_Insurance/Nbr_PO_Household/Nbr_PO_Loan/Nbr_PO_Other:*** Calculated as the number of permanent orders placed per distinct categories made from 01/01/93 to 31/12/97.
- ***Monthly_Amount_PO_Leasing/Monthly_Amount_PO_Insurance/Monthly_Amount_PO_Household/Monthly_Amount_PO_Loan/Monthly_Amount_PO_Other:*** Calculated as the monthly amount to be paid for the permanent orders placed per distinct categories made from 01/01/93 to 31/12/97.
- ***Theoritical_Amount_Loan_Due:*** Calculated as the total theorical amount of the loan to be paid by the client. This amount is the same from the amount of Loan Transaction is the client has totally paid off the loan but differentiate if the client is either still paying off the loan, run out of contracts, or is in debt.
- ***Loan_Duration:*** Calculated as the total duration of the loan.
- ***Status_Loan_Owner:*** Status of the loan owner.

1.3 DEMOGRAPHICS

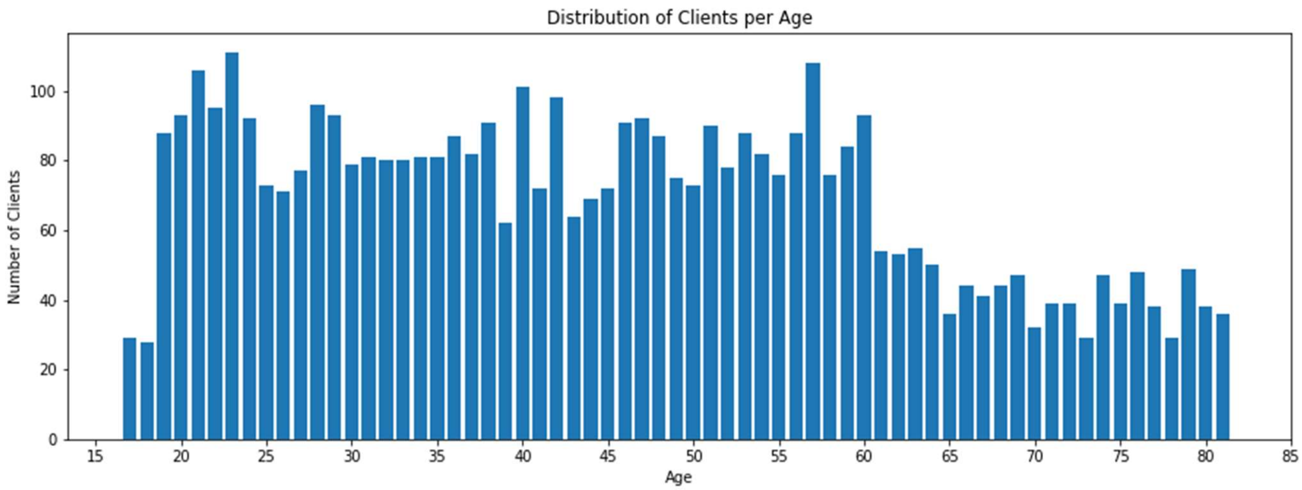
1.3.1 Account opened per year and month

The graph below shows the spread of accounts opened along the period. Distributed by months, it gives good insights on which are the months when most accounts were created, and which had lowest turn-in. According to the trend, December seems to be driest in terms of new account openings while mid-year months like May - July register good number of new clients in general.

