Rudwaan Vankar

EDA CASE STUDY

BANK LOAN ANALYSIS

Problem Statement

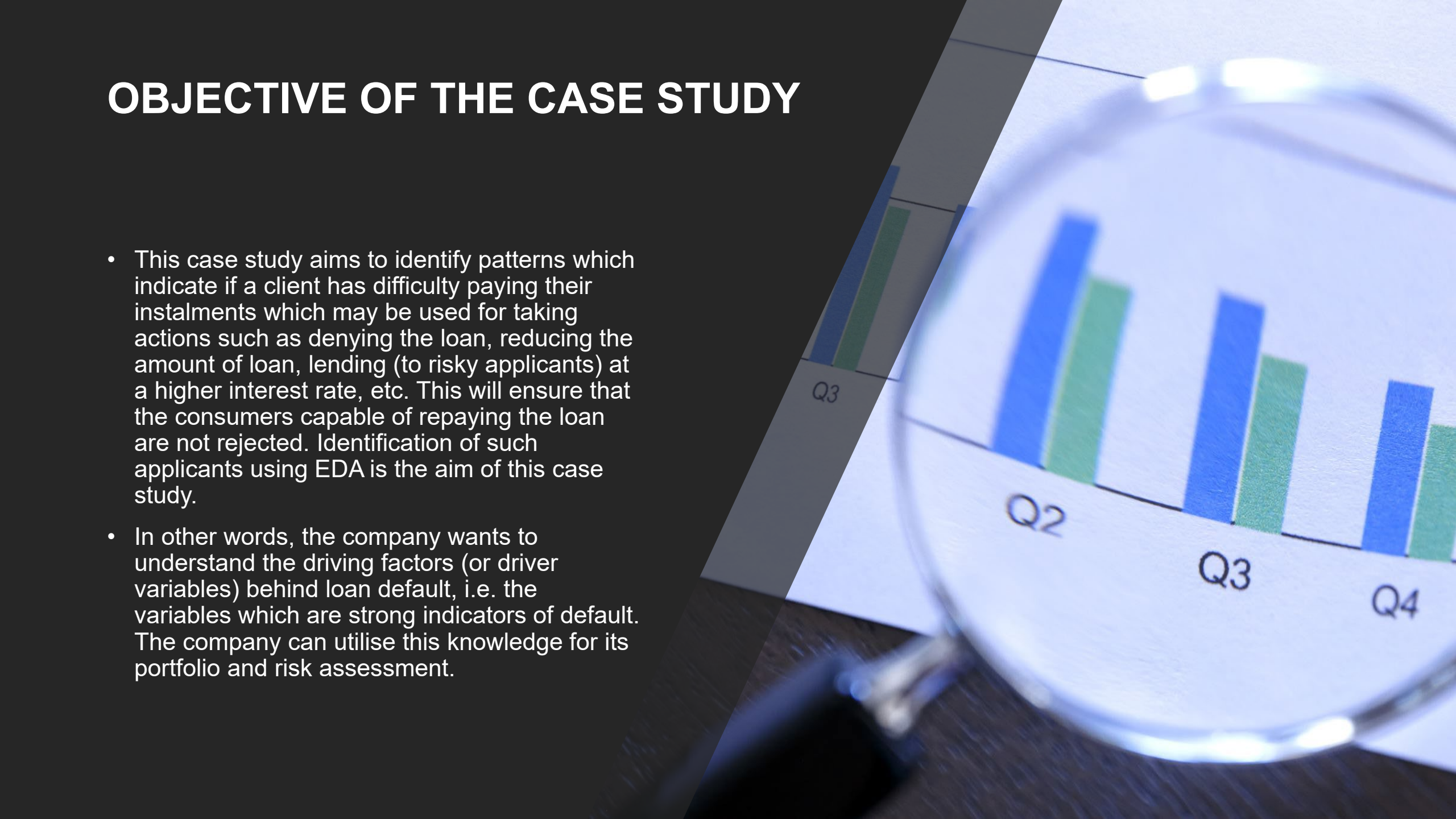
When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

1- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

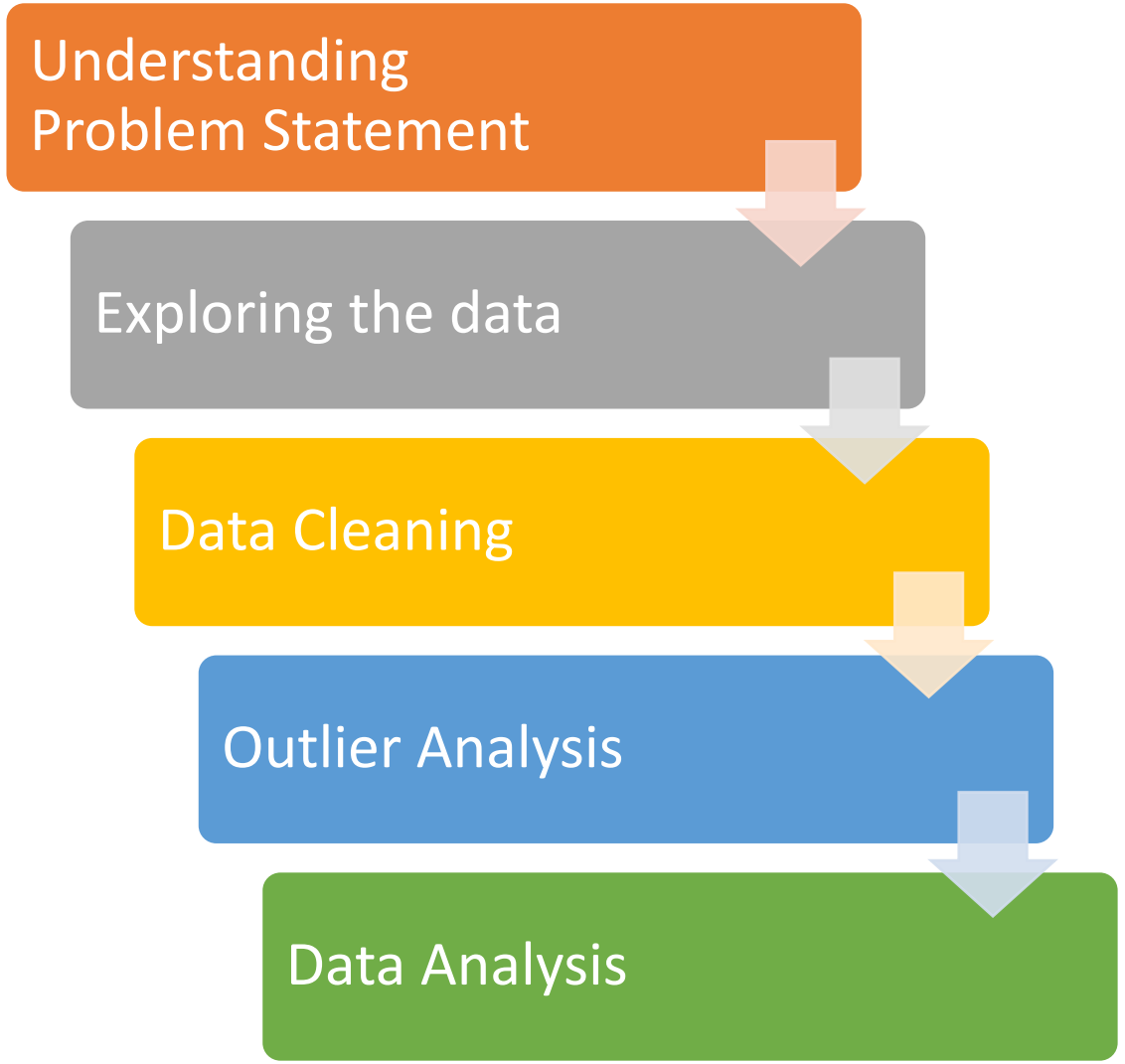
2- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

OBJECTIVE OF THE CASE STUDY

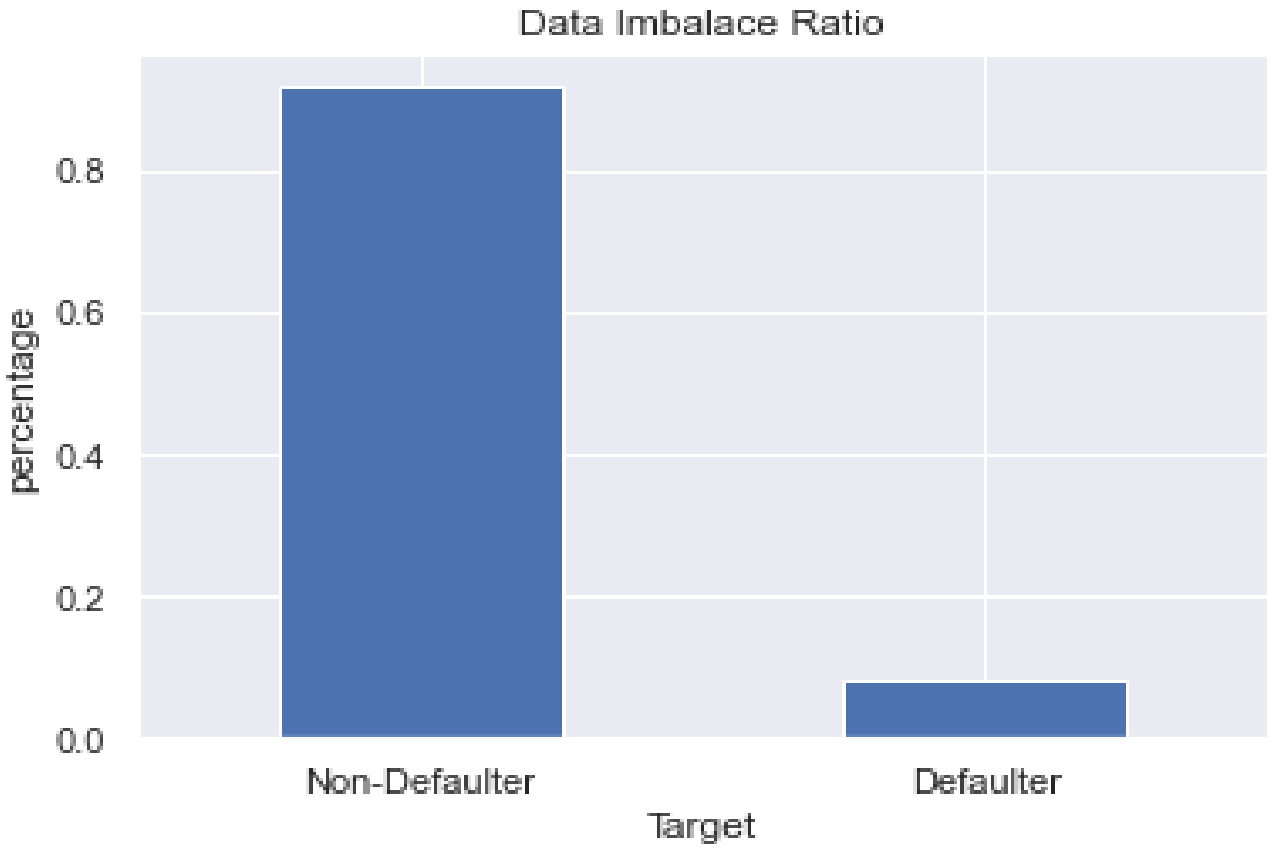
- This case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.
- In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.



Analysis Steps and Approach



Data imbalance Ratio

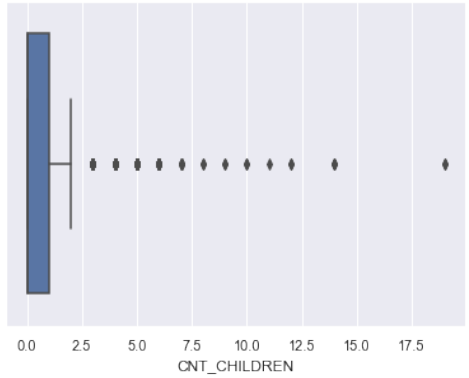


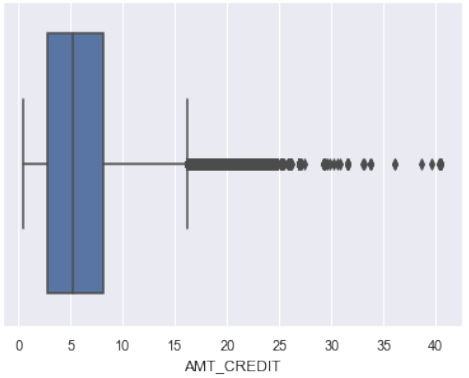
The Data imbalance Ratio Between Non- Defaulters and Defaulters In target Variable is 91.91 and 8.07

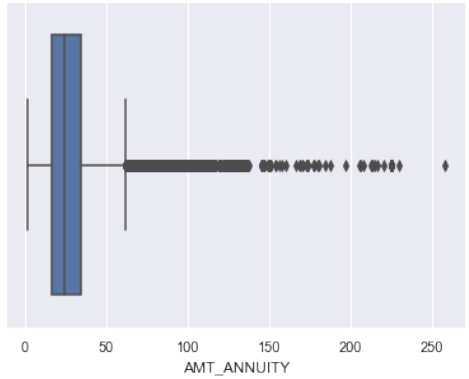
Making a ratio of 11.38

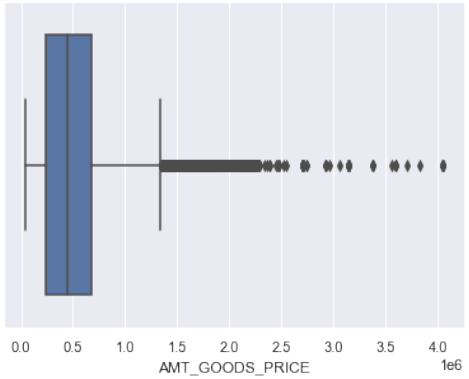
Outliers Analysis

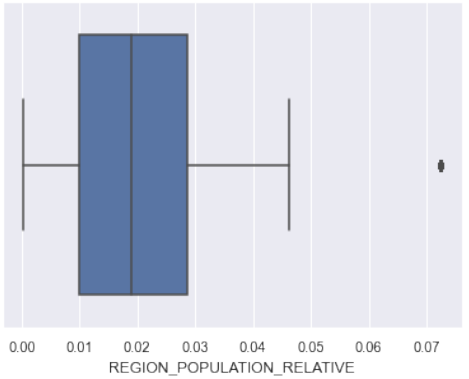
- From the above diagrams we can conclude that:
- Amt_Income_Total has a very high outlier, but it does not mean it could be an error. Some people have very high income
- Amt_Credit, Amt_Annuity, Amt_Goodprice, CNT_Fam_members, Cnt_Children have outliers which are normal considering some people have many kids or many family members,or high credit/annuity value
- Year employed has an outlier error where it shows some people have been employed for 1000 years
- Age Column has no outliers which means the data is proper

