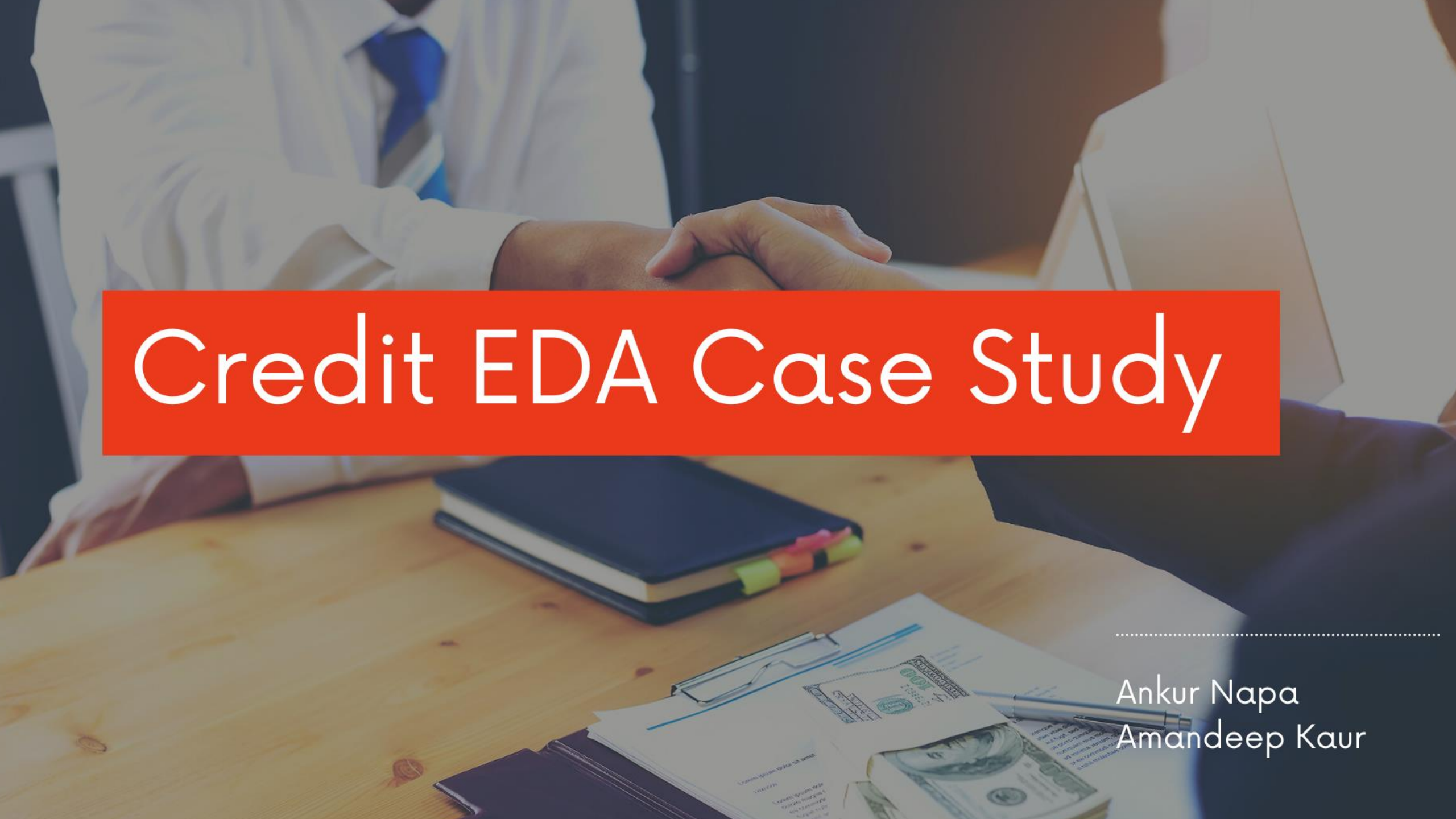


Credit EDA Case Study

Ankur Napa
Amandeep Kaur

Problem Statement

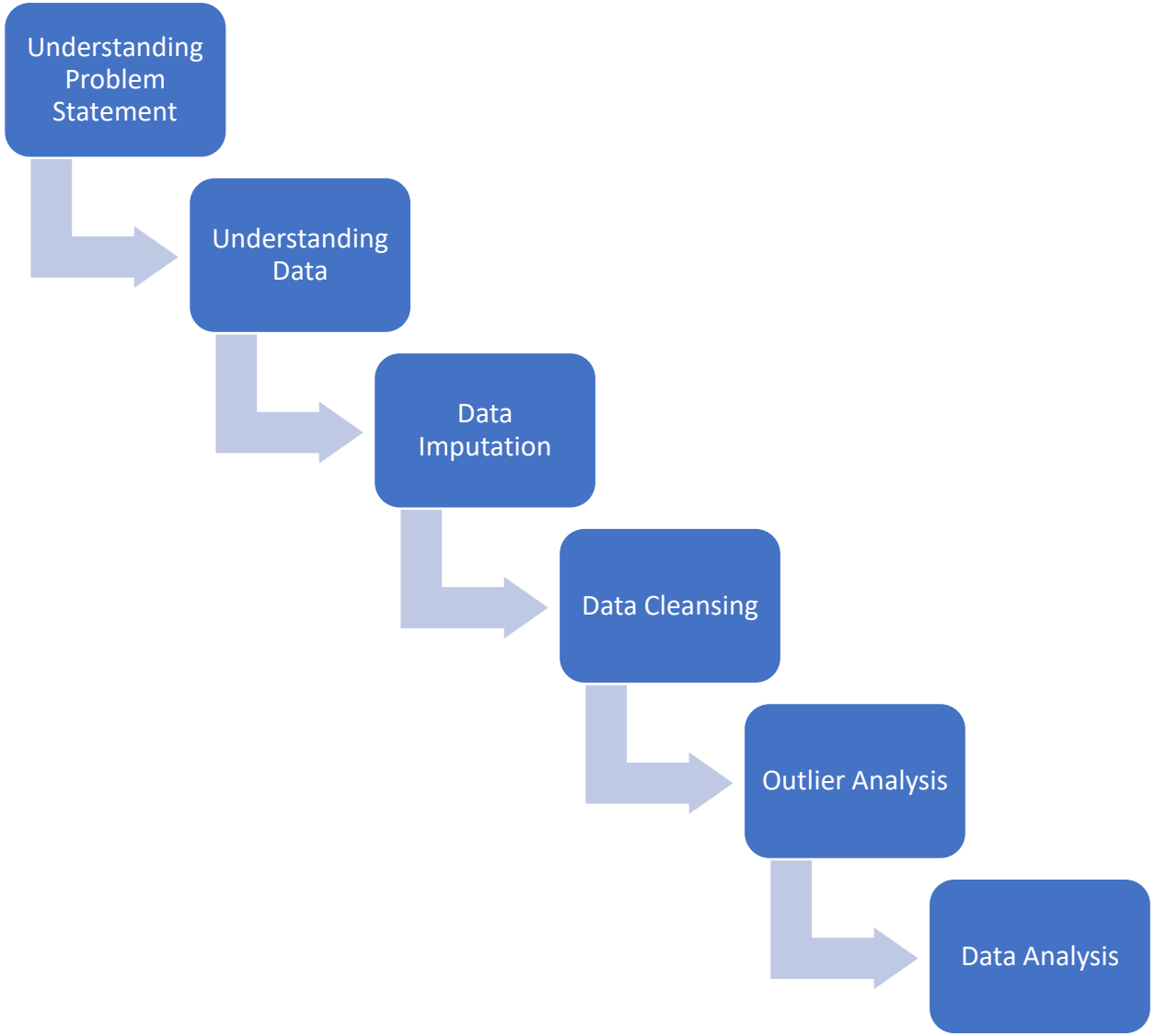
A consumer finance company specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

Two types of risks are associated with the bank's decision:

