CREDIT EDA CASE STUDY

By: Tushar Prasad

This is an EDA of past and present Application data of the Loan Applicants

The Business Objective :

To understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

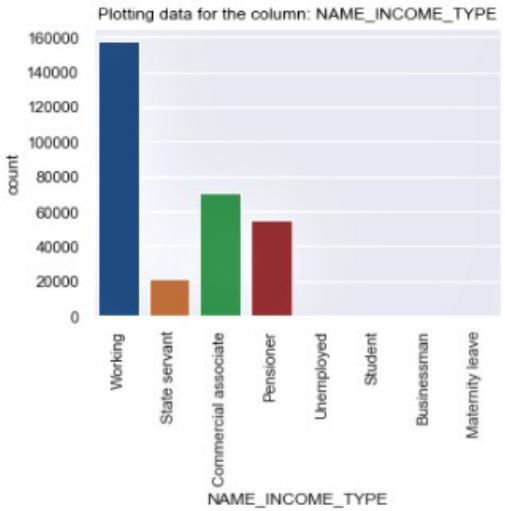
Prominent Actions

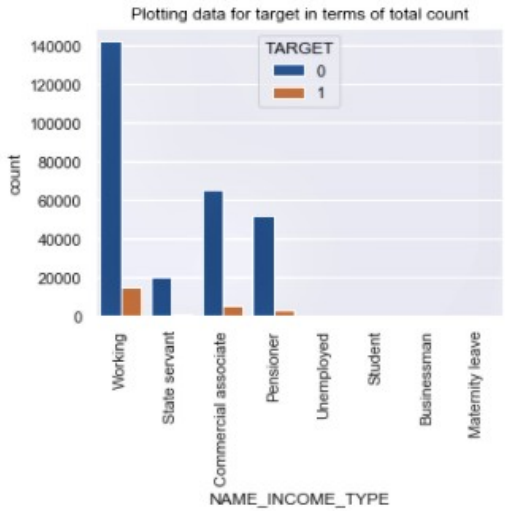
We have used EDA to analyse the patterns in the
background profile data of the loan applicants.
For this we have divided the data into two target variables i.e.
Target 1 (clients with payment difficulty) and
Target 0 (all other clients).

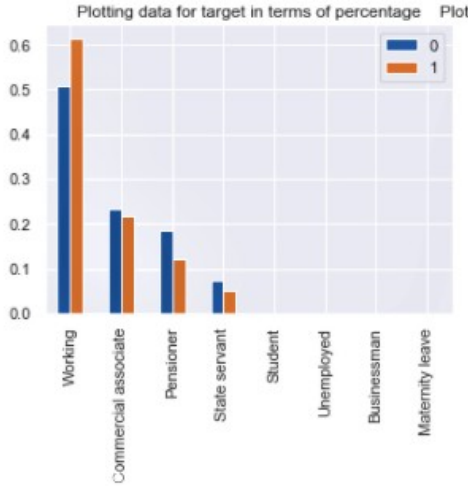
1

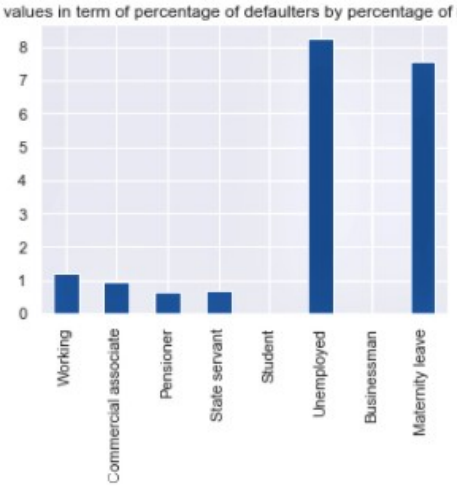
Univariate Analysis of Categorical Variables

Income Type of Loan Applicants





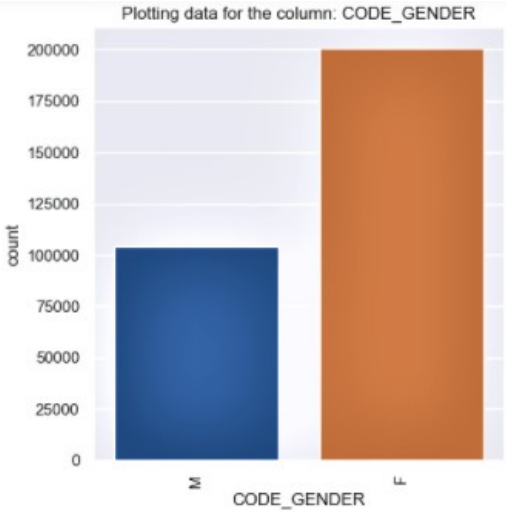


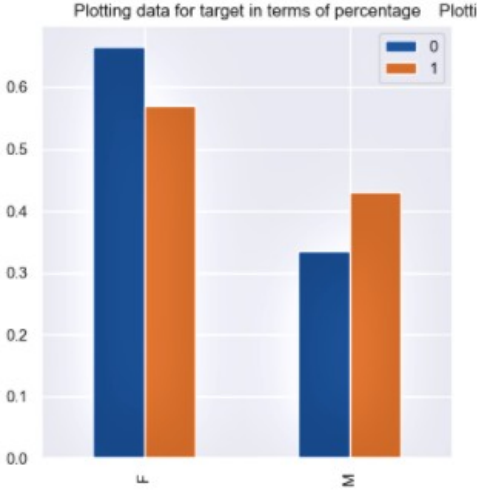


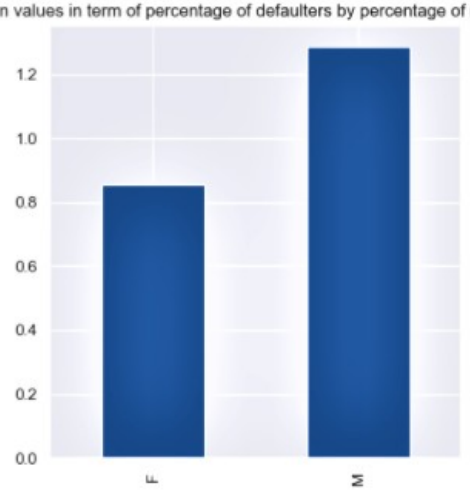
Observations from Income Type of Loan Applicants

