FINANCIAL PROGRAMMING

MBDA 2020 - Group Project

KEDA Ndjaidoukmi
SHAH Pooja
VARNELLI Alessandro

November 29, 2020

INTRODUCTION

As a part of this project, our goal was interpreting the provided client database of a Czech bank, extracted in 1999 and create a rather relevant and constructive data mart for bank managers to conveniently use data to understand their clientele. Following the creation of data mart, we have used analysed the business data to derive meaningful insights with some visualisations that can enlighten the bank managers for appropriate future strategies.

PREPARATION OF DATA MART

To combine only informative columns in our data mart, we studied the eight tables in the database – Account, Client, Demographics, Disposition, Transaction, Loan, Credit Card and Payment Orders. The data was first cleaned and processed using necessary Python libraries, for instance, extracting dates and gender from a particular column format, and changing data types of numeric values.

Next, we created more appropriate calculated columns for the existing tables and ultimately selected only necessary columns from each table to be merged further in the final data mart. The list of new variables and chosen existing variables for each table are noted below (new variables in Italic):

Client Table:

client_id – Existing variable used for merging

birth_year – Variable derive from existing variable birth_number, client year of birth

birth_month – Variable derive from existing variable birth_number, client month of month

birth_day – Variable derive from existing variable birth_number, client day of birth

gender – Variable derive from existing variable birth_number

age – Variable obtain from the calculation of difference between 1999 and birth year

age_group – Classification of client in different group base on the age

district_code – Variable used for merging, not present in the data mart

District table (demographic data):

district_code – Variable used for merging, not present in the data mart

district_name – Existing variable, A2

region – Existing variable, A3

Disposition table:

account_id – Variable used for merging

client_id - Variable used for merging

disponents_for_accounts – Number of disponents for the account

client_type – Existing variable, type of disposition

Card table

account_id – Variable used for merging

card_type – Existing variable, type of card

Years_Since_Card – variable obtain from calculation of difference between 1999 and year of issued

Account:

account_id – A meta variable, used for merging with order table and disposition table while generating the data mart.

acc_open_year – The year in which the client opened an account with the bank. This is processed and obtained from the 'date' variable in the existing table.

lor_with_bank – Determining the length of relationship with the bank. Derived by subtracting account opening year from 1999 (the year when data was extracted).

Transaction table:

account_id – Variable used for merging

min_trans_amount – Minimum single transaction amount for each account

max_trans_amount – Maximum single transaction amount for each account

mean_trans_amount – Average amount of all transaction for each account

Nbr_trans – Number of transactions for each account

Loan Table:

client_id – Variable used for merging

account_id – Variable used for merging

total_loan – Total amount of loans per client

Nbr_loan – Number of loans per client

Loan_duration – Minimum duration of a loan

last_balance – Last balance (most recent) for each account

loan_status – Loan status for each account

Permanent Order:

account_id – Identifier variable for merging with account table.

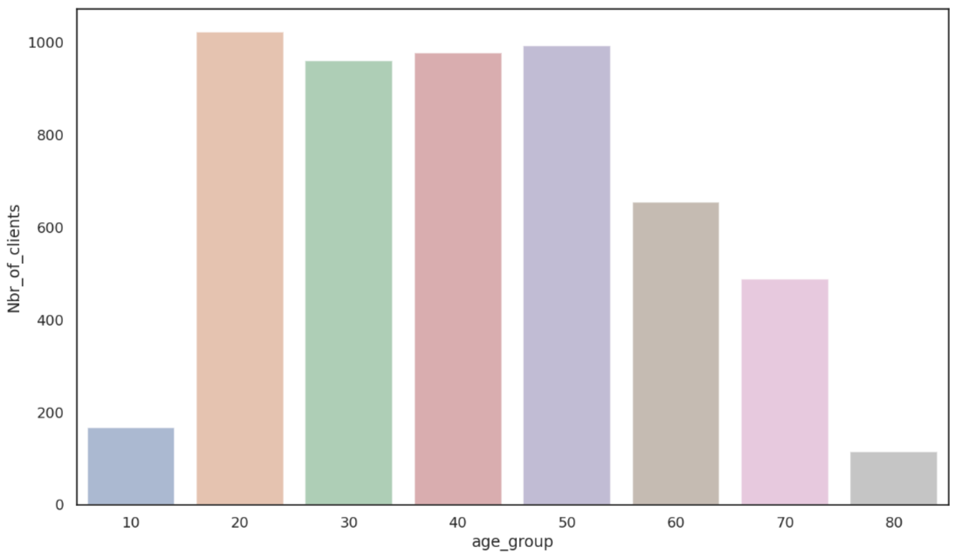
nbr_of_orders – A count of total number of payment orders per client computed by grouping the client ids.

