FINANCIAL DATAMART REPORT

Group Members

1. Betsy Varghese
2. Xinghao Zong
3. Kaveh Jalilian

Project Definition

1. To create a consolidated database that contains a unique observation for each client based on the static and dynamic information provided.
2. To analyze the information to help the bank identify its customers as assets or liabilities to the bank and consequently make targeted decisions.

Customer Segmentation

1. The 8 features from the base table that were to be used for segmentation analysis were first winsorized in order to remove any outliers that may influence the model.
2. These features were standardized using the Standard Scaler from the sklearn library.
3. A segmentation model was then created using the k-means non-hierarchical clustering approach.
4. The entire data set was divided into 6 clusters based on the 8 features seen in the table below.
5. The table below has also been sorted based on features like avg. salary, client-bank relationship, etc.

	Cluster 2	Cluster 1	Cluster 3
Age	34.71	63.33	38.42
Avg. Salary	9538.54	9474.03	9452.3
Client Relationship with the bank (in days)	1076.87	1032.67	1034.18
Total No. of Transactions	254.75	215.81	113.3
Avg. Value of Transactions	6949.83	3805.04	4951.1
Days since last Transaction	1.03	1.01	29.06
Total No. of Orders	1.56	1.43	0.31
Avg. Value of Orders	3494.97	2719.16	1244.36
Total clients in each cluster	3119	2056	194

6. Analysis based on Segmentation
   i. On an average, the number of transactions over the years for 96% of the database is 235.
   ii. Clients with a higher period of inactivity (based on their last transaction) also seem to be the clients with fewer numbers of transactions and lesser number of orders. These clients belong to the third cluster.
   iii. Clients in the third cluster also have a much smaller average value of orders.

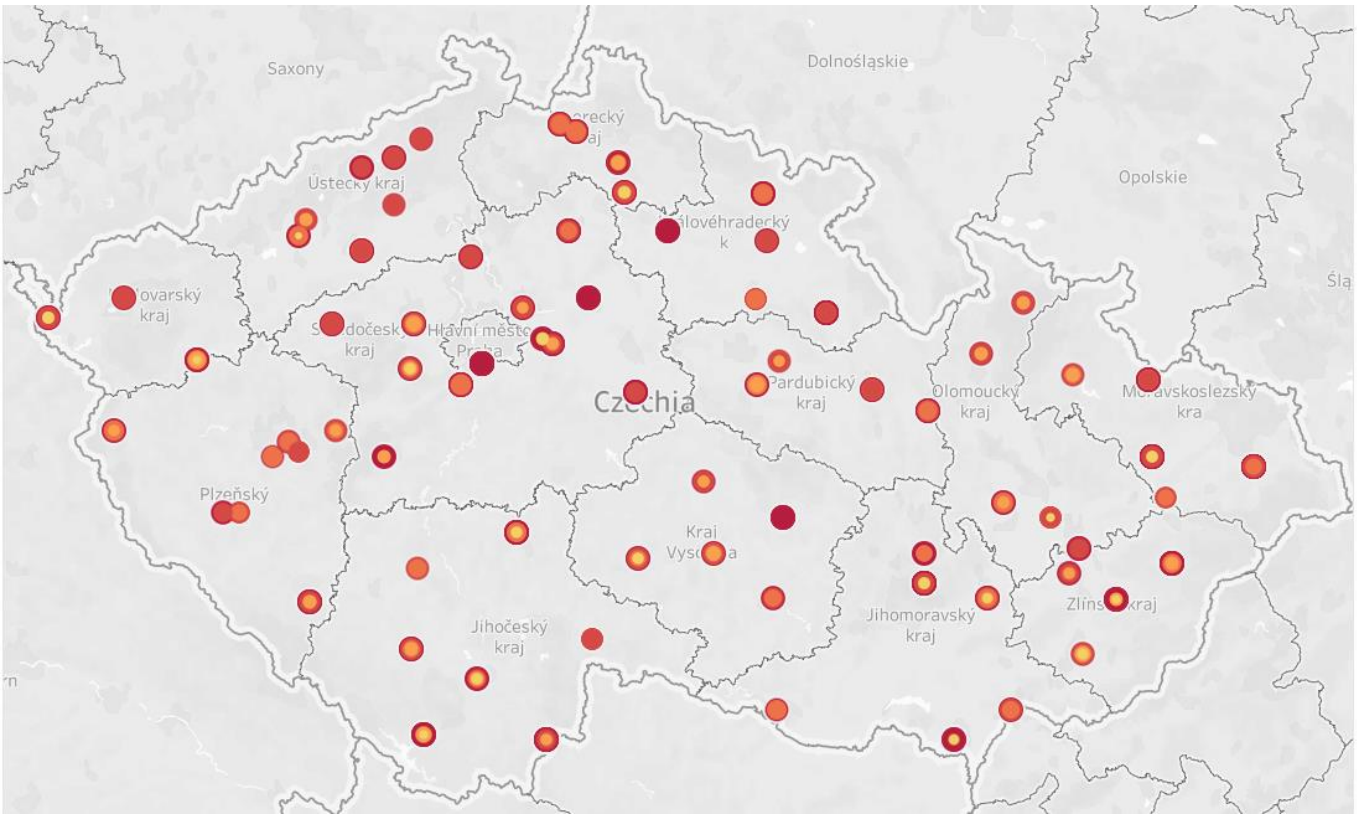
iv. Besides the age variable, the differences between clusters 1 & 2 is only marginal. Considering the high level of activity in these clusters, the clients belonging in these clusters could be considered as dear to the bank.

