FINANCIAL PROGRAMMING - IÉSEG, Lille - MBD 2018-2019

Group Project By: Nitesh Shanbhag, Manju Nagaraj Rudrappa, Lucas Bonnett

- ➢ This Project is to create a base table for the PKDD'99 Discovery Challenge and do Visualization and Descriptive Analyses on the Datamart
- ➢ Basically, the bank wants to improve the Services and distinguish more closely on who is a good client (to offer additional services) & who is a bad client (whom to watch carefully to minimize the bank loses).
- ➢ The bank has provided data about their clients, the accounts (transactions within several months), the loans already granted, the credit cards issued. The below Base table will help bank managers improve their understanding of customers and seek specific actions to improve services.
- ➢ Final Client Bastable: Client Characteristics which includes 1 row per Client

DEMOGRAPHIC DATASET

- ➢ Probability of the Client being an Entrepreneur was calculated
- ➢ Probability of the client committing a Crime in year 1995/1996 was calculated
- ➢ Special Characters in the values were replaced with mean of the other valid values of that columns.
- ➢ Objects were converted to numeric.

CLIENT DATASET

We have assumed the current year as 1999.
- ➢ Transformed the Birth Number into Birth Date and Gender was derived
- ➢ Age and Age group was calculated.

CREDIT CARD DATASET

- ➢ Date Issued was converted to Date Object.

LOAN DATASET
- ➢ Date Issued was converted to Date Object.
- ➢ Status values were changed into understandable Values.

ORDER DATASET

- ➢ Date Issued was converted to Date Object.
- ➢ K-symbol with nan or spaces were converted into Others
- ➢ K-symbol was pivoted as columns(pivot) with Amount Values filled in it

Overall One row per account id was maintained

ACCOUNT DATASET

- ➢ Length of Relationship was Derived as of 1999 with account opened year
- ➢ Frequency values were mapped with understandable values

TRANSACTION DATASET

For the transaction Dataset the Values were converted in such a way to understand monthly behavior of the Customer

- ➢ Type and Operation was converted into Understandable Values, VYBER Values from Type were converted into
- ➢ Debit:
  Type and Operation was taken as Column Values
- ➢ Average Month End Balances:
- ➢ Average Month End Balance of the account over the period was calculated.
  Logic: Balance at each month end was found and Mean and Median of those values were found to get as values per account id.
- ➢ Here the Monthly all the different amounts are summed and after we get 1 row per month for each account id then Month wise averages are calculated. So, we get an idea about his Monthly Income, Monthly expenditure.
- ➢ Avg Monthly Savings were calculated as Avg_Credit - Avg_Debit
- ➢ Average Month End Balances**:** Logic is for each month end to find the account Balances so from that we can get an idea about what are his savings at month end for that month end row is only retained along with its balance and for that the Group by condition has been applied to get Mean and Median Balance at Month end.

After all the above Data Cleaning and Preparation:
One Final Client_Basetable with 1 row per Client_Id was created.

Important Metrics of the Base table

Base table Metrics	Description
Age	Assuming the current year as 1999, Age of each client was calculated by subtracting with birth year.
Age_Group	Age group of each client was calculated by taking the modulus of the age and multiplying by 10.
Gender	Gender of each client was found. if the birth month > 50, then Female else Male.
Length_of_Relation	Length of relationship of each client was calculated by subtracting current year (1999) and Account opened year.
Prob_enterprenuer	Probability of client being an entrepreneur was calculated.
Prob_committed_crimes	Probability of client being committed the crimes was calculated.
Avg_Monthly_Credit	This determines the average Monthly Credit of the Client.
Avg_Monthly_Debit	This determines the average Monthly Debit Amount of the Client.
Collection from another Bank	Mean Amount per month transacted as collection of another bank of the client.
Credit Card Withdrawal	Mean Credit Card Withdrawal Amount of the Client per month.
Credit in Cash	Mean Amount of Credit in Cash of the Client per month.
Other_Transactions	Mean Amount from all other transactions of the Client per month.
Remittance to another Bank	Mean Amount remitted to another bank from Client per month.
Withdrawal in Cash	Mean Amount Withdrew in Cash till now per month.
Avg_Monthly_Balance	Mean Monthly Balance of the Client.
Balance_Median_per_Month	Median Balance Per month of the Client.
Avg_Monthly_Savings	Avg Monthly Savings of the Client: Credit - Debit.
LEASING	Total Amount Transacted by Leasing K_Type.
Other Transactions	Total Amount Collected by Other K_Type.
POJISTNE	Total Amount Collected from POJISTINE.
SIPO	Total SIPO Amount transacted per Client.
UVER	Total UVER Amount Transacted by the Client.

