CREDIT CARD AND BANK LOAN ANALYSIS

Report from IESEG Consultants to the Bank

MA Mui Han

GHOSLYA Aazad

**COLOMBIER Remi**

Table of Contents

1. BACKGOUND

The objective of this report is to analyze the situation of credit card and bank loan for the Bank in Czech Republic to help the Bank making a better decision for promoting its credit card and bank loan services.

In this report, we studied the data provided by the Bank, made the basetable and analyzed by using graphs with explanations according to the following time window.
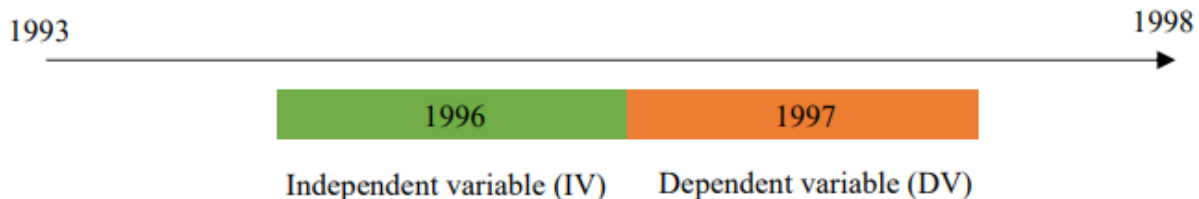
The following data tables are given by the Bank:

1. Client table
2. Account table
3. Disposition table
4. Credit card table
5. Transaction table
6. Loan table
7. Permanent order table
8. Demographic data table

2. DATA CORRECTION AND TRANSFORMATION

To carry out the analyses and to be as relevant as possible, we carried out a thorough cleaning of the various tables.

a. Client table

First, it is essential to check the null values in the client table with the function *client.isna.sum()*
Then, in order to carry out relevant analysis, we created the following variables:
- *birth_year:* calculates the age of the client from the date number
- *birth_month :* calculates the month of birth of the customer, this will help us to identify the gender of the customer.
- *gender:* Male or female, when the month of the date number is greater than 50 then it is a woman and less than 50 it is a man.
- *age_group:* groups customers according to their age range every 10 years.
- *DOB:* the date of birth of the customer formatted DD-MM-YYYY.

Once all the variables are created, we can delete the date number column which is no longer useful. Then we check for age outliers to make sure we do not have any erroneous data with the formula *client["client_age"].max()* and *client["client_age"].min(). As the client age range is from 9 to 85 years old, it represents that there is low risk of having age outliers.*

b. Account table

The first thing to do is to check for null values. Then create the interesting variables from the existing one:
- *acc_open_year :* year the customer opened his account, this will help us to find the number of years he has been a customer since his registration
- *client_age_opening:* age when the client opened his account
- *lor:* length of relationship of the client

Then we delete the date variable as it is no longer useful.

c. Disposition table

We check the null values to start with and then create 2 table to have all the owner data and all the disponent data. To do this we perform a *pd.merge()* to pool the data from disposition table with account table.
And we add a condition to collect only the date data from the dependent variable time window (1997). These two data frames will be useful for locating target variable.

d. Credit card table

Important to check the null values of the credit card table. Then we create different variables useful for our analysis:
- *issued_year*: year when the card was issued.

The following table are created from selected different period from the existing credit card table:
- A. card_new_1996
  Information of the new credit card in 1996.
- B. card_new_1997
  New credit card in 1997, value 1 is added to indicate these are new in 1997.

They are then being merged to basetable which mention in the later section.

e. Transaction table

In this table we have null values. It is imperative to process them so that they do not distort all our analysis later on. So we replace the null values of operation and k_symbol by 0 with the function *.fillna(0)*

Then we create different variables useful for our analysis:

- trans_year: in order to have the year of the transaction and to be able to select the desired year
- amount_withdrawal and amount_credit: this is the sum of credits and withdrawals per account.

We first created the table to get the withdrawal information and then we merged the two table to the trans table in order to gather the variables in the table with only the owners.

- *Credit_card_withdrawl:* Here we have extracted the transactions containing the word 'VYBER KARTOU' meaning that it is a withdrawal with a credit card.
- *Credit_in_cash:* Here we have extracted the operations containing the word 'VKLAD' meaning that it is a credit in cash.

- *Collection_from_another_bank:* Here we have extracted operations containing the word 'PREVOD Z UCTU' meaning that it is a collection from another bank.
- *Cash_withdrawl:* Here we have extracted operations containing the word 'VYBER' meaning that it is a cash withdrawal.
- *Remittance_to_another_bank :* Here we have extracted the operations containing the word 'PREVOD NA UCET' meaning a remittance to another bank.