- Working class people take the maximum loans among all income types.
- Student, Unemployed, Businessmen and people on maternity leave hardly take any loans.
- Among the people who take loans, the % default by % non- default ratio of working people is the highest and greater

Gender factor in Loan Applicants Data



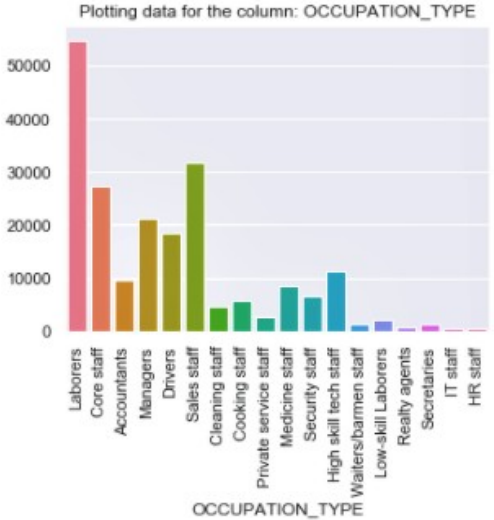


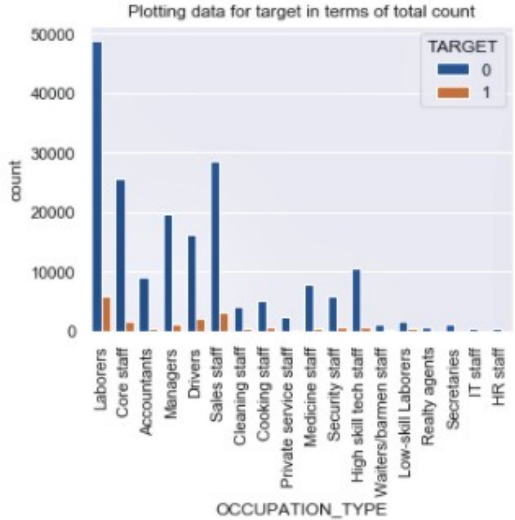


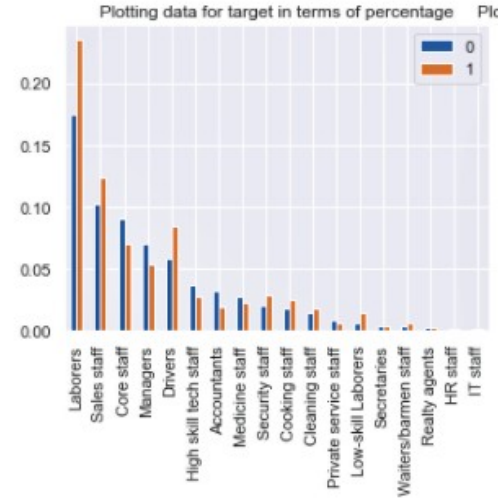
Observations from Gender based graphs

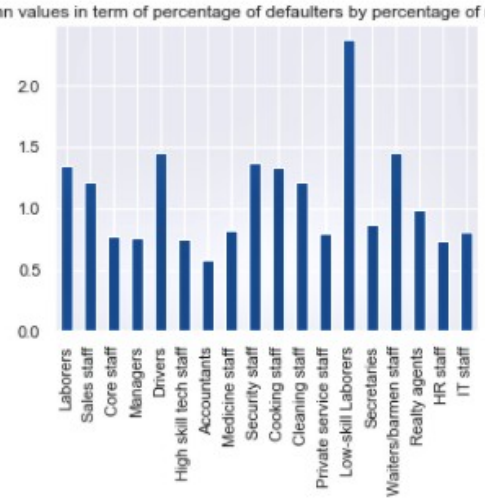
- Females (66%) seek more loans when compared to males (34%).
- This includes both Defaulters and non-defaulters.
- Males have a higher ratio of % defaulters to % non-defaulters than females.

Occupation of Loan Applicants





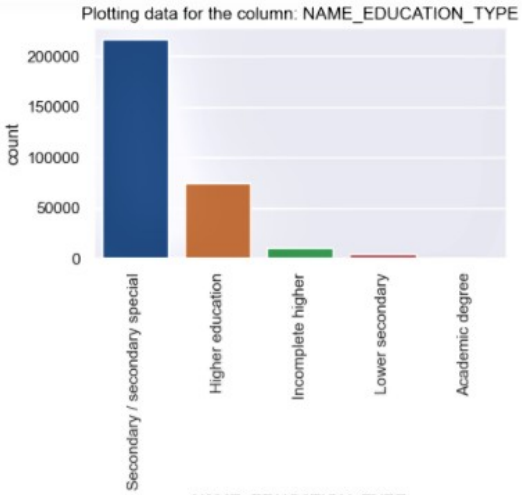


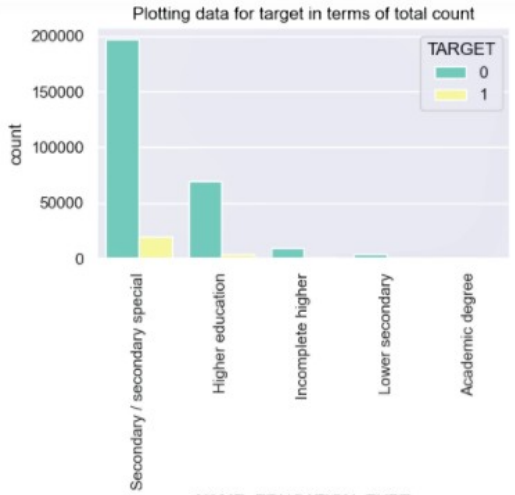


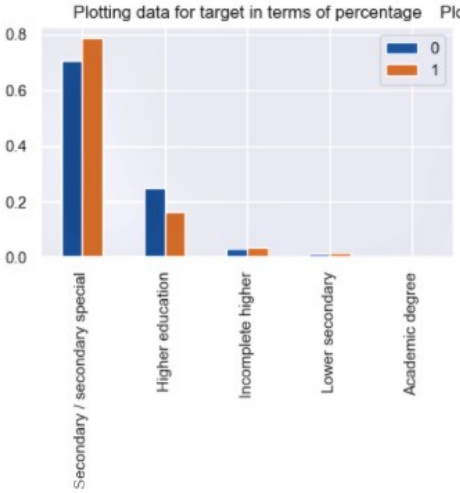
Observations from Occupational pattern of Loan Applicants

