PKDD'99 Discovery Challenge

Omar Abdelgelil (omar.abdelgelil@ieseg.fr)

&

Manoj Kumar (manojkumar.purushothaman@ieseg.fr)

&

Lucas Olivier (lucas.olivier@ieseg.fr)

Course: Financial Programming

Prof: Minh Phan

Task description

The bank wants to improve their services. For instance, the bank managers have only vague idea, who is a good client (whom to offer some additional services) and who is a bad client (whom to watch carefully to minimize the bank loses). Fortunately, the bank stores data about their clients, the accounts (transactions within several months), the loans already granted, the credit cards issued. The bank managers hope to improve their understanding of customers and seek specific actions to improve services. A mere application of a discovery tool will not be convincing for them.

Introduction

The purpose of this report is to analyze the clients based on their financial status between 1st January 1993 to 31st December 1998. We were provided initially with the following datasets :

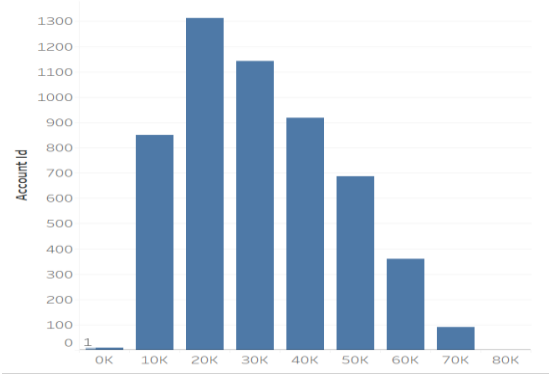
1. Credit Card
2. Client
3. Account
4. Transaction
5. Orders
6. Loans
7. Demographic
8. Disposition

Transaction Table

Transaction Table: 4500 accounts

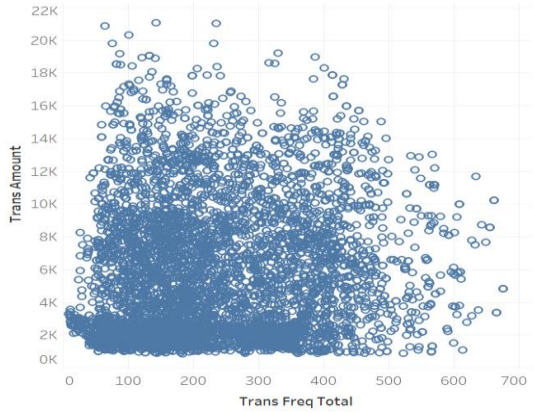
Distribution of account balance after transaction

Average balance is 36K, the distribution seems to be right skewed



Average Transaction vs Number of Transaction

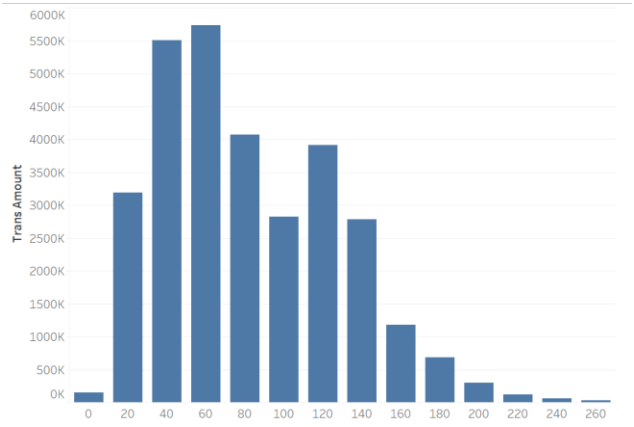
The density at the bottom show that most of the transactions happen below the range of 4K



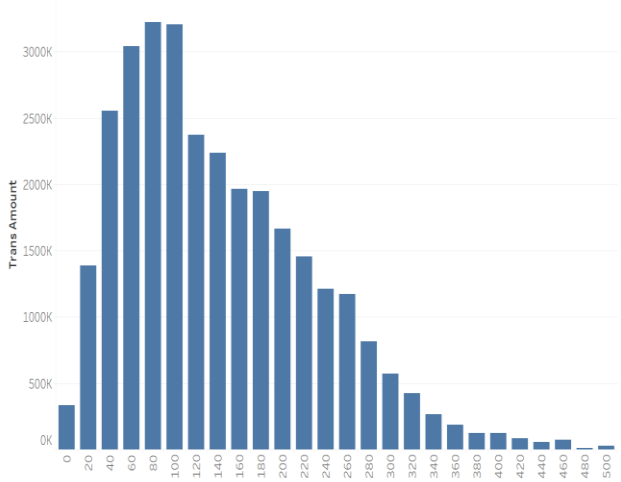
Average Transaction by Transaction Type (Credit, Withdraw)

