Financial data set analysis

Services improvement is a strategic component, services that satisfy customer desires or exceed service expectations can help create and maintain long-term relationships with loyal customers

As part of the efforts to know better the customers their behavior, characteristics and as consequence improve bank products and services; support the decision-making process and so on, the bank is looking forward to build a customer data mart including all the available information about each one of their clients and the history of their relationship with the bank in terms of transactions, payments, products, loans, demographic information, etc.

This report includes a general explanation of the procedure followed to build the customer data mart with the raw data provided, we also include the description of the variables created during the work.

Finally, we performed some descriptive analysis to get some insights about our customers base, their transactional behavior, etc. We also wanted to understand the historic data of the bank and products performance for the period of analysis.

Explanation of the procedure:

The data used contains records for six years from January 1st, 1993 to December 31st, 1998.

It is important to mention that in this bank, one account can be managed by different customers, when this is the case, we will have two types of dispositions: Owner of the account and user. This last one cannot issue permanent orders on the account or ask for a loan.

As stated before, the main goal of the work was to build a single data mart on customer level, meaning one record per customer. For this reason, some transformations in some of the relations was needed to have in every table a key variable with unique values; having this we will be able to merge all the relations into one single data mart.

We are going to briefly explain the transformations and new variables created for each one of them:

Permanent Orders: We calculated for each account number the number of orders per each type of payment (insurance, household, leasing or loan); and, the amount payed for the same type of payments.

Transactions: each record describes one transaction on an account, starting from this we calculated the following variables:

- Sum of the amount of credit transactions and withdrawals transactions.
- Sum of the amount of transactions for each mode of transaction as follows: Credit card withdrawals, credit in cash, collection from another bank, withdrawal in cash and remittance to another bank.
- We created new variables to analyze the transactional behavior of the customers in year's quarters. These variables contain the number of transactions made in every account for a specific quarter and year.

Client: Date of birth and gender were extracted from birth number in raw data. The number is in the form YYMMDD for men and YYMM+50DD for women. Where YYMMDD is the date of birth.

Having date of birth, we calculated the age for each customer on January 1st, 2000; one year later to our period of analysis.

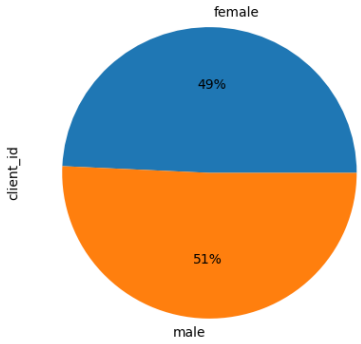
Loan, Account, disposition and demographics: These relations were already with a key variable with unique values, so basically, we just used them to merge and keep on complementing the information we were building for our customers to have the complete view for each one of them.

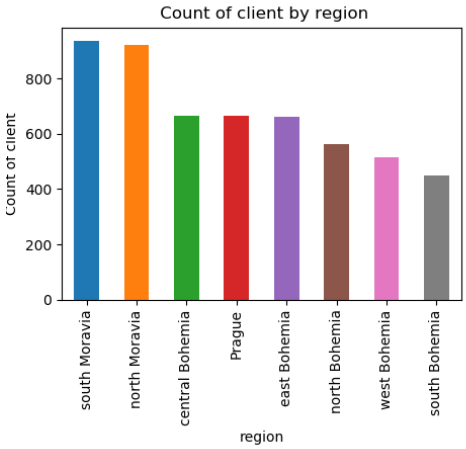
Descriptive analysis:

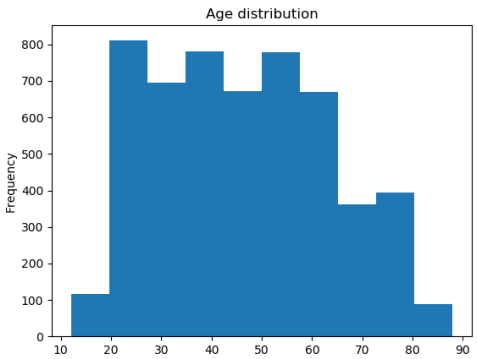
Demographics:

Clients are located in eight different regions being South Moravia the region with the highest concentration of clients. Regarding gender, we have a balanced separation between male and female as we can see in the graph below.