Visualization

Number of clients (Accounts opened) to the bank each year:

The graph shows that, there is a sudden drop of clients to the bank in the period of one year from 1993 to 1994 after that we can see the increasing trend from 1994 till 1996, but in the recent years again the clients number is decreasing.

account_created_year	Number of clients
1993	1385
1994	513
1995	768
1996	1624
1997	1079
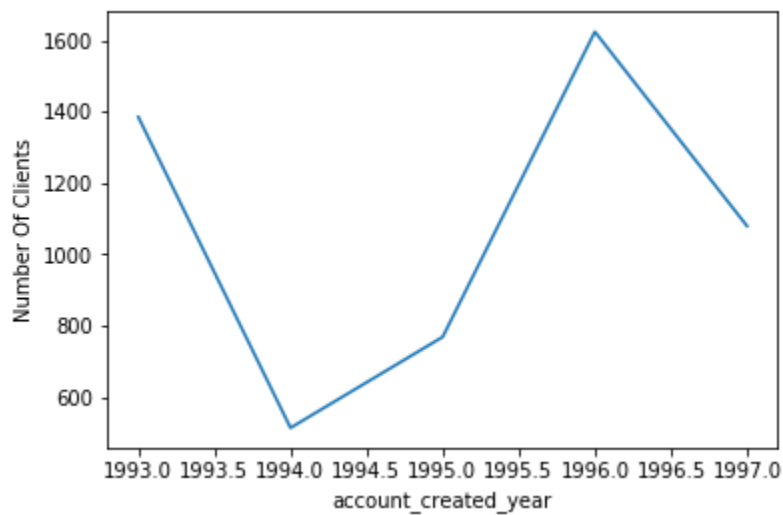
Average Monthly Savings Region Wise:

The graph shows that all the regions have almost equal Average Monthly Savings

Region	Average Monthly Savings
Prague	1287.563395
central Bohemia	1304.344262
east Bohemia	1309.064818
north Bohemia	1250.967951
north Moravia	1264.243609
south Bohemia	1254.348071
south Moravia	1271.920807
west Bohemia	1218.366009
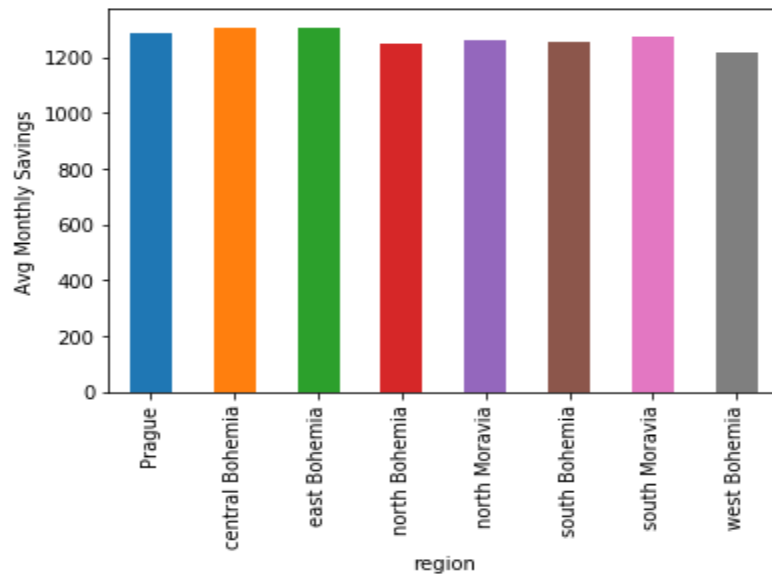
Average monthly balance Vs credit card type