Demographic Analysis

The graph below show the distribution of clients across the Czech Republic. Clients have been color coded on the map based on the highest number of transactions.

i. Based on the results of the map, the following four customers were observed to have the highest number of transactions as well as the highest average salary.

ii. Customers with a relatively higher number of transactions were concentrated in the H.L.M. Praha district of Czech Republic and all of them were owners of the bank account.



client_id	Owner	age	Gender	region	avg_salary	No. of Trans	total_sum_trans
3548	Y	24	M	Prague	12541	567	6226227.3
7299	Y	61	M	Prague	12541	567	3268769.1
11126	Y	34	F	Prague	12541	567	3290859.6
12859	Y	23	F	Prague	12541	567	2244846.5

iii. The following tables show the regions and district with the highest number of transactions.

Distribution of Customers by Region

Region	
south Moravia	937
north Moravia	920
central Bohemia	664
Prague	663
east Bohemia	660
north Bohemia	561
west Bohemia	515
south Bohemia	449

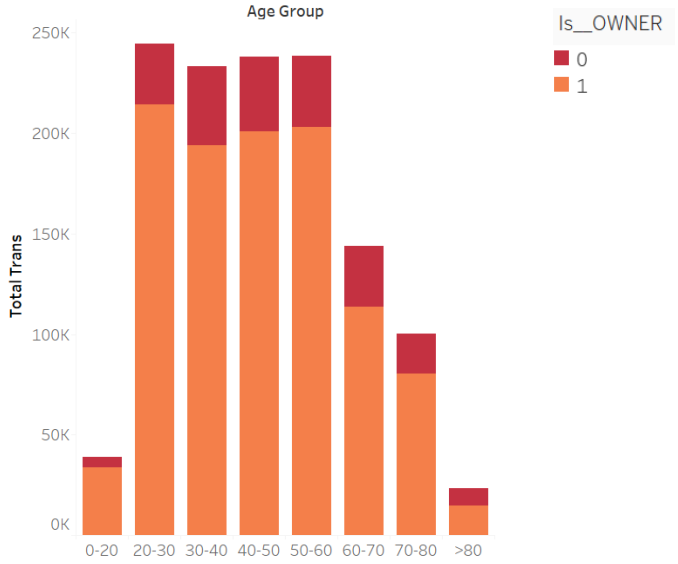
Distribution of Customers by District

District Name	
Hl.m. Praha	663
Ostrava - mesto	180
Karvina	169
Brno - mesto	155
Zlin	109
Olomouc	104
Frydek - Mistek	86
Nachod	76
Usti nad Orlici	73
Kolin	71