sum_of_all_orders – The total monetary value of payment orders per client, calculated from the existing amount variable.

BUSINESS INSIGHTS

1)Demographic Analysis of Customers

Number of clients having an account with the bank per age-group:

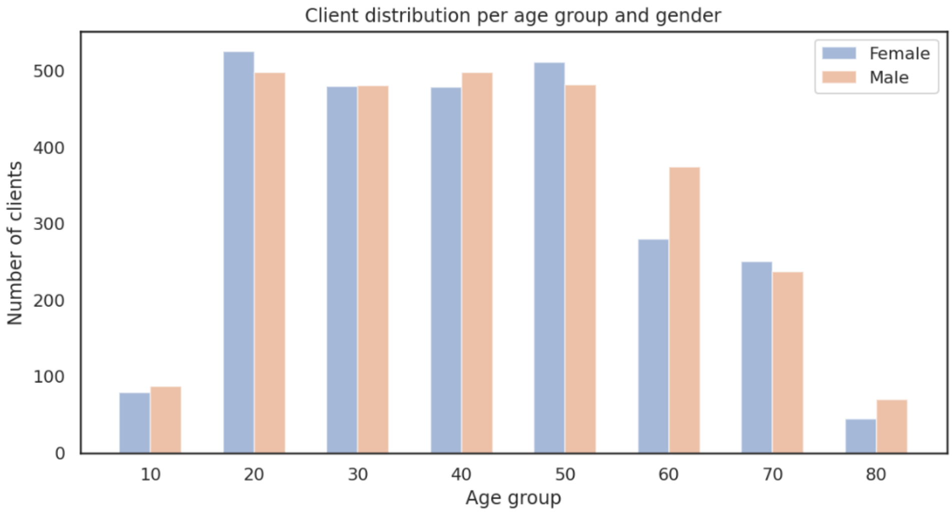


From the clients table, we know that there are 5369 clients.

The bar graph suggests that majority of the clients fall under the age group of 20 to 50 years, close to 1000 clients per group. However, there is a decline in number of clients for the age-group of 60 and onwards. Age-group of 20 has the highest number of clients while 80 years has the lowest.

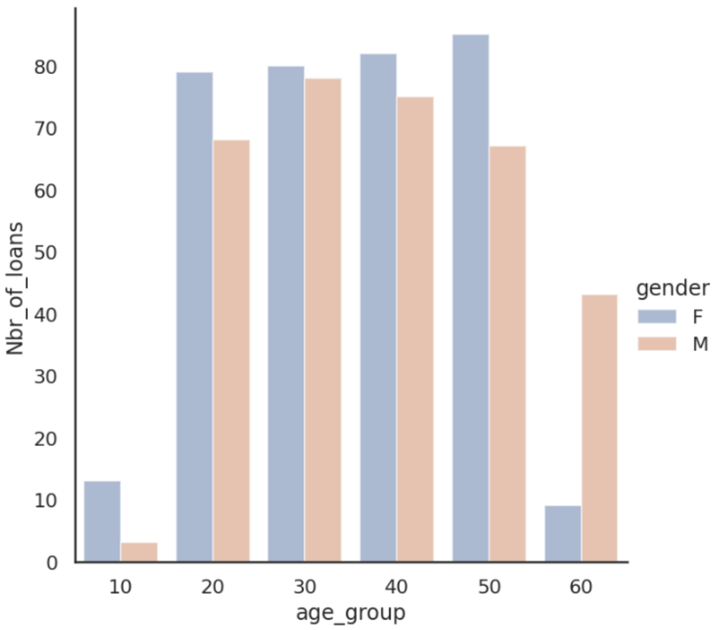
In the next few graphs, we also notice that the age-group of 20-50 years is also crucial for the bank for its clientele of loans and credit cards. Hence, they are also the most revenue generating group. Bank managers can adopt a strategy of targeted campaigns for new customer acquisition for young to middle-aged population and try to retain them for lifetime to eventually have a sizable number of older population in the future.

