Financial Programming

Professor Minh Phan
Group Members:
Stephanie Beyer Diaz
Claire Hutin
Xuejie Ma

Executive Summary

The purpose of this project was to improve the understanding of a bank's customers with the ultimate goal of seeking specific actions for achieving better services. To do this, data was cleaned and aggregated by customer, to later explore different variables that made sense with finding an adequate segmentation.

An initial broad analysis was carried out, looking at time series to explore growth patterns for transactions, loans and account metrics. No clear evidence for seasonality was found and the data suggests there is no strong growth strategy in place, as account openings are relatively volatile while transactions show a higher concentration in lower average amounts.

For a more detailed analysis, customers were divided into two groups: those who shared accounts and those who do not share accounts (which we refer to as exclusive accounts). In cases of customers with shared accounts, data was looked at account level in order to avoid duplicating information.

For both groups, gender divide is close to 50%, with shared accounts most likely belonging to couples. Both groups mostly come from South or North Moravia, with a similar distribution for other regions.

Customers with exclusive accounts are slightly faster when getting credit cards issued, with 17% of card holders having a gold card. This number is of 11% for shared accounts, but there is vast room for improvement for both groups, as both have an average time of over 2 years for getting a card issued.

In contrast to this, exclusive accounts are less likely to take out a loan, but display more risky behavior, while shared accounts do not display any debt in running or finished loan contracts. Shared accounts are also slightly more active in transaction frequency, but exclusive accounts have a higher average transaction amount.

However, although differences exist between these two groups it cannot be said for certain if they are significant without realizing a deeper statistical analysis, which we suggest as a future step.

Finally, it may be worthwhile to reexamine the relationship between interest rate and different banking products (such as housing, loan and insurance payments) as there does not seem to be in incentive for growth from a banking perspective.

Data Description

In this project, the first step was to have a look at the 8 tables we were given (loan, order, trans, card, disp, account, client, district).

The next step was to create additional variables and clean the tables. The aim of this step being to aggregate all tables by account and client id to avoid duplicating information. Here are the changes we made to each table:

Client table: we created an age variable and an age group variable, as well as an object to determine the client's gender. We then dropped client's birth number, year, day and month. The final client table thus contains client id, district id, gender, age and age group.

Account table: we replaced the Frequency variable with English translation to better read the data and also formatted the date column.

Card table: we formatted the date as datetime and added a variable card_owner to later know whether clients are card owners or not.

Order table: first thing we did was to replace the order k_symbols with English text to better understand and read the data. Then we created for variables representing these four k_symbols in which we put the amount per account for each order k_symbol. We then dropped the order id, account_to and k_symbol variables, summing amounts to group by account id.

District table: We renamed variables and then dropped A5,A6,A7,A8,nb_cities as we assumed this data would not be relevant for our analysis.

Transaction table: the first step was to turn the transaction into a dummy. Then we replaced the k_symbol to English words and created variables for each of these transactions and have them turned into dummies. We also cleaned the transaction dates and added amounts to each transaction type and operation. We also created, for each account id, an average transaction frequency per account. We then grouped the table by accounts id in order to have 16 variables: account id; date difference between each transaction; first and last transactions; number of transactions; sum and average amounts; credit and withdrawal; interest rate (overall interest amount over total transaction amount); and all the k_symbols.

Loan table: In this table, we cleaned loan dates.

We then added a prefix corresponding to each table in order to avoid confusion, as some columns from different tables had the same names.

We then merged our tables:

- First, we merged account and loan (account_loans)
- Then, we merged account_loans with transaction (al_trans)
- Then, we merged al_trans with disp (alt_disp)

- Then, we merged alt_disp wit card (altd_card)
- Then, we merged altd_card with district (altdc_district)
- Then we merged altdc_district with client (altdcd_client)
- Finally, we merged altdcd_client with order (final_merge)

Once we obtained the final_merge tables, we dropped the redundant ID columns (we kept only one account_id, client_id, disp_id and district_id columns).