Overall, the number of transactions for withdrawal is 20% more than the number of transactions for credit

Credit Transaction

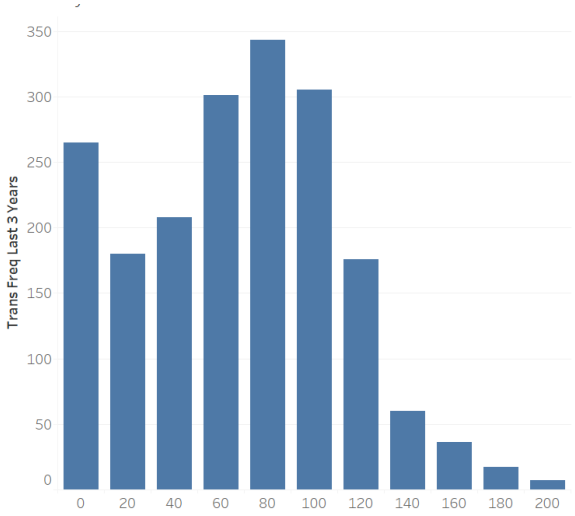


Withdrawal Transactions



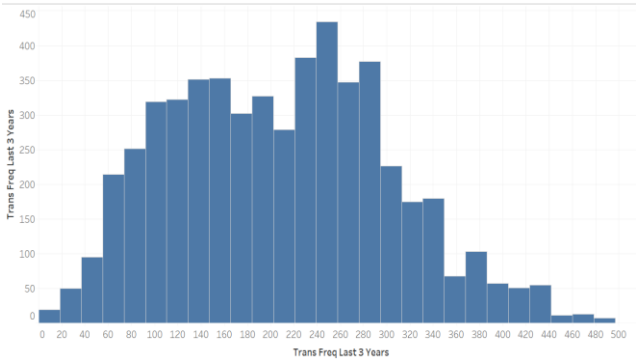
First 3 Years Transaction

There has been no transactions in some accounts in the first 3 years, 60 to 80 transactions on average



Last 3 years Transaction

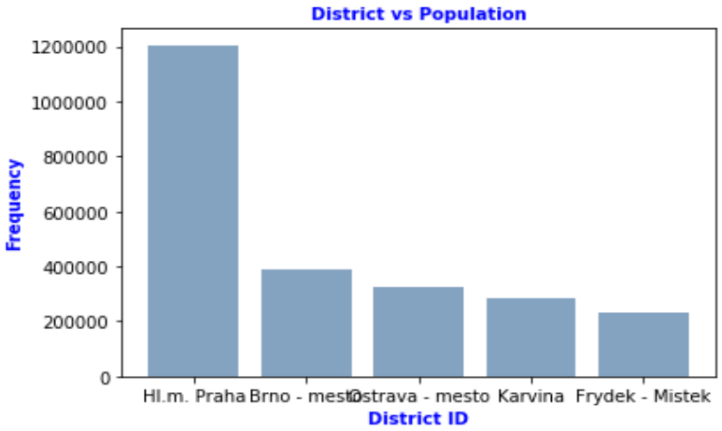
The number of accounts has incread in the last three years, on average 240 to 260 transactions being done in 3 years