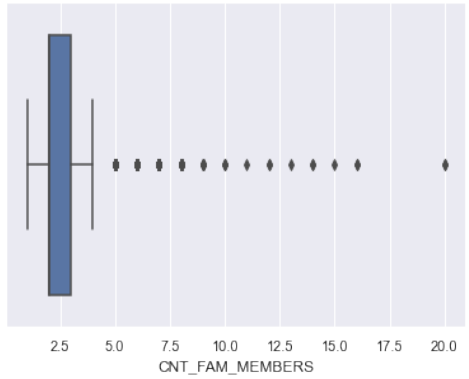


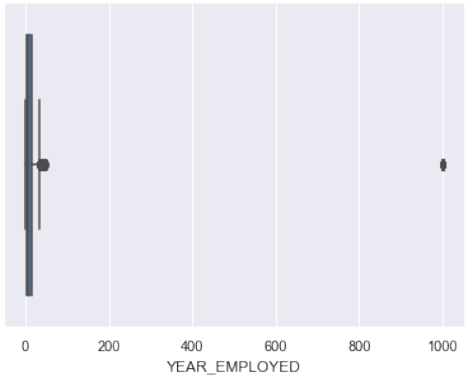


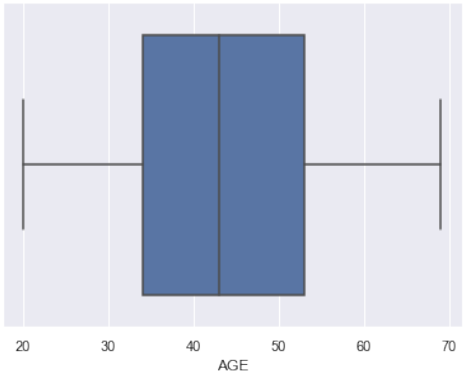






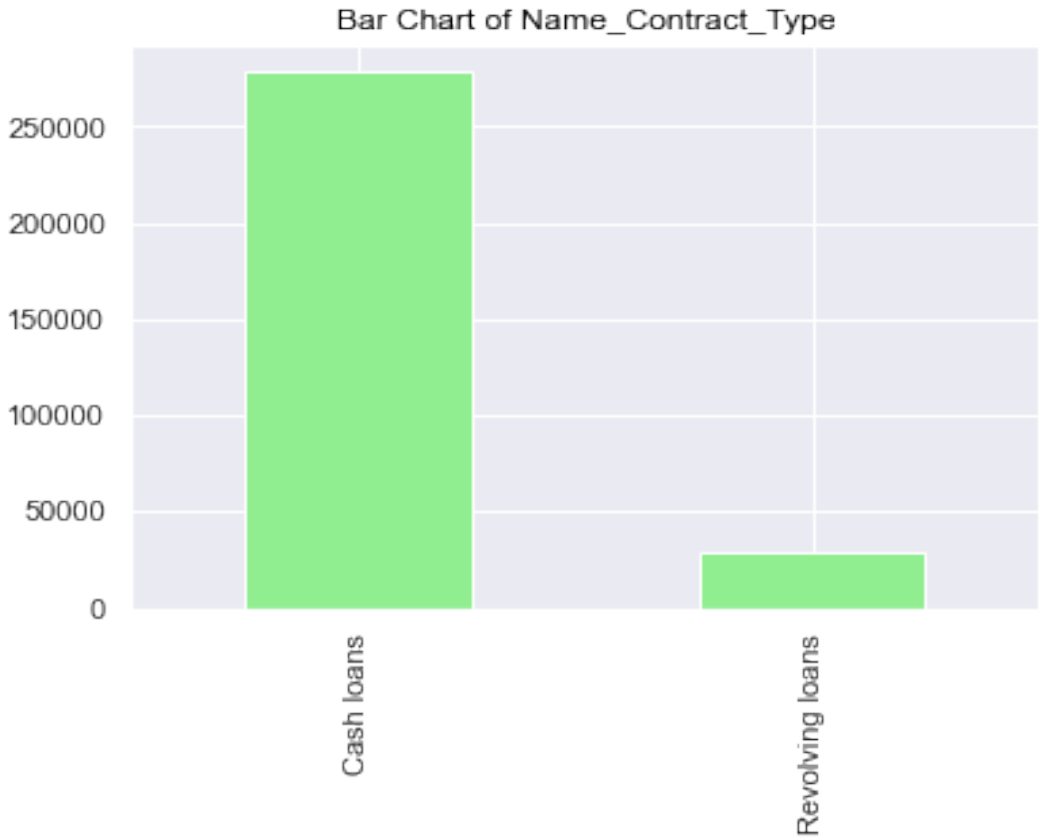


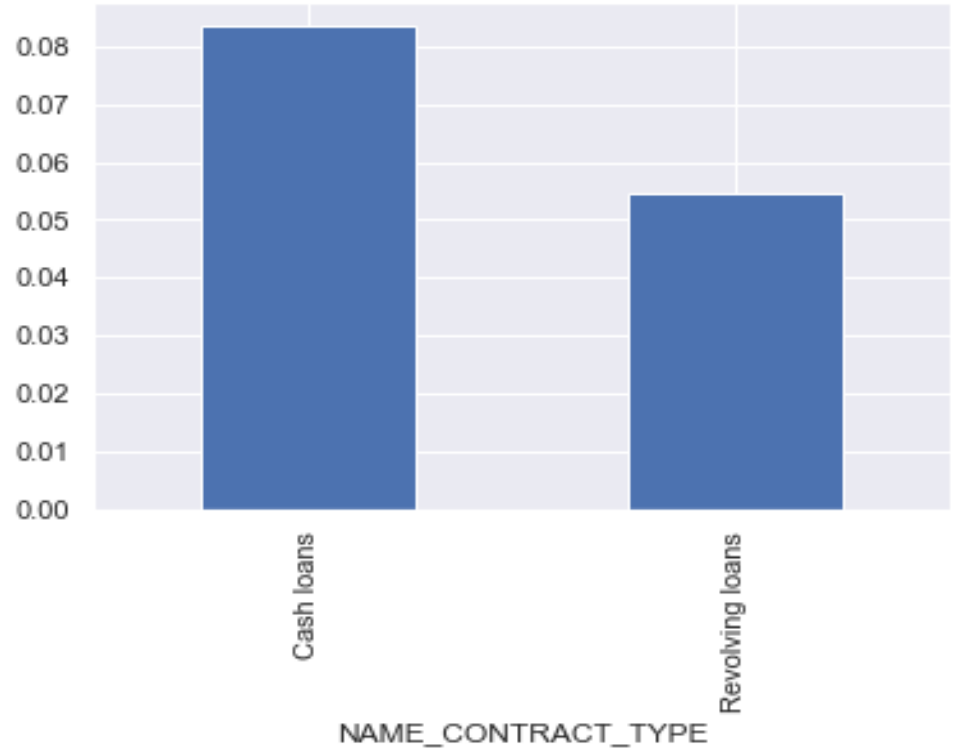




UNIVARIATE/ BIVARIATE ANALYSIS

NAME CONTRACT TYPE

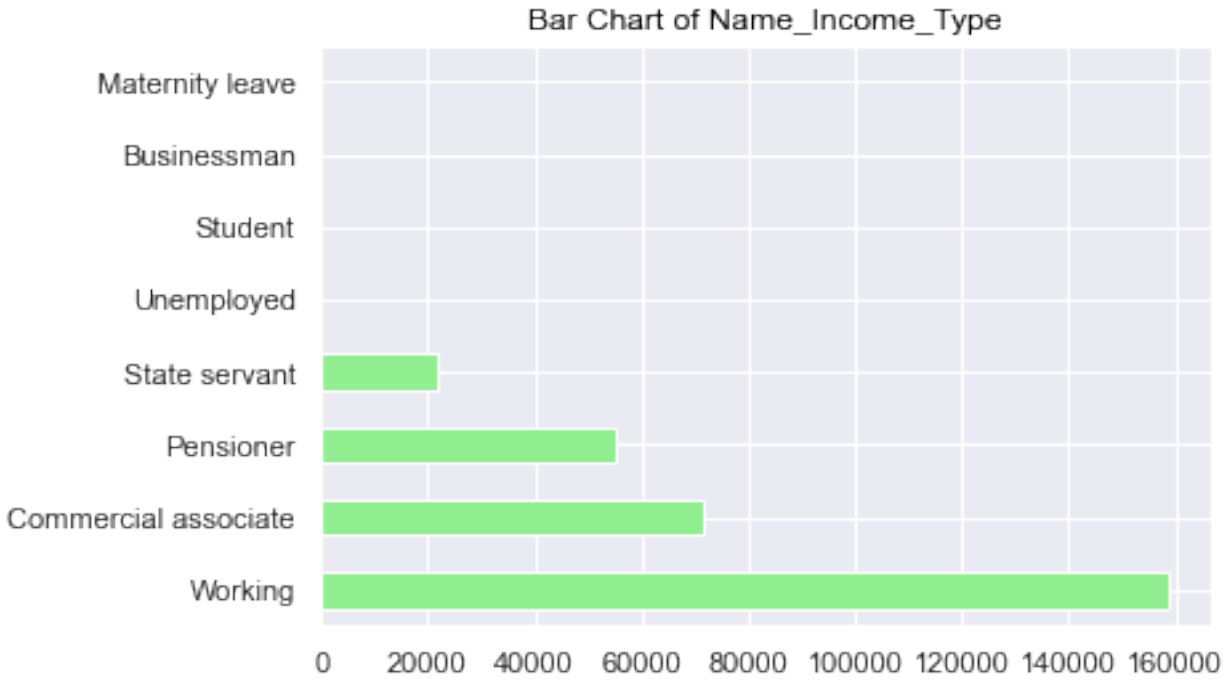


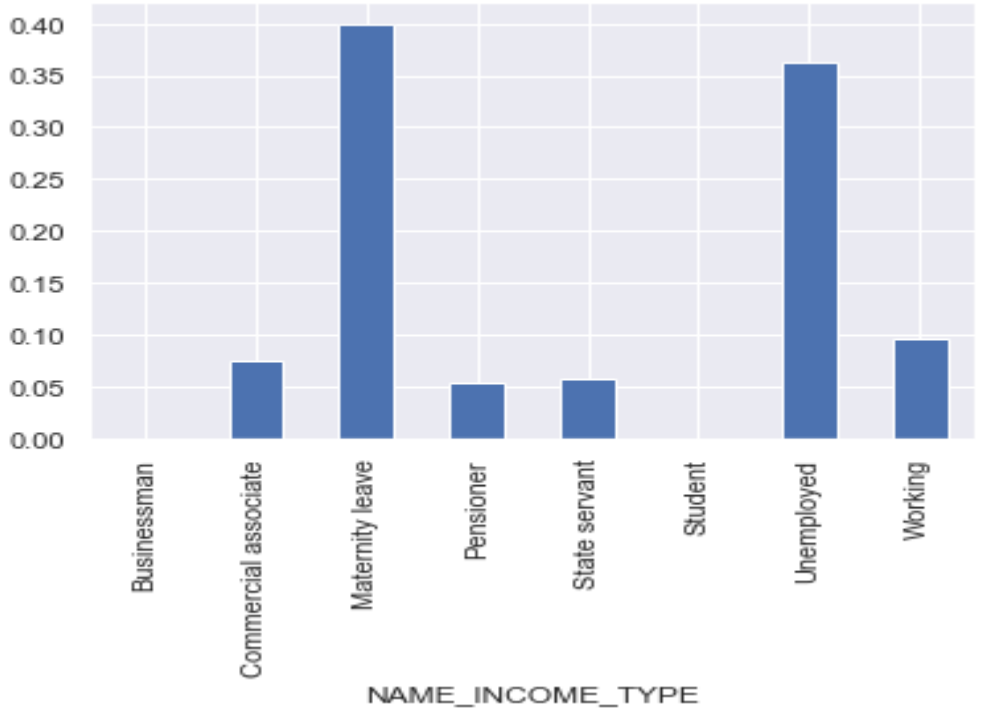


The Contract_type graph shows that CASH loans were taken by 90.47% of the people! and only 9.52% people took Revolving Loans

•Cash Loans had the highest percentage of defaulters
•whereas revolving loans had lesser percentage of defaulters

NAME INCOME TYPE





•Income_type graph shows clients income type. The majority of the clients who applied for loans were working class professionals

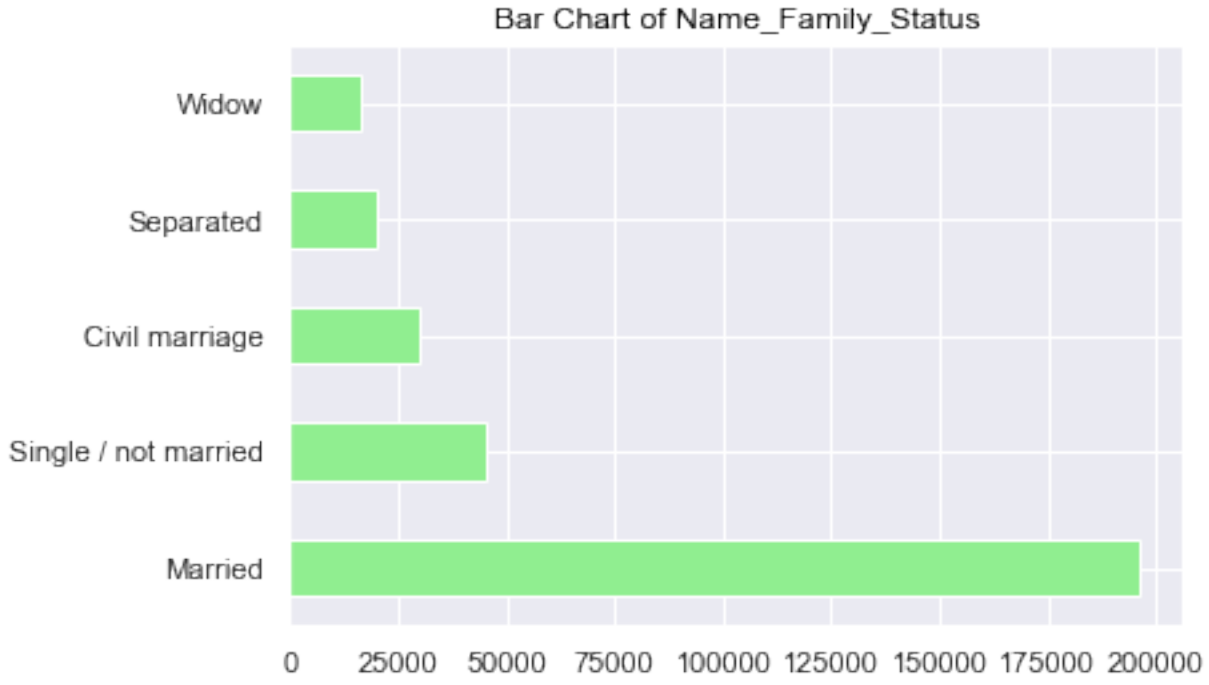
•Students, Bussinessman, Maternity leave people were the least amount of people who applied for loan