Clients per gender and age-group:



A significant difference is observed in the age group of 60 and 80 years where the number of females is quite low compared to the males of the same category. Otherwise, the ratio of total male and total female clients is quite similar (females being slightly above the average) for other age groups. As both genders play an equivalent role in bank's revenue system, the managers should target both and ensure good service and offers to both.

Loans per age group and gender:



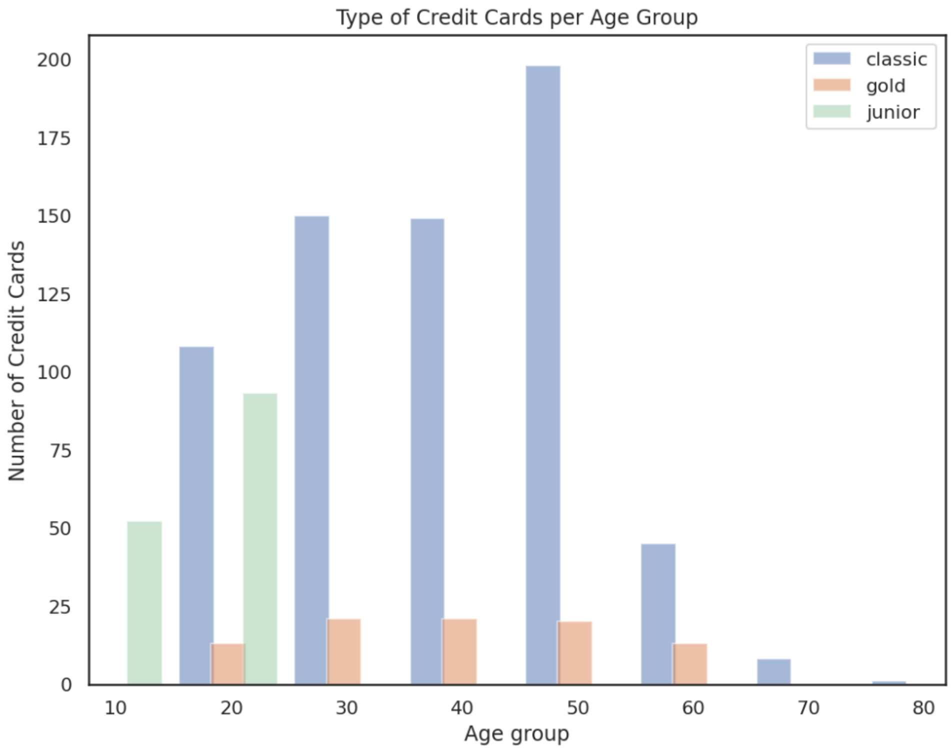
From the loan table we know that 682 clients have borrowed loan from us so far of the total 5369 clients which is 12.7%.

The above graph is interesting. We see that from age 10 to 50, more females have borrowed loans compared to males. On the contrary, for age group 60, suddenly the number of loans taken by females have gone drastically down to around 10% from ~85%. There are no loan borrowers after the age-group of 60 which makes sense. Looking at the total count of loans (males & females), age-group of

30 and 40 have maximum borrowers which also seems logical because most home, auto and personal loans would be required at this stage of life.