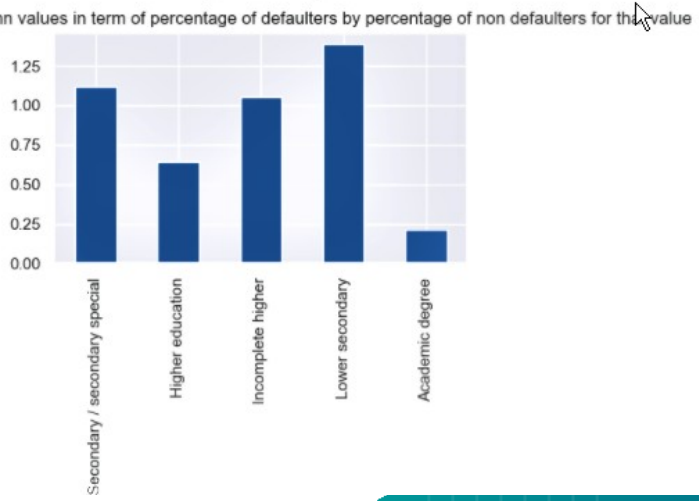
- Labourers, Sales staff and core staff comprise of over 50% of loans.
- Occupations like Managers, Accountants and high level staff members(Core, High Skill Tech, Medicine) have a lower default % than occupations like Labourers, sales staff, drivers and other lower staff members.

Education Type of Loan Applicants









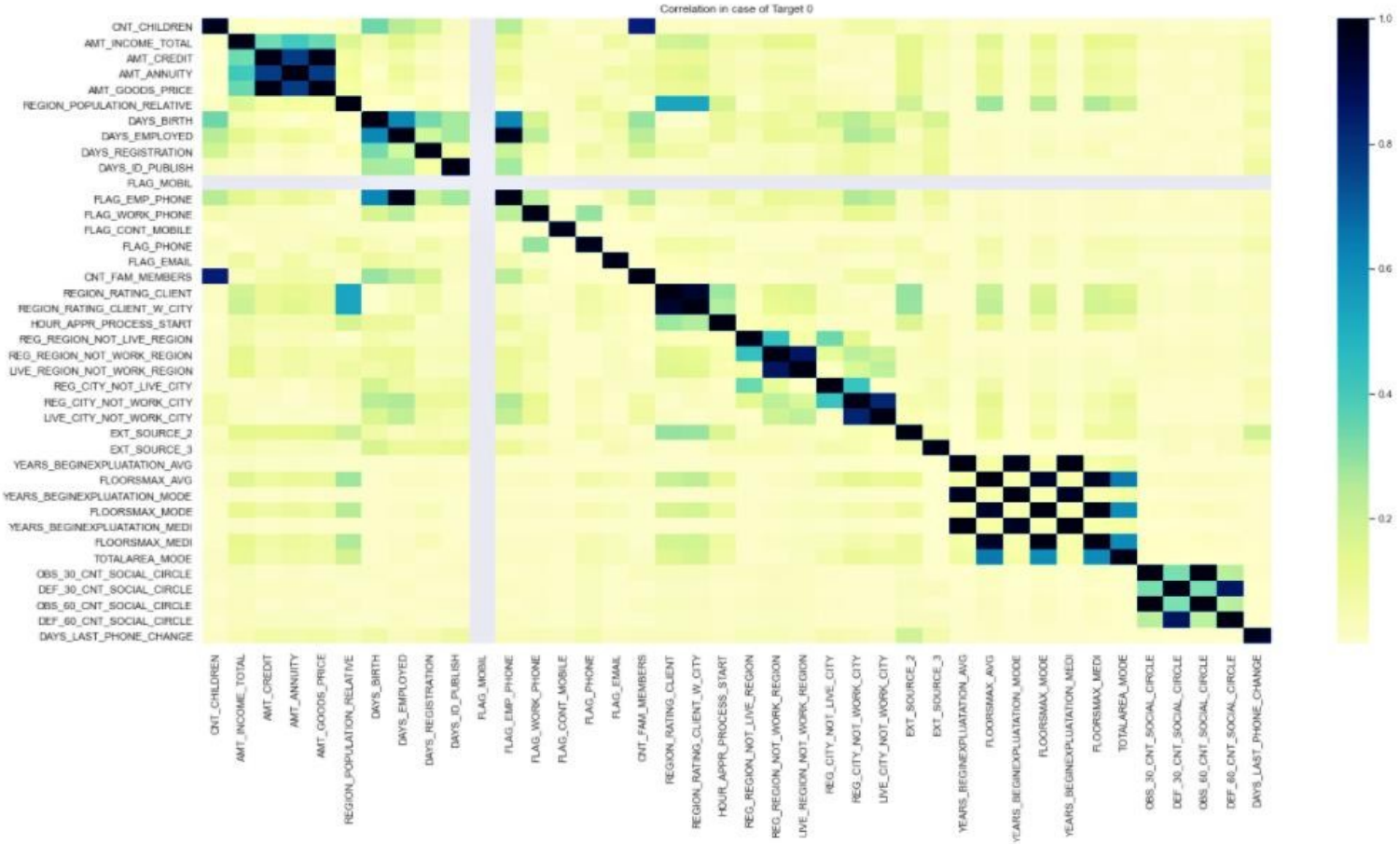
Observations from Education Pattern of Loan Applicants

- People having education to the level of Secondary / secondary special take the highest amount of loans.
- People having higher education is a lower value of % defaulter by % non-defaulter ratio than other levels of education.