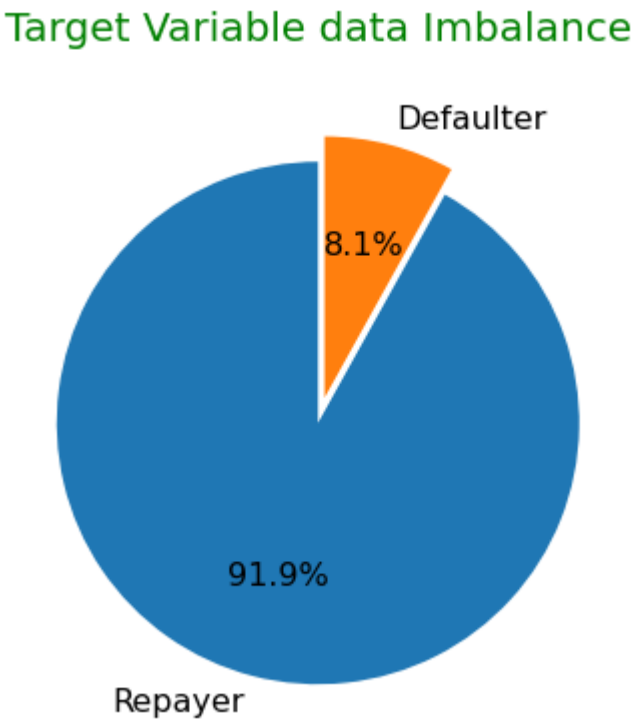
1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
2. If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

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 utilize this knowledge for its portfolio and risk assessment.