Then we merge these variables with the df_owner table. These variables are explained in the first part of this document, which is the characterisation of the transaction: *insurrance_payment_payment, statement_payment, intrest_credited, sanction intrest, household_payment, pension received, loan_payment.*

We also merge these variables from the transaction characteristics to the df_owner table. Then we create the variable *trans_last_date* and *last_balance* to get accurate information about the date of the last transaction and the last balance of the customer. We then replace all null values with 0 so as not to distort our analysis.

f. Loan table
We calculate the *variables loan_year, loan_months, loan_day* and *loan_enddate*. The following table are created from selected different period from the existing credit card table:

A. loan_new_1996
   Information of the new bank loan in 1996.
B. loan_new_1997
   New bank loan in 1997, value 1 is added to indicate these are new in 1997.

They are then being merged to basetable which mention in the later section.

g. Permanent order table
We perform the check for null values.

h. Demographic data table
We perform the check for null values.

i. Basetable
Finally, we create our final basetable and add all the tables to each other with left joins. Doing a quick inspection for missing values and replacing them with 0.

3. VISUALIZATION

Today the Bank only have a vague idea of who their customers are. But fortunately, the Bank stores all data from customers. This is rich and complete data of more than 5,000 customers with more than 1,000,000 transactions. And thanks to this data we will be able to give some business insights on the good customers and areas of improvement on the less good ones. The below study is going to illustrate more on the identity and behavior of the bank account owners and this will allow the Bank to offer more adapted services to each one.

a. Background Analysis

Firstly, take a look at the gender distribution for account owner who opened their bank account before 1996. we can see that 48.4% of the customers are women and 51.6% are men. It is important to keep a good parity in these customers, it allows to have a good image with the customers and a diversity. Analyzing the transactions of men and women makes it possible to detect, for example, whether two people live under the same roof, or have 1 child. All this information is essential for targeting customers and offering them the best possible service.
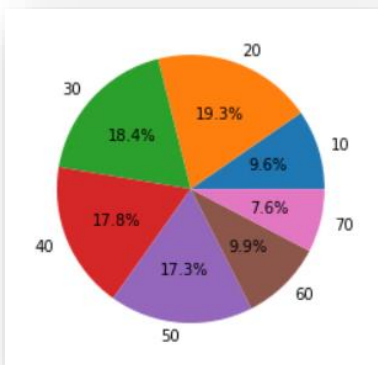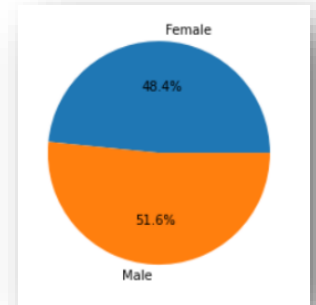
But what is interesting to look at is the age range for account owner who open their bank account before 1996. For example, here on this chart, in blue, we can see that 9.6% of the customers are between 10 and 20 years old. The Bank should identify who their parents are and offer them a personalized service to manage their children's accounts so that they remain loyal to the Bank in the years to come. The bank have a wide spread of the age range which is from 9 to 85. And the service offered to 18-year-olds and 80-year-old pensioners is not the same so it is important to tailor the service to the customers' needs.