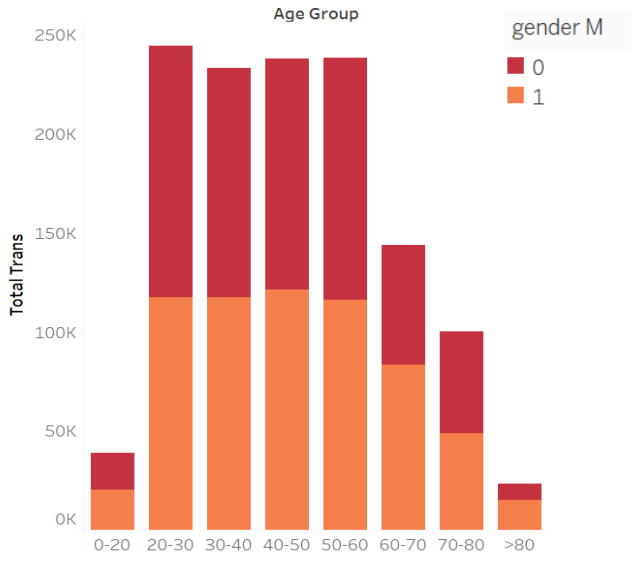
Base Table Analysis

1. Typically, we see that the number of transactions made by the actual owner of the account is much greater than those made by the dependants.
2. We also see that the number of transactions is relatively higher for people between the age groups 20-60 years. This holds true with reality as people below the age of 20 depend on their family for their immediate needs thus resulting in lower transactions. Similar is the case for people beyond the working age group (usually > 60 years)
3. When we analyse gender, we observe that the number of transactions made by females is only marginally greater than those of males (with the exception of the age group > 80)

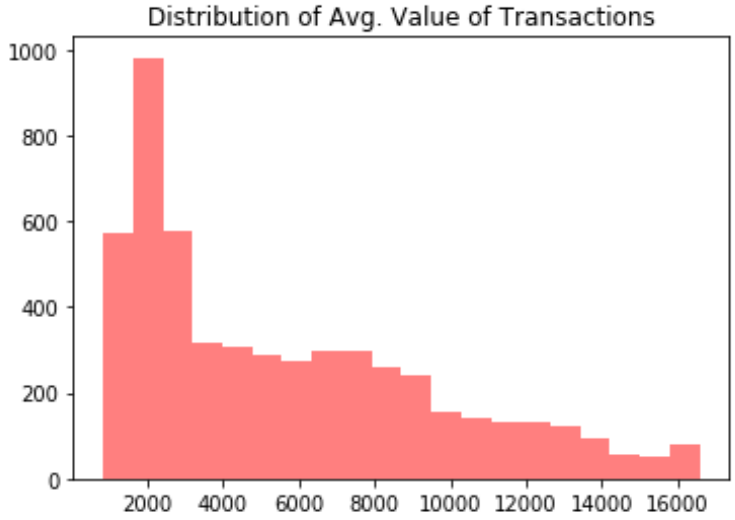
Total No. of Transactions by Age Group and Type of Disposition