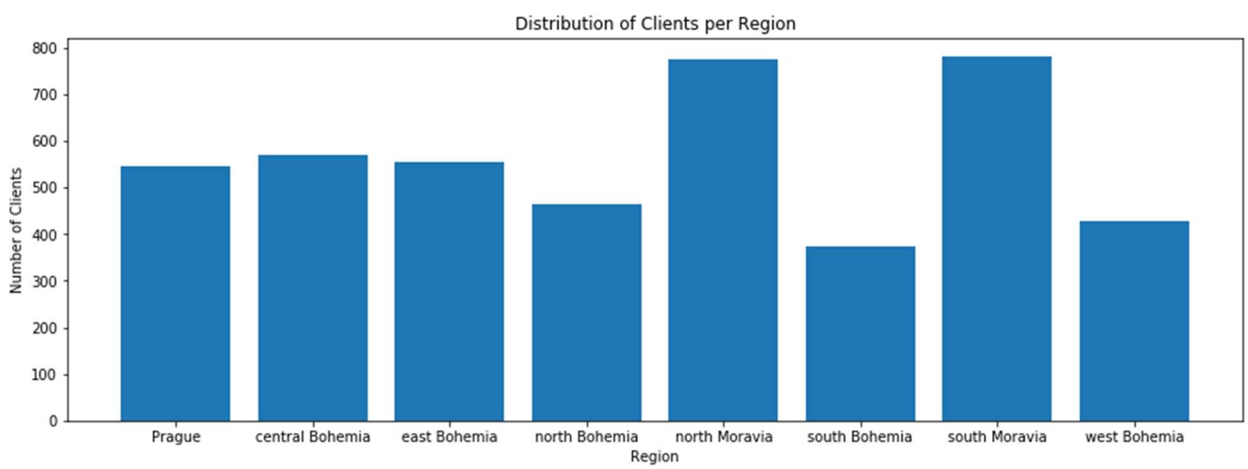


1.3.2 Overall distribution of the clients

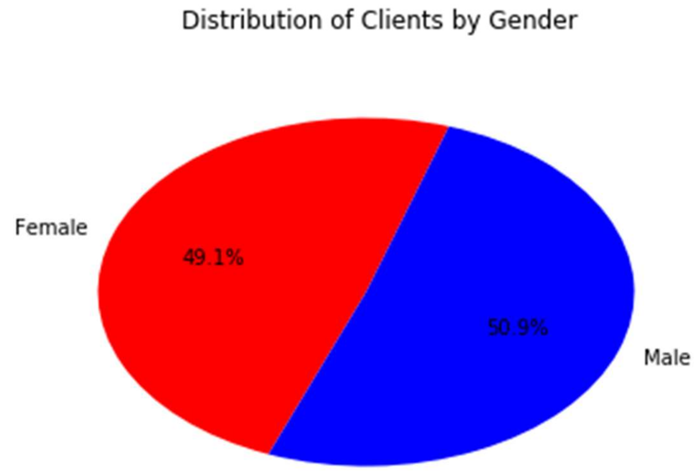
The graph below shows the distribution of the clients per age. Some concrete trend can be inferred from this plot, like, many clients are from early 20s and mid 20s, the number of clients falls rapidly beyond 60s, which can be a matter of concern and possibly an opportunity to draw in more clients for this target group.



The graph below illustrates the distribution of the clients per region. From this distribution, we can infer the geographical spread of clients which would help us understand the client behavior and situation based on the district he/she comes from. According to the dataset, majority of clients are from North Moravia and South Moravia region, while, least encountered district is South Bohemia.

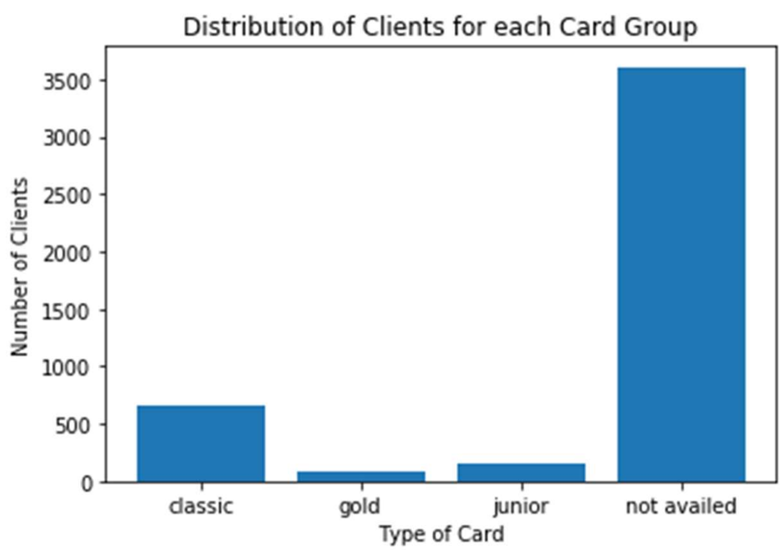


The graph below depicts the gender distribution of the account owners in the Datamart. Evident from the plot, one can say that the share of male and female owner is quite balanced with around 2200 females and 2300 males.



The graph below represents the population distribution of clients with respect to the type of credit card they have availed. There are 3 possible categories of cards including Classic, Gold and Junior. The 4th category depicts all the clients who have not availed any cards.