2

Univariate Analysis of Numerical Variables

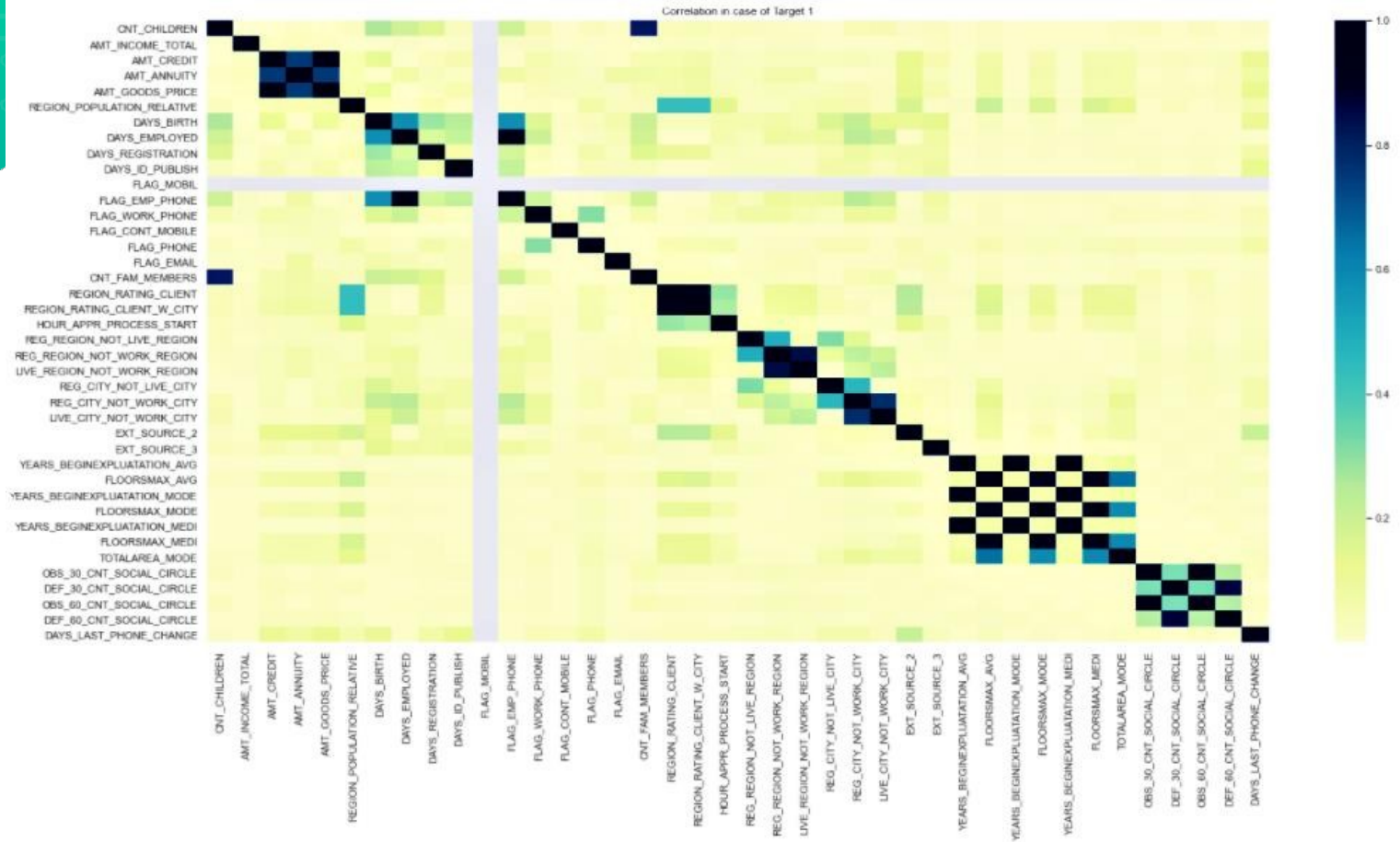
Correlation of Target 0



Observation from Correlation Heatmap of Target 0

- Credit amount of loan and annuity is higher for clients with higher income.
- Credit amount of loan is inversely proportion to children count. For a client with more children credit amount is lesser.
- Work phones are mostly found with clients of higher age.
- Region population relative is inversely proportional to clients age which means that mostly young people live densely populated regions.
- Family member counts is directly proportional to children count.
- Days employed is inversely proportional to clients age. There is higher probablity of recently changing the job in younger clients.