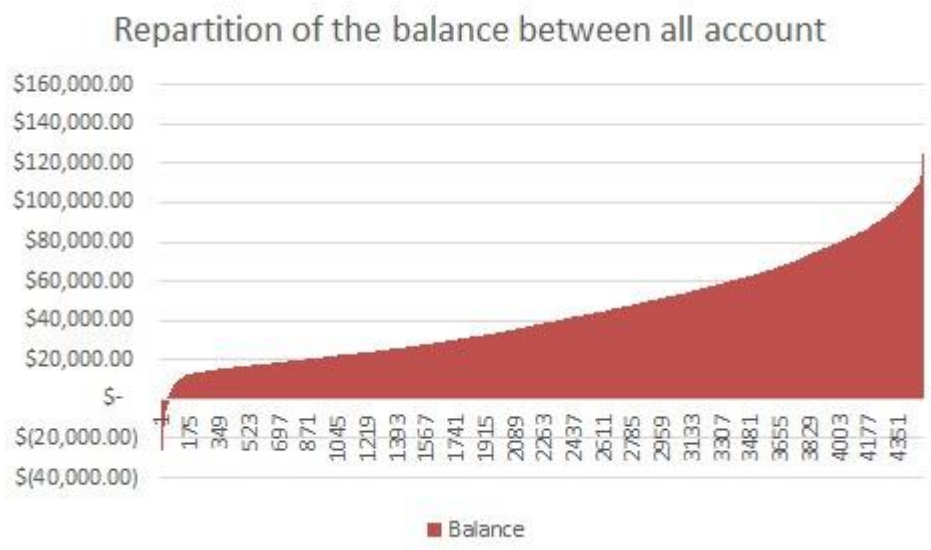


Demographics

Top 5 Populated Districts

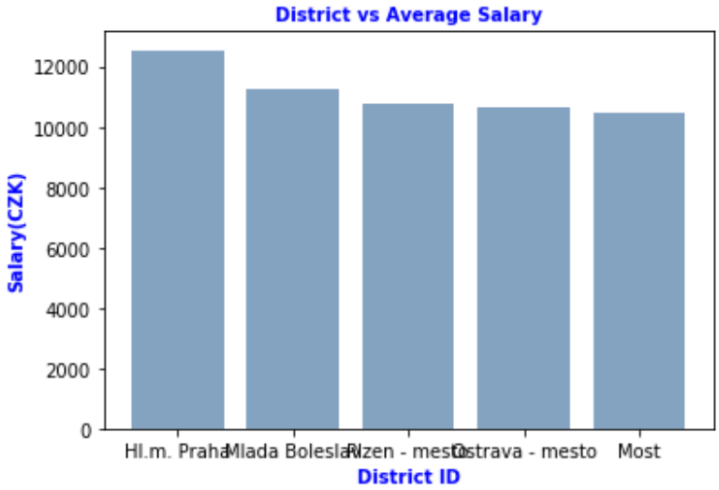
```
0         Hl.m. Praha
53       Brno - mesto
73    Ostrava - mesto
69            Karvina
67    Frydek - Mistek
Name: district_name, dtype: object
```





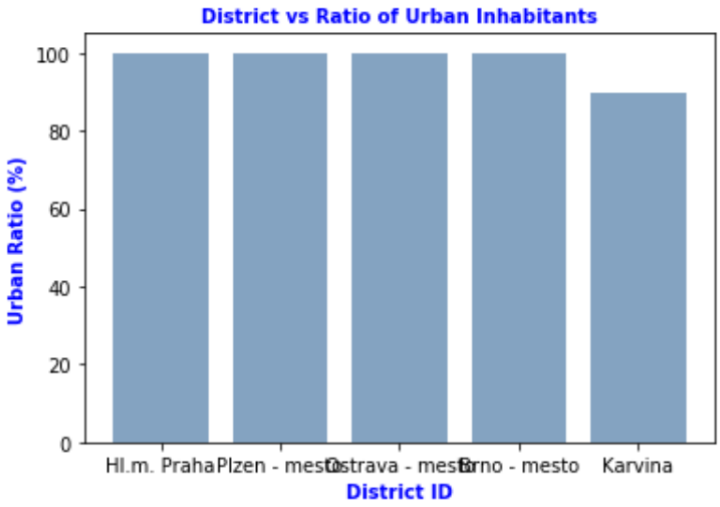
Top 5 Districts with Average salary

```
0          Hl.m. Praha
7       Mlada Boleslav
25       Plzen - mesto
73     Ostrava - mesto
38                Most
```