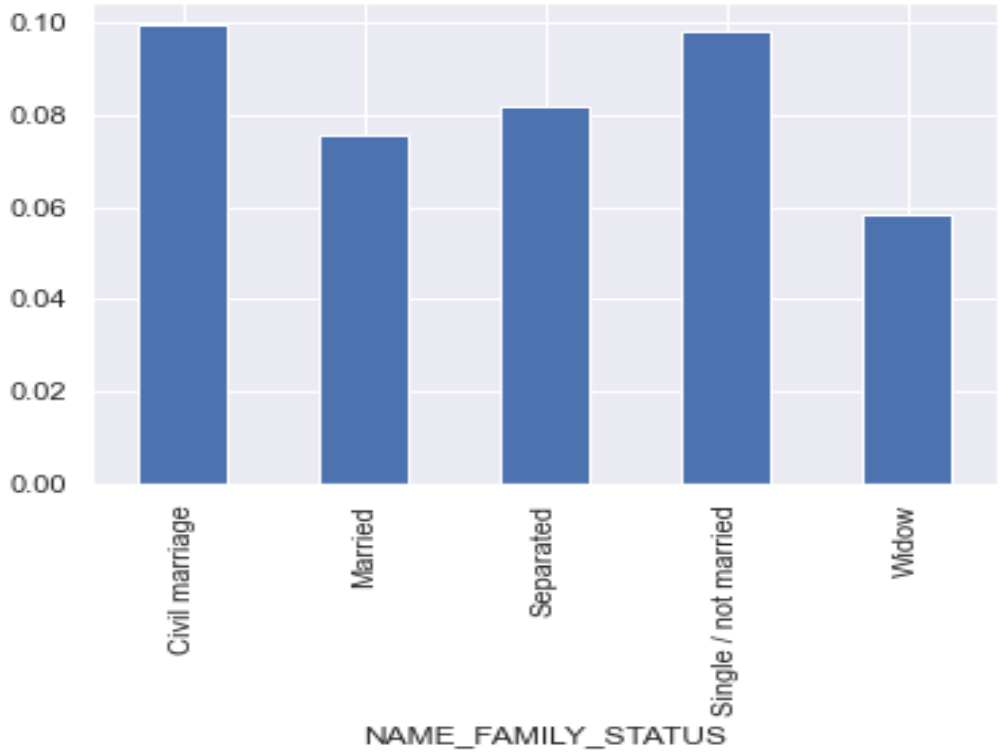
•Unemployed clients had high default rate

•Maternity Leave clients were also higher in default rate

•Students and Businessman were very less likely to Default

NAME FAMILY STATUS

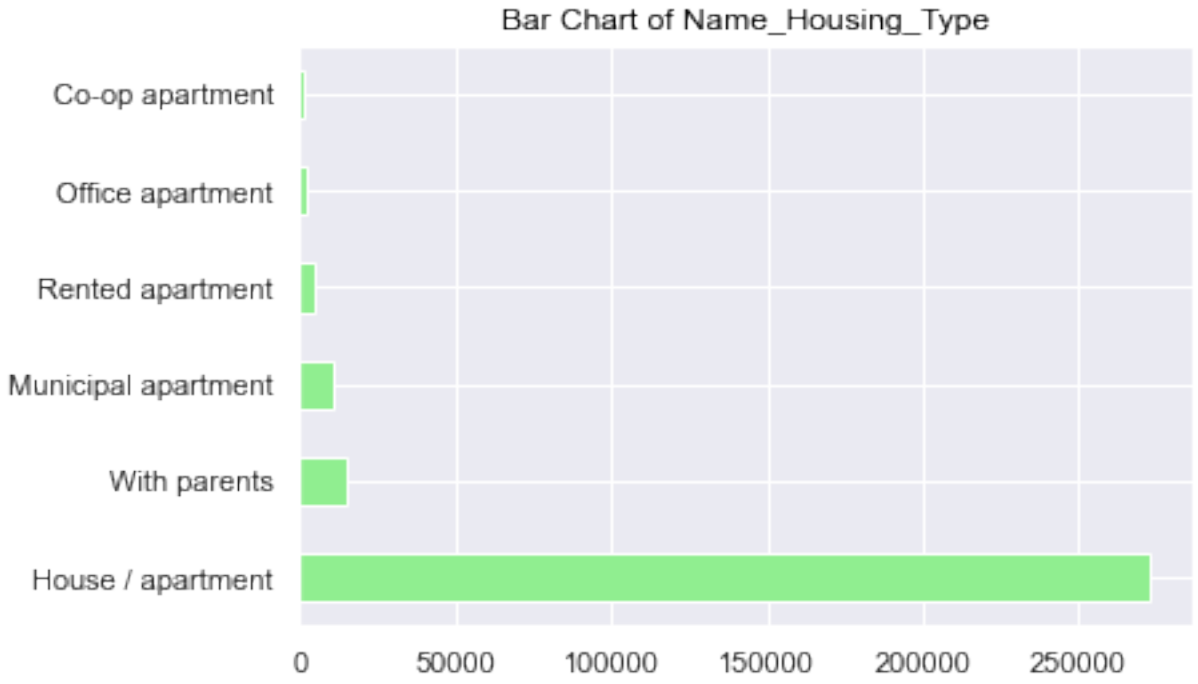


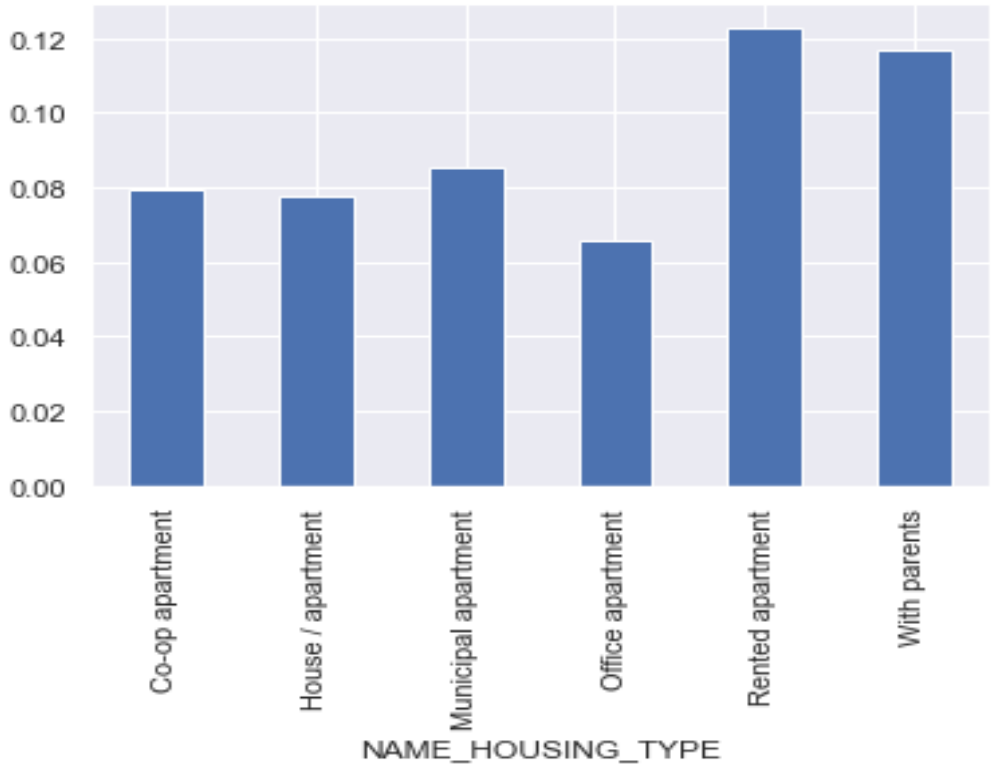


- Married people were the higher percentage of loan applicants with 63.87%. Followed by Single people having 14.77%
- Civil marriage, Seperate, Widow were among the lower percentage category who applied for loans

- Civil marriage and Unmarried Clients had the most default ratio
- Widows were the least likely to default

NAME HOUSING TYPE

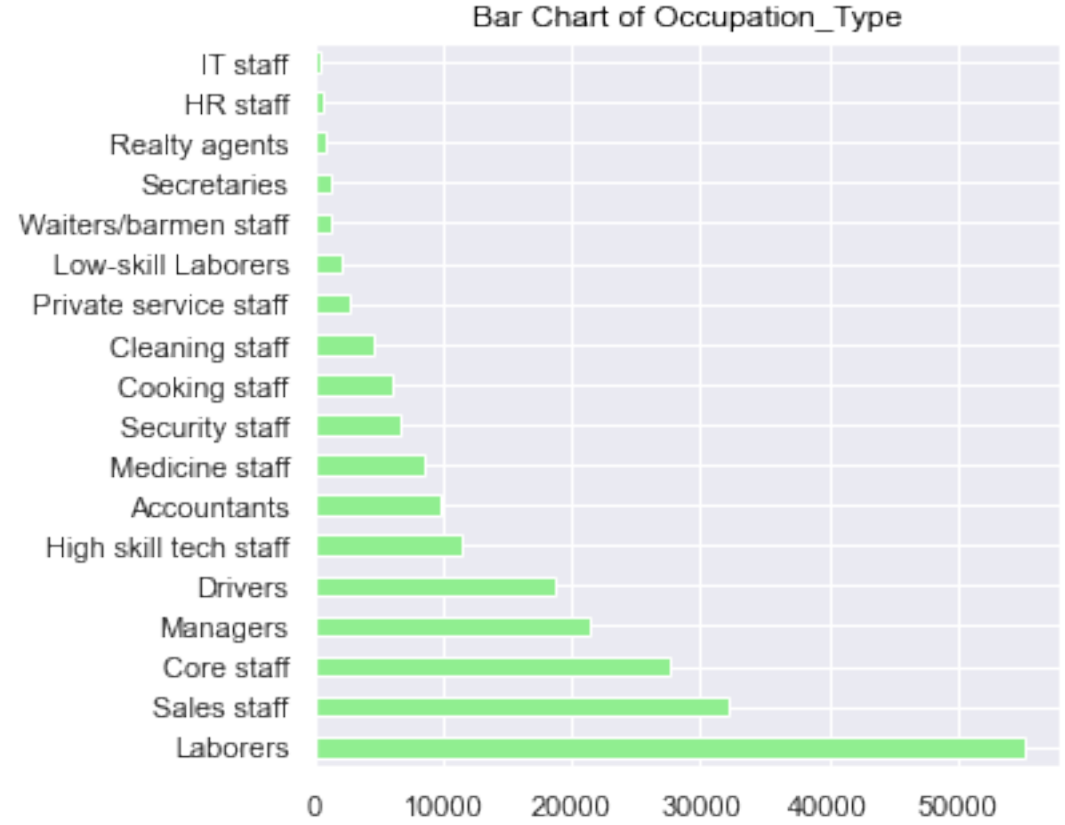


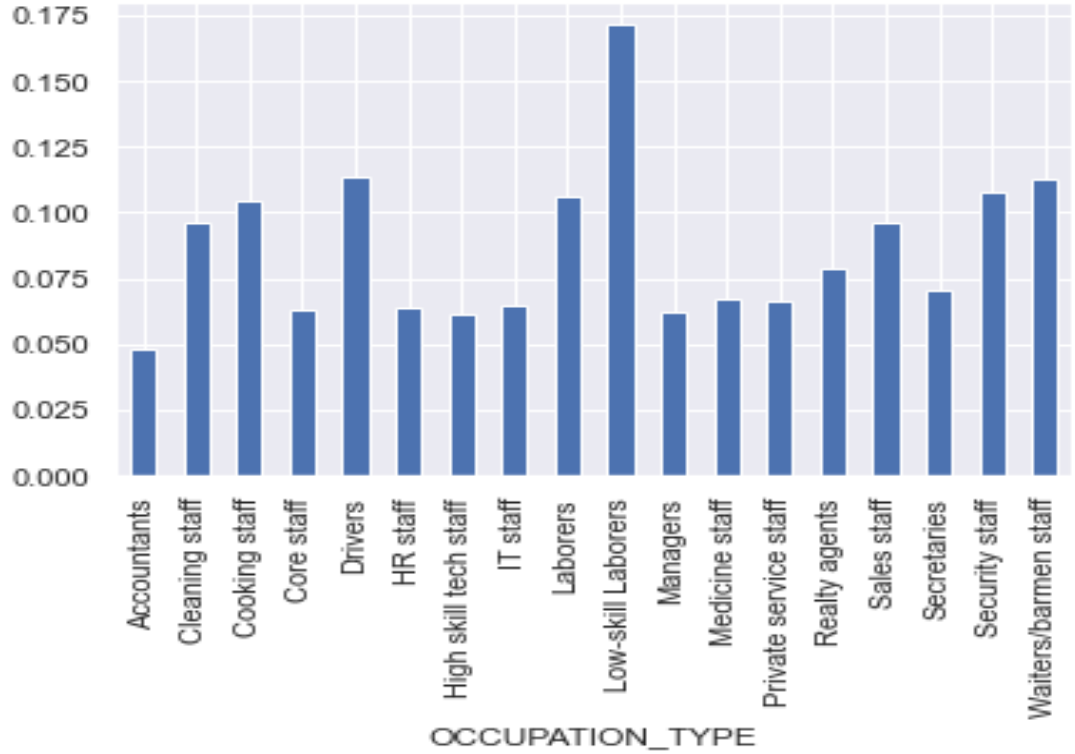


- 88.73% of the applicants were living in a House / Apartment
- 4.82% were living with their parents followed by 3.63% living in Muncipal Apartment