Analysis Approach



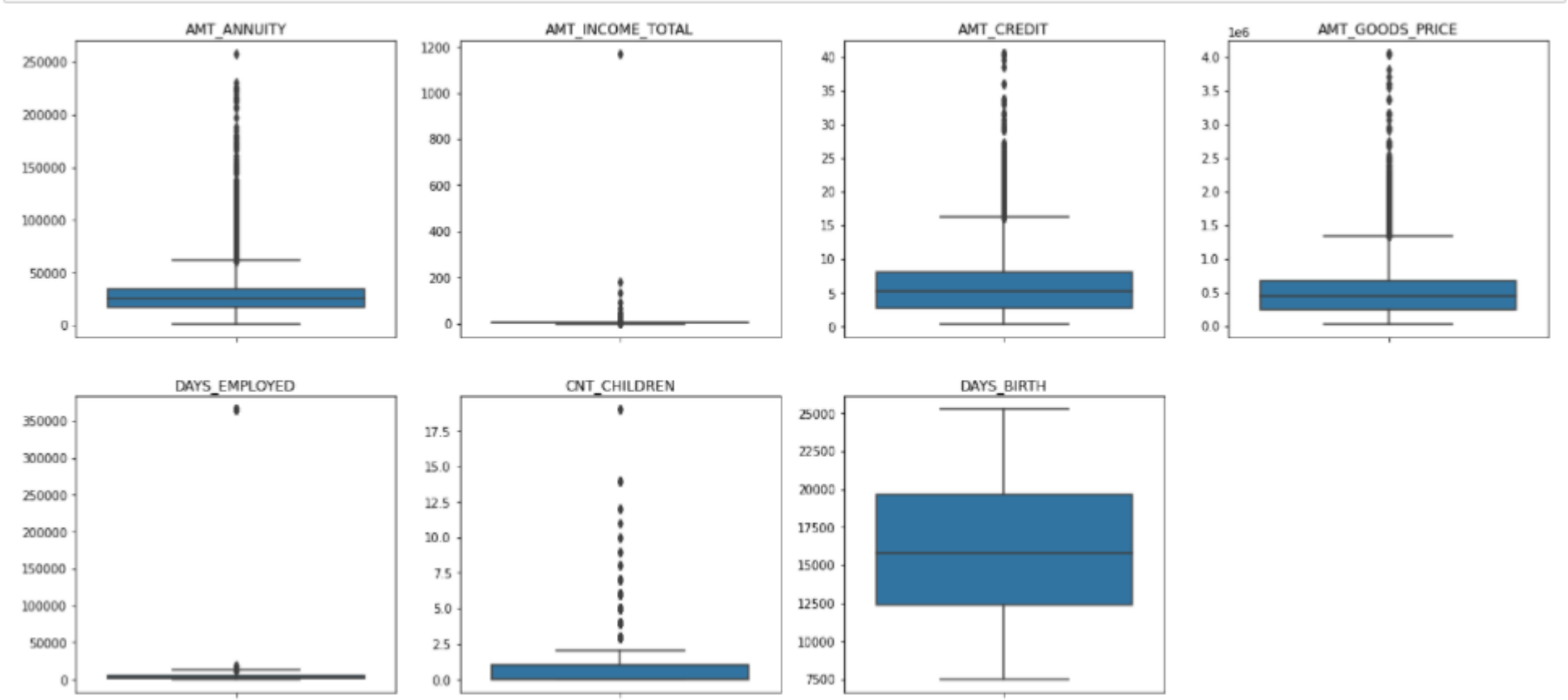
Data Imbalance



Ratios of imbalance for Repayer and Defaulter in Percentage is: 91.93 and 8.07

Ratios of imbalance for Repayer Vs Defaulter is: 11.39 :1 (approx.)

OUTLIERS ANALYSIS



INFERENCE:-

- AMT_ANNUITY, AMT_CREDIT, AMT_GOODS_PRICE,CNT_CHILDREN have some number of outliers.

- AMT_INCOME_TOTAL has huge number of outliers which indicate that few of the loan applicants have high income when compared to the others.

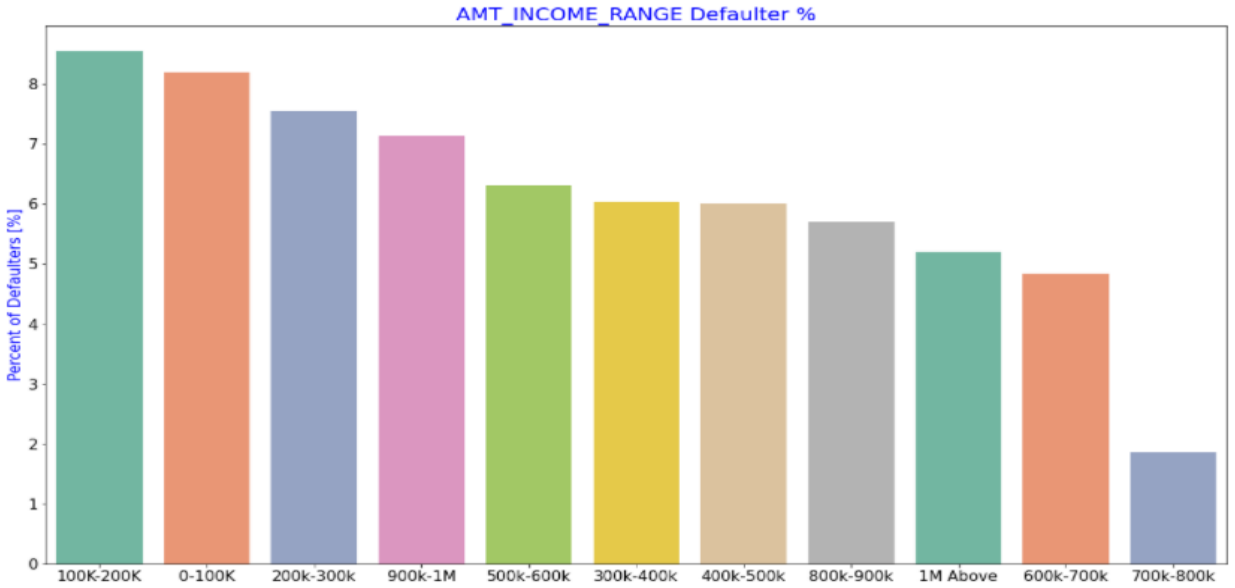
- DAYS_BIRTH has no outliers which means the data available is reliable.