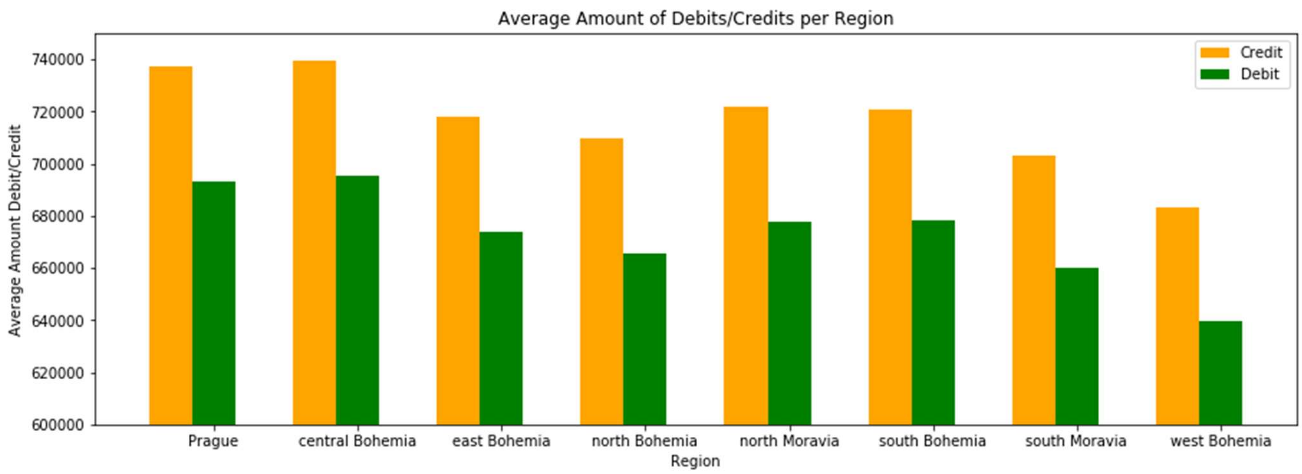
The plot clearly shows that around 80% of the population has not availed any type of card. Among the people who have received a card for their account, they opted for Classic in general with few occurrences of Junior and Gold cards.



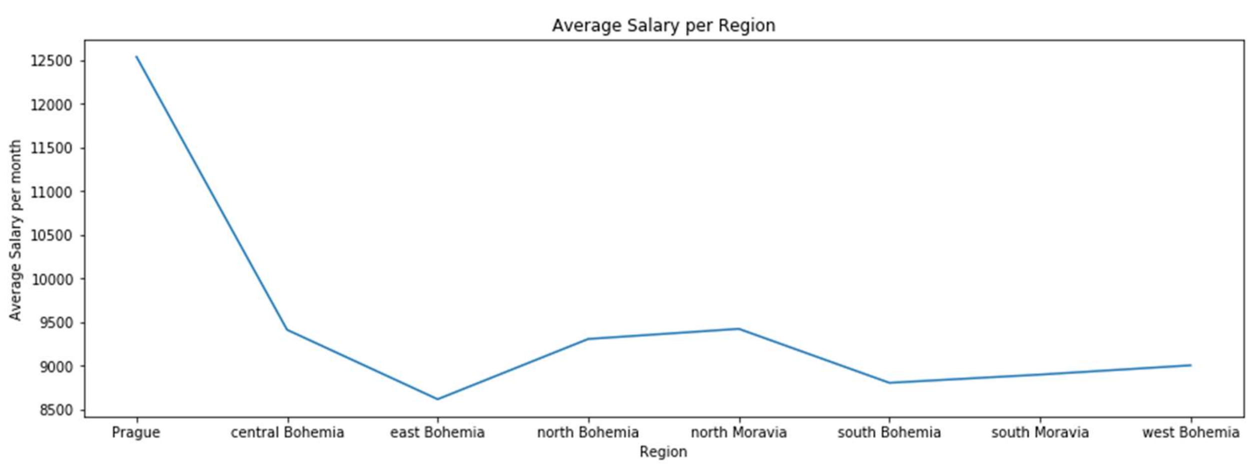
1.4 VARIABLES

1.4.1 Spending Behavior

The graph below shows the average amount of credit and debit transactions per region. As expected, every region is having a higher average amount of credits compared to debits transactions.