- Clients living in Rented Apartments and with parents had the Highest default percentage
- Office apartment, House, and Co-op had the least default percentage

NAME OCCUPATION TYPE

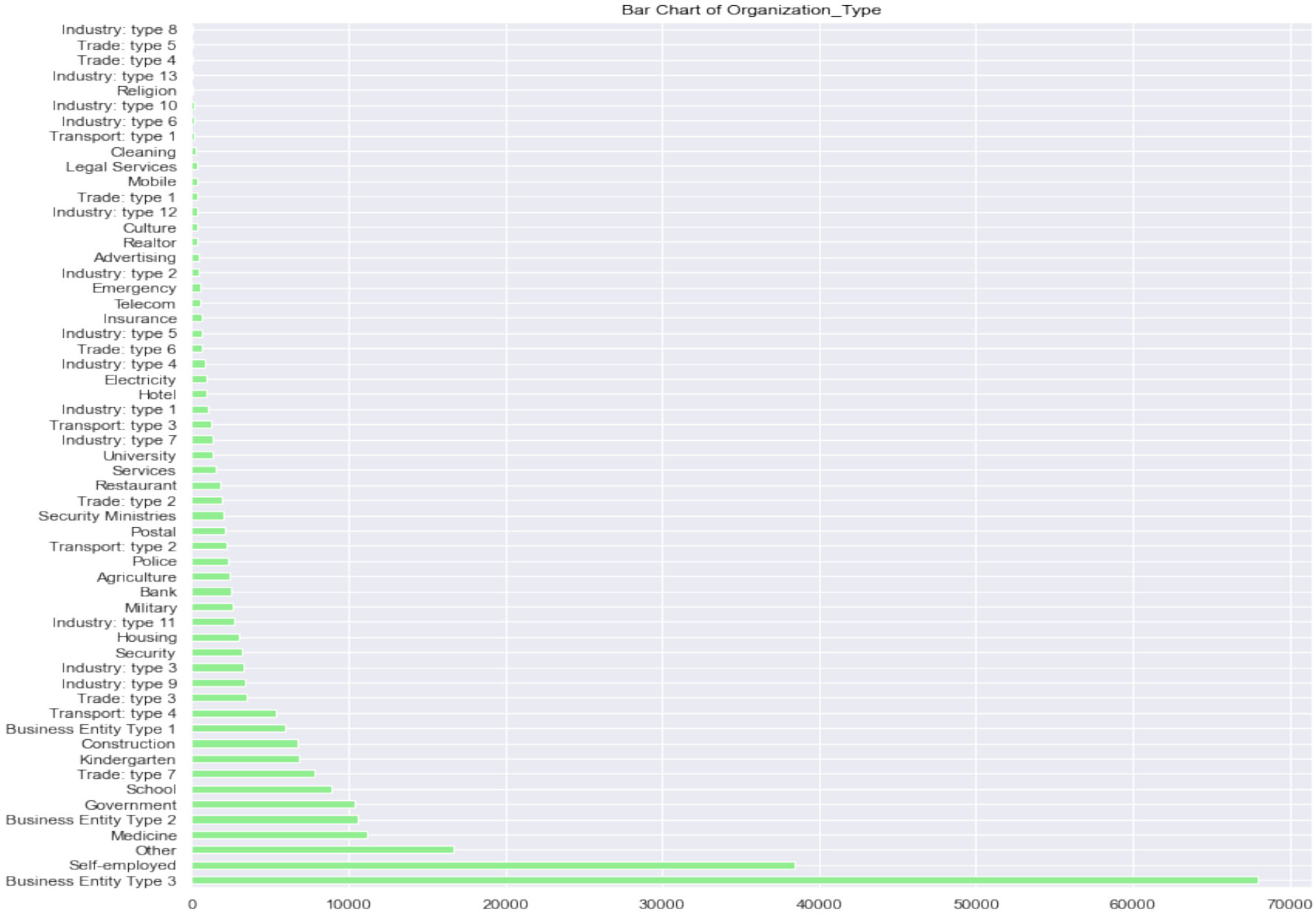




Laborers were the highest percentage of people who applied for a loan having the value of 26.13%, Followed by Sales staff (15.20%), Core staff (13.05%), Managers (10.12%), Drivers (8.81%) etc

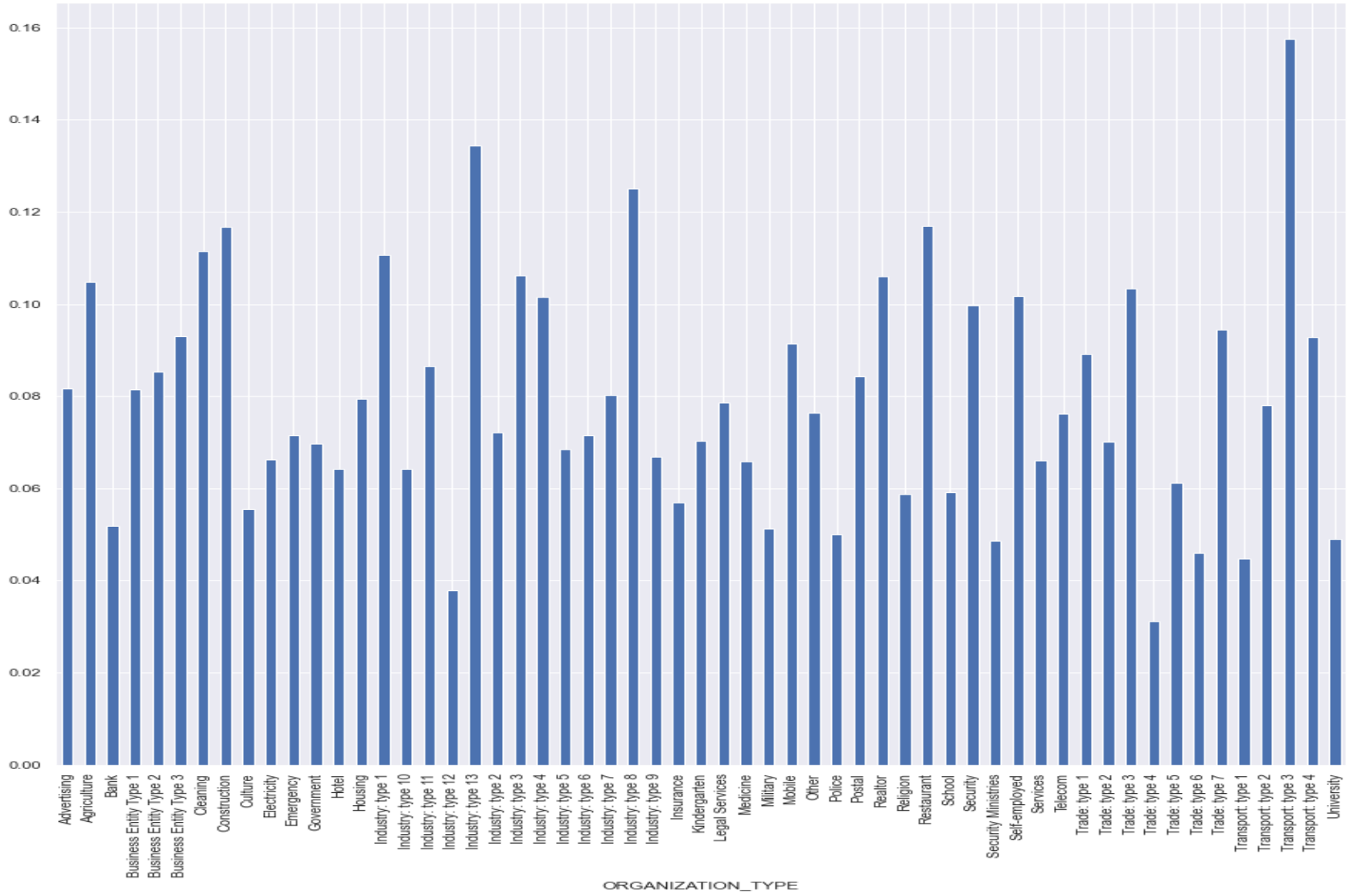
- Low skill laborers had the most Default rate
- Followed by Waiters, Drivers
- Least default rate was from Accountants and IT staff, HR staff

ORGANIZATION TYPE



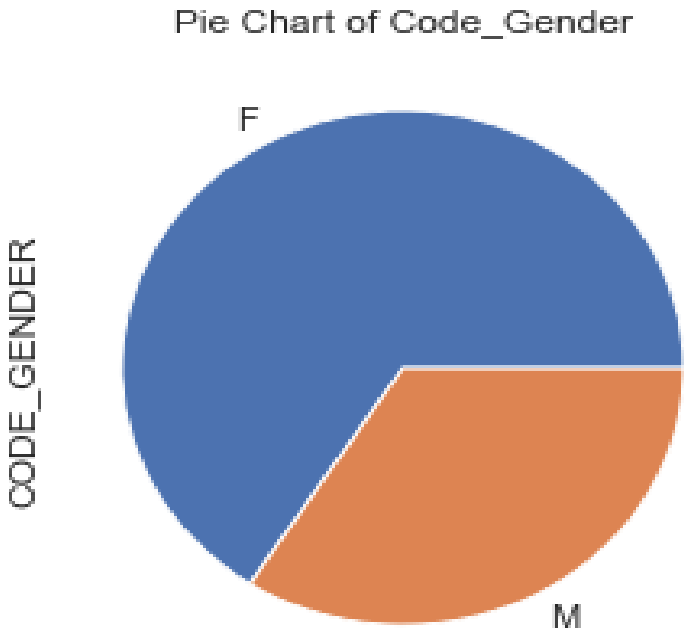
We Can tell that Business, Self employed, Medicine, Government were the major organizations the loan applicants worked for

ORGANIZATION TYPE PAGE 2



- Transport Least being Trade Type 4, Industry type 12
- Type 3, Industry type 13 organizations had the highest default rate

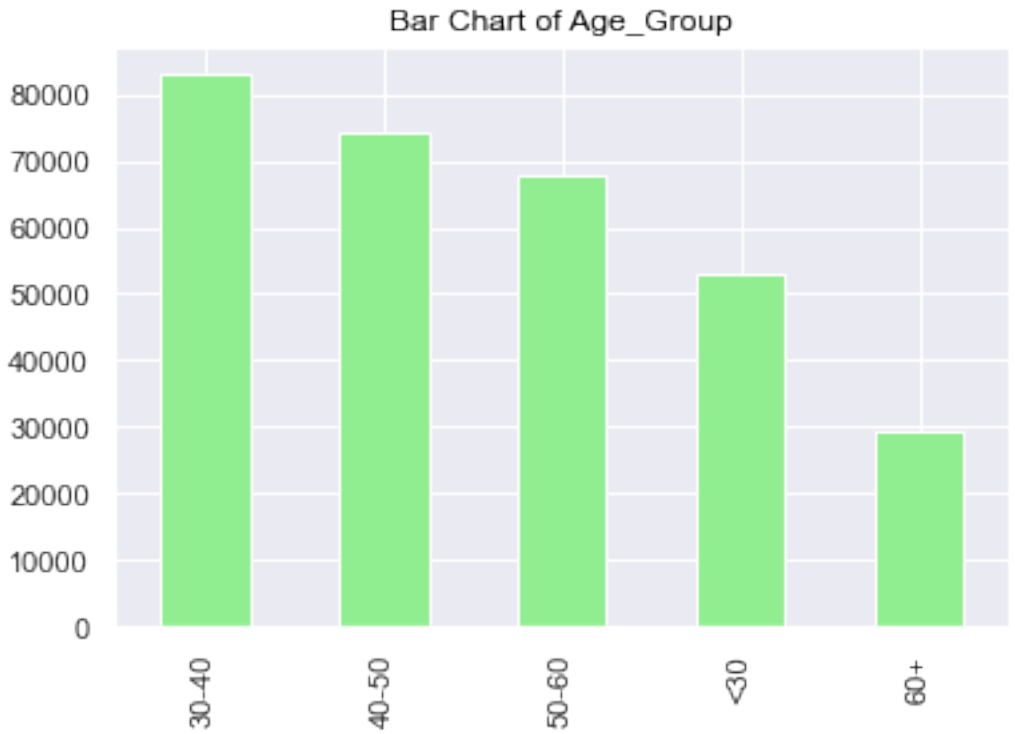
CODE GENDER

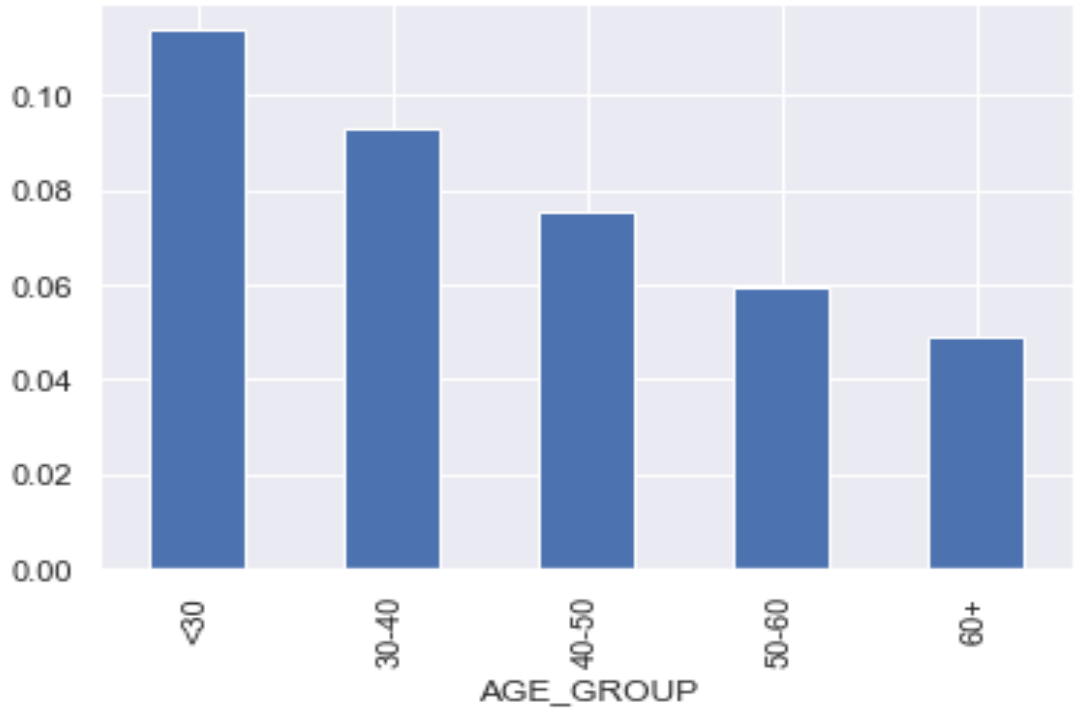


- 65.83% of loan appliers were infact females
- rest 34.16% being males

- Males had a higher percentage of Defaulting in comparison to Females
- Females were very less likely to default compared to males