This bar chart provides a clear idea of the number of account owner who open their bank account before 1996 based on their age group. With orange indicating the number of female account owner and blue the number of male account owner.
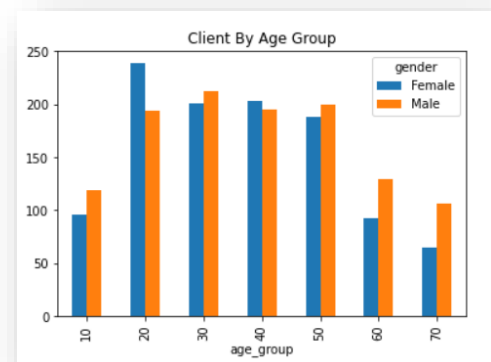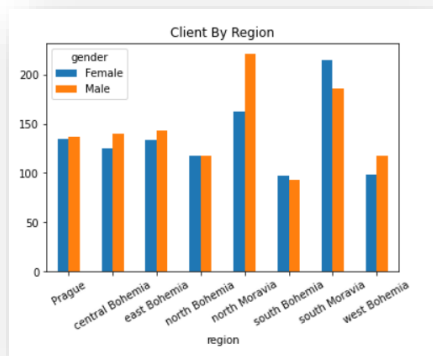
This bar chart provides an overview of the number of account owners who opened their accounts before 1996. This shows the Bank should better target the regions where there is more accounts opened to optimize the various services to them.
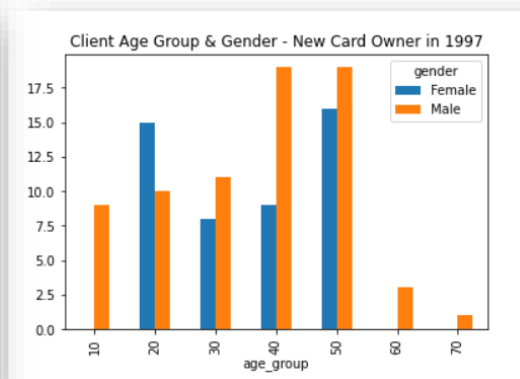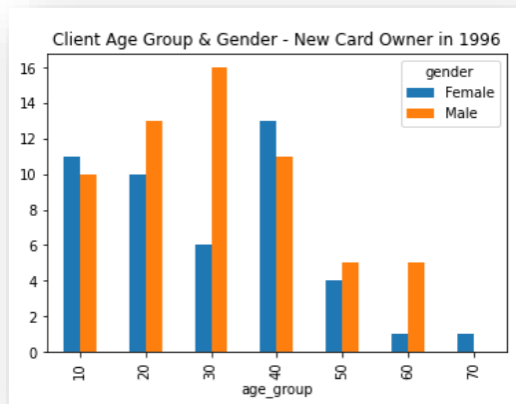
b. Credit Card analysis

The Bank have 191 clients who are new owner of credit card in 1996. Of these, 43.4% are women and 57.1% are men. Of these customers more 50% are between 20 and 40 years old. It is this age group that the Bank should tailor their credit card services to. People over 70 represent only 0.9% of the customers who have new cards, therefore a senior card is not very significant. In 1997, there were 120 new customers with a credit card. Fewer women than last year, since women now represent only 40% of new cardholders.
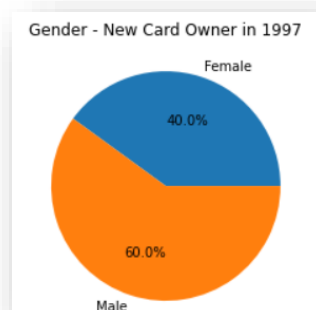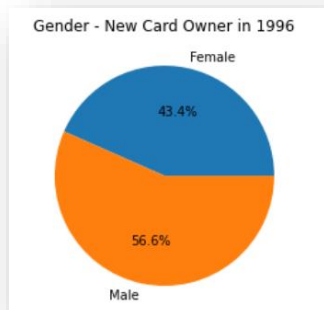
In 1996, only 14.2% of cardholders were between 50 and 60 years old. And yet the following year more than 30% of customers will have their first credit card. The Bank should have targeted this type of customer to achieve this growth. As for customers in the 10-year-old age group, 19.8% of them already have a credit card, but only 7.5% are new cardholders the following year.

Age group - New Card Owner in 1996

Age group - New Card Owner in 1997
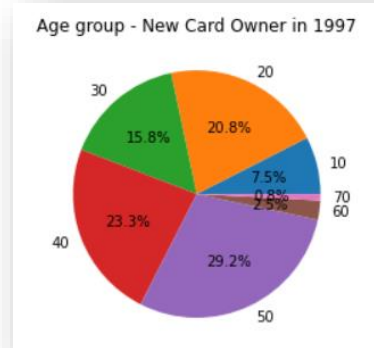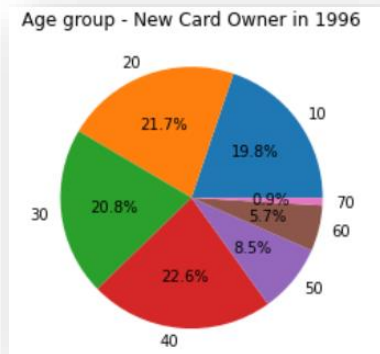
The two graphs below show the distribution of ages and genders. The first graph shows that generally it is more men who have a new credit card in 1996 and the vast majority are between 20 and 40 years old. Whereas in 1997, the majority of new credit card holders will be between 20 and 59 years old. And there is no female customers for age over 60.