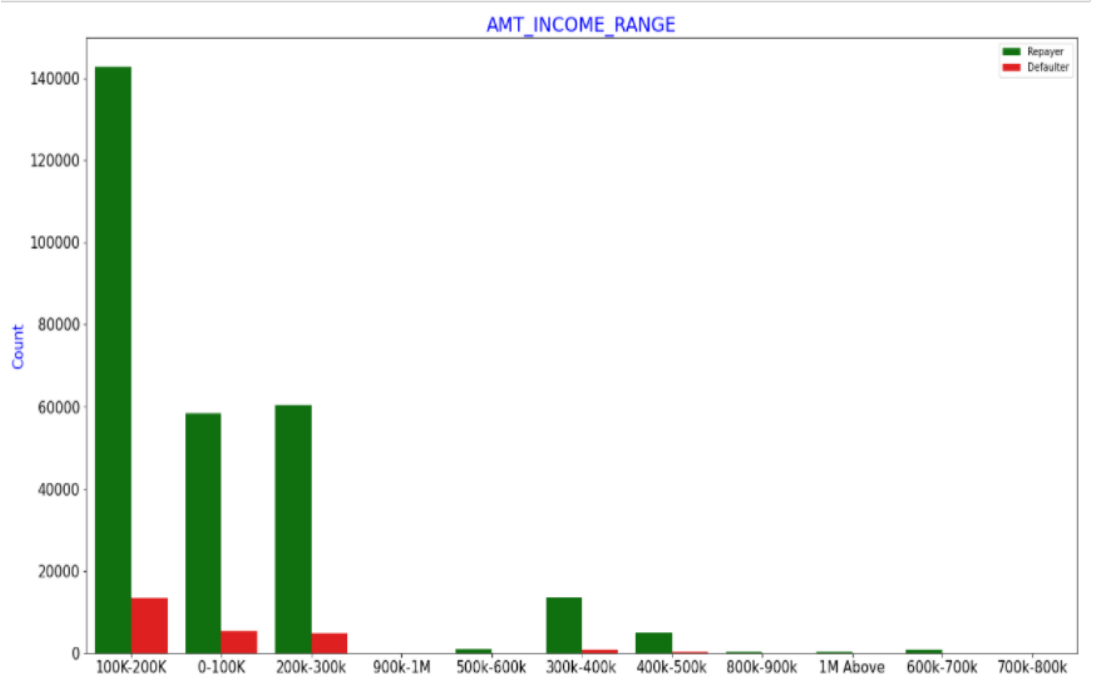
- DAYS_EMPLOYED has outlier values around 350000(days) which is around 958 years which is impossible and hence this has to be incorrect entry.

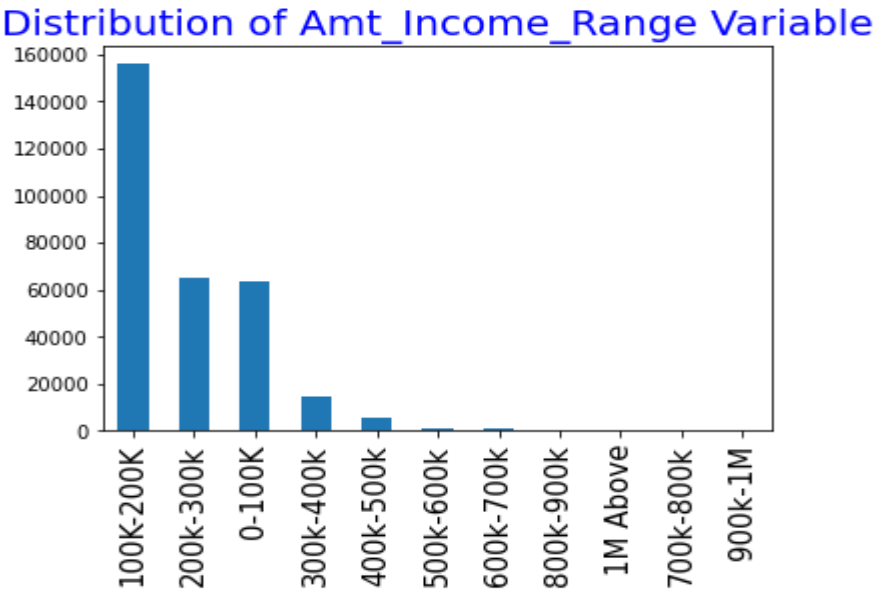
- We can see the stats for these columns below as well.


Univariate
3.32 Pages/Visit
Traffic Sources Overview
Direct Traffic
3,097.00 (40.49%)
Search Engines
2,910.00 (38.04%)
Referring Sites
1,642.00 (21.47%)
Visitors Overview
Apr 19
Apr 26
May 3
Visitors
2,958

Amount Income Range





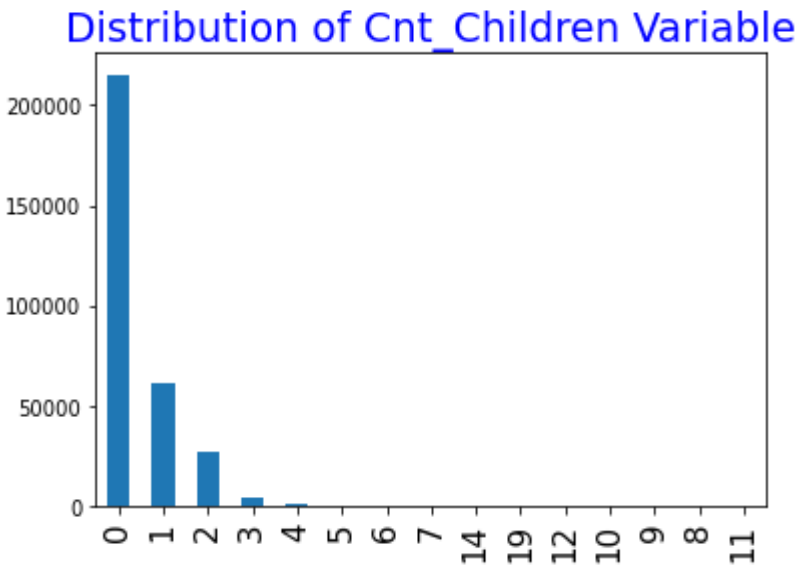


- Majority of the applicants have salary between 100-200K
- Application with Income less than 300,000 has high probability of defaulting
- Applicant with Income between 700-800k are less likely to default