AGE GROUP

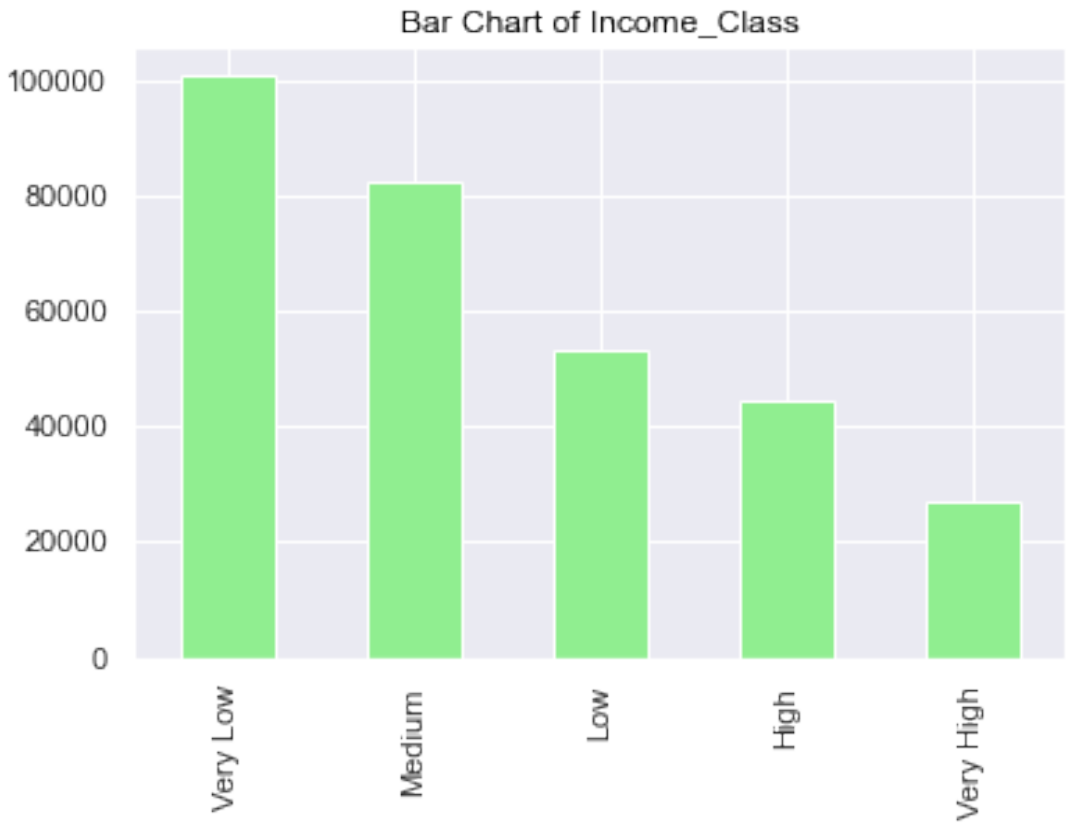




- The middle aged people ranging from 30-40 and 40-50 had the highest amount of loan applications
- Less than 30 and 60+ aged group applicants were lower

- Clients with age less than 30 had the most default percentage, followed by middle aged clients (30-40)
- 60+ aged clients had the least default percentage

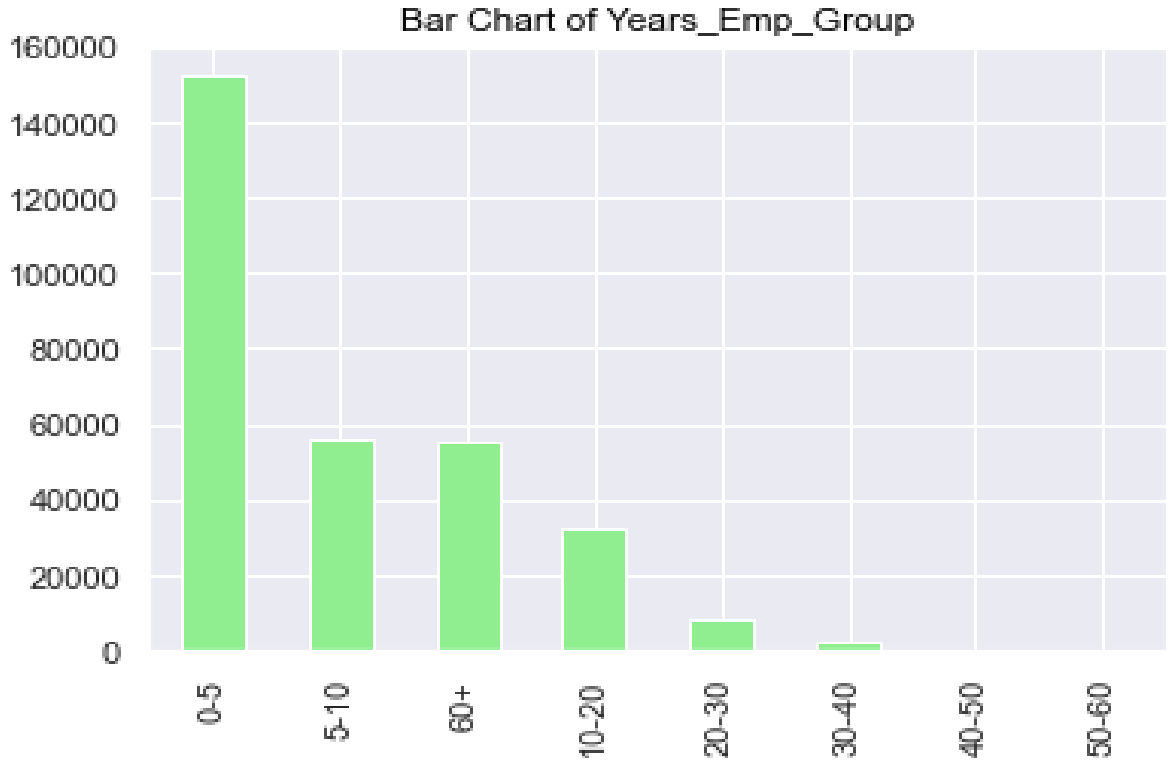
INCOME CLASS

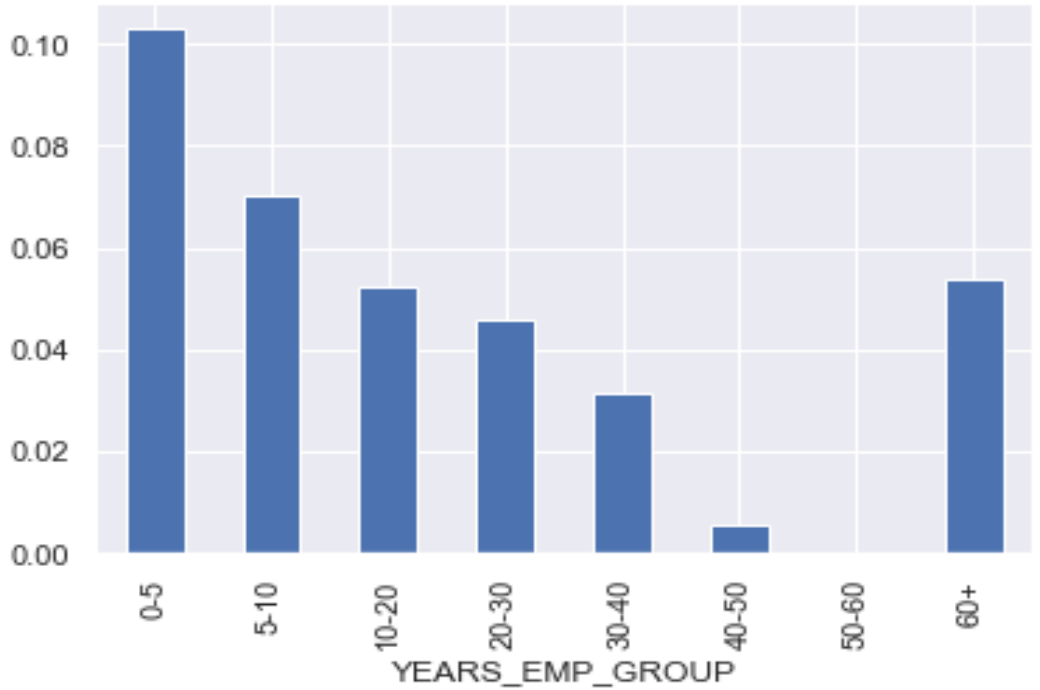


•Majority of the loan appliers were from Very Low wage Income being 32.70% total
•followed by Medium class income being 26.73%
•Somehow Medium class people were ranked 2nd who applied for loans and not Low income class

•Very Low to Medium Class waged clients had the most default percentage
•High to Very High waged Clients were less likely to default

YEARS EMPLOYED

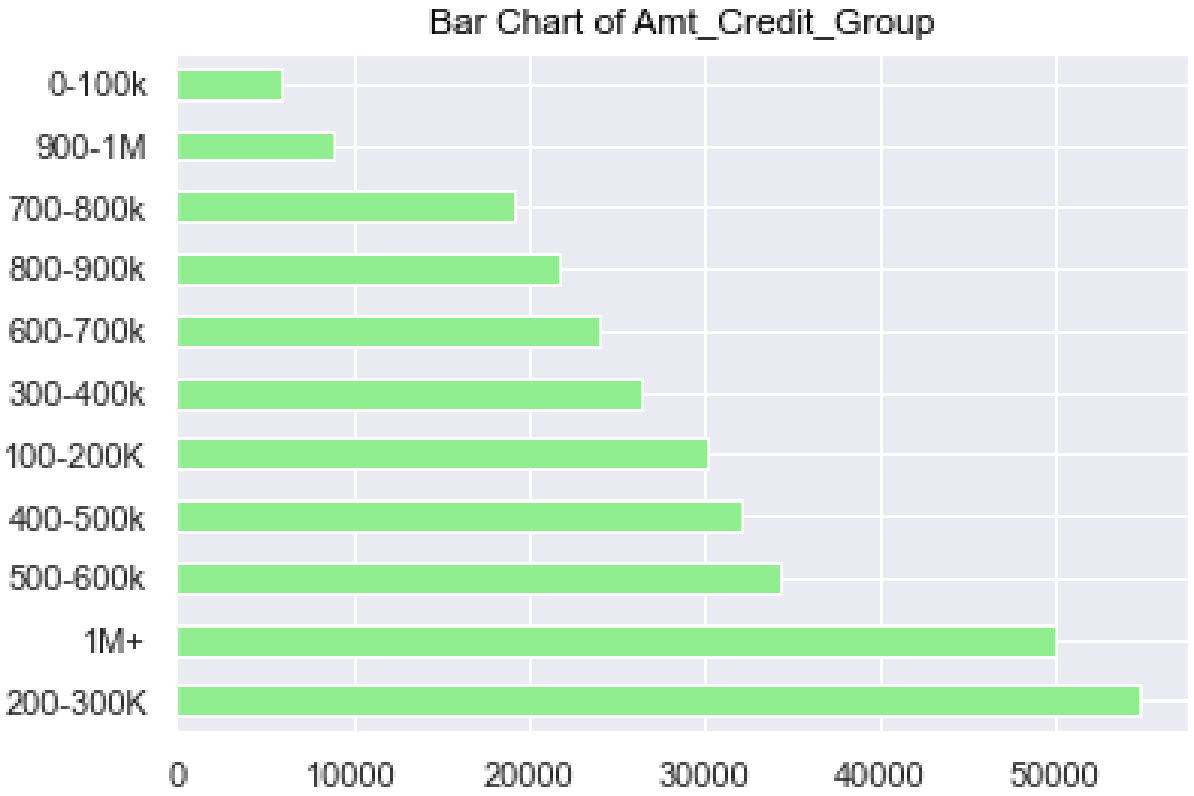


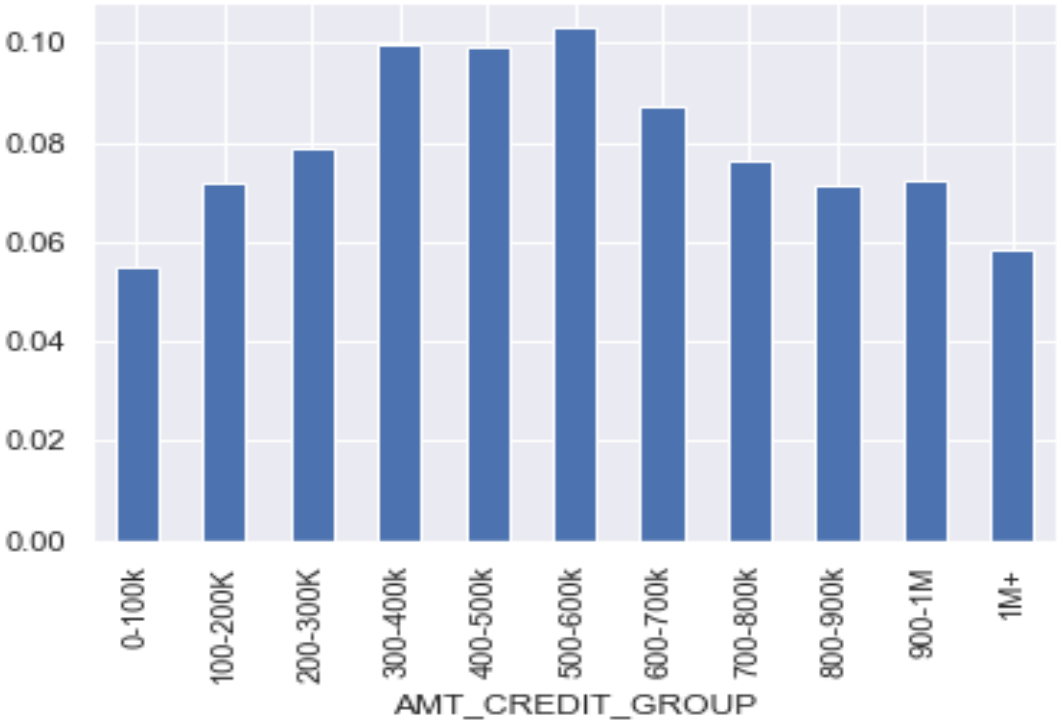


•0-5 years of experience people topped the chart with 49.60%
•with none being from 50-60 years , and some being from 40-50 years