Client Age Group & Gender - New Card Owner in 1996

Client Age Group & Gender - New Card Owner in 1997

Below is an overview of credit card clients according to regions and gender. The clients who already have a card in 1996 are in almost all regions. Whereas in 1997 the new cardholders are very few in South Bohemia. The Bank should investigate further to identify the reason.
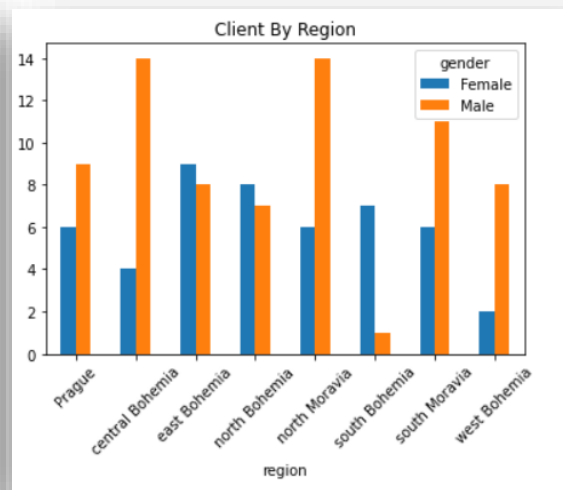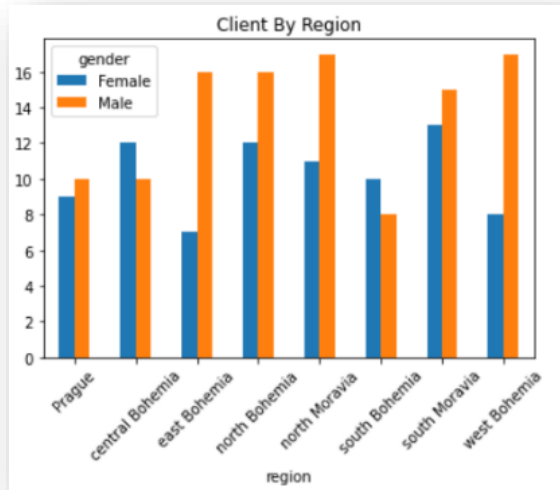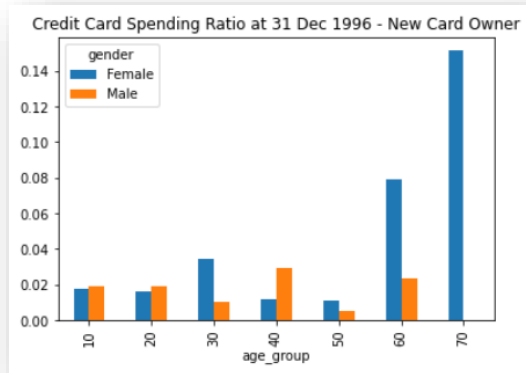

c. New Credit Card Owner in 1996 and 1997

As far as customer consumption is concerned, it is homogeneous between customers who have a new card in 1996 and customers who have a new card in 1997. However, it can be noted that the new customers in 1997 aged over 60 are only men and they spent quite a lot in 1996.
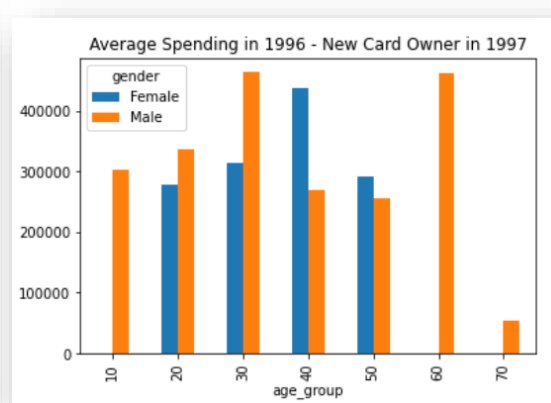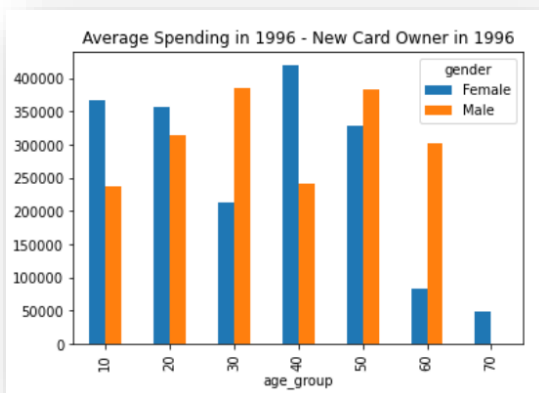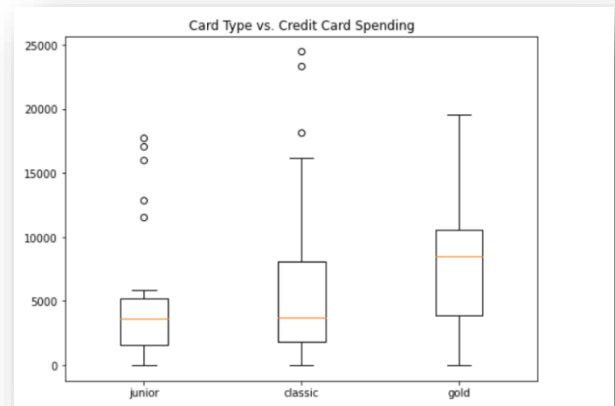
This graph shows the ratio credit card spending to total expenses who already had a card in 1996, according to their age and gender. From the graph, it shows the older the customer, the higher the ratio.