Total No. of Transactions by Age Group and Gender

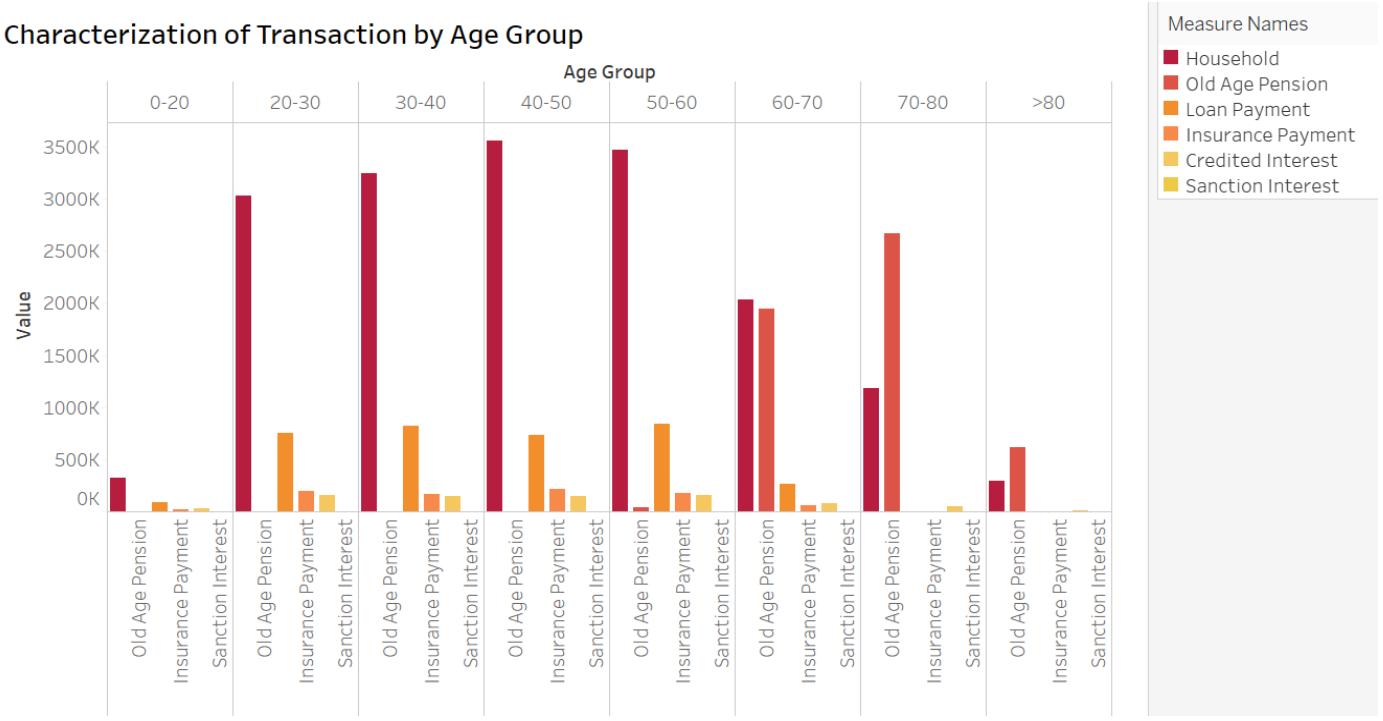


4. Below is a histogram showing the distribution of the mean value of transactions for each client. We see that the distribution is right-skewed and that approximately 50% of the client database has an average transaction value less than 4595.87 CZK.

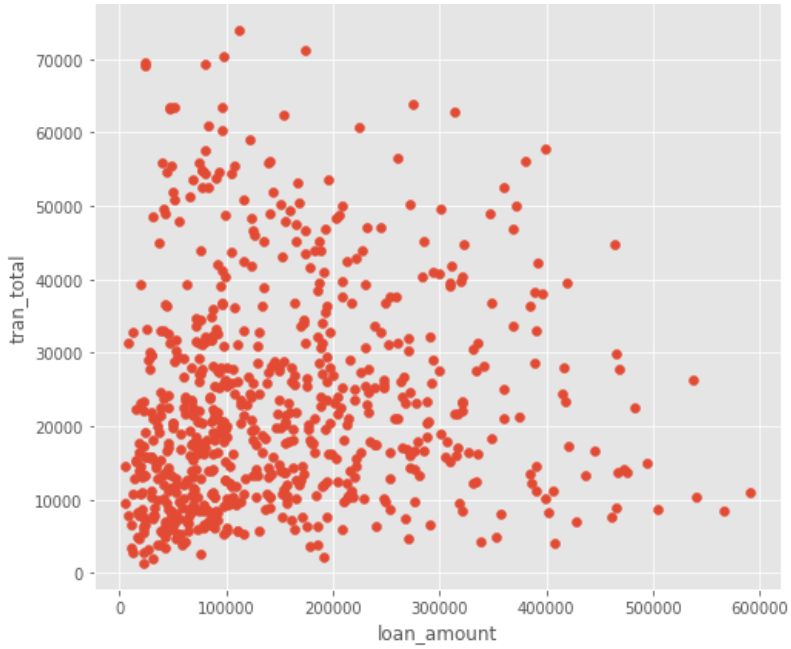


Distribution Description	
count	5369
mean	5675.104
Std. dev	3972.598
min	843.4355
max	16573.03
25%	2271.899
50%	4595.87
75%	8272.47

5. Through the graph below, we see that for people between the ages 20-60 years the greatest proportion of their transactions revolved around household payments. Whereas, most of the transaction activities for people greater than 60 years of age were related to their old age pension.