•0-5 Years Employed people were the most likely to default followed by 5-10 years
•30-40 and 40-50 years employed clients had the least default percentage
- 60+ age is an outlier and to be ignored

AMOUNT CREDIT

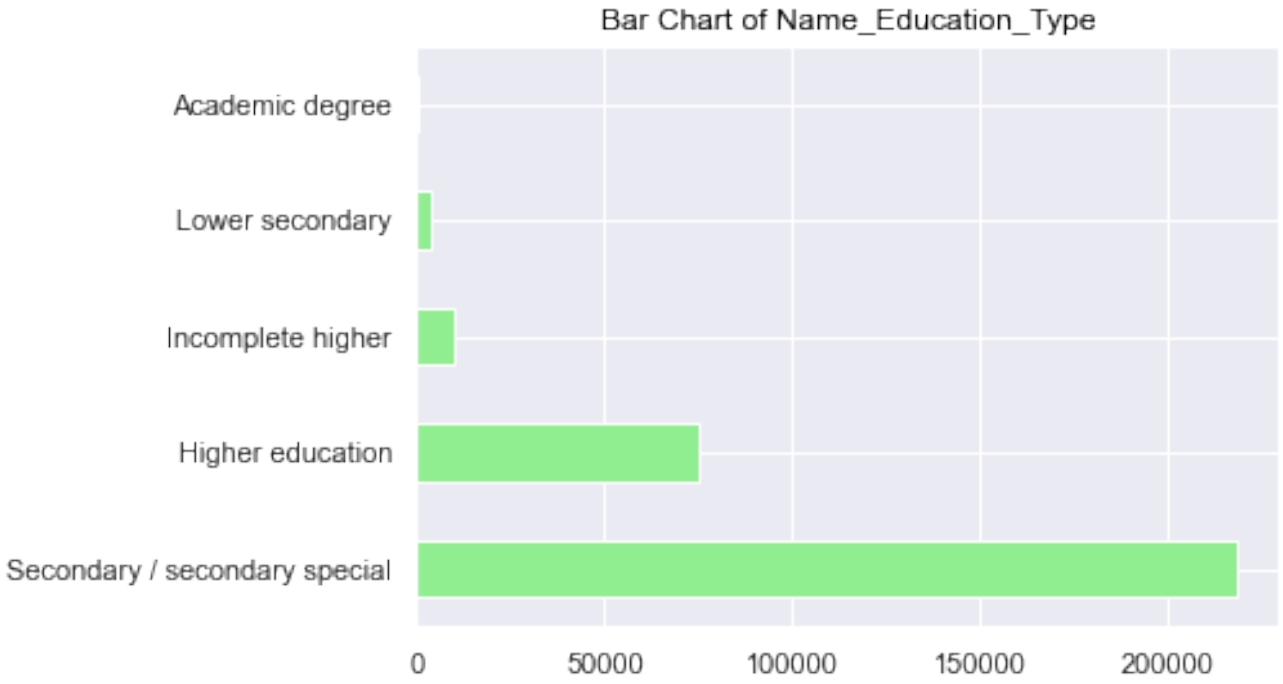


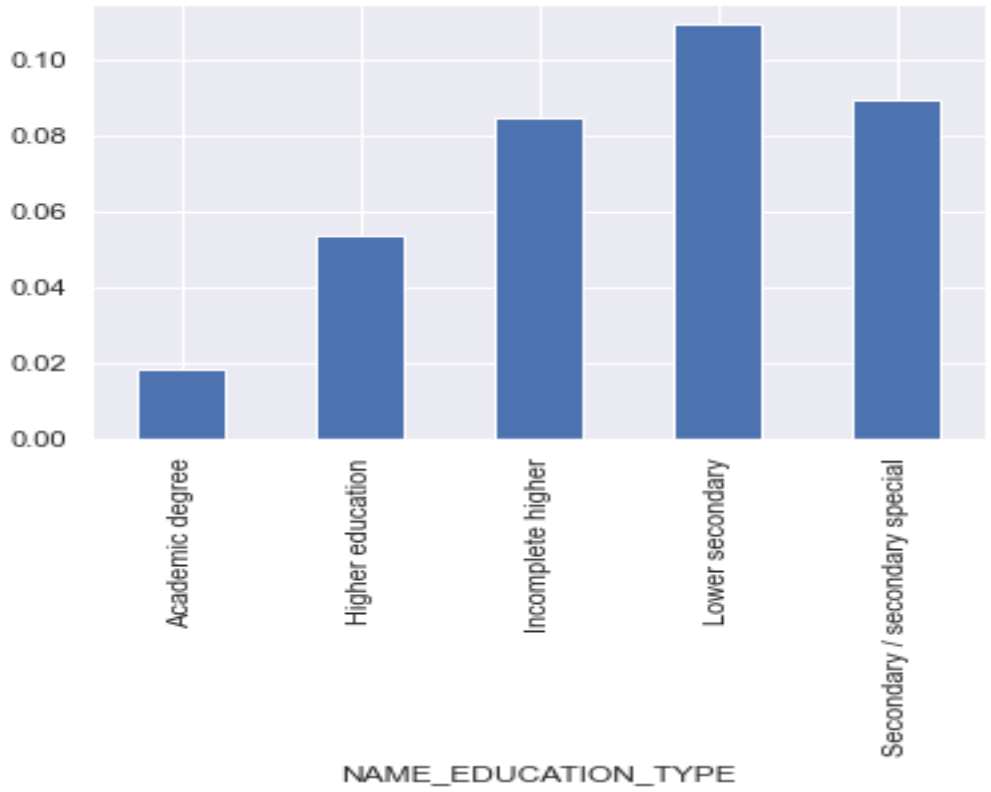


Mainly the credit amount of loan ranged from 200-300K followed by 1M+ and followed by 500-600k

- 300K- 700K credit group had the highest default percentage
- whereas 0-100k credit group had the least default percentage

EDUCATION TYPE

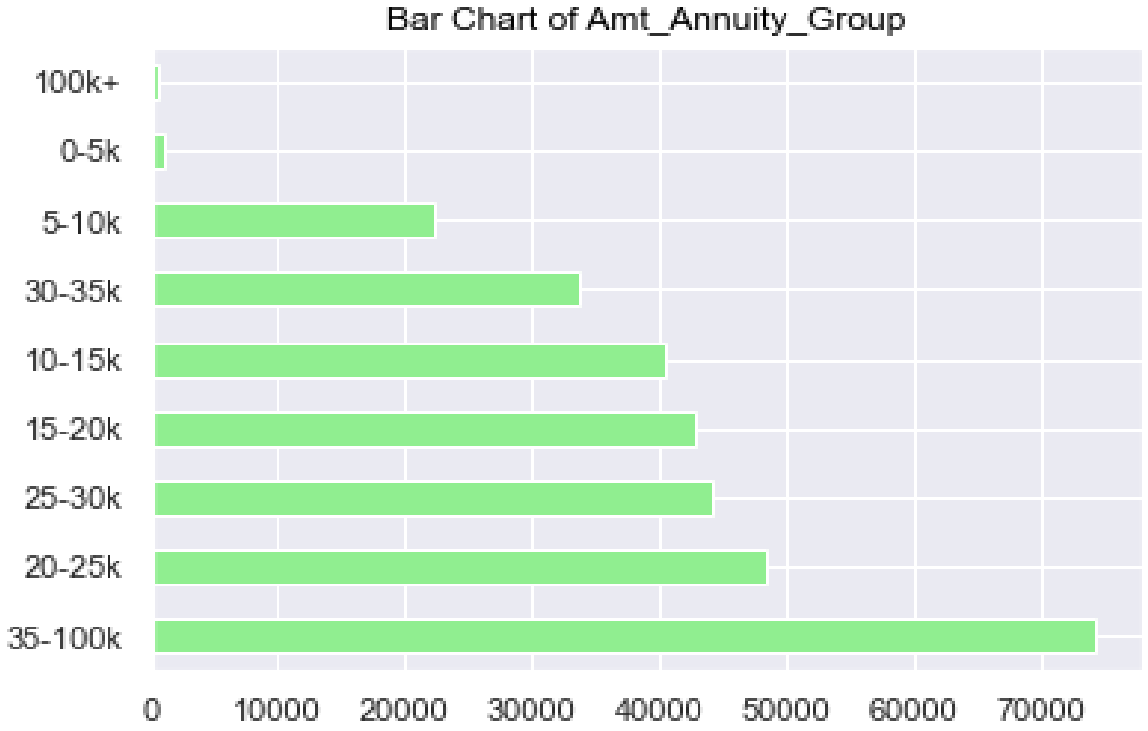


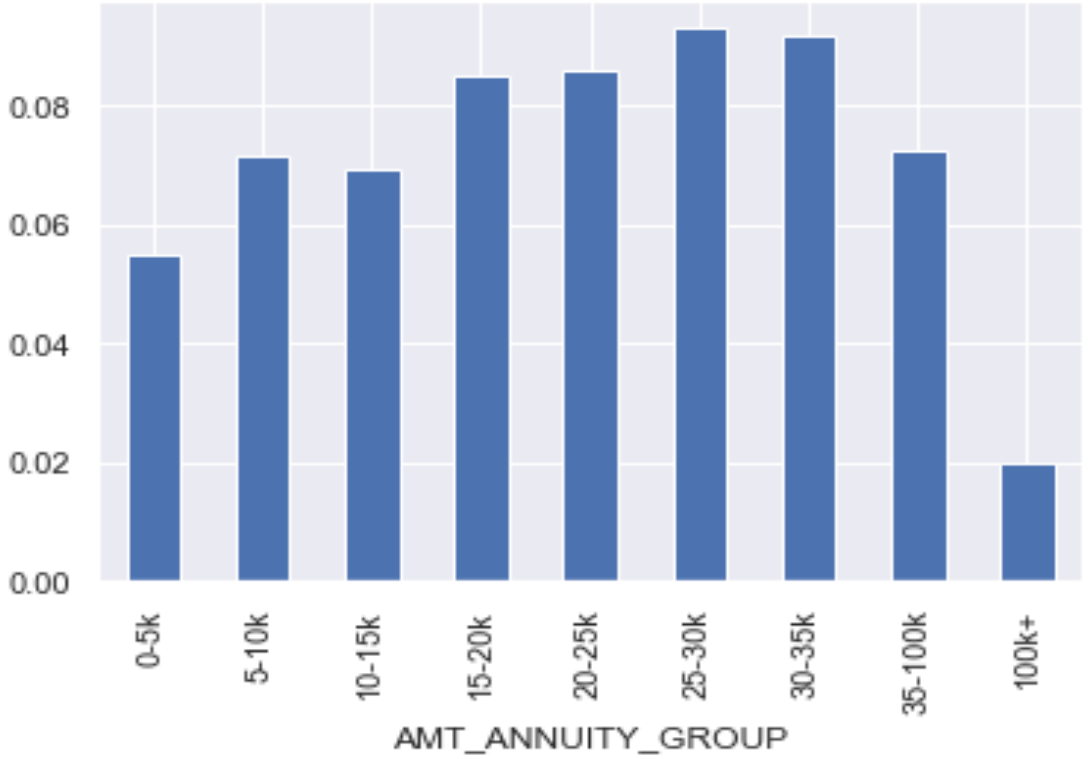


Majority of the applicants had a Secondary education (71.01%) followed by Higher education (24.34%). Academic degree had the lowest percentage

- Lower Secondary, Secondary, Incomplete higher education clients had the bigger hand in defaulting
- Academic Degree holders had the least default ratio