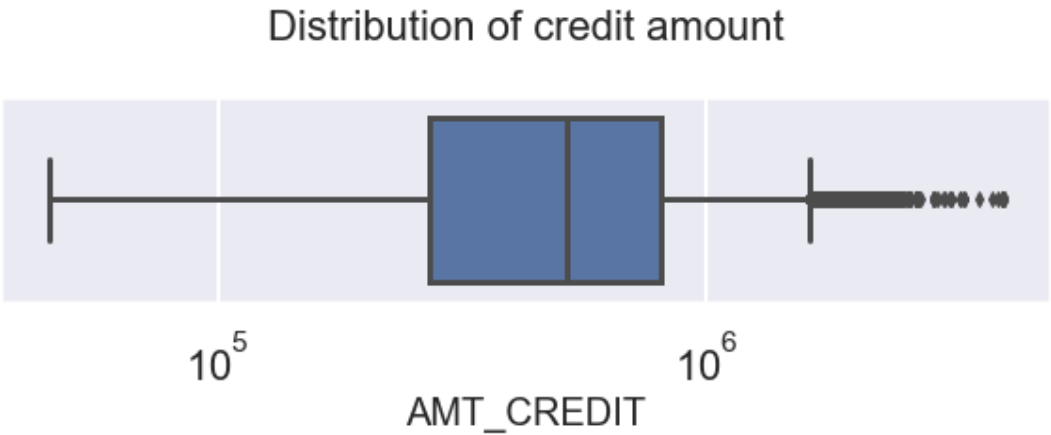
Correlation of Target 1



Observation from Correlation Heatmap of Target 1

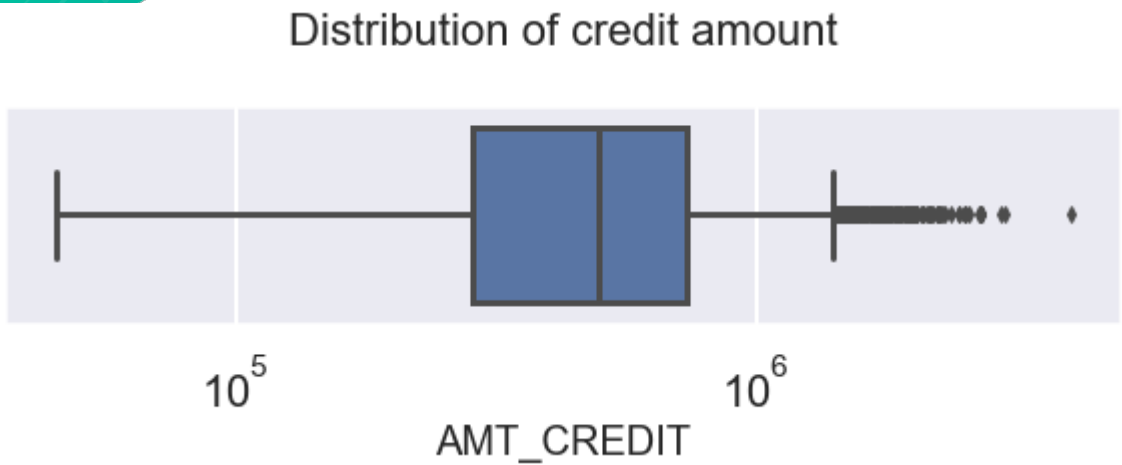
- Total annual income is inversely proportional to region population relative. Highly populated region have low income clients.
- Annuity is inversely proportional to clients age. Young clients pay higher annuity.
- Credit amount of the loan is inversely proportional to the number of children the client has.
- Credit amount of the loan is inversely proportional to the client's age.
- DAYS_EMPLOYED is inversely proportional to region population relative. For a region with high population density, there is higher propbablity of clients with a record of recently changing their job.

Distribution of Credit Amount for Target 0



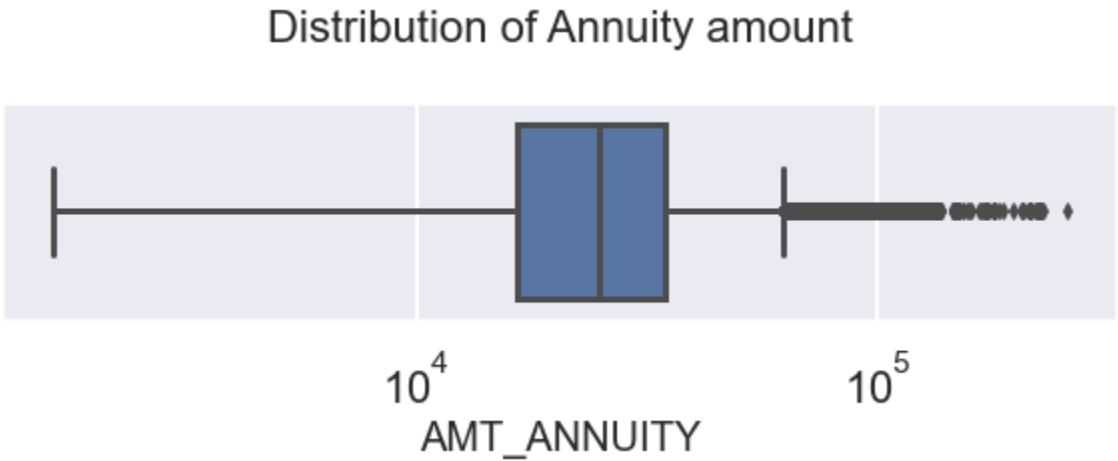
- The credit amount of loan is usually between 1 Lakh to 10 Lakhs.
- There are some outliers beyond 10 Lakhs but not spread for too larger amounts.
- The first and second quartiles have most of the loans.

Distribution of Credit Amount for Target 1



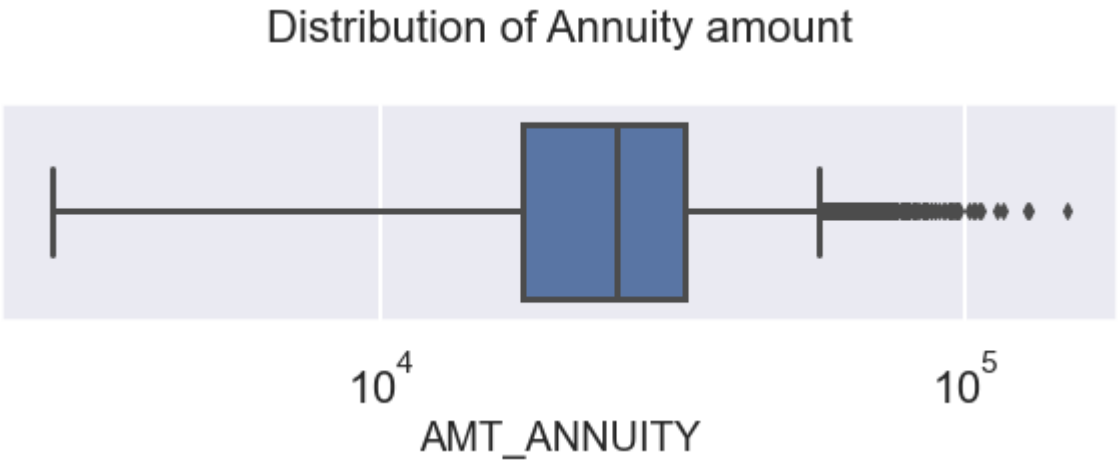
- The credit amount of loan is mostly between 100000 to 1000000.
- There are some outliers beyond 1000000 but not spread for too larger amounts.
- The second quartile have most of the loans followed by first