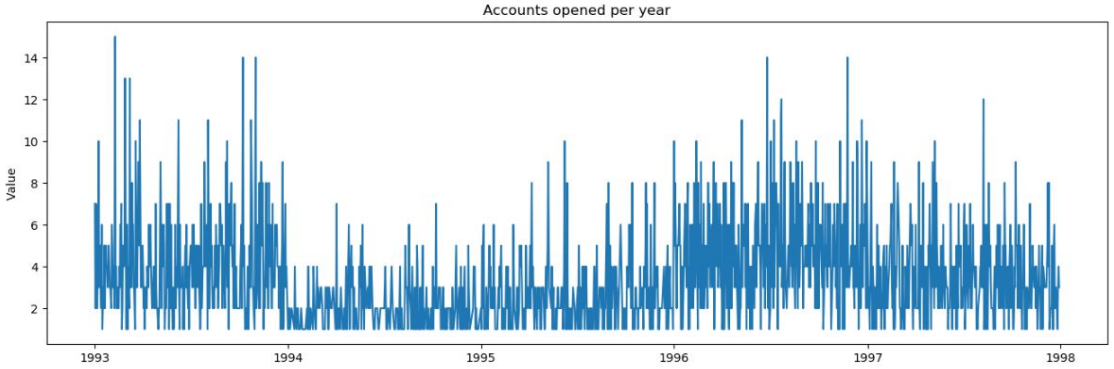
We then replaced all missing values that appeared due to the merge.
At the end, our 51 variables with their respective type and definitions are as follows:

Column Name	**Type**	**Definition**
acc_account_id	int64	account id
acc_frequency	object	frequency of issuance of statements
acc_date	datetime64	date of creating account
loan_loan_id	int32	loan id
loan_date	object	date when loan was granted
loan_amount	int32	amount of loan money
loan_duration	int32	duration of loan
loan_payments	int32	monthly payments
loan_status	object	status of paying off loan
trans_datediff	float64	average difference between transaction dates
trans_last_trans	object	last transaction date
trans_n_trans	int64	count of number of transactions
trans_sum_amount	float64	total amount from trans table
trans_avg_amount	float64	average amount from trans table
trans_credit	float64	amount for transaction type "VYDAJ"
trans_withdrawal	float64	amount for transaction type "PRIJEM"
trans_household_amount	float64	"SIPO" amount
trans_insurance_amount	float64	"POJISTNE" amount
trans_interest_amount	float64	"UROK" amount
trans_loan_amount	float64	"UVER" amount
trans_pension_amount	float64	"DUCHOD" amount
trans_sanction_amount	float64	"SANKC. UROK" amount
trans_statement_amount	float64	"SLUZBY" amount
trans_int_rate	float64	interest rate (interest over transaction amounts)
disp_disp_id	int64	disposition id
disp_owner	int64	flag for account disposition (1 for owner, else 0)
cc_card_id	int32	card id
cc_type	object	card type (gold, junior, classic)
cc_date	object	card open date
cc_owner	int32	card owner flag (1 for owner, else 0)
dist_dist_name	object	district name
dist_dist_region	object	region
dist_nb_habitants	int64	inhabitants
dist_ratio_urban	float64	ratio of urban inhabitants

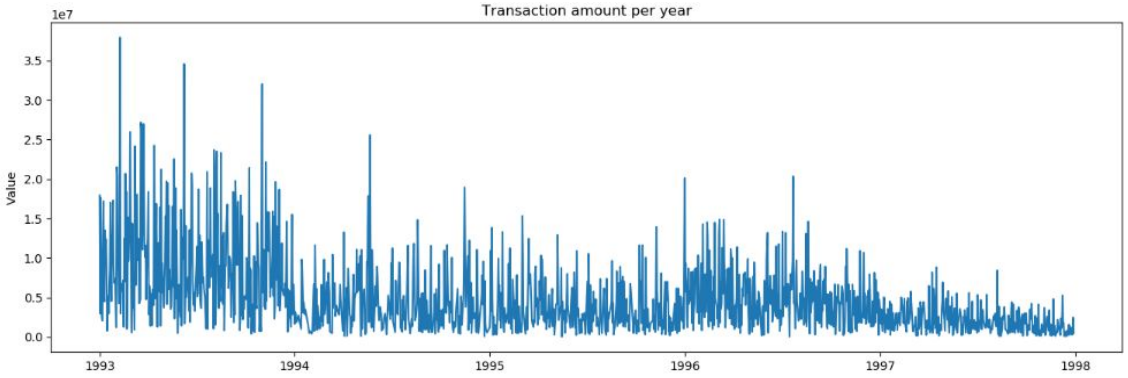
dist_avg_salary	int64	average salary
dist_unemp_95	object	unemployment rate ‘95
dist_unemp_96	float64	unemployment rate ‘96
dist_nb_entrepreneurs	int64	entrepreneurs per 1000 inhabitants
dist_crimes_95	object	number of committed crimes '95
dist_crimes_96	int64	number of committed crimes '96
cust_client_id	int64	customer client id
cust_g	object	customer gender
cust_age	int64	customer age with 31 Dec 1999 as cut off date
cust_age_group	int64	customer age floored
ord_amount	int32	amount from table order
ord_leasing	int32	amount for k_symbol="LEASING"
ord_hh_payment	int32	amount for k_symbol="SIPO"
ord_loan_payment	int32	amount for k_symbol="UVER"
ord_insr_payment	int32	amount for k_symbol="POJISTNE"
cc_create_time	int32	days between account and card creation
loan_create_time	int32	days between account and loan creation