2) Credit Card Analysis

Credit cards per age group:

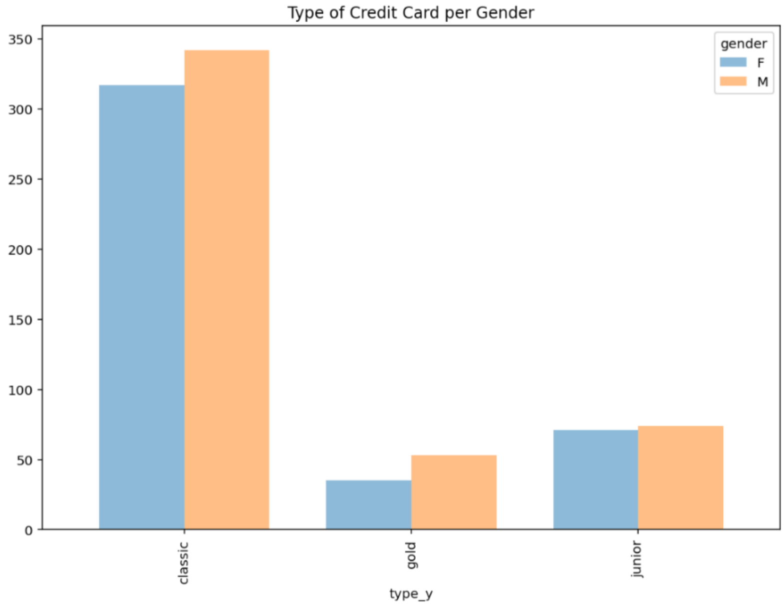


As we have been observing from the previous graphs, clients from 30-50 years participate the most in different bank services. This graph is no exception. Nevertheless, maximum card holders are in the age group of 50. There are hardly any credit card holders above the age of 70.

The credit card table from original dataset shows that 892 clients hold a credit card which is 16.6% of the total clients.

Looking at the distribution between 3 types of credit cards, we notice that most of the clients have a Classic card. The ratio of junior card holders (145) against the total number of clients below 20 years (1188) is very less ~12%. Gold card holders (88) are less than 10% of the total card holders (892).

Distribution of credit cards between gender:



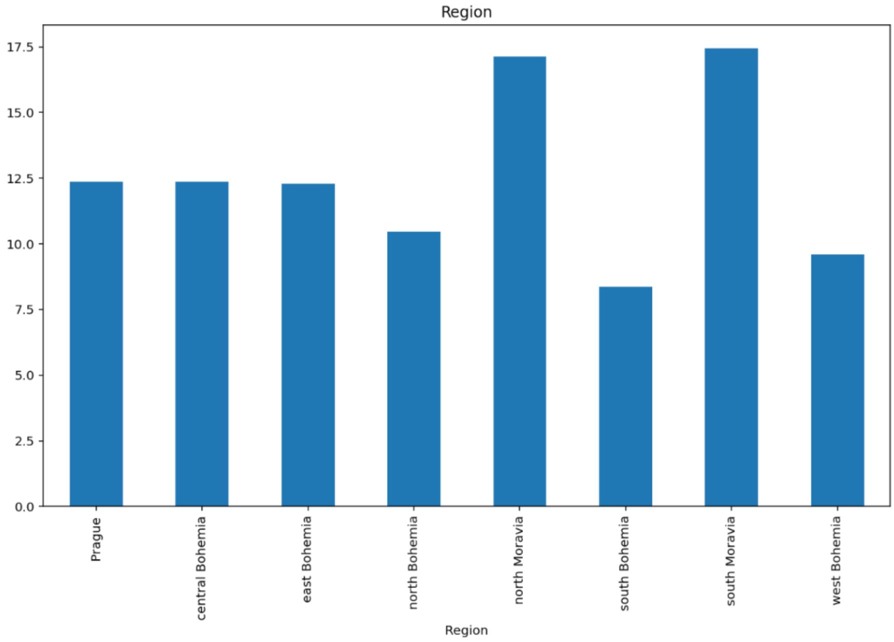
In contrast with distribution of loans where females seemed to lead in most age-groups, for credit cards, it is the males who have a greater % share. However, the difference in numbers between genders is not remarkably high.