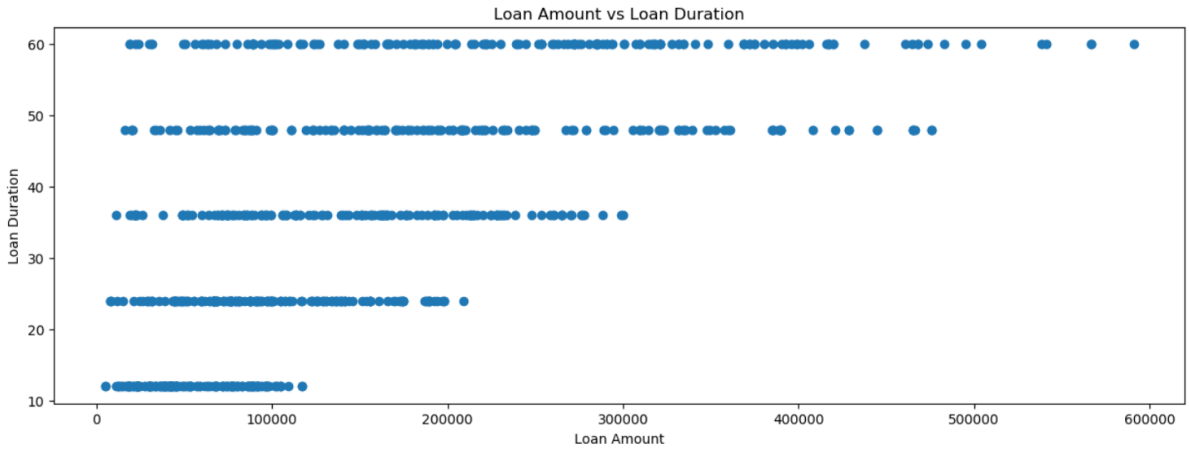
Urban Regions

```
0          Hl.m. Praha
25       Plzen - mesto
73     Ostrava - mesto
53        Brno - mesto
69              Karvina
Name: district_name, dtype: object
```



Cities (urban ratio 100%) and rural regions have a small difference in salary range, it is hard to say that urban residents always get paid more.

Loan Amount vs Loan Duration



Clients with higher amount of loans tends to have longer loan duration

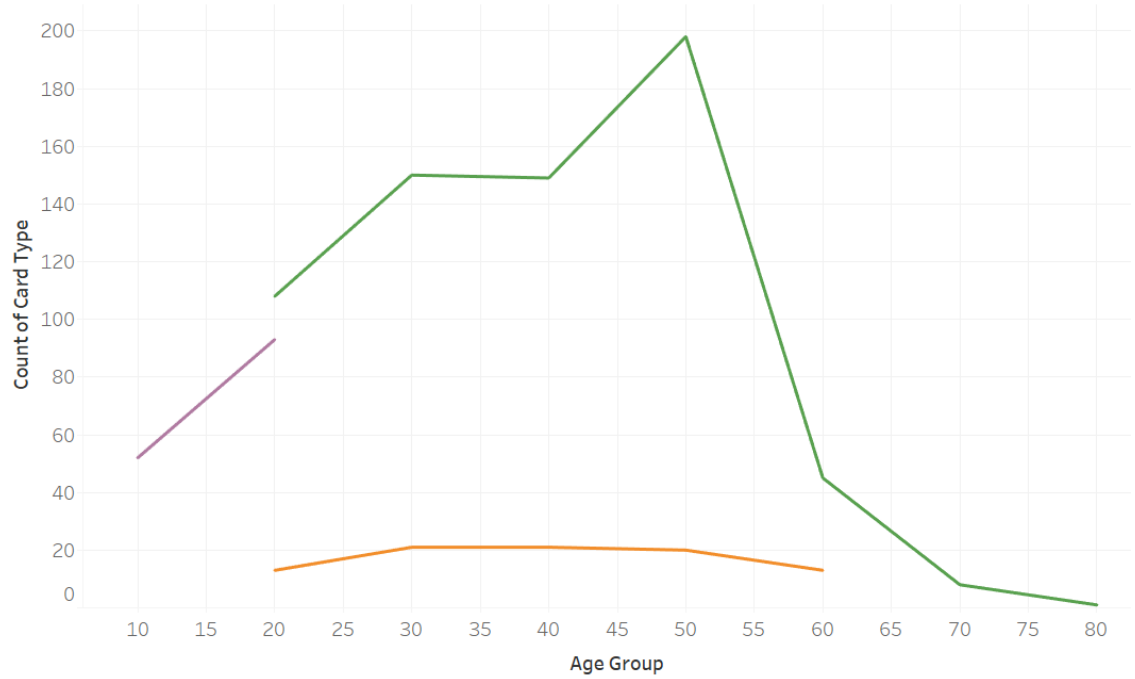
Loan Amount vs Monthly Payment

Clients with low amount of loans do not necessarily make low amount of monthly payments.