General Overview

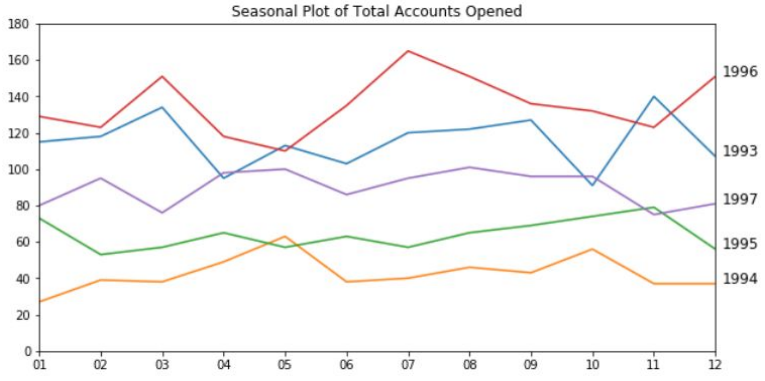
Growth: when looking at accounts opened per year, no evidence of a clear growth pattern is found. This suggests an absence of strategic decisions in place.



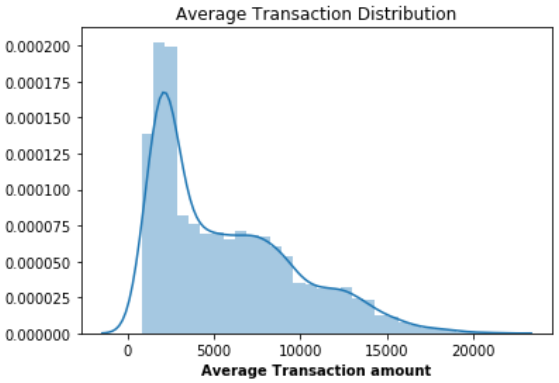
The previous statement is also supported by the fact that transaction amounts are decreasing over the years.



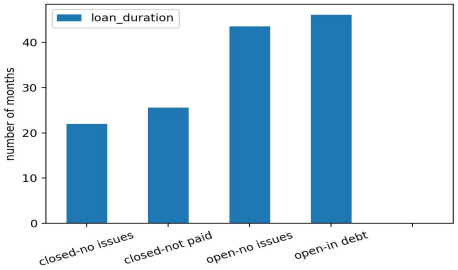
Seasonality: when looking for evidence of seasonality on a monthly basis throughout the years, no clear patterns emerge, supporting the thesis of a lack of a clear strategic direction.



Average Transaction: When looking at transaction distribution, there is a right skew in the data, showing higher concentration in lower average transaction amounts. This may also be a reflection of the fact that transaction amount has been steadily decreasing throughout the years, as seen before.

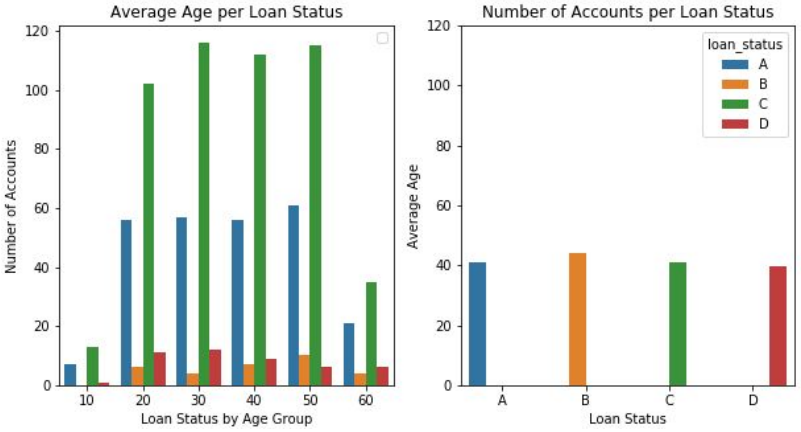


Loan Status:

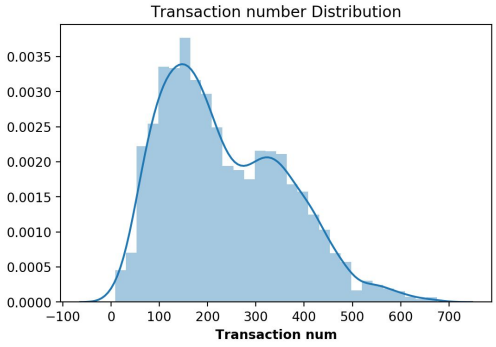


We can see that the open-no issues and the open-in debt type account take a longer loan duration, which has a big differences with ones of the closed-no issues and closed-not paid.

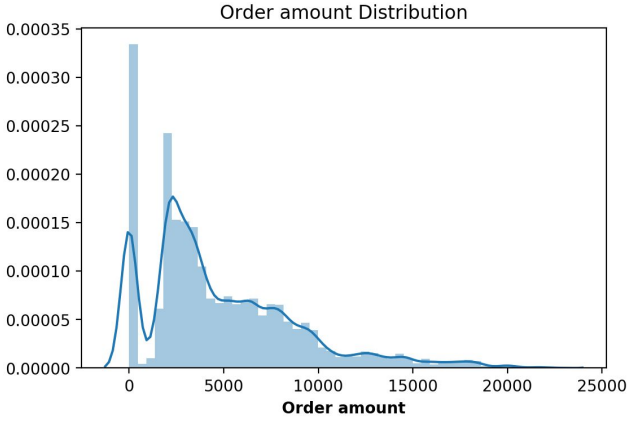
Similar trend shows in the plot of loan status and loan amount. However, when crossing loan status with age data, no clear risk groups pop up.



Distribution



As for the distribution of the number of transactions, it has two peaks. This indicates that transaction number near 150 and 350 takes the biggest proportion.

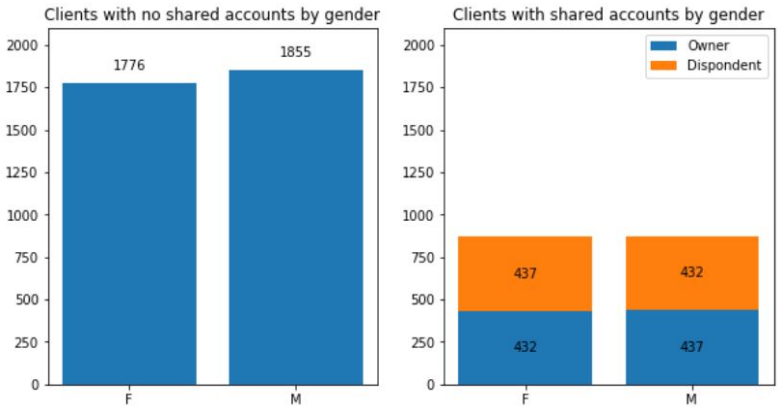


The order amount concentrate around 3000, however, has a great distribution of a slight amount.

Customer Profiles

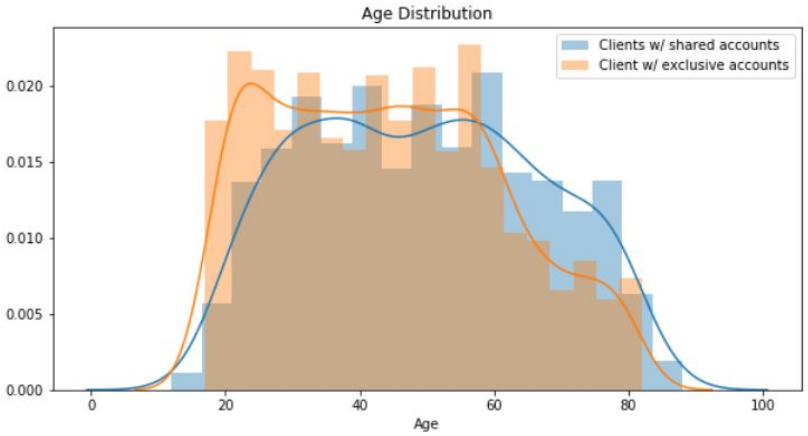
To better understand the bank's clients, the first steps taken were building profiles with demographic data. For this, clients were grouped depending on if they shared accounts or not, to later compare variables such as gender, age and regional location. After this, their behavior within the bank was examined by looking at frequencies and amounts of transactions, loans and other products.