The graph shows that the clients who has maintained the highest Average Monthly Balance are issued with gold credit card and then follows classic and junior
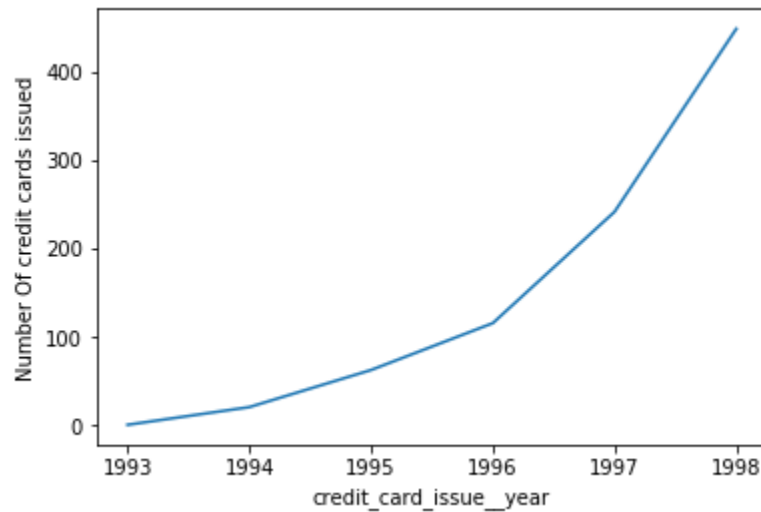
credit_card_type	Average Monthly Balance
classic	49240.64544
gold	56839.39631
junior	45932.23293
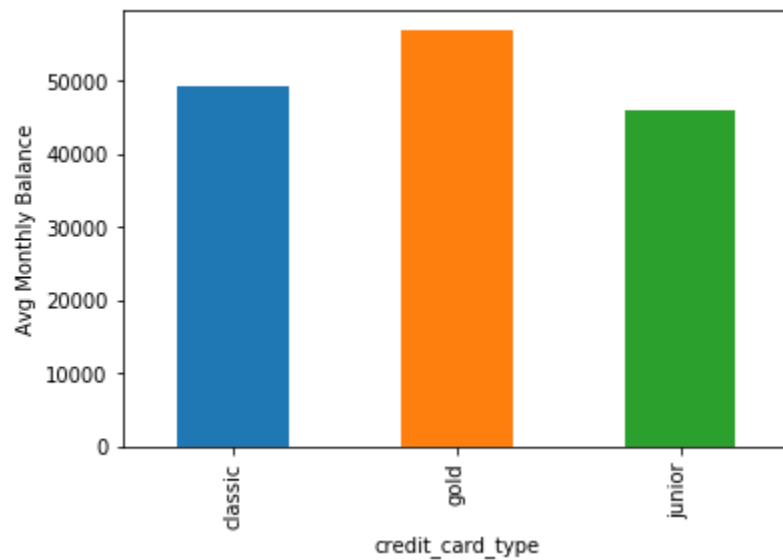
Number of credit card issued Yearly wise

The graph clearly shows that, the credit card users are increasing each year.

credit_card_issue_year	Number of credit card issued
1993	1
1994	21
1995	63
1996	116
1997	242
1998	449

Age group that asked for loans the most :

The age group between 20 to 50 have asked the loans for the most than the teens and elderly clients.