AMOUNT ANNUITY

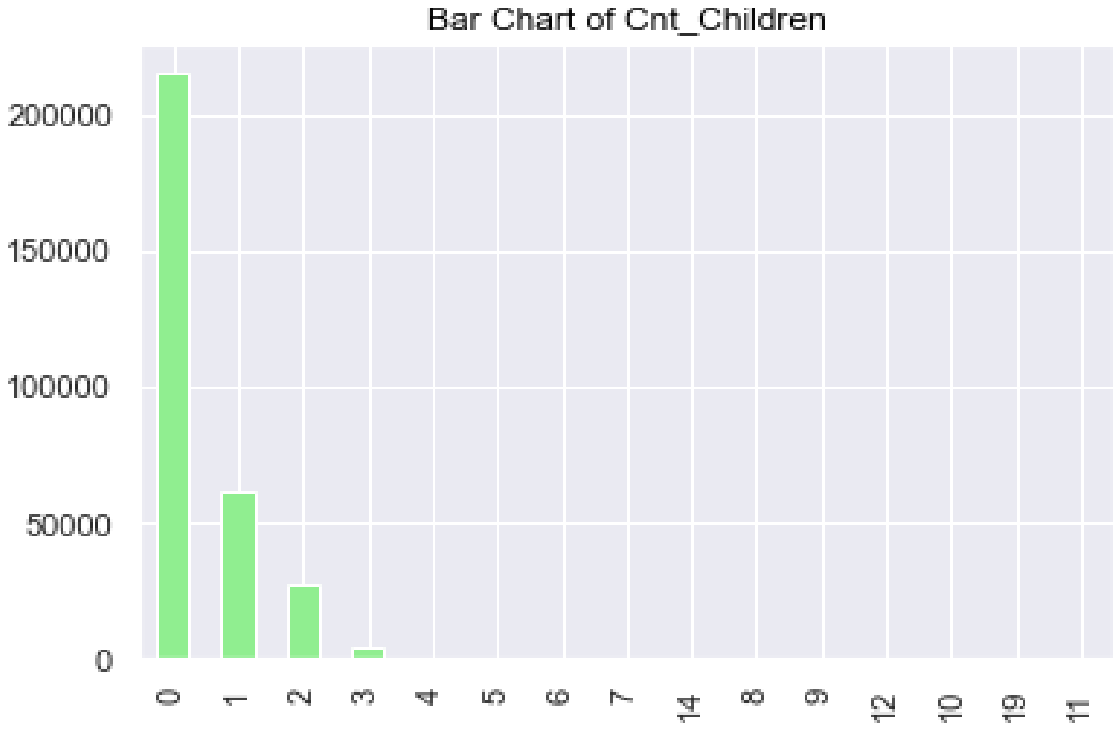


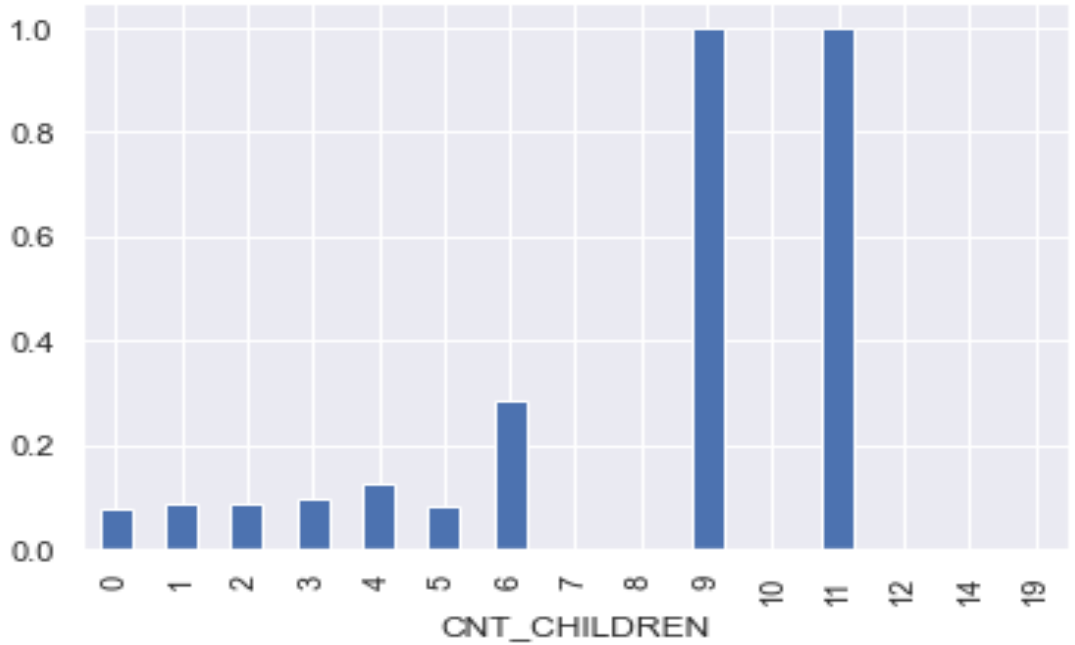


- Applicants took Annuities ranging from 35-100K which topped the graph with 24.17%
- Followed by 20-25k (15.72%), 25-30K (14.34%) etc.

- Annuity ranging from 15K-35K had the highest Default rate
- 0-5K annuity was lower in default
- 100k+ Annuity was the lowest in default rate

CHILDREN COUNT

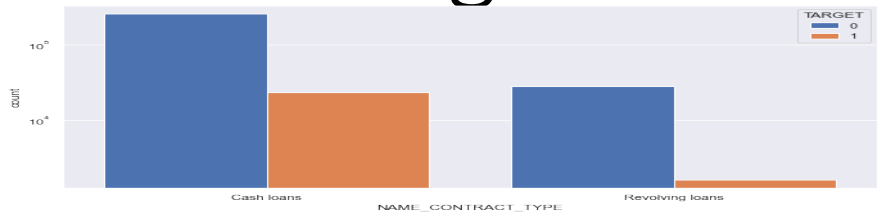


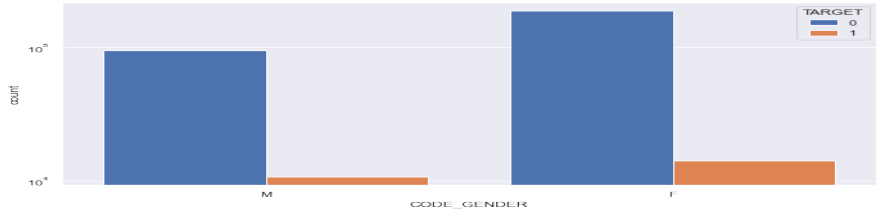


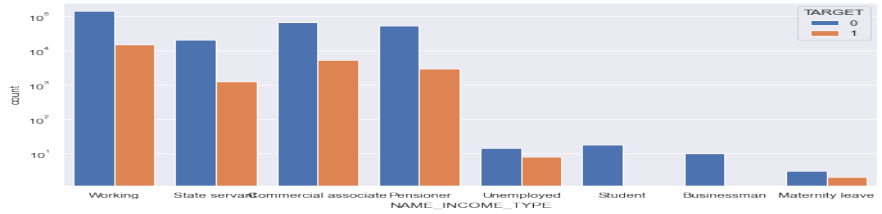
applicants who had childrens count with 0 was the highest

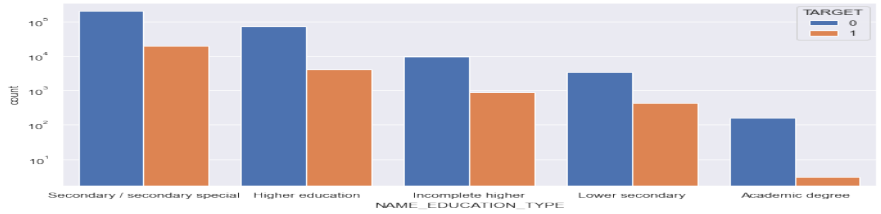
- Clients who had high amount of Children were the most likely to default, ranging from 9-11
- Lower number of Children allowed the Clients to Repay the loan