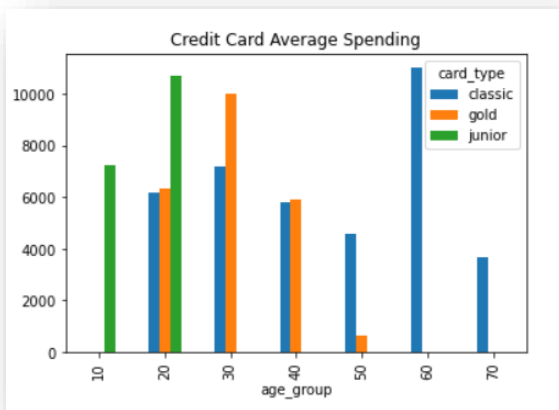Credit Card Spending Ratio at 31 Dec 1996 - New Card Owner

### d. Spending Habit by card type (credit card owner in 1996)

A large majority of customers do not have a card. The Bank should put more promotion for credit card. Indeed, having a card multiplies banking transactions and it is a service that customers appreciate. Then, once the customer has a card, the Bank can offer different card types to the right age group.

On the three types of card, this boxplot shows the spending by card type on customers who already had a card in 1996. The median for the junior card is around 4000 with a 25th and 75th percentile of 2500 and 5000, the classic card has the same median but a much higher 75th percentile at 7500. As for the gold card, it has a median of 8000 and a 25th and 75th percentile of 4000 and 1000. In short, gold cards spend more. The Bank offers 3 types of card, the classic, gold and junior card. Only a few customers use gold card.

Card Type vs. Credit Card Spending

Credit Card Average Spending

From the age of 50 onwards the gold card has little or no effect, so it may be worthwhile to put more resources and capital into the gold card for customers between 20 and 50. The junior card works very well with young people, it is a strength that should be built on. For those over 60, only one type of card is issued. It is important for the Bank to review the policy of getting a gold card and to investigate more on the reason of no client gets a gold card for the age more than 60.

This scatter plot shows the amplitude of credit card spending according to the type of card held by the customer and their age group. This graph explains a lot of things, especially the interest for the classic card which makes the most expenses among the 3 cards. And the junior card which works well for the under 20 years old.

Client Age vs. Credit Card Spending

Client Age vs. Credit Card Spending Ratio
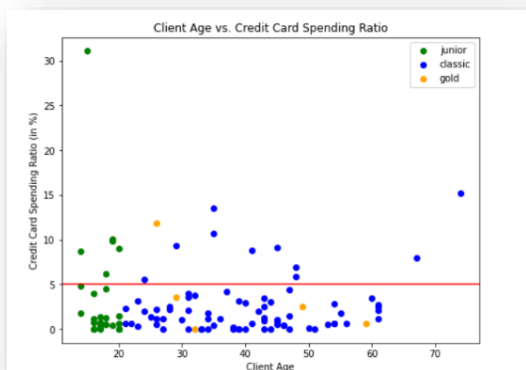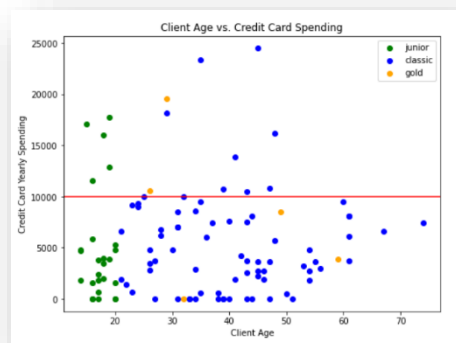
When focusing more on the ages with these two scatter plots, in the bottom right-hand corner that customers with a junior card are maximum 20 years old and their spending can range from 0 to 18000. The age started to use classic card is 20. However, there is only 5 gold credit card holder. In addition, the Bank could also consider to upgrade for those client with a classic credit card with the yearly credit card spending is over EUR10,000, so as to attract more client to use the credit card from the Bank. It is a very low credit card spending ratio to its total expense from the bank account. The majority having a ratio below 5% of their spending.