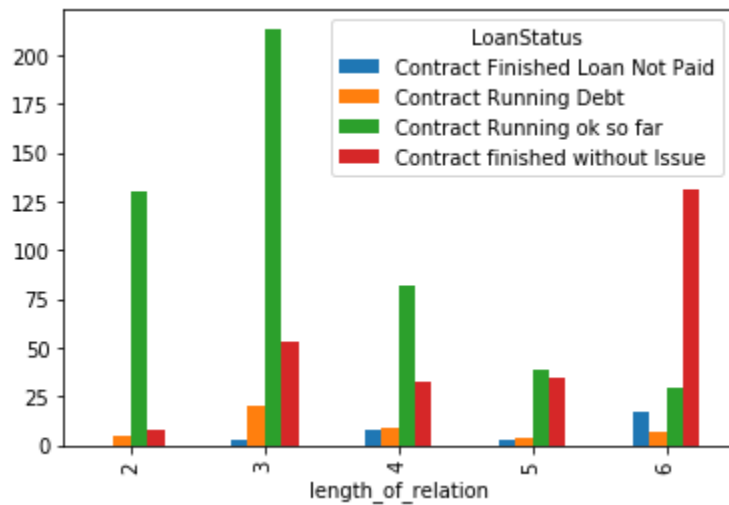| age_group	Number of time Loan asked
10	21
20	176
30	188
40	186
50	190
60	66
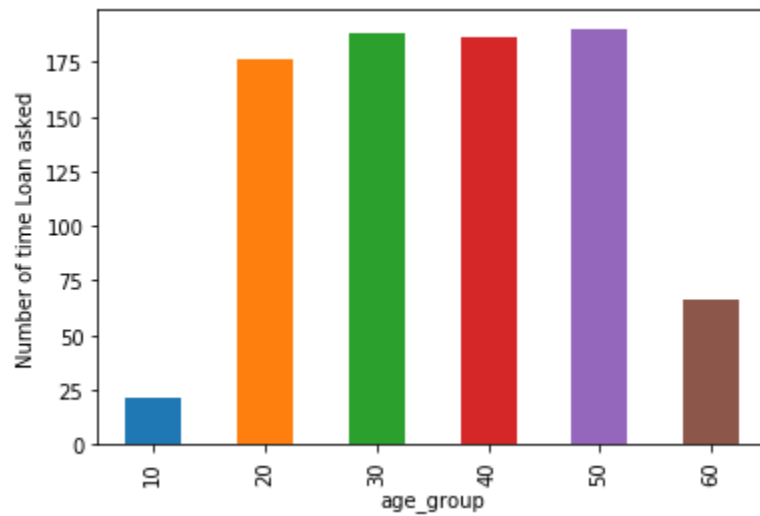
Loan status in each gender

Interestingly both Male and Female have same figures of loan status , having more number of loan contracts running smoothly so far, and only few clients have finished their contracts without paying the loan amount in which majority are females.

Loan Status	Contract finished without Issue	Contract Finished Loan Not Paid	Contract Running ok so far	Contract Running Debt
Gender				
F	130	17	246	24
M	128	14	247	21
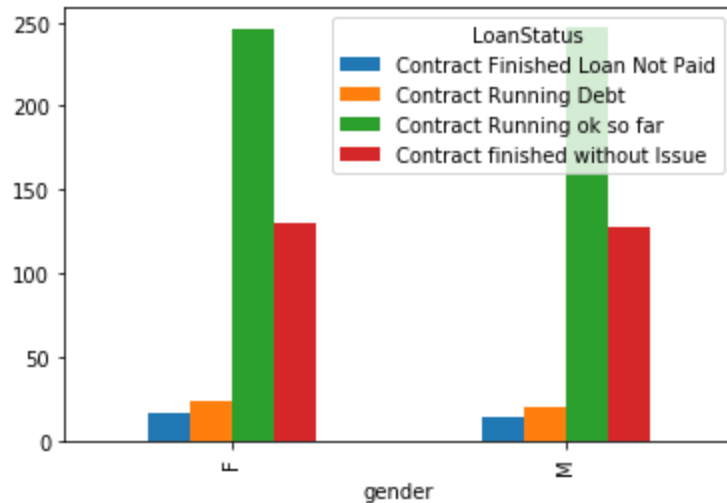
loan status in each group of Length of relationship