Client's age goes from 12 to 88 years old, however the most par of them are inside the middle age ranges.





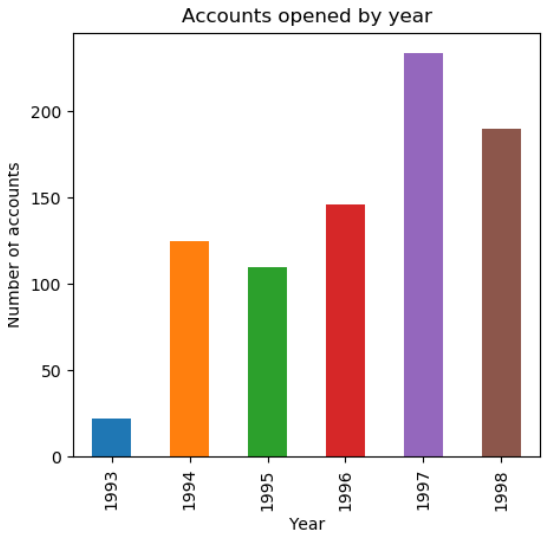


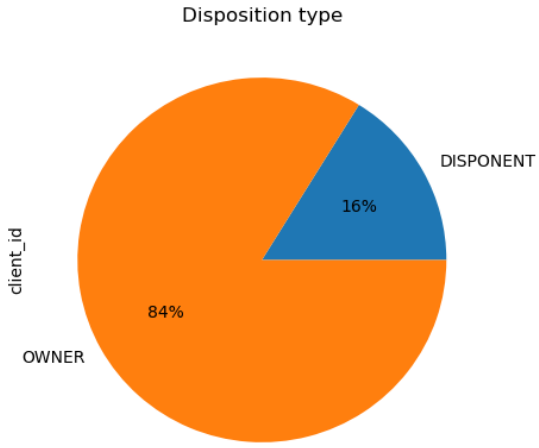
Accounts:

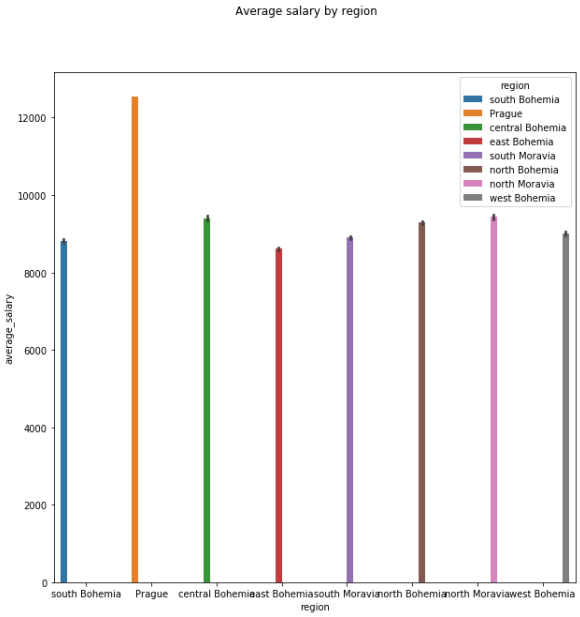
Accounts opened has increase during the years, where 1997 has the most account opened. With a decline as result in 1998.

Within chart disposition type, 84% of the client are owners, 16% are disponent. With this result, although majority has their own account but still there's 16% of client uses the account of others, they are not allowed to ask for a loan and issue permanent order.

Prague has the highest average salary.