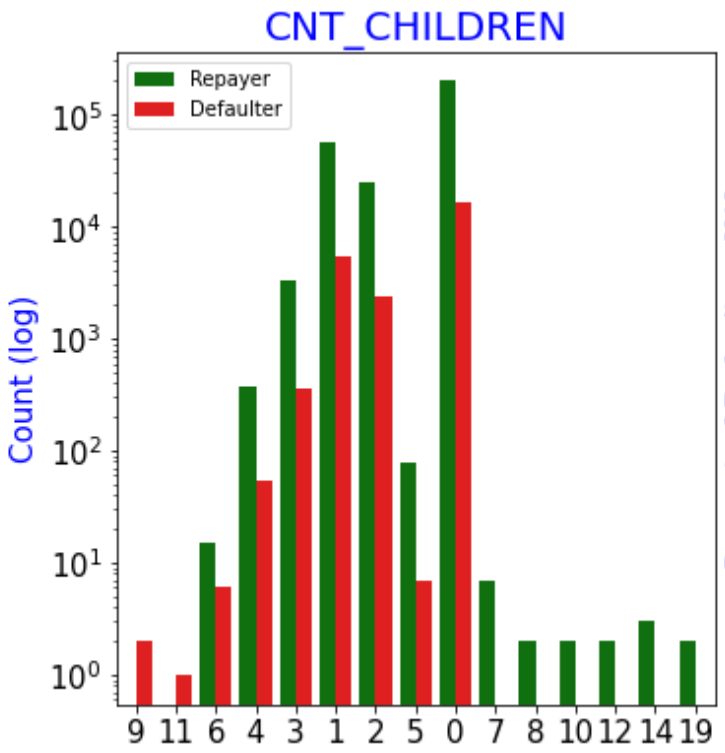
INFERENCE:-

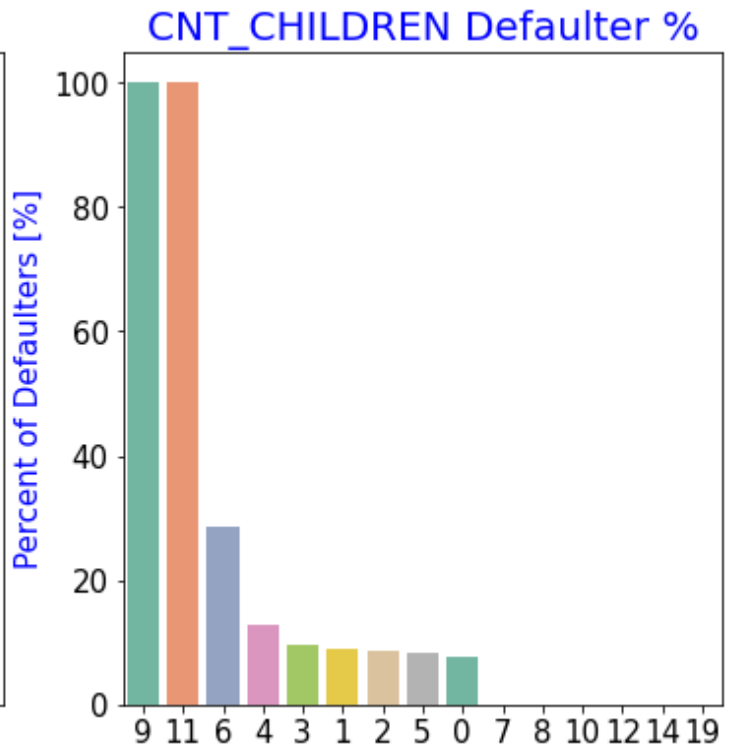
Applicants with Income more than 700,000 are less likely to default

Count of Children



Most of the applicants do not have children.
Very few clients have more than 3 children