| Loan Status	Contract finished without Issue	Contract Finished Loan Not Paid	Contract Running ok so far	Contract Running Debt
LOR				
2	8	0	130	5
3	53	3	213	20
4	32	8	82	9
5	34	3	39	4
6	131	17	29	7

Average Amounts Transacted Per Type of Transactions per Month:

In the Below Graph:

The Total Amount Transacted from each Type of Operation has been Calculated. And we can see that Withdrawal in Cash and Credit in Cash stands amounts to max percentage of the total.

| Type of Transactions:	Total Amount Transacted in Euros
Another_Bank	EUR 23,082,353.36
CreditCardWithdrawal	EUR 479,686.29
Credit_In_Cash	EUR 68,524,987.91
Other_Transactions	EUR 770,976.10
Remittance_Other_Banks	EUR 19,315,754.93
Withdrawal_In_Cash	EUR 65,751,806.52
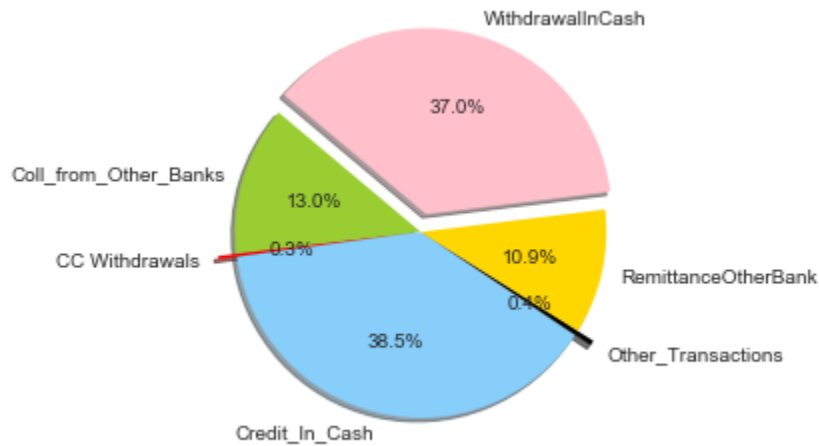
Plotting Average Monthly Savings of all the Customers:

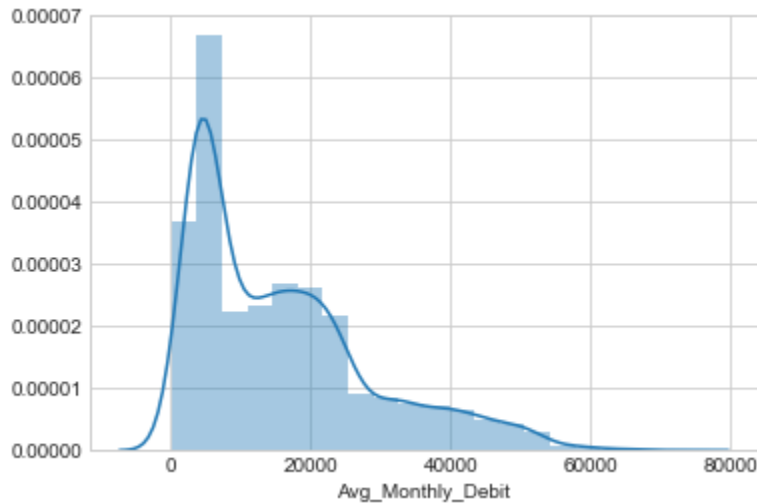Here we can see that, Among the customers, majority of them are able to save just 300 to 1000 Euros per month
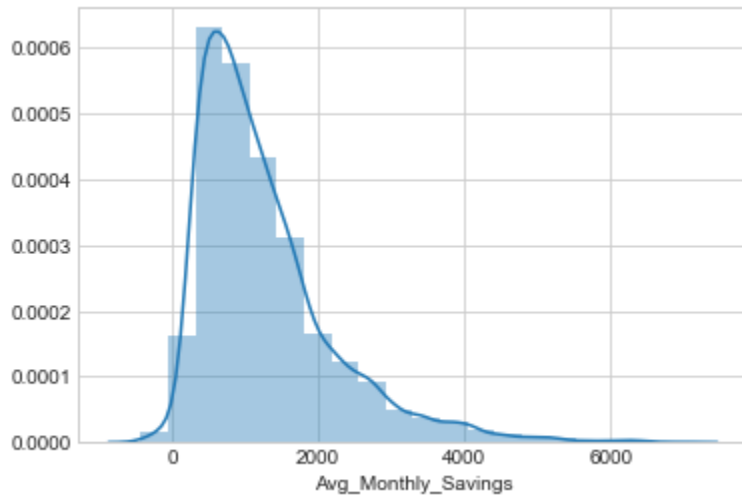
Plotting Avg Monthly Credit :