Distribution of Annuity Amount for Target 0



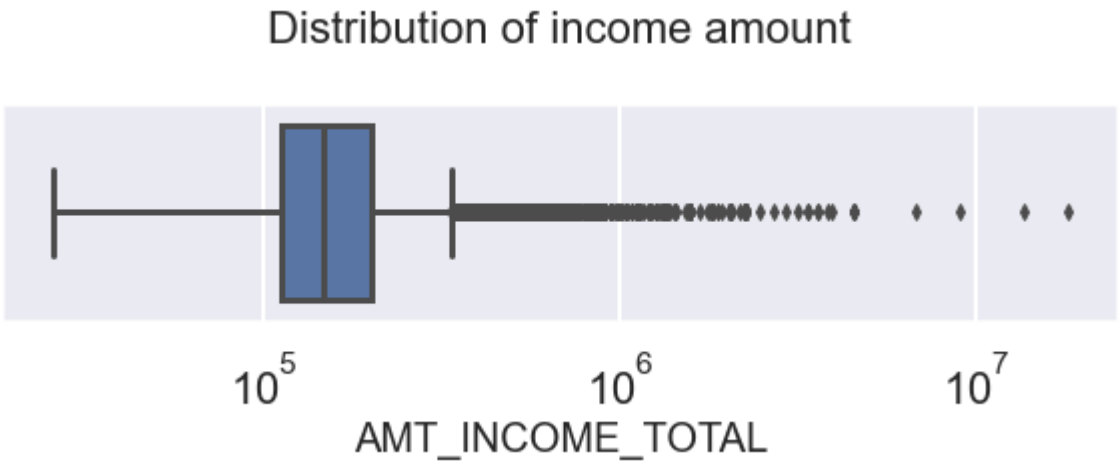
- Most of the annuities value fall in the first quartiles and followed by second quartile.
- Outliers are there and more sporadic than in case of credit amount.

Distribution of Annuity Amount for Target 1



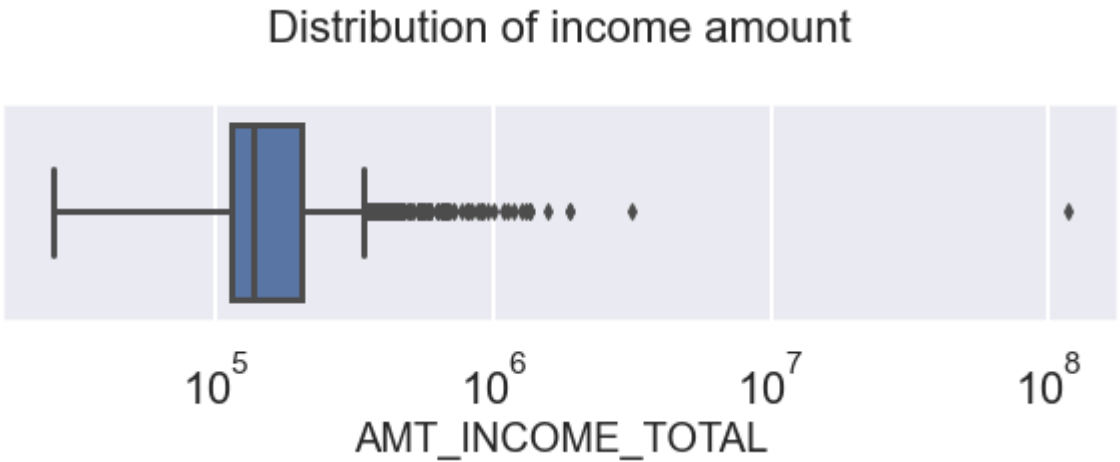
- Most of the annuities value fall in the first quartiles and followed by second quartile.
- Outliers are there and more spread than in case of credit amount.

Distribution of Income for Target 0



- Most of the clients fall in second quartile of the income range
- The income range is spread over a broad range with many outliers.

Distribution of Income for Target 1



- Most of the clients fall in first quartile of the income range
- The income range is spread over a vast range with many outliers.

3

Outlier Analysis

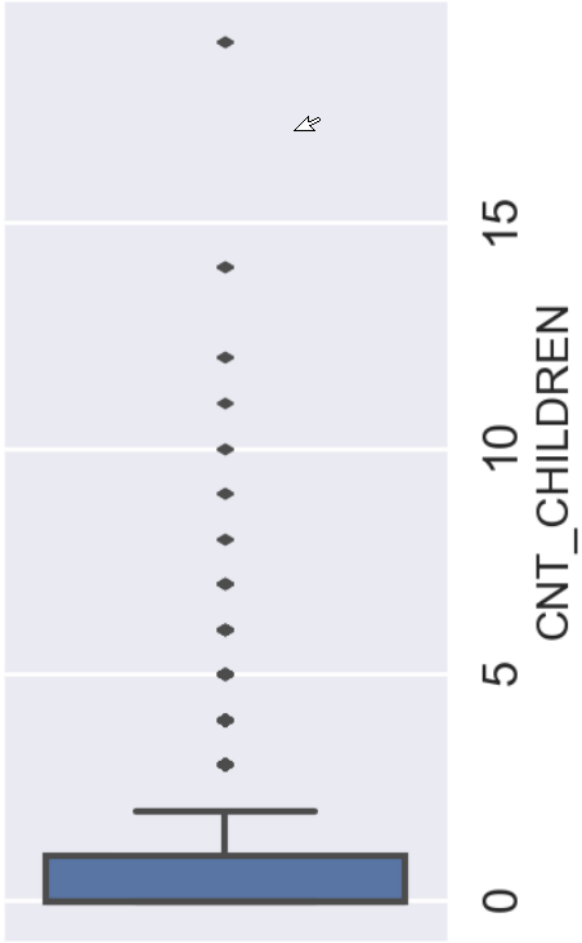
Outlier Analysis

Outliers in Children Count

There are some data points where the value

is above 10 and even above 15.