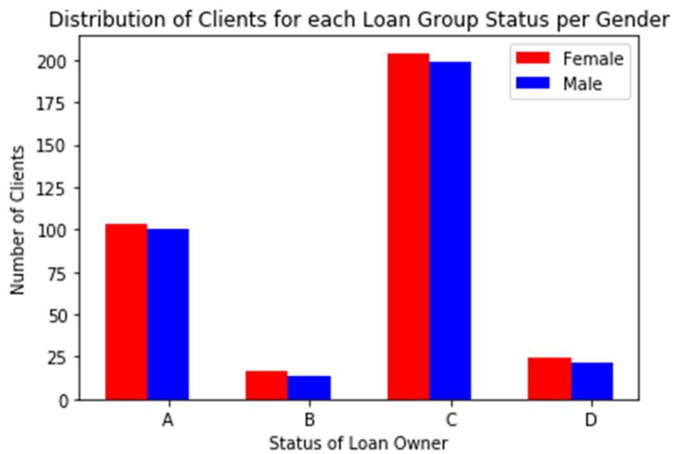


Moreover, from the graph above, we can see that the region of Prague and Central Bohemia have clients with higher average amount of debits and credits. Those clients tend to have higher spending behaviors and higher income. It is also seen in the graph below where the average salary in Prague is way higher than for the other regions. We can infer that targeting clients coming from those regions might be a good strategy.



1.4.2 Loan

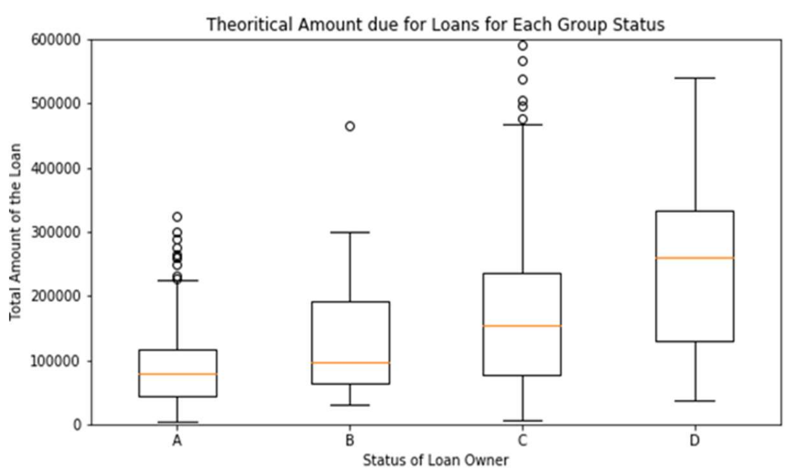
We identified 682 clients that have taken a loan from the bank. Those clients are distinguished into 4 categories: 'A' stands for clients that have paid off the full amount of the loan without problems, 'B' for clients that have contracts that are finished but have not paid the full amount of the loan, 'C' for clients that are currently enrolled into a loan and everything is going alright so far, and 'D' for clients that still have contracts but are running into debt.



As seen in the graph, we plotted the distribution of our 682 clients that have loans with the bank to see the number of clients that can be considered as potential good clients for the bank and eligible for loans (Categories A and C) and potential bad clients (Categories B and D).

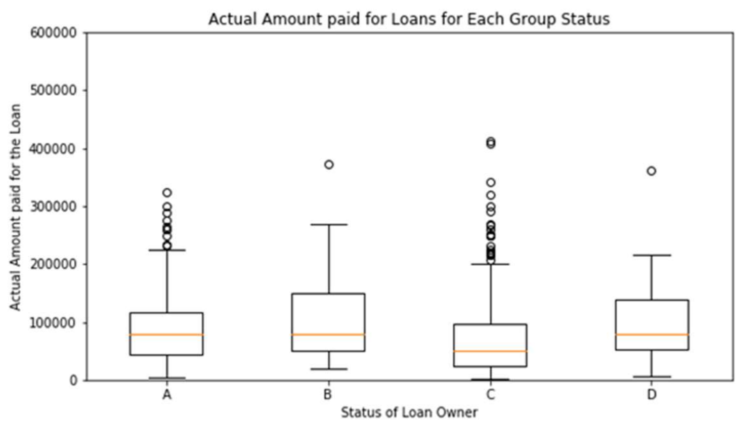
To go further on the analysis of our clients, we decided to check for the theoretical amount of the loan for each categories of loan owners.