6. An upward trend between the loan amount and the total value of transactions for each account id can be visualized through a scatter plot. This indicates that accounts and clients with a high total value of transactions across the years are also the ones who take loans of a greater amount. Using the base table to identify these customers and checking their status of loan payment would help distinguish good clients from bad. Monitoring and maintaining the relationship with clients with a good credit score would make for a lucrative business decision.



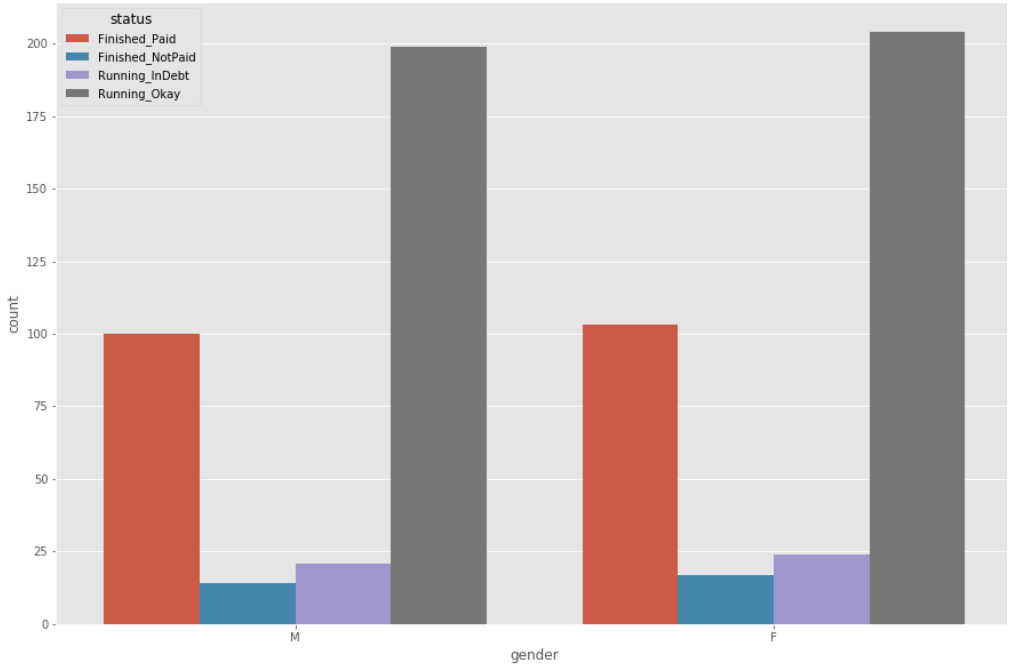
7. The histogram of average value of orders, is also skewed to the right. Approximately 50% of the clients have an average order value less than 2648 CZK.

Distribution Description	
count	5369
mean	3116.992
Std. Dev.	2455.155
min	0
max	10054.85
25%	1464
50%	2648
75%	4359.5