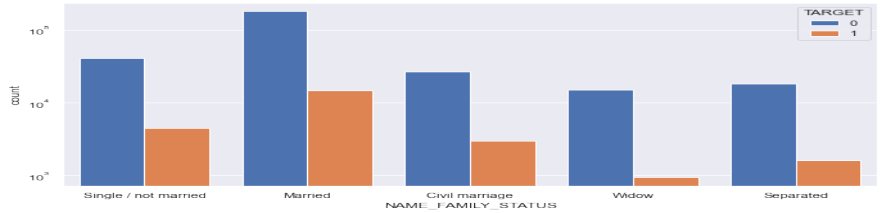
Analysis w.r.t Target variable

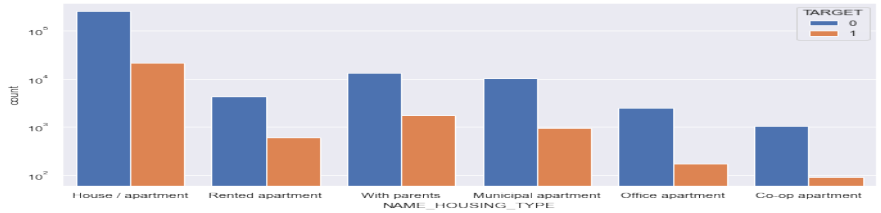


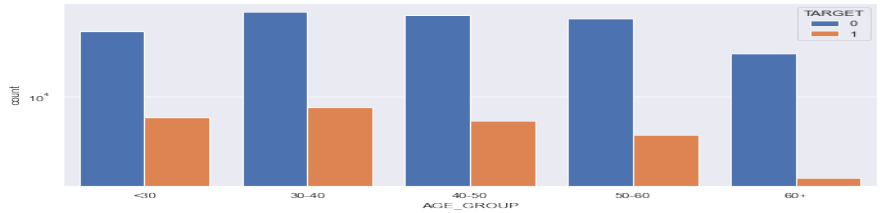


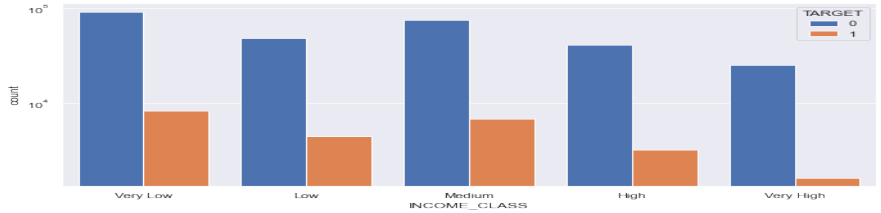


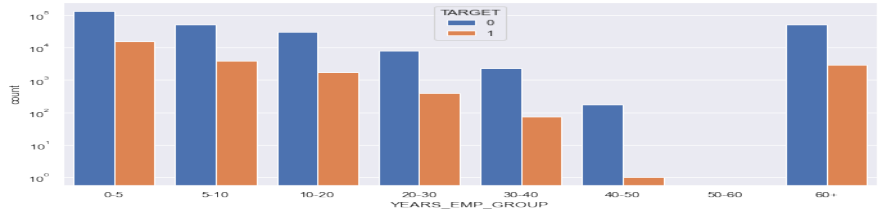


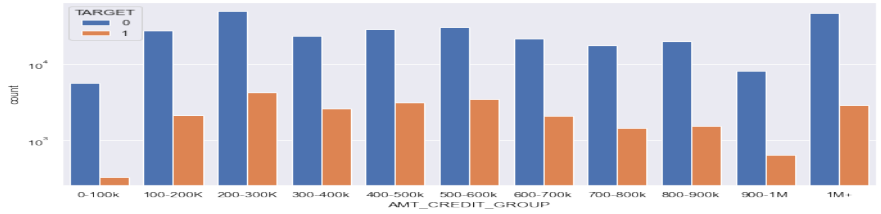


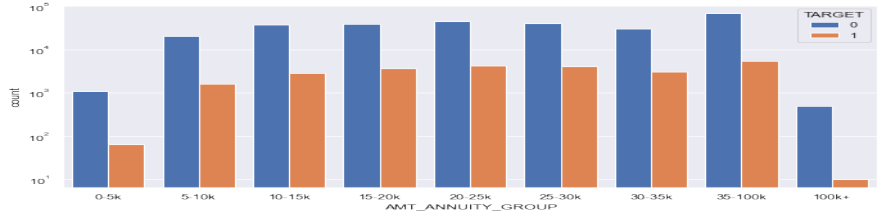


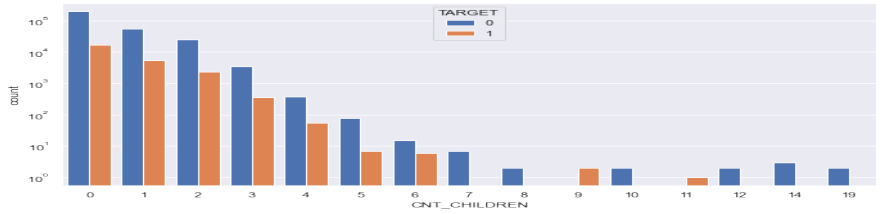




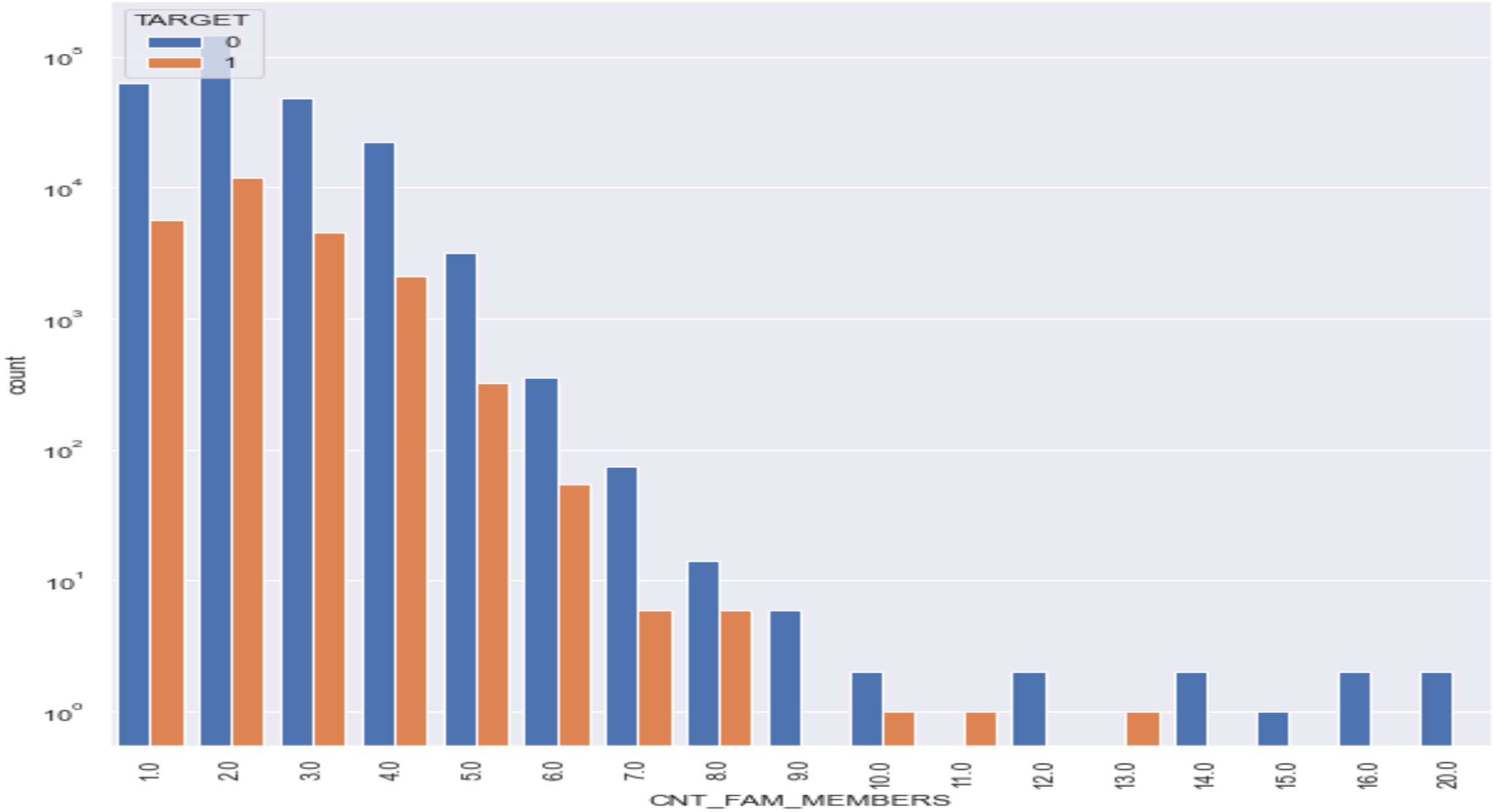






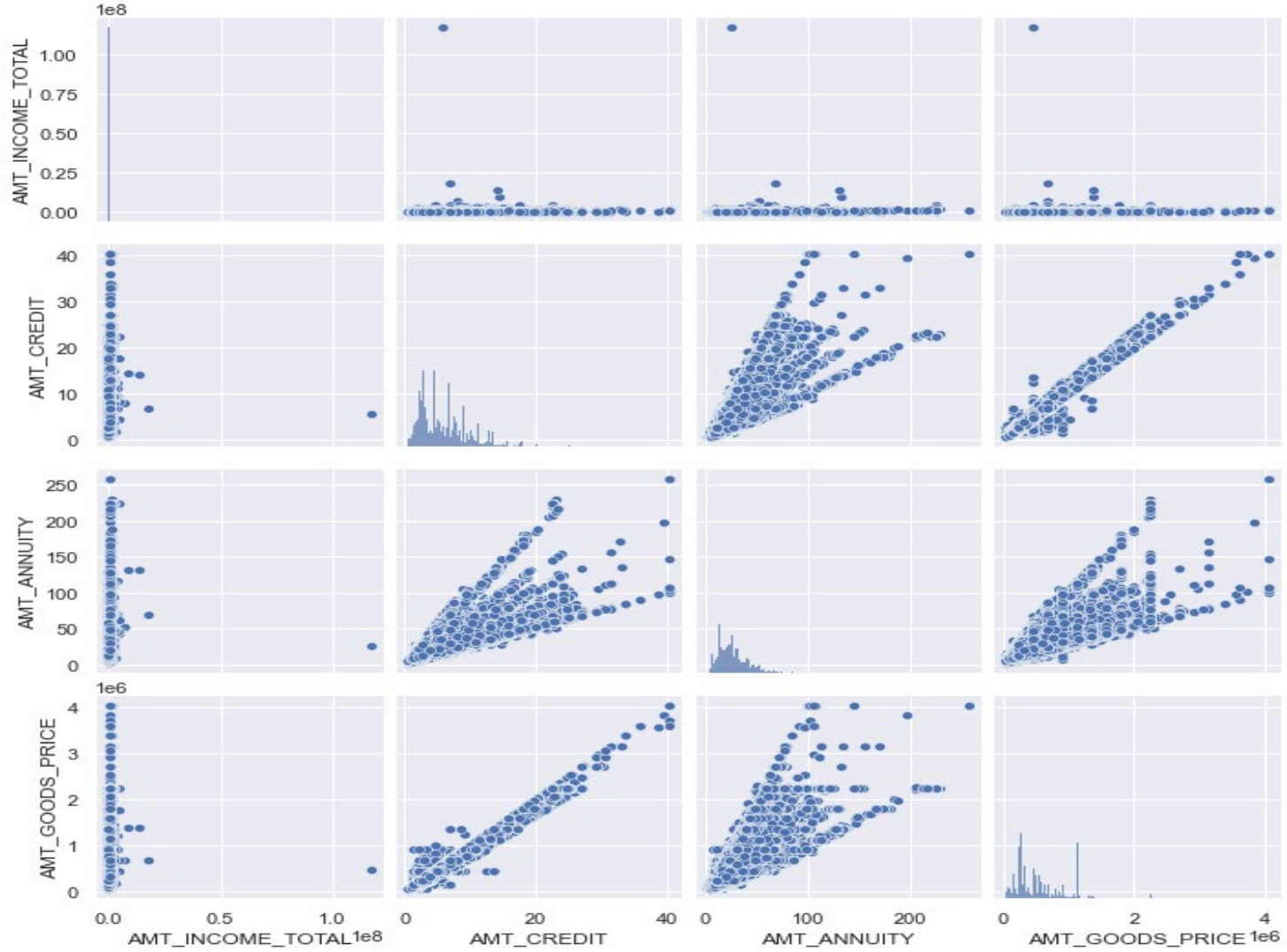


FAMILY COUNT WRT TARGET



MULTIVARIATE ANALYSIS

PAIRPLOT OF NUMERICAL VARIABLES

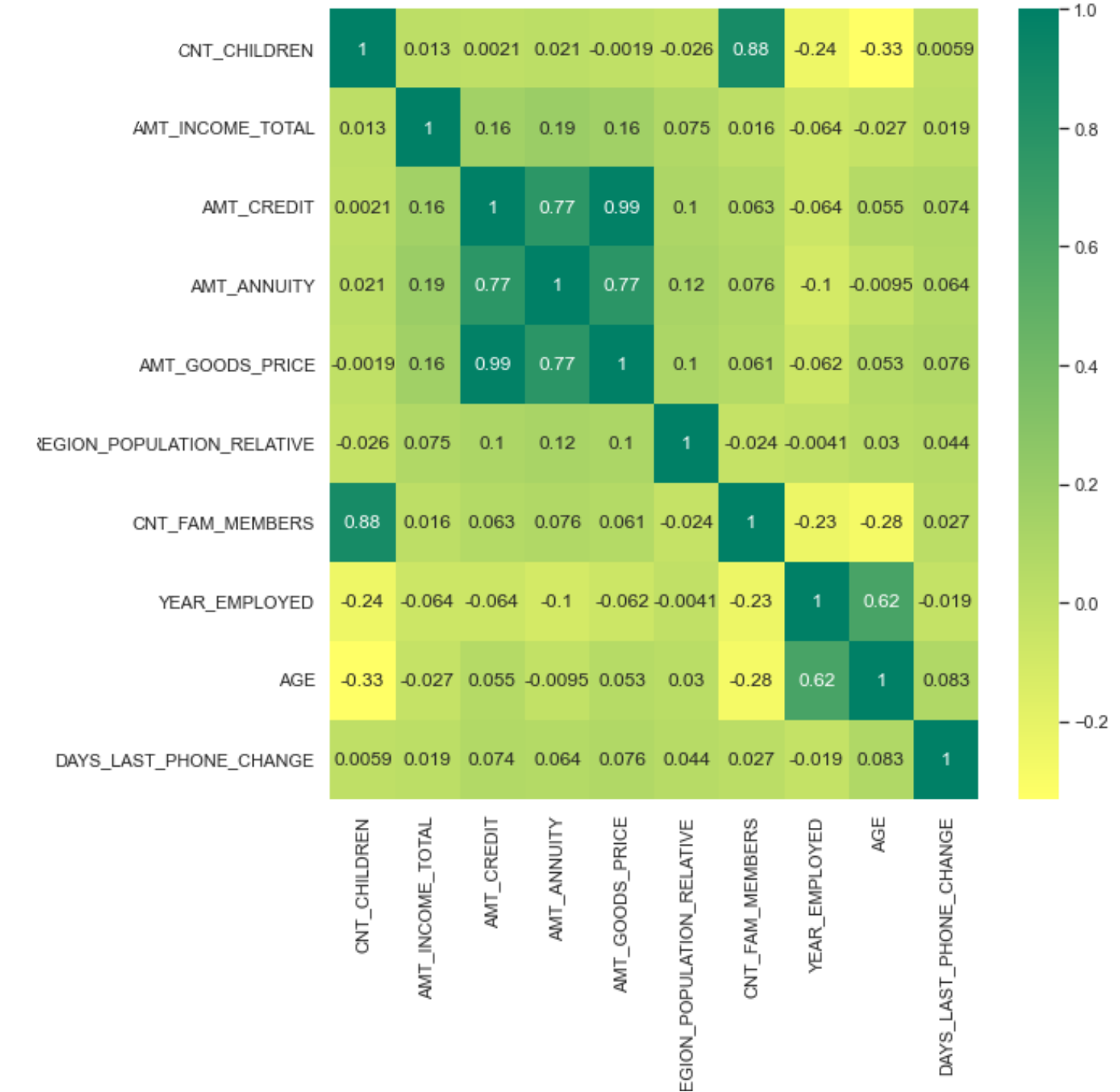


•There is a linear Correlation between:

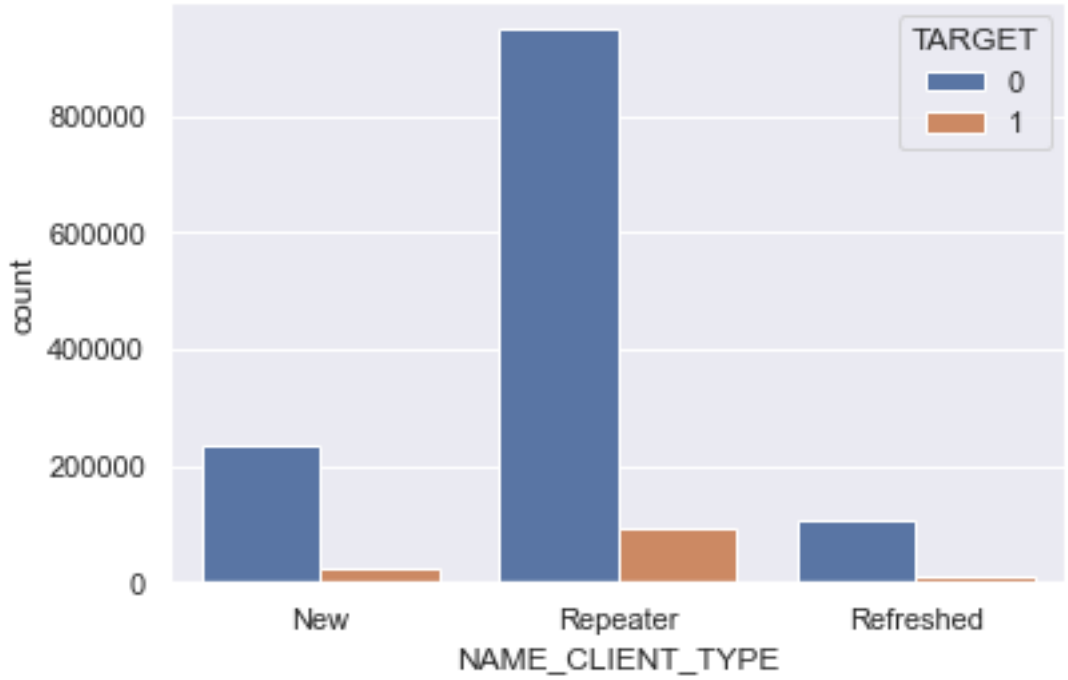
- AMT_GOODS_PRICE and AMT_CREDIT
- AMT_ANNUITY AND AMT_GOODS_PRICE
- AMT_ANNUITY AND AMT_CREDIT

CATEGORICAL-NUMERICAL CORRELATION

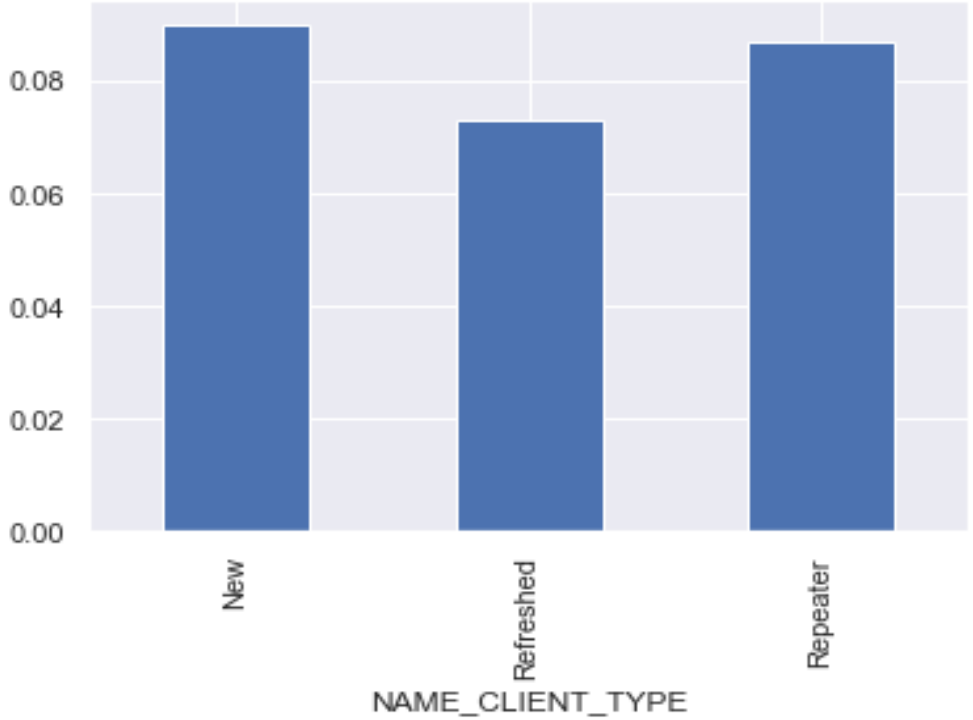
We can see the Correlation between different Categorical and numerical variables from the given heatmap



MERGED DATA ANALYSIS



Repeater Clients who came back we higher in percentage
Compared to new clients



New clients were higher in ratio who had defaulted,
whereas repeater were less defaulting

SUMMARY

Clients who potentially wont default

Clients having 0-2 children were more likely to repay the loan

Academic degree clients have very less defaults

Applicants with 40+ years experience had less defaults

Clients above 50+ age had very less defaults

Clients with income more than 700K+ were less likely to default

Students and businessmen have very less default rate

Organizations like Trade Type 4 and Industry type 12 have very less default

Office apartment, House, and Co-op had the least default percentage

Clients who potentially wont default

Least default rate was from Accountants and IT staff, HR staff

Clients having Family members with 1-5 were able to repay the loan

Clients who potentially could default

Male Clients have higher default rate

Applicants who have 9 to 11 children are more likely to default.

Low skill laborers had the most Default rate

Family members with higher (9-12)number of members were likely to Default

Annuity ranging from 15K-35K had the highest Default rate

300K- 700K credit group had the highest default percentage

0-5 Years Employed people were the most likely to default

Very Low to Medium Class waged clients had the most default percentage

Clients who potentially could default

Clients with age less than 30 had the most default percentage

Clients living in Rented Apartments and with parents had the Highest default percentage

Civil marriage and Unmarried Clients had the most default ratio

Lower Secondary, Secondary, Incomplete higher education clients were high defaulters

Unemployed clients had high default rate

Cash Loans had the highest percentage of defaulters

Transport Type 3, Industry type 13 organizations had the highest default rate