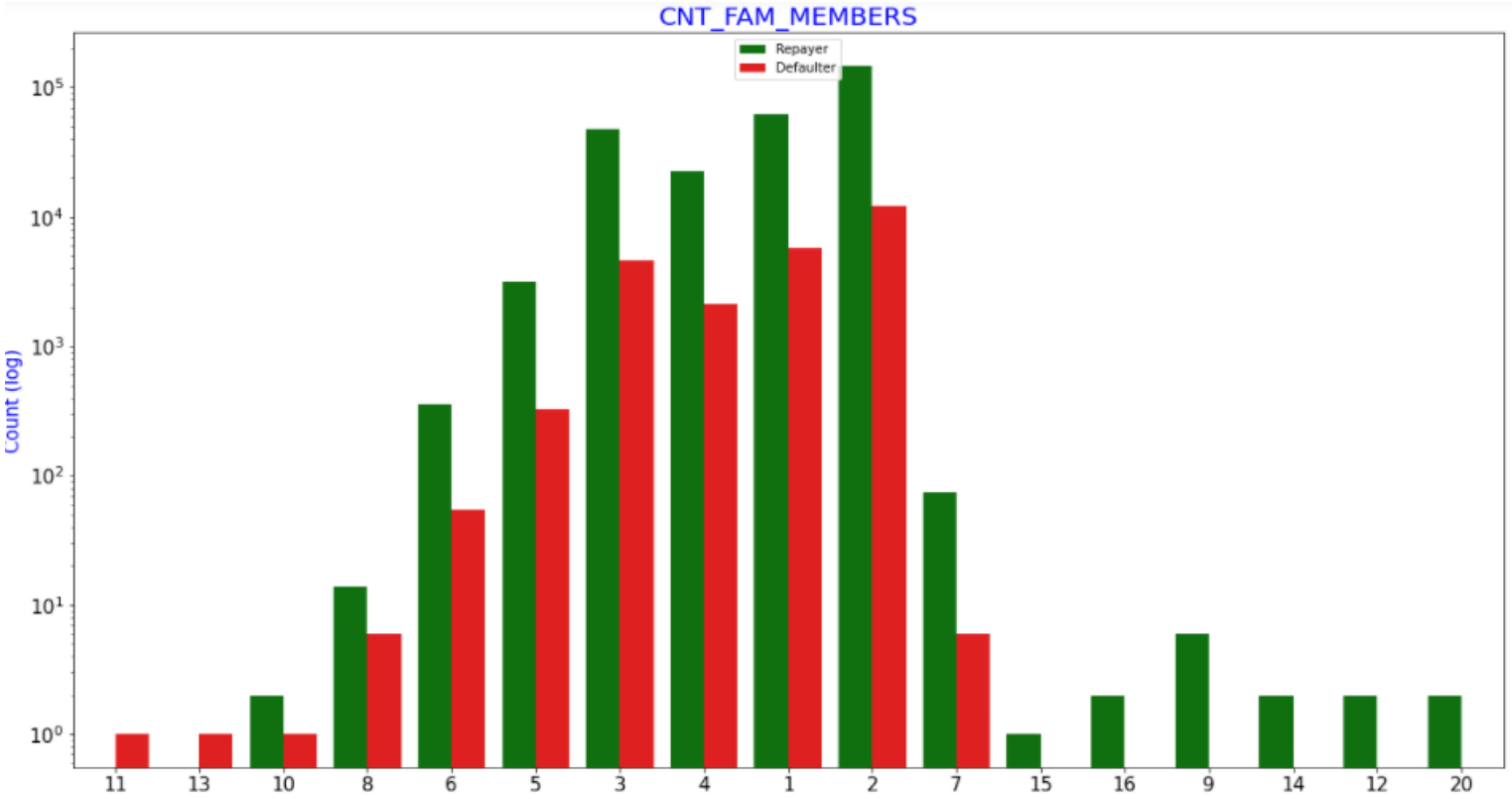


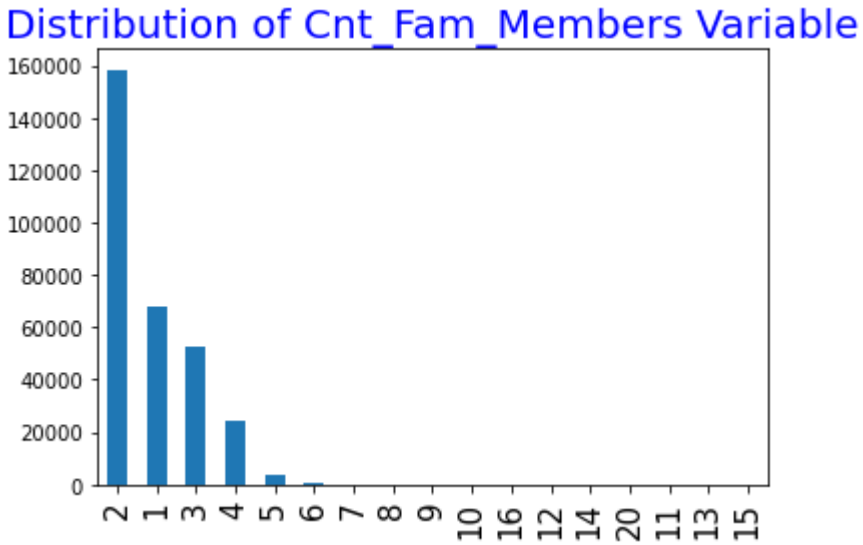
Most of the applicants no children. As applicants in this group are more the no. of defaulters are also more in this group.
Client who have more than 4 children have a very high default rate with child count 9 and 11 showing 100% default rate

INFERENCE:-

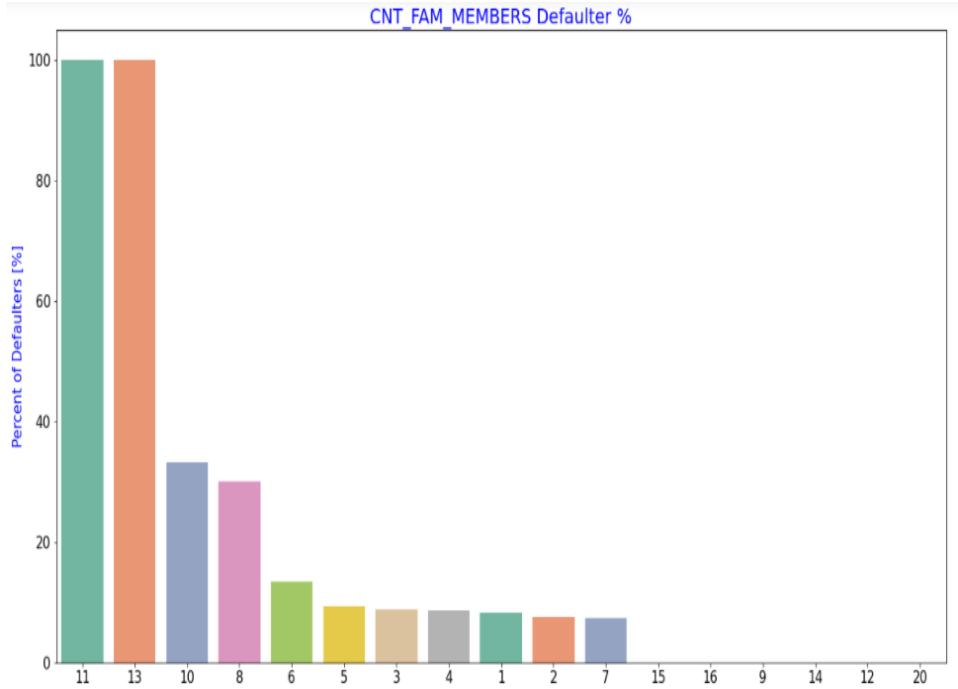
Applicants with zero to two children tend to repay the loans