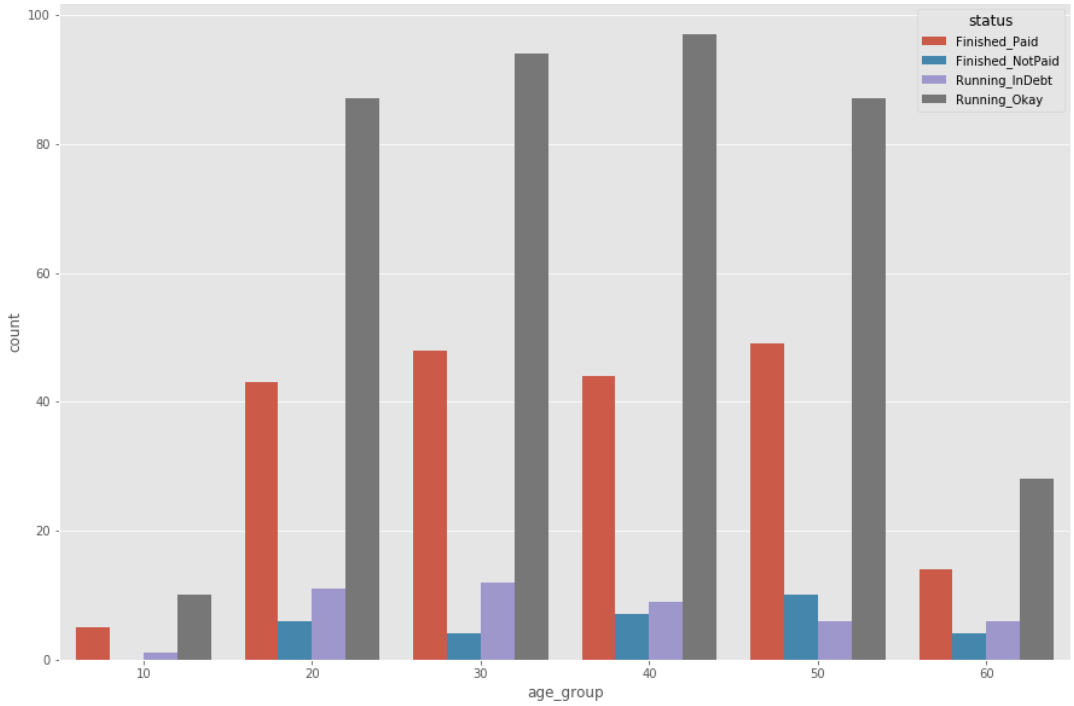
8. Results of the analysis between age group and card type issued by the bank is useful to understand customer transaction behavior. It is also useful in sending targeted messages to each client. Based on the table below, we see that a majority of the bank's clients own a classic card. And the people who use the classic card lie between 20-60 years of age.

	Type		
Age Group	junior	classic	gold
0-20	52		
20-30	93	108	13
30-40		150	21
40-50		149	21
50-60		198	20
60-70		45	13
70-80		8	
>80		1	

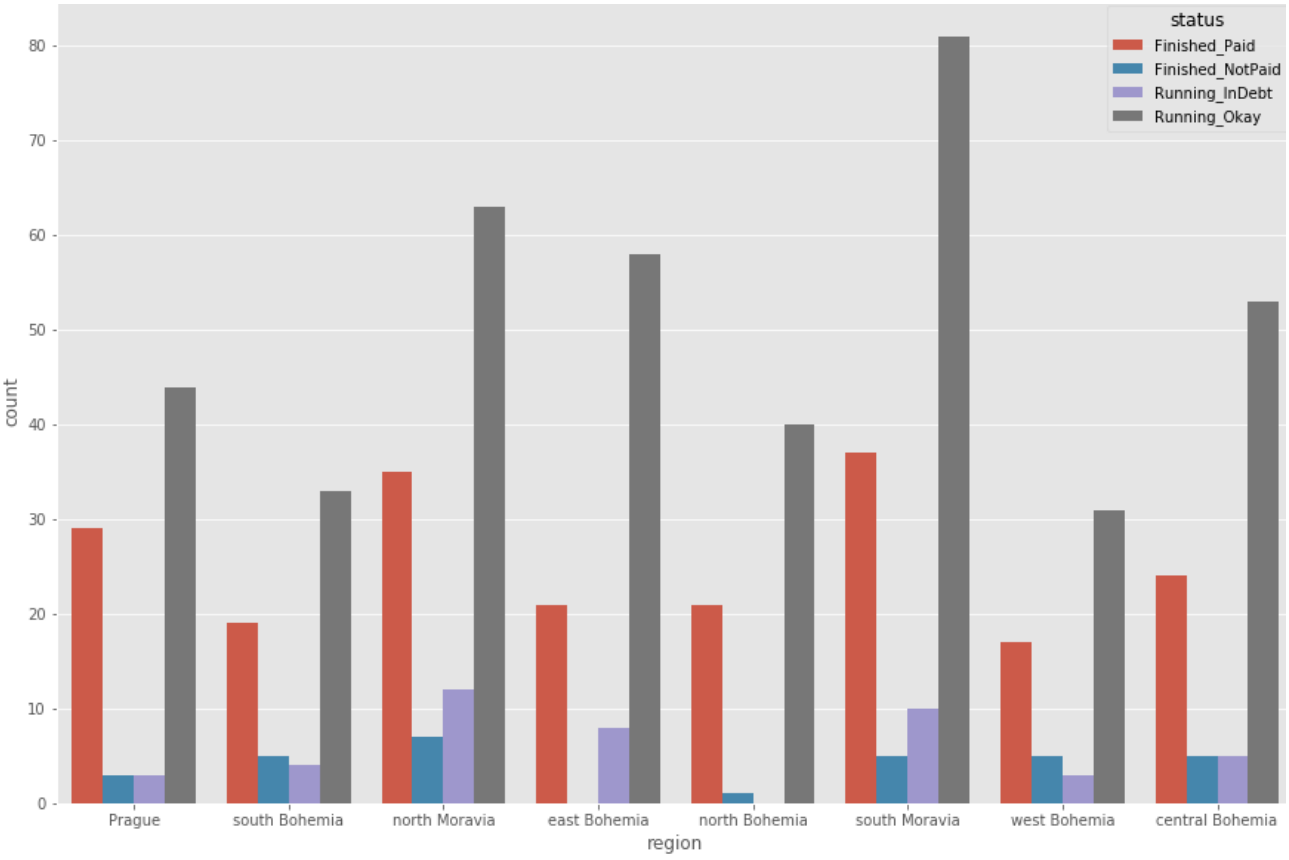
9. Analysis of loan status against different demographic parameters gave us the following results
   i. Loans Status vs. Gender: There is not much difference between males and females w.r.t. the status of their loans.