Card type issuance distribution over the years:

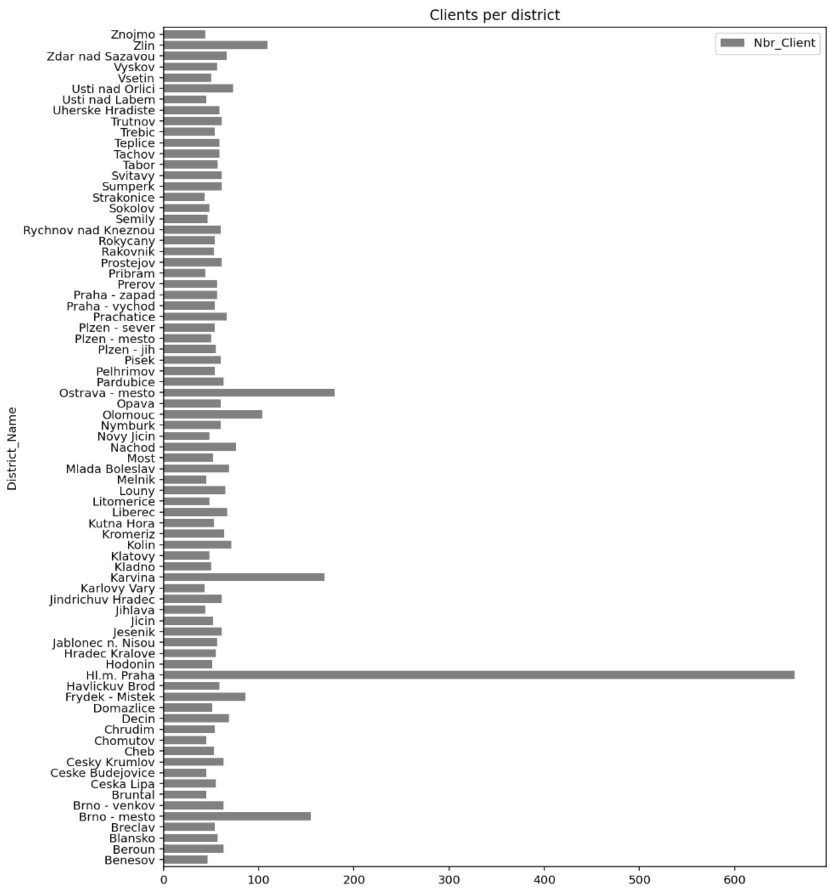
type	issue_year	card_id
classic	**1993**	1
	1994	17
	1995	42
	1996	83
	1997	190
	1998	326
gold	**1995**	4
	1996	5
	1997	26
	1998	53
junior	**1994**	4
	1995	17
	1996	28
	1997	26
	1998	70

From this aggregation by year and by type we can see the positive trend in card issuance for all types during the years. As clearly seen from previous graphs too, ”classic” type is the most issued by the bank.

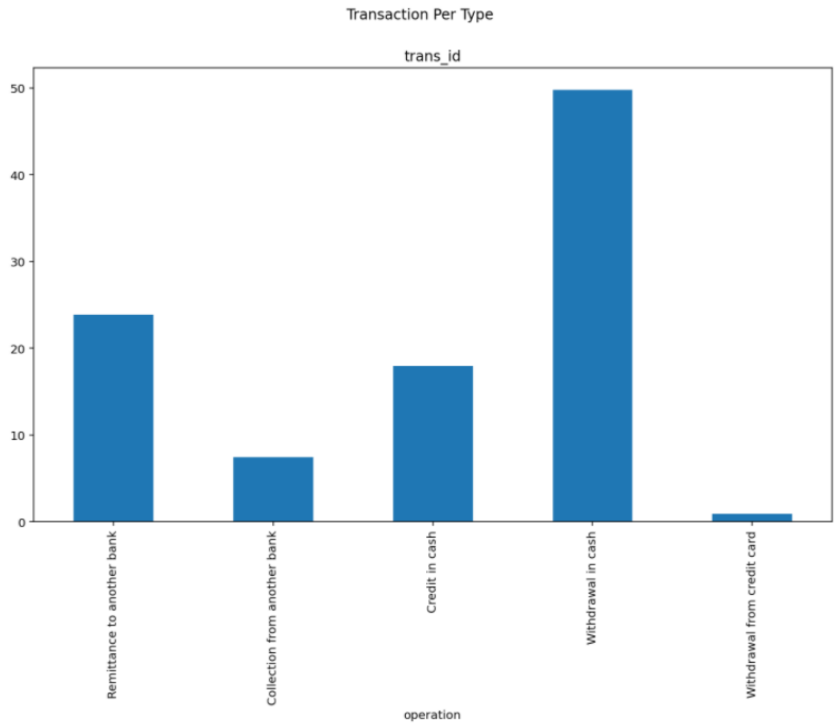
Client repartition per region:



Most of the bank client are in the region of north Moravia and south Moravia. However, Prague that will need more attention from the bank because he has the most attractive district as shown below



3) Transaction analysis



Most of the transaction done in the bank are withdrawal in cash. It represents around 47% of all transaction meanwhile withdrawal from credit card is less than 1%.

4) Loan Status Analysis

loan_status	client_id
A	258
B	31
C	493
D	45

There are 289 finished contracts (A and B) with 10.72% bad debts and 538 running contracts (C and D) with 8.36% potential bad debts. Possibly, the lending of money has stricter background check now than earlier.

loan_status	% of total loans	avg_lor	avg_age	min_ending_balance	avg_ending_balance	max_ending_balance
Perfect (A)	31.20	4.88	41.16	5567.8	56246.65	124542.5
Bad Debt (B)	3.75	5.10	43.97	-6132.5	34477.33	108204.4
Good So Far (C)	59.61	3.24	41.04	7610.9	50442.14	125279.8
Possible Trouble (D)	5.44	3.73	39.51	-5845.8	34769.68	108857.0

Around 91% of the loans can be considered safe (A & C). The below calculation of negative balance reassures that no client with loan status A or C have overdrafts. We notice that the % of D (with running contract) is more than % of B (finished contract). While the average age of A and C borrowers are quite similar, we cannot say the same for B and D.

It is also interesting to know that borrowers whose loan contract have finished but have not yet repaid have an average length of relationship with bank of over 5 years which is more than for any other loan status. This may hint that bank's risk analysis and actuarial calculations have improved overtime. Earlier, they may have been liberal in lending money.

loan_status	nbr_clients_with_negative_bal
B	3
D	7

loan_status	nbr_clients_with_add_bal
A	91
B	2
C	58
D	3

The left table displays a count of clients who have a loan and have negative ending balance before the extraction of data. Only the clients with potential bad loan status have negative balances.

The table on the right shows the number of clients with loans who have an additional balance over the loan amount at the time of closing.

Around 19% of the loan borrower have last balance more than the borrowed amount. This could be a positive thing, but no concrete conclusions can be drawn as the last balance does not indicate future cashflows from client accounts.

The number of borrowers with running contract who have current balance less than total loan amount are – C: 435 and D: 42. Majority of the current borrowers have a virtual deficit balance in their account. This again does not indicate anything substantial but can be valuable information for the bank to keep a track.

loan_status	acc_open_year	client_id
A	1993	131
	1994	34
	1995	32
	1996	53
	1997	8
B	1993	17
	1994	3
	1995	8
	1996	3
C	1993	29
	1994	39
	1995	82
	1996	213
	1997	130
D	1993	7
	1994	4
	1995	9
	1996	20
	1997	5

Here we grouped the datamart by loan_status and age_group to get the count of client_ids for each category.

The higher value in the table is 115 for C status related to people between 30 and 40 years.

It is really realistic as a situation and it reflects the idea of an healthy customer base for loans. We expect from this category of people a higher certainty to pay as they have a higher Life Expectancy/Income Growth ratio in comparison to other categories.