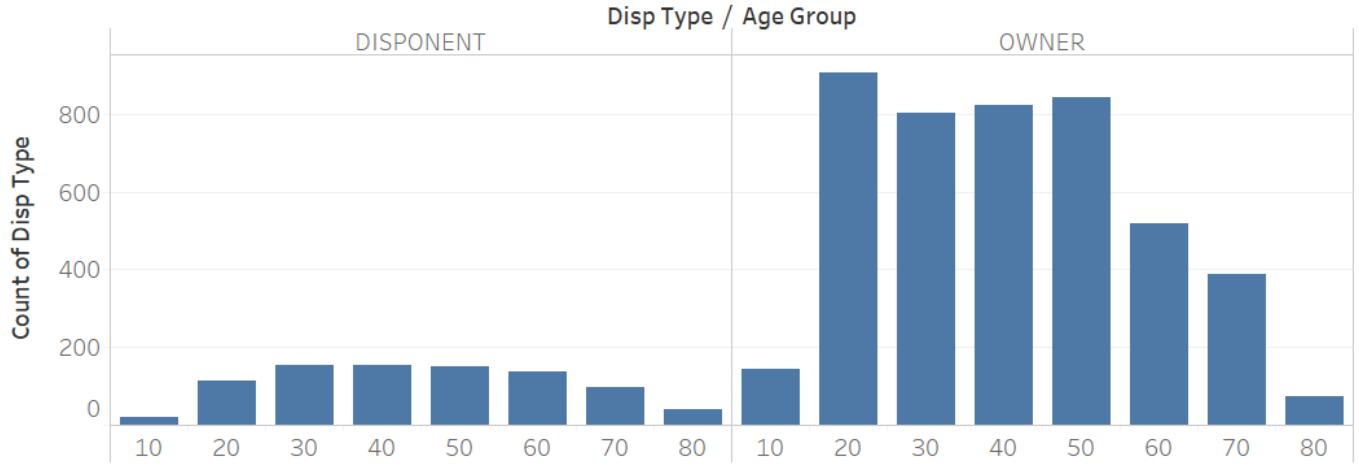
Loan Amount vs AGE

Clients regardless of the age seem to borrow loans between 50000 and 150000.