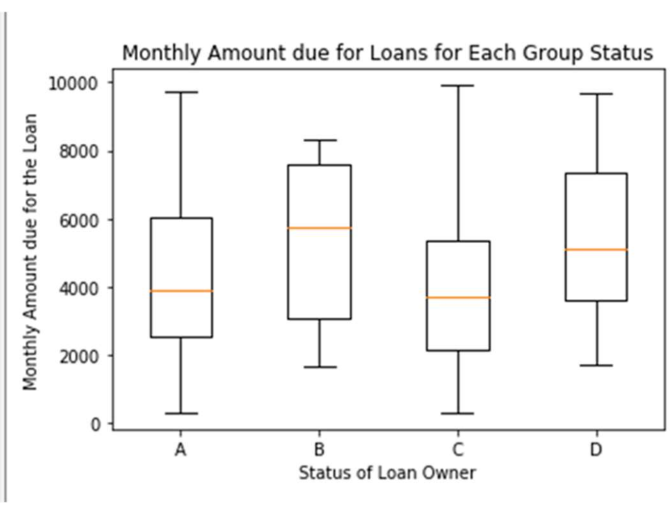
As per the graph below, we can see that people in the 'A' group tend to have lower loan amount with a smaller spread around the median value whereas clients in group B and D tend to have higher loan amount to pay off, which lead to deficit in payments of the loan.



We did the same analysis for the Actual amount paid for the loans in each group categories (graph below). We see that that clients in group B and D paid less that their expected (Theoretical) amount of loans. With regards to the actual amount spend on loan, the four groups of people have very close amount.

From those two graphs (Theoretical Amount due Loans for Each Group Status and Actual Amount paid for Loans for Each Group Status), we can draw the conclusion that clients B and D tend to take out higher loans than clients belonging to group A. However, they are usually not able to pay off the loan, therefore their actual spending is much lower than expected.





In addition, in this graph, we are analyzing the monthly amount due for each group of loan owners. Following the patterns that we saw earlier, clients belonging to categories B and D tend to have higher monthly payments but are not able to fully pay off the loan. Clients A and C have lower monthly payments but are more trustworthy clients.

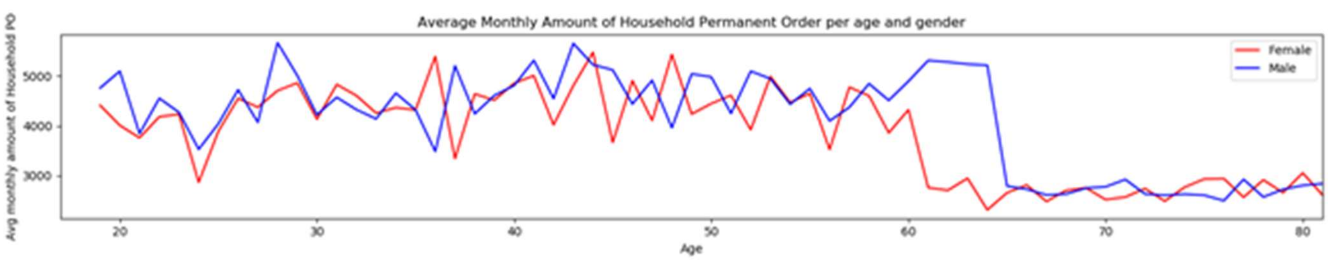
1.4.3 Permanent Order

According to the permanent order table, we noticed that there were 5 different types of permanent order: insurance, household, leasing, loan and others. We already discussed about the loan in the previous point.

On the two graphs below, one can observe the average amount and the average monthly amount of Household Permanent Order per age and per gender. The first thing we can observe is that there is almost no difference between male and female.

Moreover, the average amount of Household permanent order is quite steady according to the age of the account owner even if a small decrease can be observed after 60 years old.

Finally, one can also notice that this drop is sharper for the monthly amount of household permanent order.

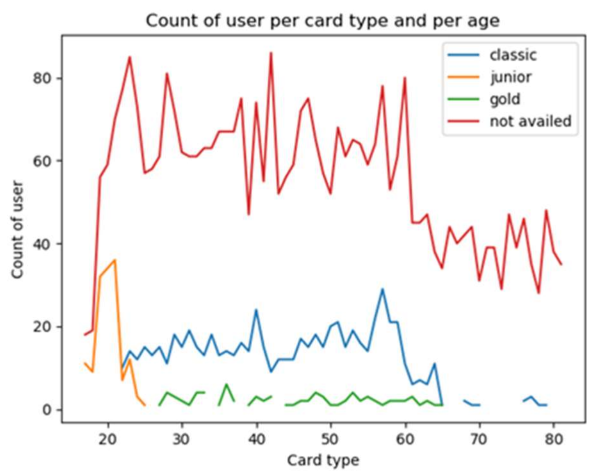


We also see, that these patterns were similar for the other type of permanent order. You can find the same two graphs for the insurance permanent order in appendix (appendix 1.5.1).