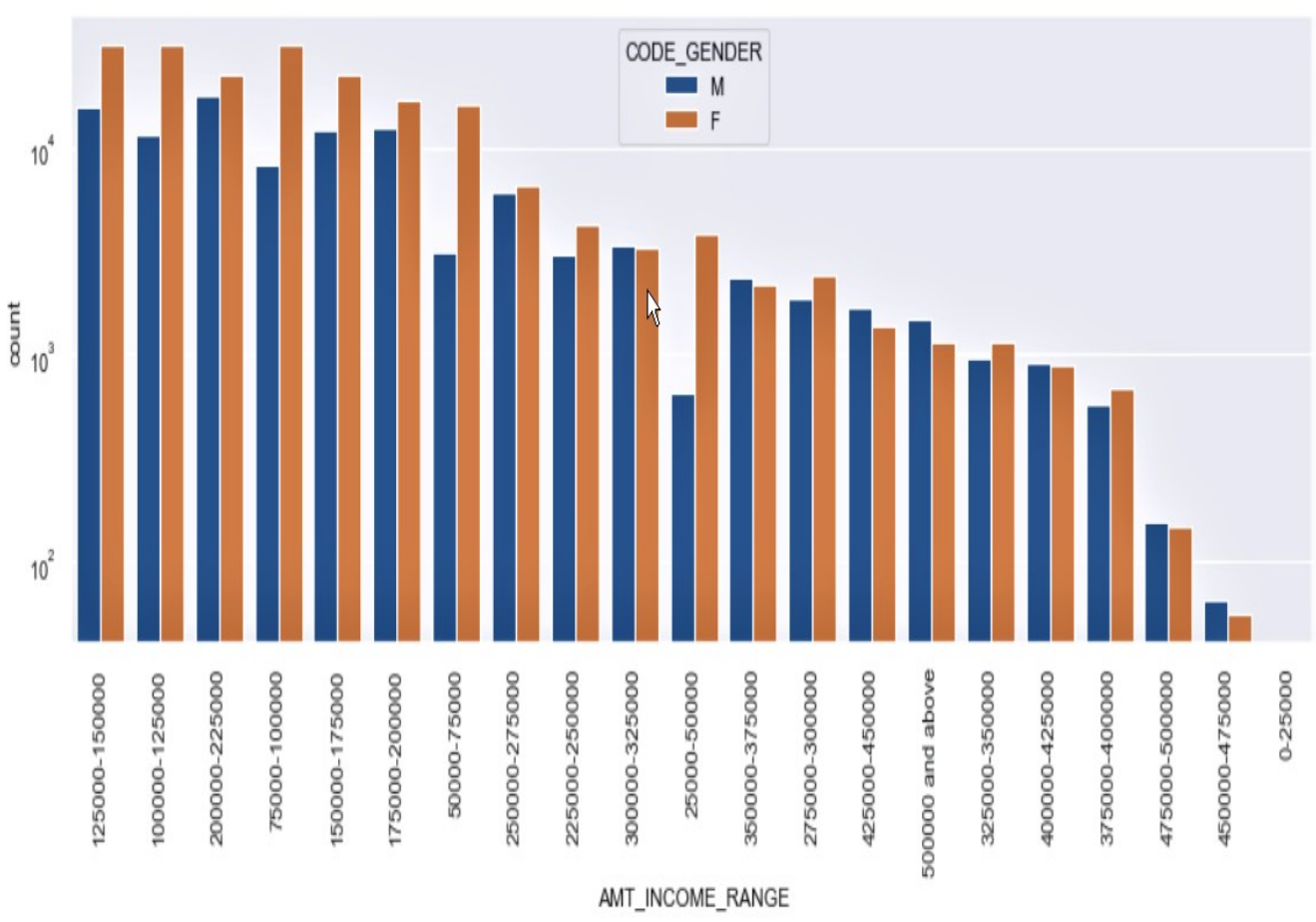
- Logically this means the number of children is above 10 which is an outlier because nowadays the chance of that happening is very less.
- We need to check other columns also before making such a decision since having more children might lead to a higher distribution of income towards them hence leading to



4

Bivariate Analysis

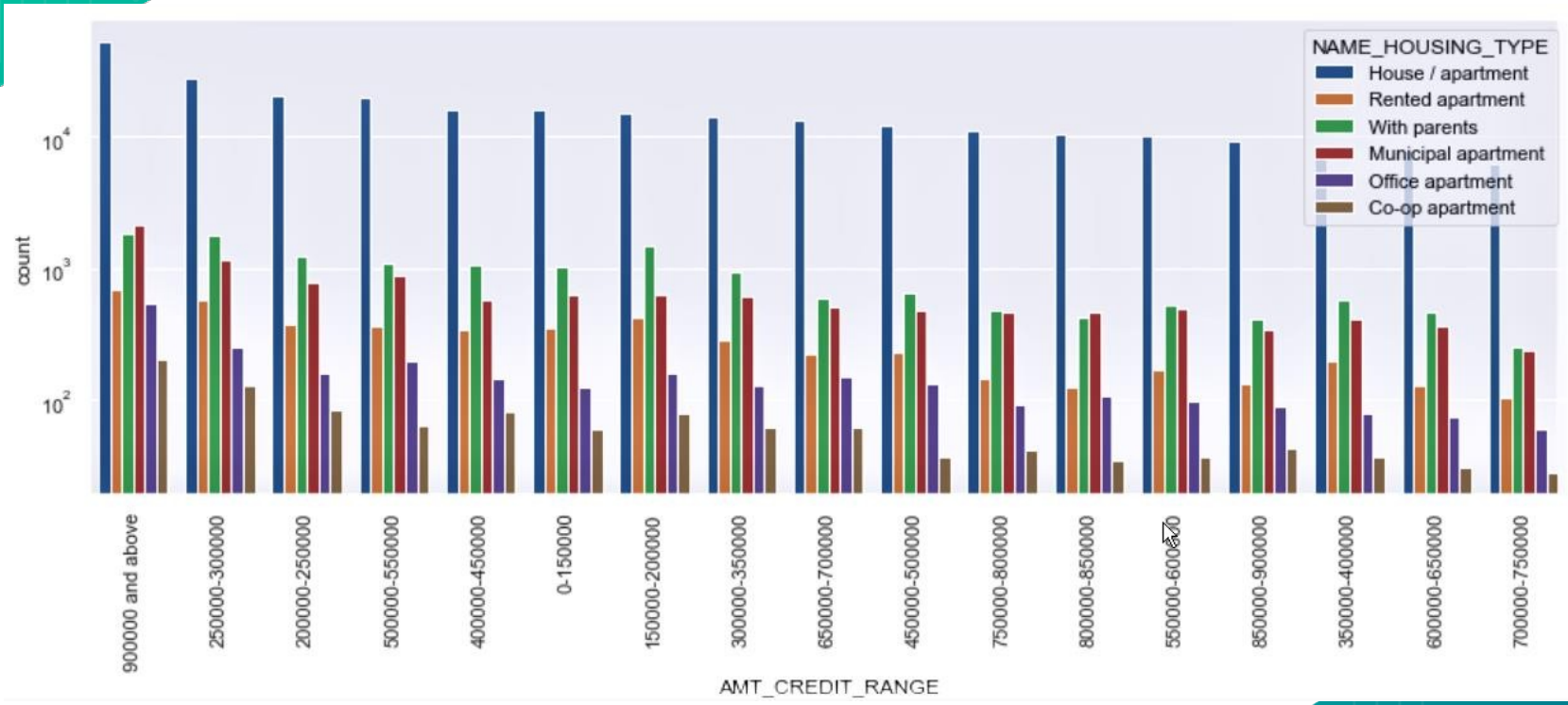
Gender wise Loan Applicants in various income ranges



Observations from Gender wise Loan Applicants in various income ranges

- Maximum amount of people who take loans lie in the mid income ranges like 75000-200000.
- As the income goes up, the amount of loan people taking decreases.
- People who earn very less don't take any loans.
- For the lower income range, there are more females loan takers than males.
- For the higher income range, their counts might almost be equal.