We can see that, most of the Avg_Monthly_Credit of Customers falls in between 1000 to 20000
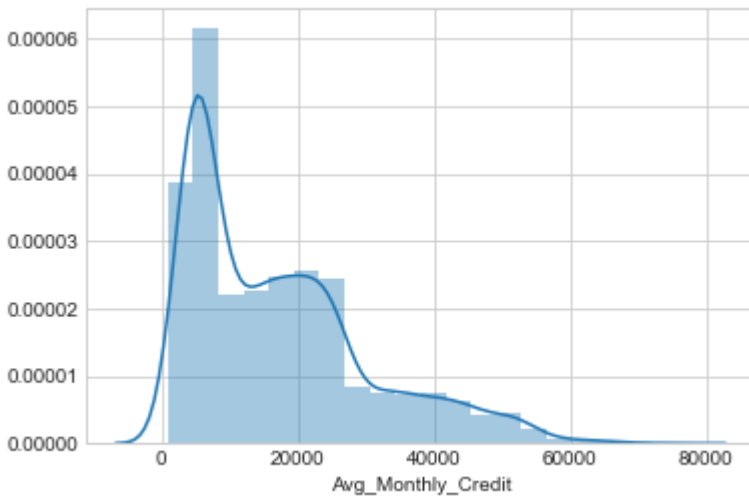
Avg Monthly Debit:

We can see that, most of the Avg_Monthly_Debit of Customers falls in between 1000 to 20000

Summary Statistics of Important Metrics

	Avg_Monthly_Credit	Avg_Monthly_Debit	Avg_Monthly_Balance	Balance_Median_per_Month	Avg_Monthly_Savings	LoanAmount
count	5369	5369	5369	5369	5369	5369
mean	17205.87	15933.553	35394.972	35004.949	1272.3169	23382.356
std	13361.538	13024.033	15283.644	15582.977	945.97192	70875.633
min	934.69388	107.69231	5522.8576	358.5	-430.6775	0
25%	5991.5333	5065.0071	22566.629	22346.4	607.57031	0
50%	14193.928	12835.013	33405.09	32453.4	1029.896	0
75%	24302.075	22591.947	45773.986	44679.8	1643.4934	0
max	75197.181	72330.919	93869.35	102374.25	7025.6188	590820

Amount of Transactions Per K-Type:

Here we plotted the total amounts transacted per K-Type of Transaction and we can see that SIPO transactions were max percentage of share.

K-Type Transaction Type	Total Amount Transacted in Euros
LEASING	EUR 877,578.20
Other	EUR 3,052,763.00
POJISTNE	EUR 834,813.00
SIPO	EUR 16,803,206.50
UVER	EUR 3,730,286.60