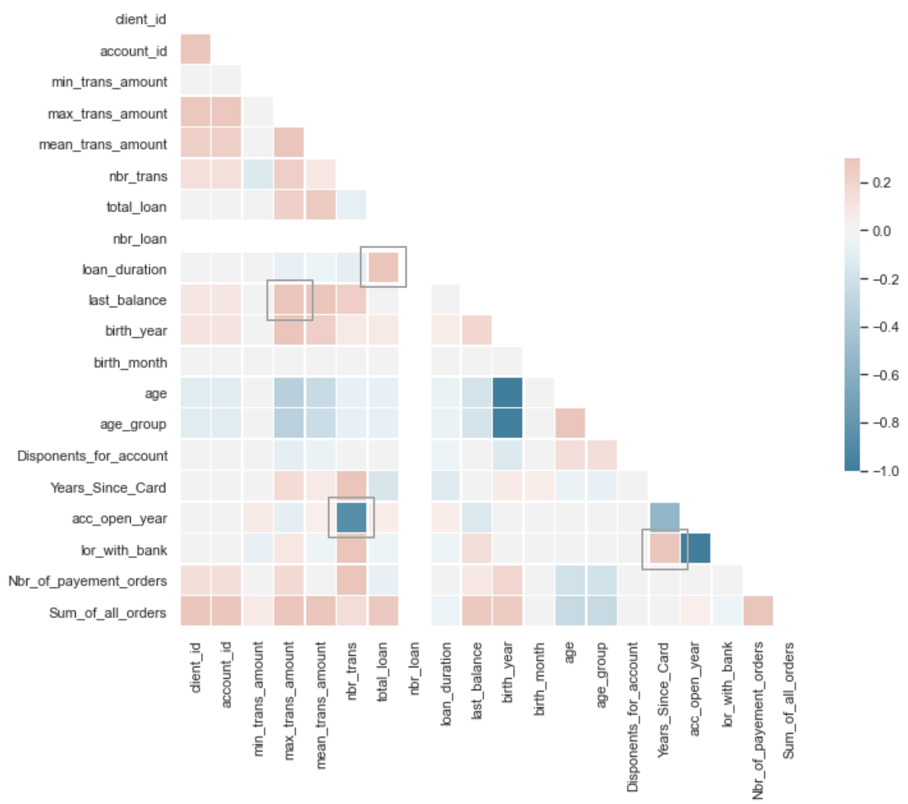
loan_status	age_group	client_id
A	10	7
	20	56
	30	57
	40	56
	50	61
	60	21
B	20	6
	30	4
	40	7
	50	10
	60	4
C	10	13
	20	103
	30	115
	40	114
	50	113
	60	35
D	10	1
	20	11
	30	12
	40	9
	50	6
	60	6

Here we grouped the datamart by loan_status and the year when the client opened her/his account.

From this we can highlight a clear trend: The trend of "A" loans (debt already repaid) decreases over the years, as it is clear that people that created their account earlier have already succeeded to repay all debt splits.

Moreover, type "C" loans (splits are still due but regular position) counts increase over the years. Customers that opened the account more recently are less likely to already have paid everything back.

5) Correlation Between Variables



The above matrix shows correlation between important variables. The blue blocks suggest a negative correlation between the two variables while the peach blocks mean there is some positive correlation.

We see four blue blocks that stand out which have a correlation close to -1. These relationships are logically correct. For example: Account open year and number of transactions – Smaller the year in numbers, greater its length of relationship with the bank, consequently, a greater number of transactions.

We also notice positive correlation between variables such as loan duration and total loan amount. This means that larger the loan amount, longer the term of loan repayment. Similarly, lor with bank and years since the card have positive correlation. This indicates that the clients get a credit card soon after opening the account.

CONCLUSION

After considering these insights, the bank can have a clearer image of the final situation of the customer base and of which are the categories on which it would be better to focus more in future campaigns.

For example, loans for younger people can be a more exploitable market, as the values for this category are quite low in comparison to other categories.

Moreover, for what concerns loan management, as half of the total amount of loans issued is still ongoing, we would suggest to focus some representatives' energies to control the ease to pay of customers with splits due in order to make sure to be able to be proactive in case of problems and to avoid delays in payments or to reschedule the current deadlines.