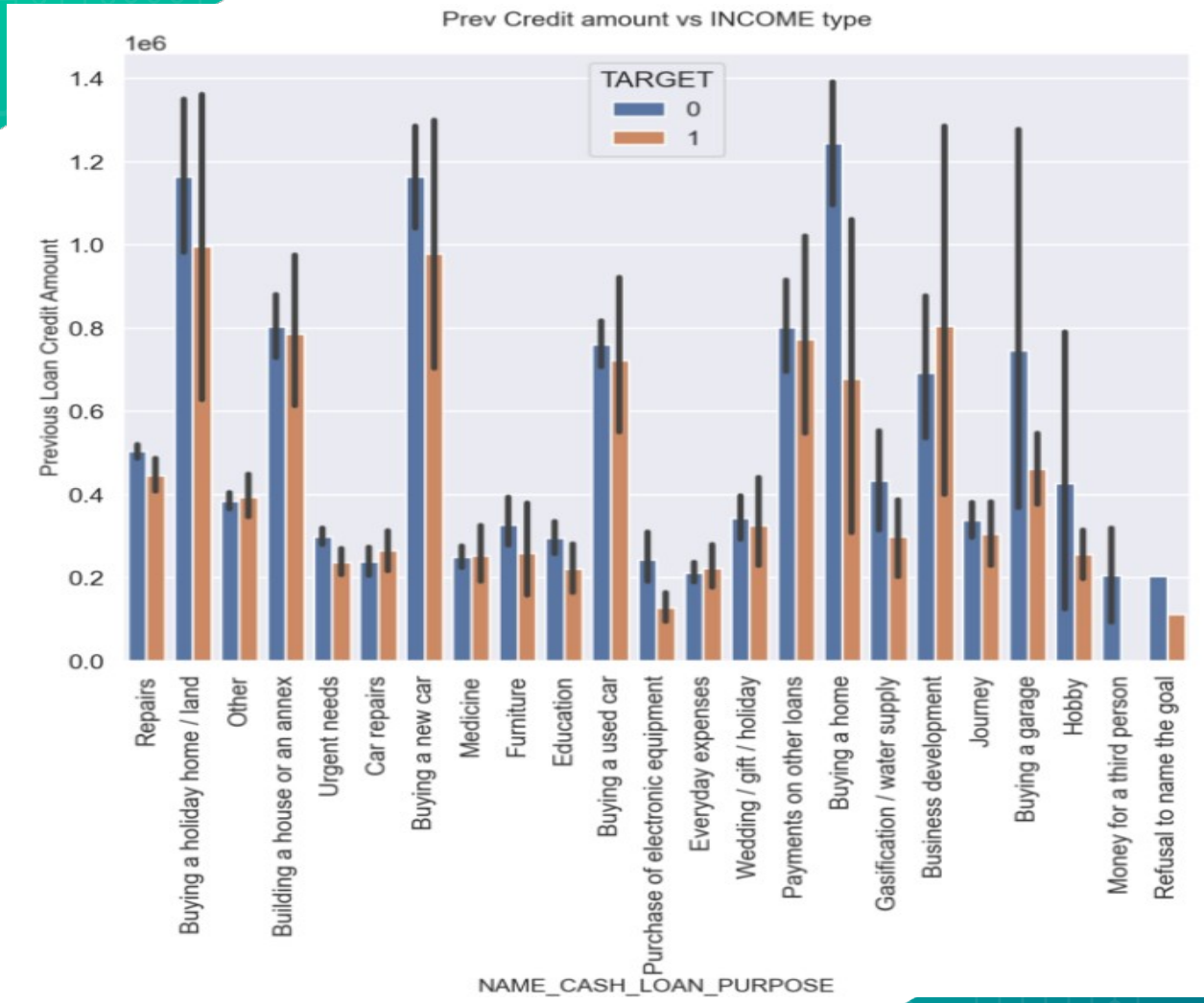
Housing Type wise Applicants in various Credit ranges



Observations from Housing Type wise Applicants in various Credit ranges

- The trend is that the loans with a relatively lower value have a higher count than people with higher loan value.
- There is one exception to the trend which is 90000 and above which has the highest count.
- Generally people living with parents have a higher count of loans than people living with Municipal apartment.
- The exceptions are people with high loan credit values. As the credit value increases, the count of people in Munipal apartment goes up relative to people living with friends.

Previous Credit Amount V/s Income type



Observations from the Previous Credit Amount vs Income type

- Loan reasons like Buying a holiday home, buying a home or a new car has a lot of high credit amount.
- Reasons like Purchase of electronic equipments, everyday expenses, medicine and car repairs have a lot credit amount.
- For most cases the credit amount of defaulters is lesser than the credit amount
for non defaulters according to previous applications.
- Reasons like business development, medicine, car repairs are areas where the
defaulters have a higher credit amount than the non defaulters.

Conclusions

- It is safer to give loans to Commercial associates, Pensioners and state servants than working people since they have a lower chance of defaulting. This may be due to their higher and more secure income.
- It is safer to give loans to females than males.
- It is safer to give loans to Managers, Accountants and high level staff members(Core, High Skill Tech, Medicine) than Labourers, sales staff, drivers and other lower staff members which makes sense that they have a higher income and better financial security.
- It is safer to give loans to people having a higher level of education than those who have secondary, lower secondary or incomplete educational background.
- For Loan purposes such as business development, medicine, car repairs, we should be careful providing a high loan amount as it might lead to a default.

THANK YOU!

Best Regards

You can contact us at

tusharprasad96@gmail.com