e. Loan Analysis

For bank loan services, it is a perfect parity between men and women for the new bank loan granted in 1996. In 1997, the distribution becomes 54.8% for women and 45.2% for men. This interest of women in loans is interesting because women were less present in credit card spending. However, for loans women are in the majority 1997.
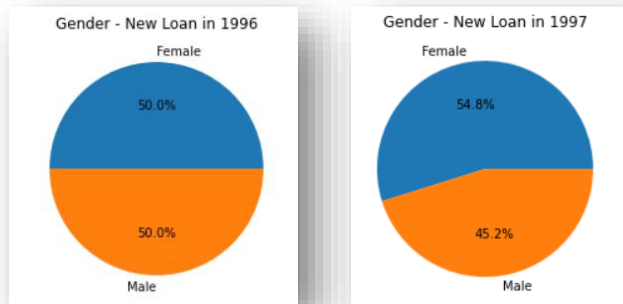

When analyzing the ages of the different clients who made a new loan in 1996 and 1997. There are no major differences.
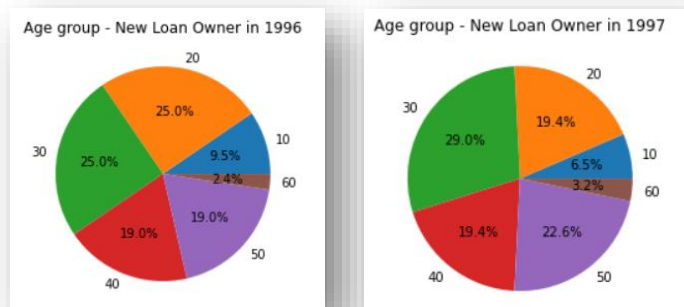
Below show bar charts for new loans in 1996 and 1997 by gender and age group. In 1997, no more client granted a new loan from the 10-year-old age group for men. The Bank should investigate more on this result. And overall, the average age for new loans in 1996 is lower than for new loans in 1997.
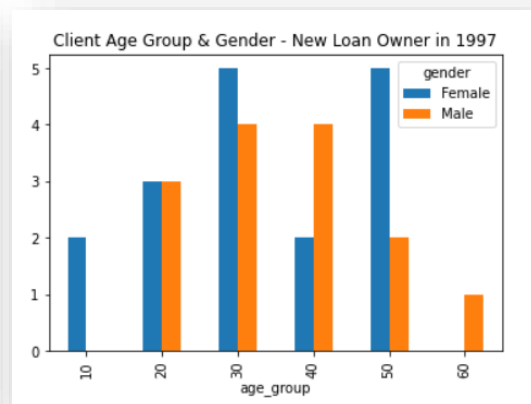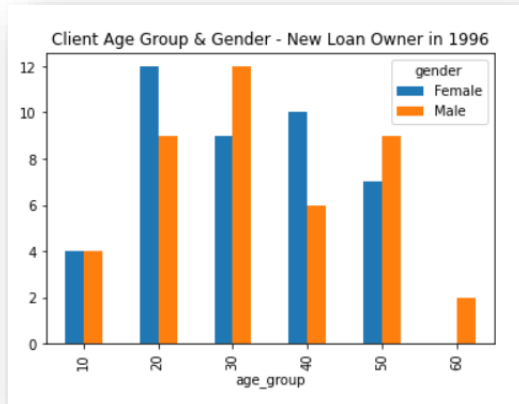
In terms of regions, all regions have new clients making loans in 1997 whereas in 1996 men made no loans in these two regions: East Bohemia and West Bohemia. The Bank should investigate the reason for no clients in 1997 for these two regions.
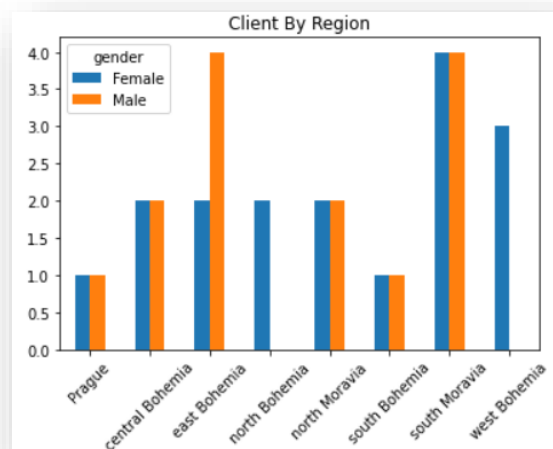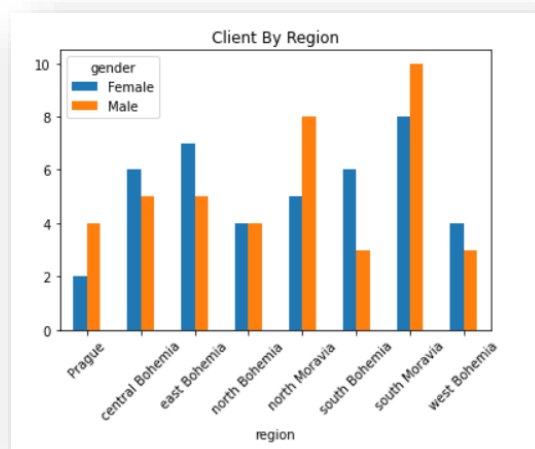
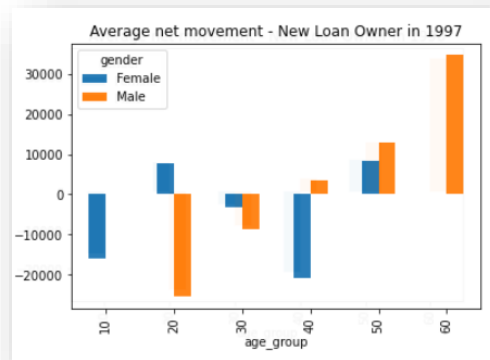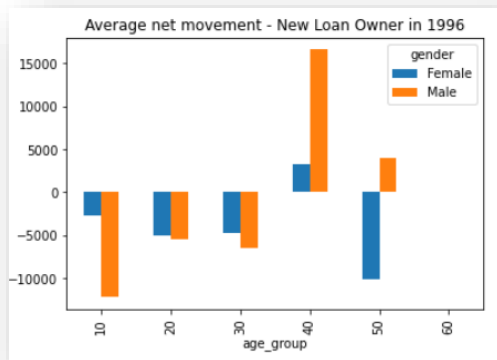## f. Net movement for New loan owner in 1996/1997