ii. Loan Status vs. Age Group: Loans not paid is the highest for people in the 50s.



iii. Loan Status vs. Region



ANNEXURES

Following is a list of all variables in the final data mart.

Variables	Description
disp_id	Disposition to an account
client_id	Unique identification for each client
account_id	Unique identification for each account
Is__OWNER	Mentions if the client is an owner or not (1: Owner 0: Dependent)
district_id	Unique identification for the 77 districts
birth_year	Year of Birth for the client
age	Client Age
age_group	Age Group the client fits into
birth_month	Month of Birth for the client
gender_M	Gender (1: Male 0: Female)
Issuance_after_transaction	Logical variable to check if statements are issued after transactions
Monthly_Issuance	Logical variable to check if statements are issued on a monthly basis
Weekly_Issuance	Logical variable to check if statements are issued on a weekly basis
account_creation_date	Date of Account Creation
days_since_acc_creation	Number of days elapsed since the client's account creation
district_name	Name of the district
region	Name of the region
nbr_inhabitants	Number of Inhabitants
num_muncipal_<499	No. of municipalities with inhabitants <499
num_muncipal_500_1999	No. of municipalities with 500-1999 inhabitants
num_muncipal_2000_9999	No. of municipalities with 2000-9999 inhabitants
num_muncipal_>10000	No. of municipalities with inhabitants >10000
num_cities	Number of Cities
urban_ratio	Urban Ratio
avg_salary	Average Salary
unemp_rate_95	Unemployment Rate '95
unemp_rate_96	Unemployment Rate '96
num_entrepreneurs_per1000	No. of entrepreneurs per 1000 inhabitants
num_crimes_95	No. of Crimes in '95
num_crimes_96	No. of Crimes in '96
first_trans	Date of First Transaction
last_trans	Date of Last Transaction
total_trans	Total No. of Transactions
total_sum_trans	Total Sum of all Transactions
mean_all_trans	Mean Amount of Transactions
num_cash_withdrawl	No. of Transactions involving Cash Withdrawls
num_trans_collections	No. of Transactions involving collections from another bank