COUNT OF FAMILY MEMBERS



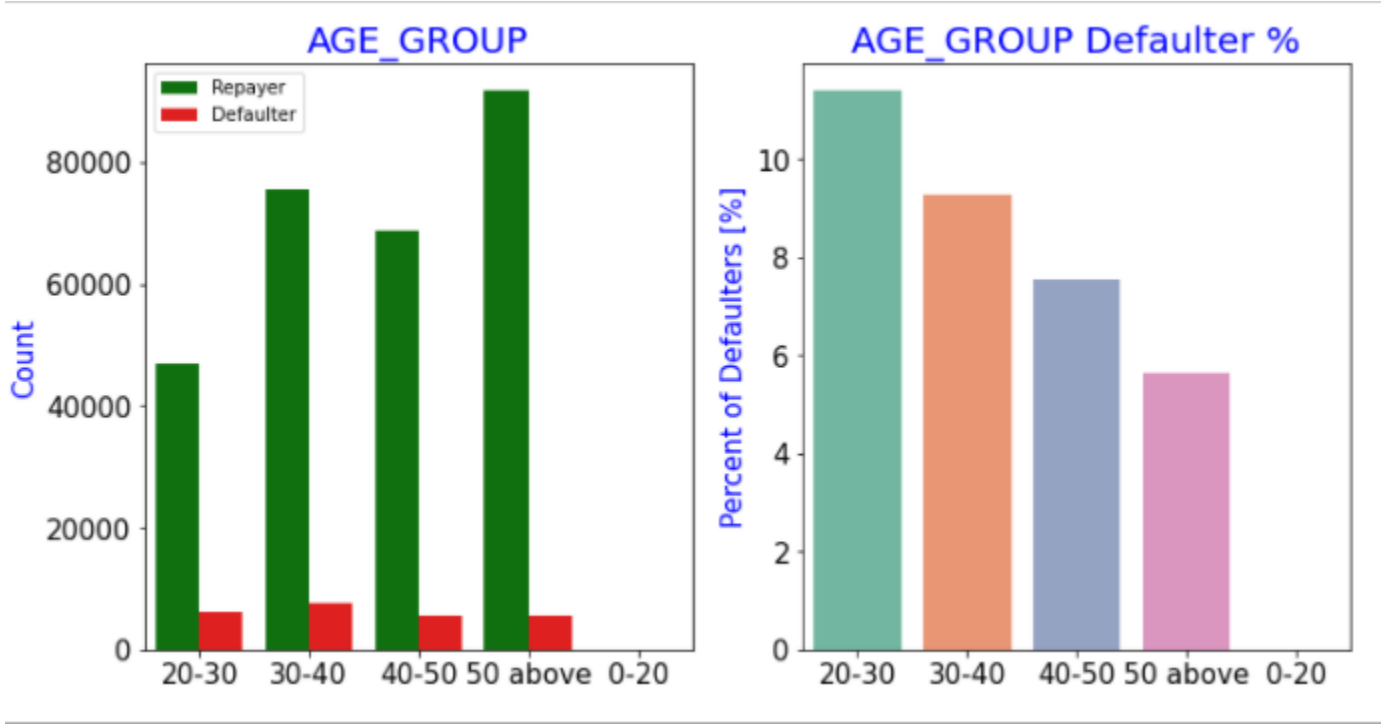


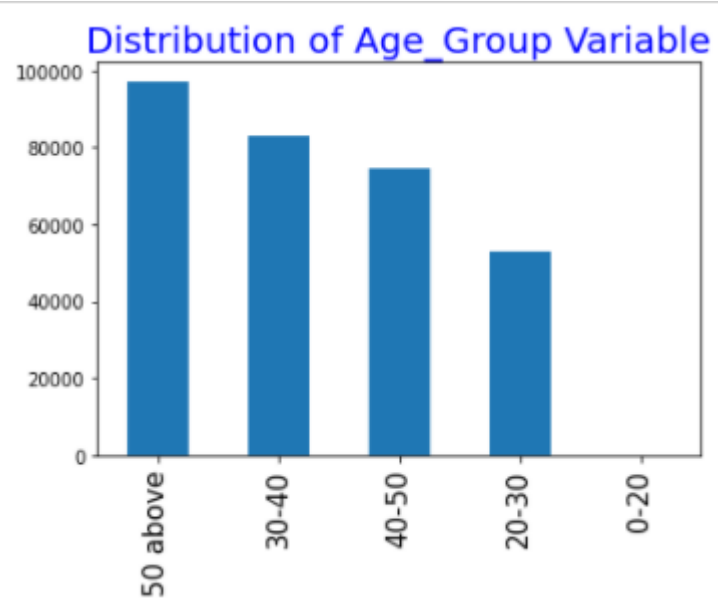
Majority of the clients have 2 family members



Inference: *Applicants who have higher family members (>=11) have higher default rate and their applications can be rejected*