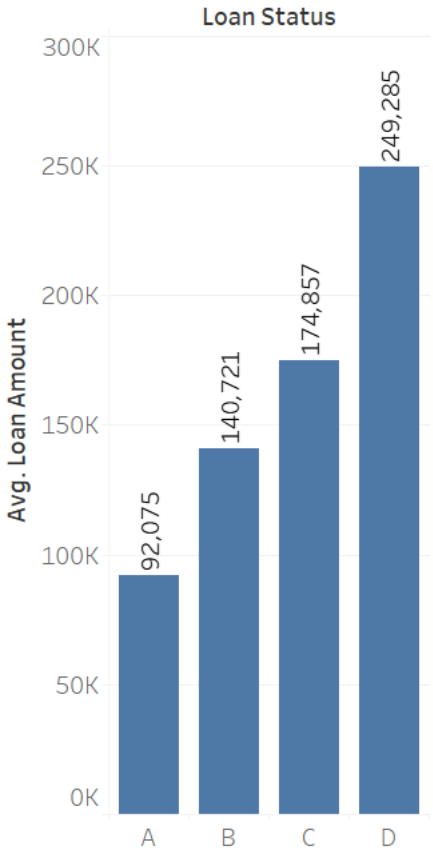
Type of card per age group



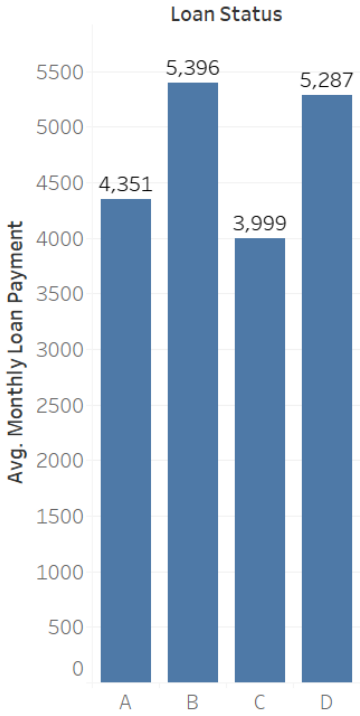
Dist type per age group



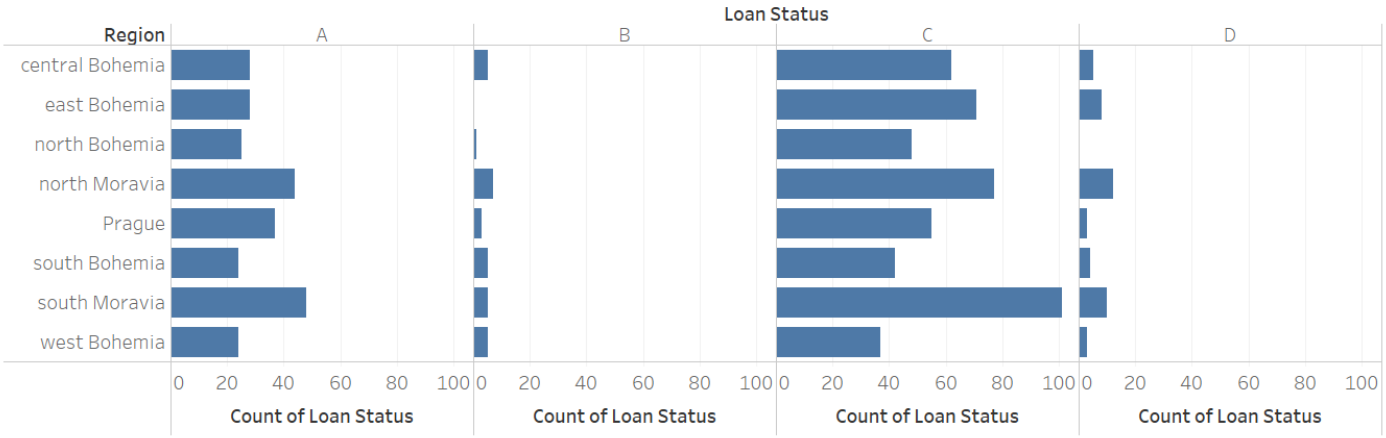
Loan Amount vs loan status



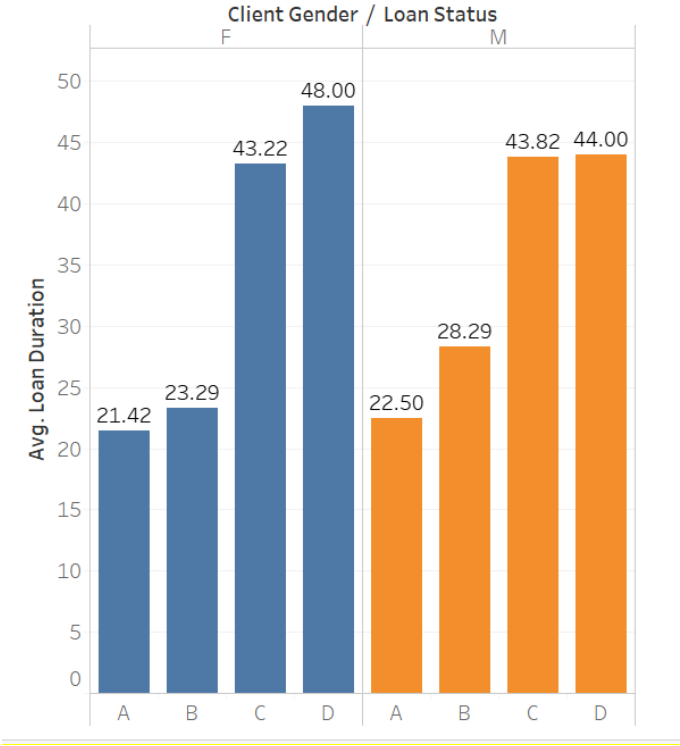
Loan payment (monthly) vs loan status



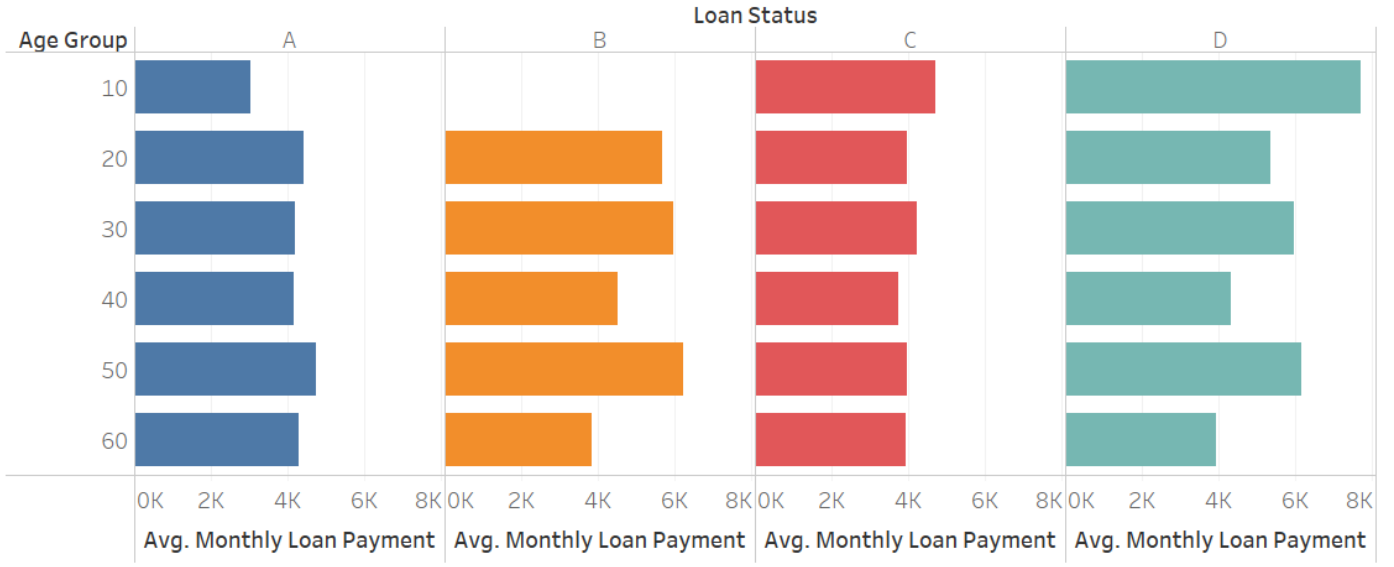
Loan type per region



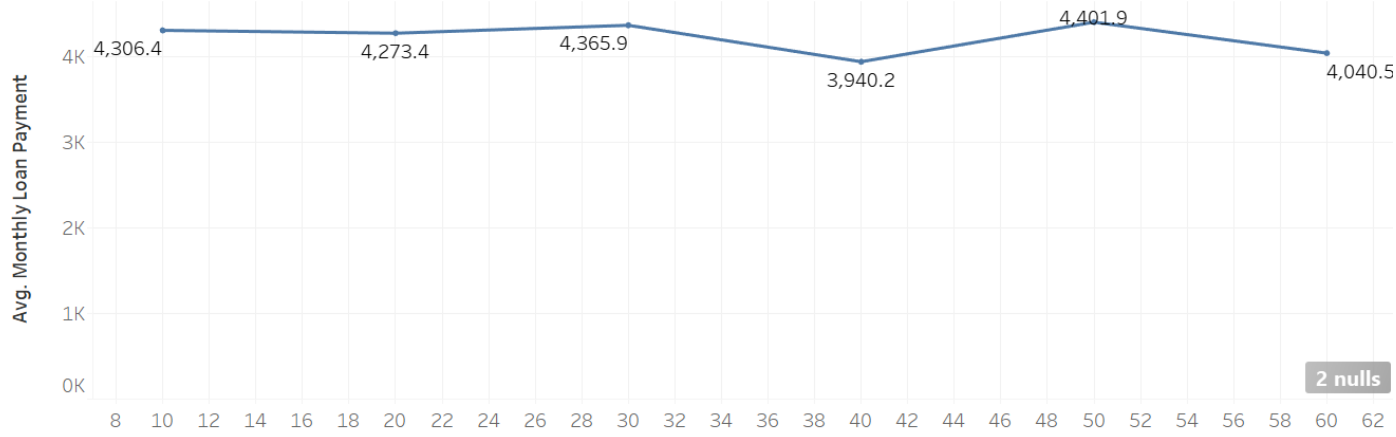
Loan duration (in month) vs loan status

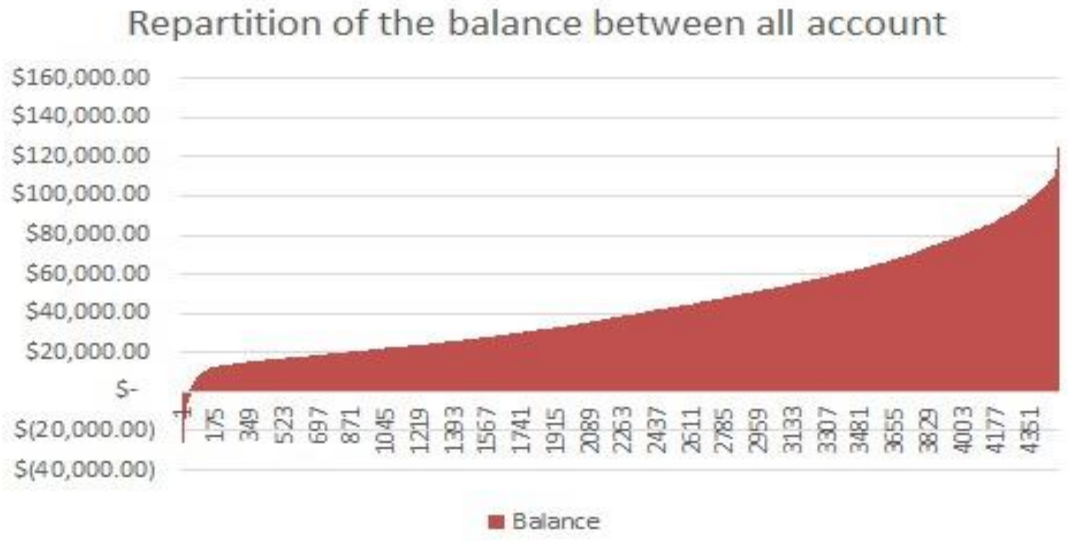


Loan type and monthly payment per age group



Age group vs the monthly loan payment (explanation)




Repartition of the balance between all account
$160,000.00
$140,000.00
$120,000.00
$100,000.00
$80,000.00
$60,000.00
$40,000.00
$20,000.00
$-
$(20,000.00)
$(40,000.00)
1
175
349
523
697
871
1045
1219
1393
1567
1741
1915
2089
2263
2437
2611
2785
2959
3133
3307
3481
3655
3829
4003
4177
4351
Balance

Datamart Table Explanation

Variable Name	Variable Description