num_trans_cc_withdrawls	No. of Transactions involving credit card withdrawls
num_trans_credit_cash	No. of Transactions involving credit in cash
num_trans_Other_Operations	No. of Transactions used for Other Operations
num_trans_remittances	No. of Transactions involving remittances to another bank
avg_trans_cash_withdrawl	Avg. Monetary Value of Transactions involving Cash Withdrawls
avg_trans_collections	Avg. Monetary Value of Transactions involving collections from another bank
avg_trans_cc_withdrawl	Avg. Monetary Value of Transactions involving credit card withdrawls
avg_trans_credit_in_cash	Avg. Monetary Value of Transactions involving credit in cash
avg_trans_Other_Operations	Avg. Monetary Value of Transactions used for Other Operations
avg_trans_remittances	Avg. Monetary Value of Transactions involving remittances to another bank
total_trans_cash_withdrawl	Total Monetary Value of Transactions involving Cash Withdrawls
total_trans_collections	Total Monetary Value of Transactions involving collections from another bank
total_trans_cc_withdrawl	Total Monetary Value of Transactions involving credit card withdrawls
total_trans_credit_in_cash	Total Monetary Value of Transactions involving credit in cash
total_trans_Other_Operations	Total Monetary Value of Transactions used for Other Operations
total_trans_remittances	Total Monetary Value of Transactions involving remittances to another bank
num_trans_credited_interest	No. of transactions for interest credited
num_trans_household	No. of transactions for household
num_trans_insurance_payment	No. of transactions for payment of insurance
num_trans_loan_payment	No. of transactions for loan payment
num_trans_old_age_pensions	No. of transactions related to old age pensions
num_trans_Other_Payments	No. of transactions for Other Payments
num_trans_sanction_interest	No. of transactions for interest sanctions
num_trans_statement_payment	No. of transactions for payment statements
avg_trans_credited_interest	Avg. Monetary Value of transactions for interest credited
avg_trans_household	Avg. Monetary Value of transactions for household
avg_trans_insurance_payment	Avg. Monetary Value of transactions for payment insurance
avg_trans_loan_payment	Avg. Monetary Value of transactions for payment loan
avg_trans_old_age_pension	Avg. Monetary Value of transactions for old age pensions
avg_trans_Other_Payments	Avg. Monetary Value of transactions for Other Payments
avg_trans_sanction_interest	Avg. Monetary Value of transactions for interest sanction
mean_trans_statement_payment	Avg. Monetary Value of transactions for payment statement
total_trans_credited_interest	Total Monetary Value of transactions for interest credited
total_trans_households	Total Monetary Value of transactions for households
total_trans_insurance_payment	Total Monetary Value of transactions for payment insurance
total_trans_loan_payment	Total Monetary Value of transactions for payment loan
total_trans_old_age_pensions	Total Monetary Value of transactions for old age pensions
total_trans_Other_Payments	Total Monetary Value of transactions for Payments Other
total_trans_sanction_interest	Total Monetary Value of transactions for interest sanction
total_trans_statement_payment	Total Monetary Value of transactions for payment statement
days_since_last_trans	Number of days elapsed since the last transaction
total_nbr_orders	Total number of orders

sum_all_orders	Total sum of all Orders
average_all_orders	Avg. Monetary Value of all Orders
num_orders_house_payment	No. of orders for house payments
num_orders_insurance_payment	No. of orders for insurance payments
num_orders_loan_payment	No. of orders for loan payments
num_other_orders	No. of orders for other activities
num_leasing_payment	No. of orders for leasing payment
avg_household_payment	Avg. value of orders for household activities
avg_insurance_payment	Avg. value of orders for payment of insurance
avg_loan_payment	Avg. value of orders of payment for loan activities
avg_other_orders	Avg. value of orders for other activities
avg_leasing_payment	Avg. value of orders for payment for leasing activities
total_household_payment	Total Value of orders for house payments
total_insurance_payment	Total Value of orders for insurance payments
total_loan_payment	Total Value of orders for loan payments
total_other_orders	Total Value of orders for other activities
total_order_leasing	Total Value of orders for leasing payment
loan_id	Unique identification for each loan linked to the account
date	Date when Loan was taken
amount	Loan Amount
duration	Loan Period
payments	Monthly Payments
Finished_NotPaid	Logical Variable indicating overdue loans
Finished_Paid	Logical Variable indicating loans paid on time
Running_InDebt	Logical Variable indicating ongoing loans running okay
Running_Okay	Logical Variable indicating ongoing loans running in debt
loan_start_date	Loan Start Date
card_id	Unique identification of Card Id
type	Type of Card
issued	Date of Issue
kmean cluster	Client Classification based on the Segmentation Model