Out of 5,369 clients, 49% are females and 51% are males. Overall, around 32% clients are sharing an account, where the gender divide is practically 50%. Additionally, the average age difference between clientes sharing accounts is around 4 years. This strongly suggests that shared accounts mostly correspond to couples.

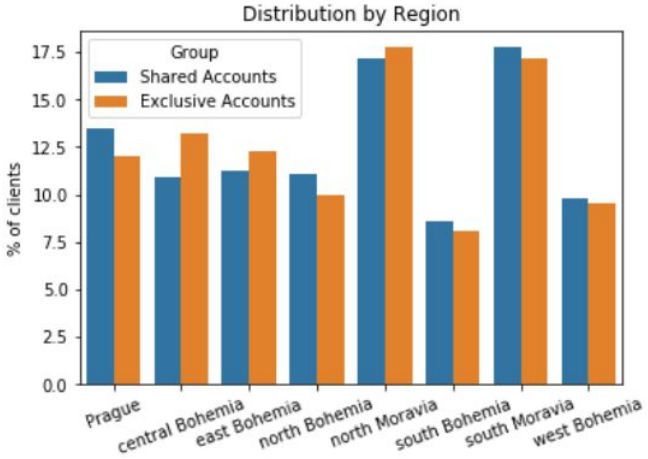


Looking at age in more detail, clients with exclusive accounts show a higher concentration in the age bucket between 20 and 30, while clients sharing accounts show a higher middle-age concentration. On average, exclusive account owners are around 44 years old,

with 75% of them being 57 or under. For shared account owners, the average age is 49, with 75% being 63 years of age or lower.

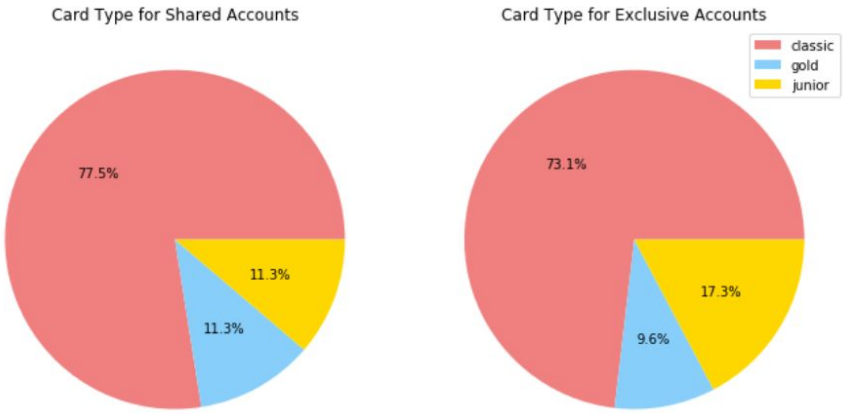


When looking at regional data, there is no significant difference between clients who share and do not share accounts. For both groups, North and South Moravia show the largest concentration of customers, which corresponds to around 17% for each region.



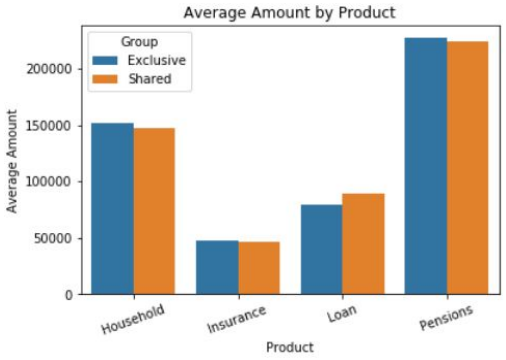
In terms of products, around 20% of customers with exclusive accounts have a credit card, which they get on average 772 days after opening their account. This is true for only 17% of shared accounts, who on average get the card 804 days after their account opening. In both groups, this period is over 2 years, which may present an opportunity for the bank to get more customers to get credit cards or to shrink the period between opening an account and getting a credit card.

Exploring card data in more detail shows that almost 78% of shared accounts are linked to a classic card, compared to 73% of exclusive accounts. Around 11% of shared accounts are tied to a junior card, in contrast to almost 10% of exclusive accounts. Finally, around 11% of shared accounts own a gold card, versus 17% of exclusive accounts. Therefore, we can conclude that in both groups there is a large untapped opportunity for customers to acquire gold cards, even if exclusive accounts may be performing slightly better in this regard.