1.4.4 Credit card

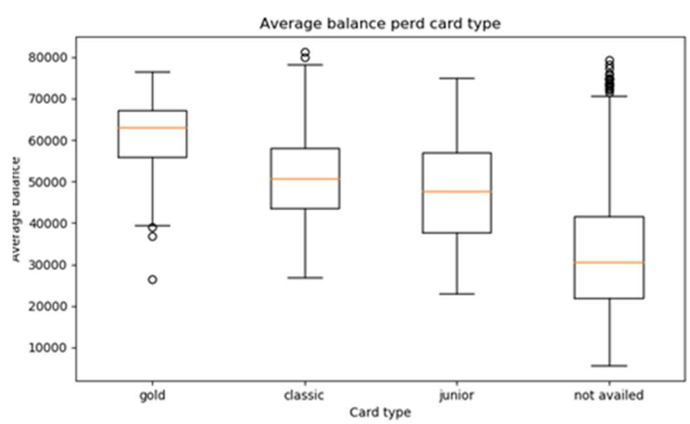
In the demographics part of the report, we noticed that the customers could be identify according to the type of card they own.

In the following graph, one can see the count of customer per card type and per age. We can observe that the junior card is only available for the customer having less than 25 years old. Moreover, we can notice that that number of customers per card type is quite steady with the age from 22 to 60 years and then suddenly drop.



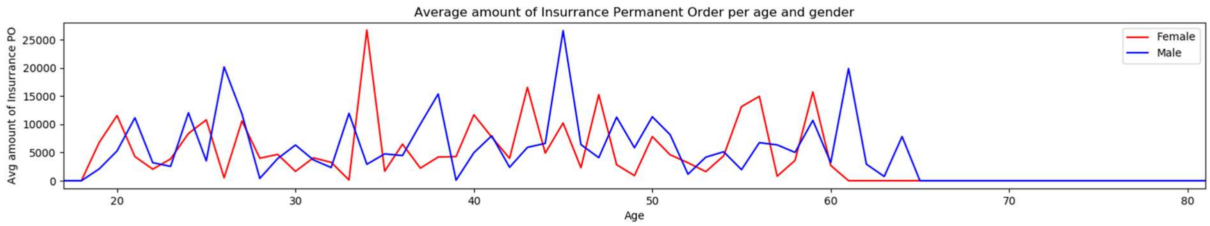
To go deeper, we created a box plot of the bank account's average balance for the customers having a 'gold' credit card, a 'junior' credit card, a 'classic' credit card and for those having no credit card.

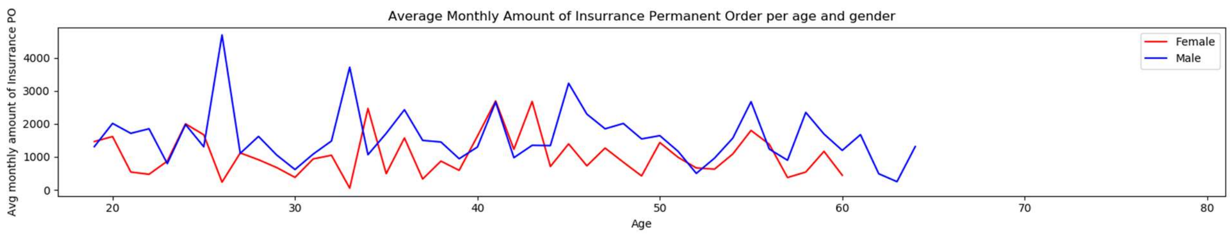
We can observe that, in average, the persons owning a gold credit card have a higher average balance on their bank account, followed by the classic card, the junior card and the others. Thus, we can conclude, that having more money on a bank account may be a good indicator to evaluate if a person is a good or a bad client. You can find a more detailed graphs by gender in appendix (appendix 1.5.2).



1.5 APPENDIX

1.5.1 Permanent Order Insurance





1.5.2 Average Balance per card type and per gender