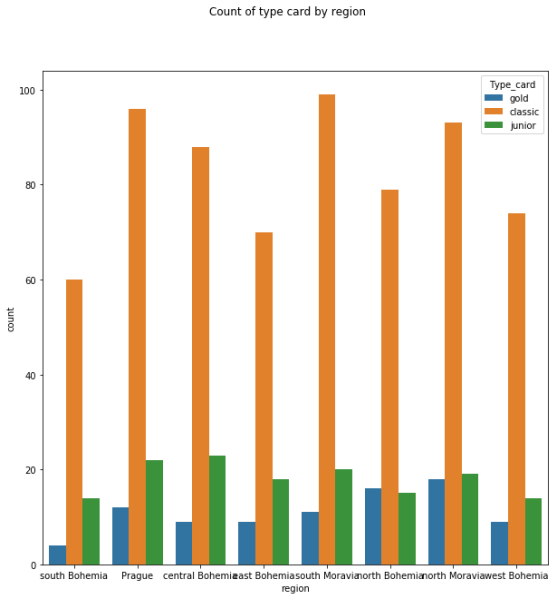


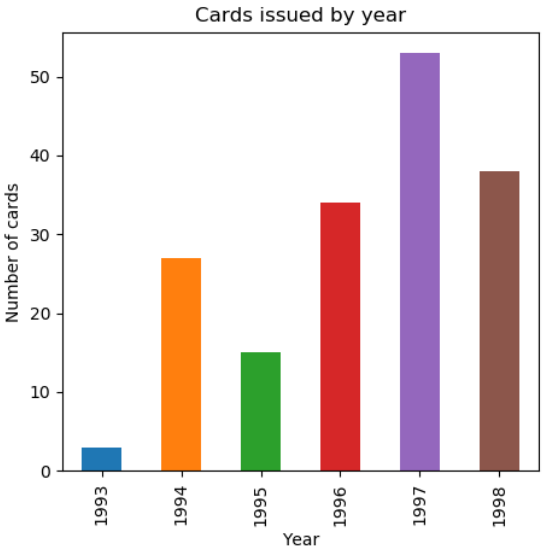




Credit cards:

In chart 1 we can see count of card type issued per region. Chart 2 number of issued cards from 1993 to 1998. Classic is the most popular type of card. In the same way, in 1997 the bank issued the highest number of cards to its customers.





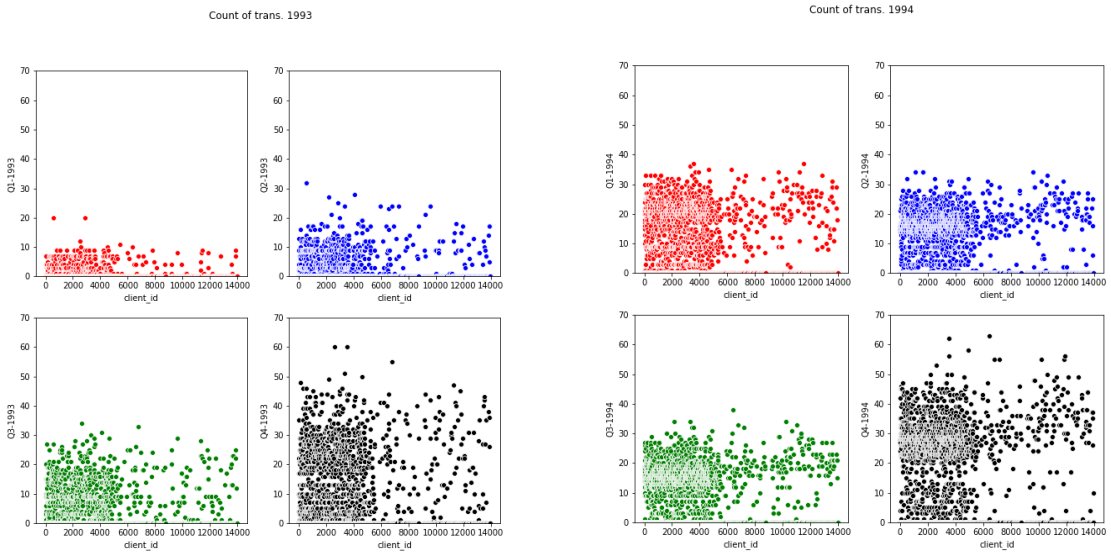
Transactions:

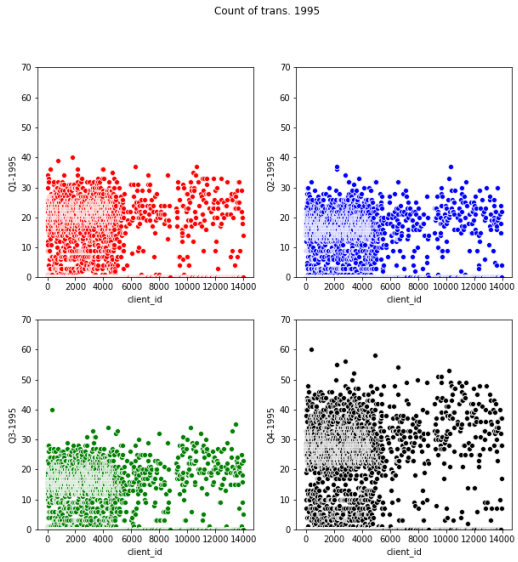
GIF

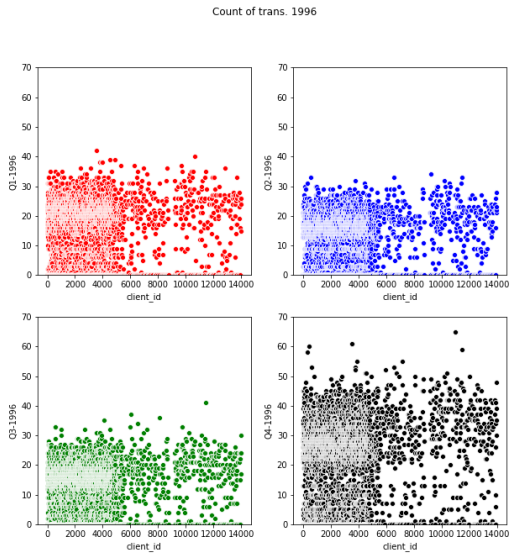
Graphs under are all count of transaction per client, from 1993 to 1998 per quarter (Q1 to Q4).

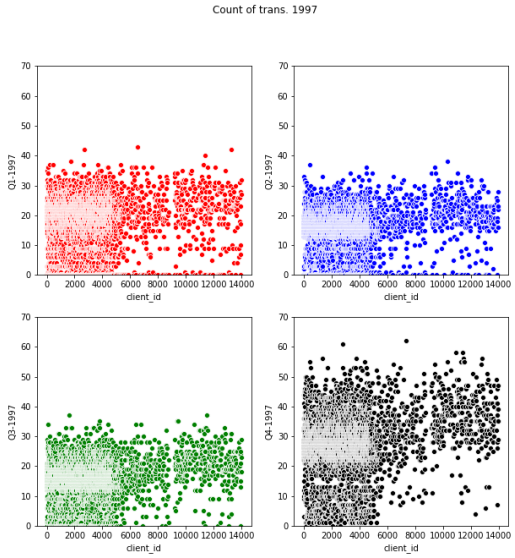
Left side density is much higher than right side is due to the number of clients, client id above 6000 has decrease.

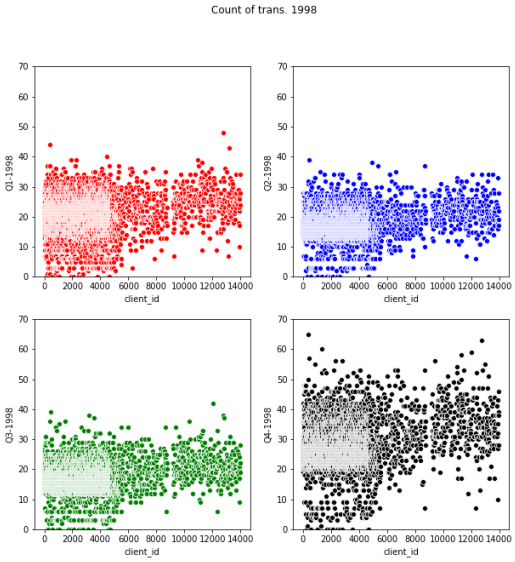
In these plots we can see that the highest active period is Q1 and Q4, Q3 is the lowest. Number of transactions per account starts to increase during the years, scatter plot result starts to shift up. On Q4 has the most amounts of clients who has high number of transaction than clients in general, seem to be outlier.