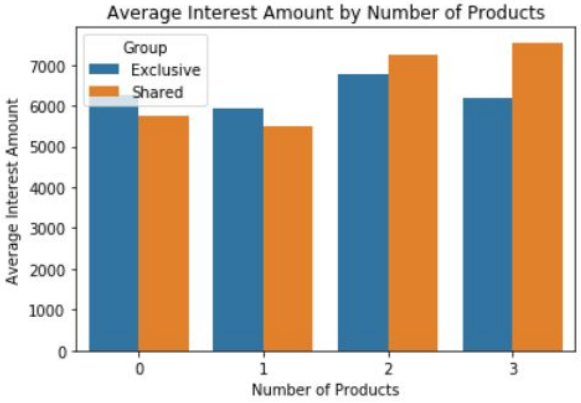


Around 17% of shared accounts taken out a loan. This is true for only around 15% of exclusive accounts but they are slightly faster at getting a loan, with an average of 395 days since opening their account, compared to 411 days for shared accounts. Furthermore, exclusive accounts take out loan amounts of an average duration of 37 months and an average amount of $150.807, compared to 35 months an $153.642 for shared accounts, which suggests shared accounts take out higher loans and are quicker to pay back. This is also supported by the fact that shared accounts only show loans with status A (contract finished with no problems) and C (running contract with no problems so far). On the other hand, exclusive accounts do show up as being in debt, as seen on status B (contract finished but unpaid) and D (contract running but in debt).

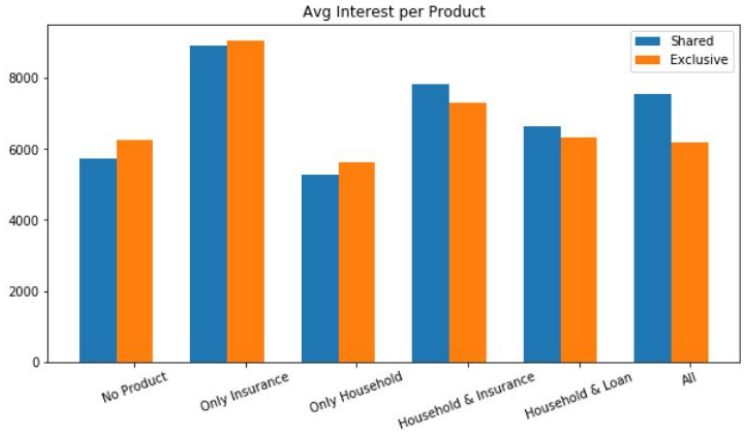
Moreover, transaction data can be useful for a closer look at client activity. Exclusive accounts showed transactions every 3 to 64 days, while shared accounts carry out transactions every 3 to 51 days. Shared accounts are also slightly more active, with an average transaction number of 237 per account, compared to 234 per exclusive account. However, exclusive accounts have a higher average transaction amount of $6.000, compared to $5.612 for shared accounts. Exclusive accounts also show slightly higher averages for household, insurance and pension amounts, while loan amounts are higher for shared accounts.



Finally, looking at interest per transactions were used as a way to understand who were the most profitable customers. To look at this in more detail, Household (or mortgage), Insurance and Loan were considered exclusively banking products as Pensions may also be public. First, average interest amount was looked at by combination of products. For example, if an account was linked to all three products, then it has a combination of 3, but if it has only Household and Insurance or Loan and Insurance, then it has a combination of 2. With this metric, the chart below shows that shared accounts are more profitable (or display a higher average interest amount) when the combination of products is higher, but exclusive accounts are more profitable when the combination of products is lower.



Specifically, when looking at products, the most profitable combination is only insurance for both groups, while having no products is slightly more profitable than having only household. This suggests that it may be worthwhile to rethink the way interest is charged per transaction as it currently does not seem to be an incentive for achieving growth.



In conclusion, although there are several differences between the segments of exclusive and shared account, it is difficult to tell if these differences are notorious. As a further step, it would be interesting to carry out statistical significance tests to determine if these segments are behaving in a significantly different way.

Additionally, it may be worth rethinking the interest rates associated to each banking product (Household, Loan and Insurance), as they do not seem to be aligned with a strong growth strategy. A lack of growth strategy is also supported by transaction amount and volume data seen in the general overview.