These two graphs provide the information on the movement of the bank account net movement for the clients with bank loan. It shows that in average, the clients with bank loan are having negative bank movement, the Bank should have done more regular check to ensure the client have low or no default risk.

g. Loan information for the new loan in 1996
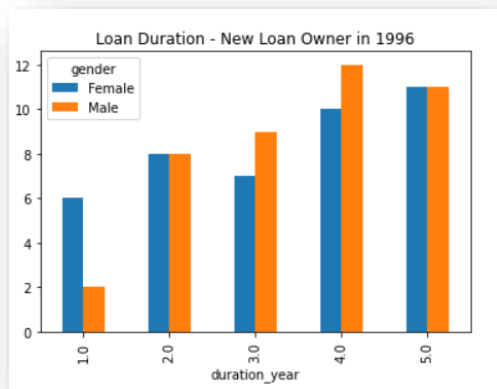
This graph informs the duration of new loans made in 1996. The longer the duration, the more loans are made. Clients tend to want to repay their loans over a longer period. This is a service to be retained but more attention needs to be paid to their ability to repay loans over time. It is fairly mixed male and female for the duration of loans except for the short 1-year loans where there are 3 times more women than men.

This graph shows the ratio of loan payments to expenses (in %). This ratio is between 0.15% and 0.40% and generally varies with age.

Loan Repayment to Spending Ratio (in %) - New Loan Owner in 1996

Loan Duration - New Loan Owner in 1996
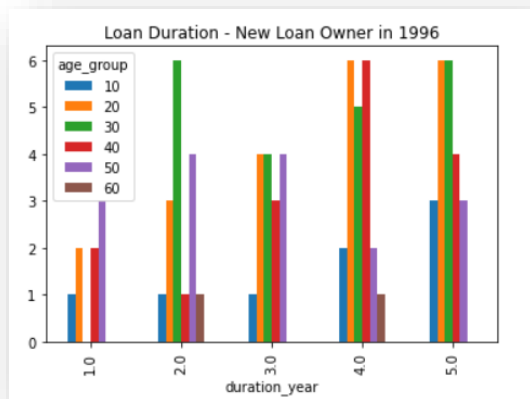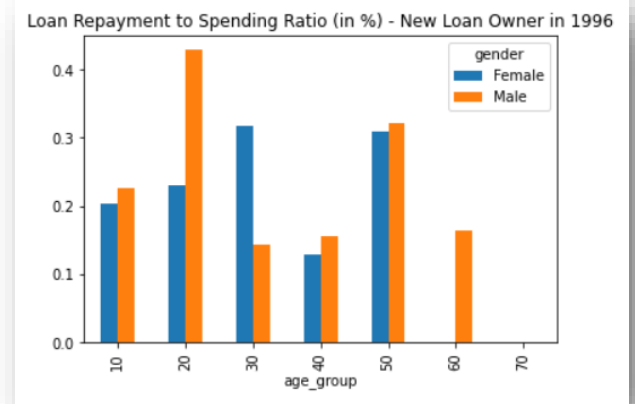
It might be worth for checking whether the interest rates of the Bank are attractive compared to the market and other competitors. The Bank may also broaden their range of loan services, as the total number of the bank loan is low.