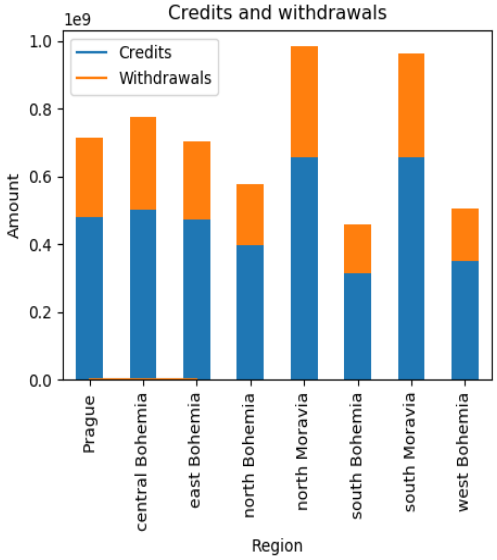


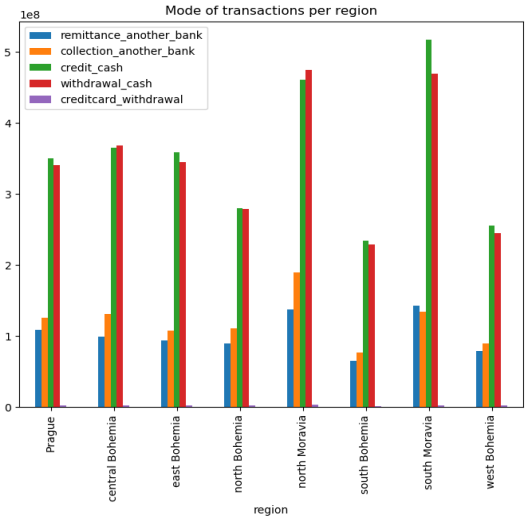




North and South Moravia are the regions with the highest amount of transactions among the eight regions represented. Overall, the proportion on credits transactions is always greater than withdrawals.

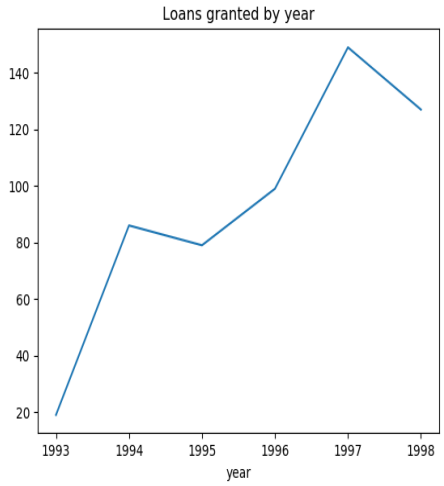
Moreover, cash is the most common way of doing transactions for clients; credit in cash and withdrawal in cash are the mode of transactions with the highest amount mobilized.

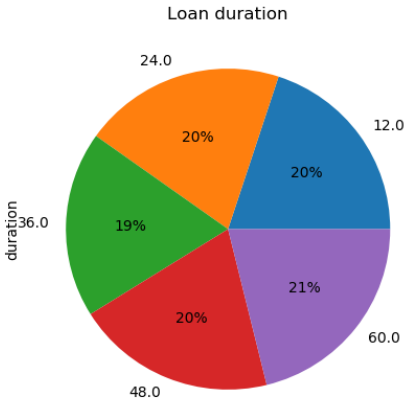




Loans:

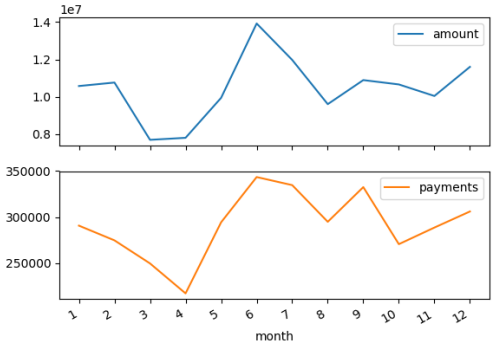
We observe an increasing tendency of loans granted every year until 1997.The total of loans granted by the bank have a duration of 12,24,36,48, or 60 months evenly divided in each group.

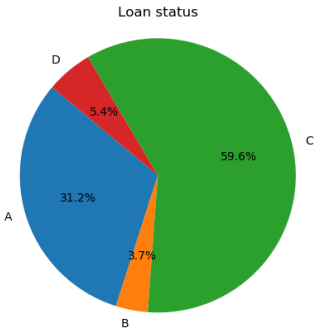




The tendency of payments and amount granted for the loans have a similar behavior during the months for the years of the period under analysis. This can lead us to think that clients are having a healthy payment schedule.

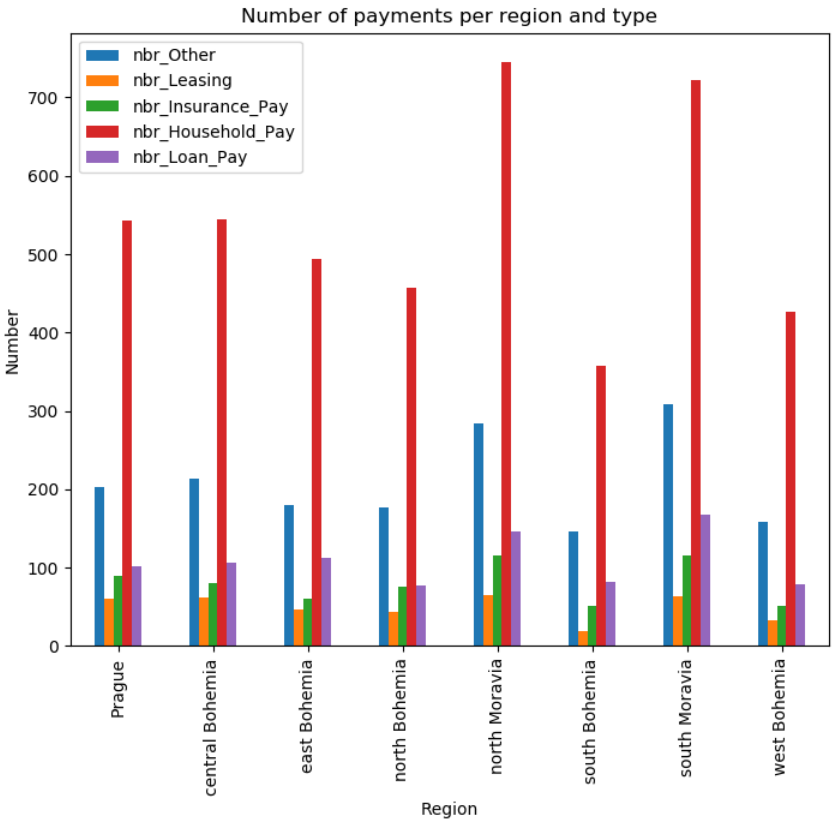
We can also confirm this affirmation by looking at the plot of status of loans, that show us that around 60% of loans have a status of *running contract, OK so far*.





Orders:

The graph below shows the number of orders grouped by type of payment per each region where the bank has presence. For each region the outstanding bars represent Household payments, followed by other kind of payments not specified in the data provided.



There is a wide range of possibilities to keep on exploring and getting more insights form our final data mart. This data mart is also the first step and input to perform some predictive analysis on the future to be able for instance to predict whether a customer will fail in loan payments, churn, which customers are more likely to acquired certain product, to which customers should I offer this specific kind of product and so on.