Client_id	
Birth_number	
Client_district_id	Address of the client
Birth_year	
Birth_day	
Birth_month	
Client_gender	
Client_age	
Age_group	Group age of clients per 10 years, so 22 is considered 20 and 35 is considered 30 and so on
Disp_id	Relation disposition record identifier
Account_id	
Disp_type	type of disposition (owner/user) only owner can issue permanent orders and ask for a loan
Account_district_id	Location of the branch
Frequency	
Account_date	
Account_Date_opened	Date when account was opened
Frequency_statement	Frequency of statement issuance
accountyear	
Account_month	
Account_day	
Loan_id	
Date	
Loan_amount	Amount of loan granted
Loan_duration	Duration of the loan
Monthly_loan_payment	
Loan_status	Status of paying off the loan
num_days_since_loan_acquired	Number of days since loan was granted
Account_id_disp	
Card_id	
Card_type	
Issued	When card was issued
Card_issued_date	
District_id	
District_name	
Region	
no_of_inhabitants	
no._of_municipalities_with_inhabitants_<_499	

no._of_municipalities_with_inhabitants_500-1999	
no._of_municipalities_with_inhabitants_2000-9999	
no._of_municipalities_with_inhabitants_>10000	
no._of_cities	
ratio_of_urban_inhabitants	
average_salary	
unemploymant_rate_"95"	
unemploymant_rate_"96"	
no._of_enterpreneurs_per_1000_inhabitants	
no._of_commited_crimes_"95"	
no._of_commited_crimes_"96"	
trans_balance	
trans_freq_total	
trans_amount	
trans_freq_first_3_years	
trans_freq_last_3_years	
trans_freq_credit	
trans_freq_withdraw	
trans_last_in_days	
order_SUM_amount	
order_count	
order_bank_Is_AB	
order_bank_Is_CD	

order_bank_Is_EF	
order_bank_Is_GH	
order_bank_Is_IJ	
order_bank_Is_KL	
order_bank_Is_MN	
order_bank_Is_OP	
order_bank_Is_QR	
order_bank_Is_ST	
order_bank_Is_UV	
order_bank_Is_WX	
order_bank_Is_YZ	
order_k_Is_LEASING	
order_k_Is_POJISTNE	
order_k_IS_SIPO	
order_k_Is_UVER	