4. CONCLUSION

To conclude, the Bank should investigate more of the clients' spending behavior, credit card and loan trend so to make their decision, for instance, credit card policy, loan policy and so on.

In this report, all the techniques of cleaning the database and correcting it to have the most accurate basetable possible are explained. The illustration by graphs and the analysis of this database allowed the Bank to have a better view for making their coming plans to customers.

5. REFERENCE

1. Class material (jupyter notebook) from Financial Programming
2. Matplotlib documentation
   https://matplotlib.org/

APPENDIX I. DESCRIPTION OF VARIABLES

acc_open_date : Integer, bank account open date with the date format YYMMDD.

acc_open_year: Integer, client account opening year, from 1993 to 1995.

acc_owner_district: Integer, unique identifier of the ower account's district.

account_id: Integer, unique identifier for the account.

age_group: Integer, grouped age of clients by 10 years.

amount: Float, the total amount debited by the client for a transaction.

Amount_credit: Float, total credit per account.

Amount_withdrawal: Float, total withdrawal per account.

bank_district: Integer, unique identifier for the district.

birth_day: Integer, client birthday.

birth_month: Integer, client birth month.

birth_year: Integer, client birth year.

card_id: Float, unique identifier for a card.

card_type: Categorical, type of card of the client, possible value: "Junior", "classic", "gold".

Cash_withdrawl: Float, mode of transaction: "VYBER", withdrawal in cash.

client_age: Integer, age of the client, from 9 to 85 years old.

client_age_opening: Float, client age at the year they open bank account.

client_id: Integer, unique identifier for the client.

Collection_from_antoher_bank: Float, mode of transaction: "PREVOD Z UCTU", collection from another bank.

Credit_Card_withdrawal : Float, mode of transaction : "VYBER KARTOU", credit card withdrawal.

Credit_in_cash: Float, mode of transaction: "VKLAD", credit in cash.

disp_id: Integer, unique identifier for the disposition.

district_name: Character, district name of the account owner.

DOB: String, date of birth of the client in the format DD-MM-YYYY.

duration_year: Float, contract duration in year.

frequency: Strings, frequency of issuance of statements.

gender: Categorical, client gender M or F.

have_card_1996: Float, client who have card in 1996. 1 if yes, nan if no.

Houselhold_payment: characterization of the payment: 'SIPO' stands for household payment.

insurrance_payment_payment: Float, characterization of the payment: 'POJISTNE' stands for insurance payment.

Intrest_Credited: Float, characterization of the transaction: 'UROK' stands for interest credited.

issued_year: Float, year card issued.

last_balance: Float, last balance of the client.

loan_date: Integer, date number in original format which is the number of days since 1960.

loan_endate: Float, contract end date.

loan_endyear: Float, contract end year.

loan_id: Integer, unique identifier for the loan.

Loan_payment: Float, characterization of the payment: 'UVER' stands for loan payment.

loan_repayment_history: Loan classification repayment history as at 31 Dec 1996, possible value: good_record and default_record.

loan_year: Float, year of the loan.

lor: Integer, length of relationship in year.

new_card_1997: Float, client with new credit card issued in 1997 = 1.

new_loan_1997: Float, client who get a new loan in 1997 = 1.

Number_trans_credit: Float, number of credit per account.

Number_trans_withdrawal: Float, number of withdrawals per account.

payments: Float, monthly payments of the loan.

Pension_received: Float, mode of transaction: 'DUCHOD' stands for old age pension.

region: Character, region name of the account owner.

Remittance_to_another_bank: Float, mode of transaction: 'PREVOD NA UCET', remittance to another bank.

sanction_intrest: Float, characterization of the transaction: 'SANKC. UROK', sanction interest if negative balance.

statement_payment: Float, characterization of the transaction: 'SLUZBY', stands for payment for statement.

trans_last_date: Integer, last transaction date.

type_of_disp: Type of client = owner.