AGE GROUP





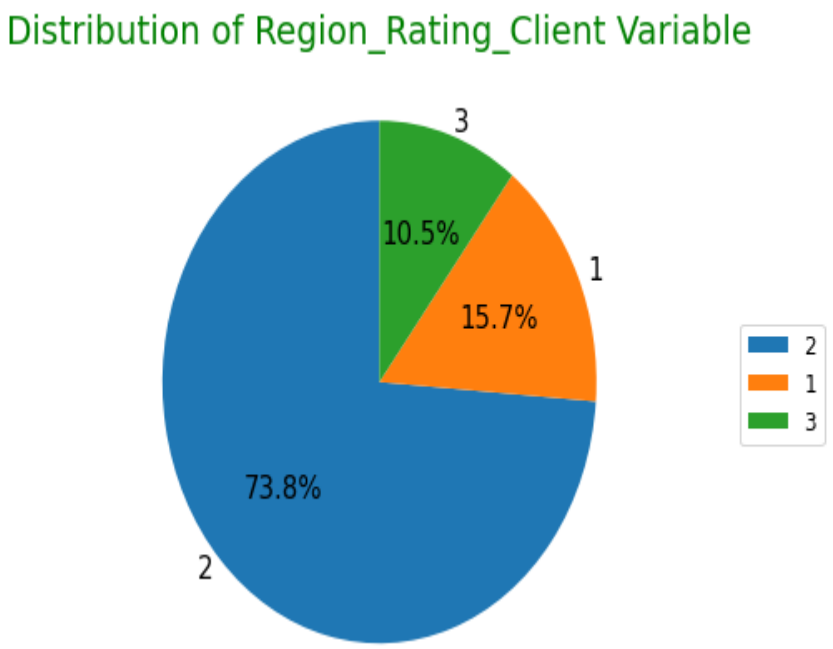
Majority of the clients are of age more than 50

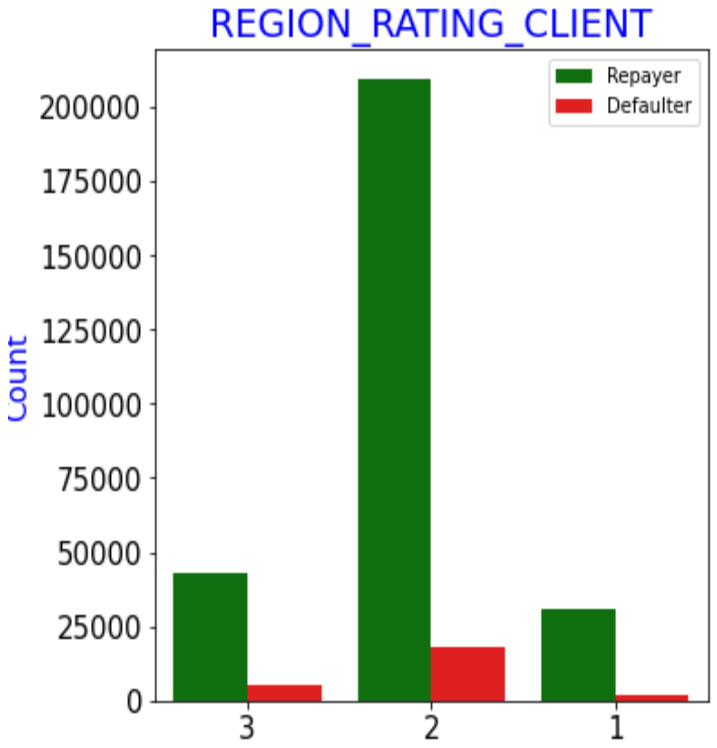
- People in the age group range 20-40 have higher probability of defaulting
- People above age of 50 have low probability of defaulting

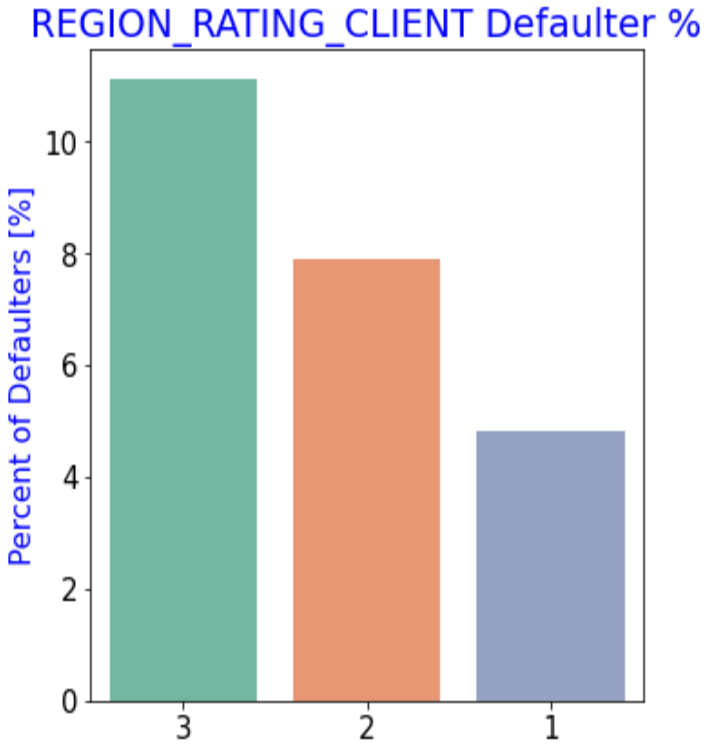
INFERENCE:-

Applicants above age of 50 have low probability of defaulting, hence there applications can be approved.
Young applicants who are in age group of 20-40 have higher probability of defaulting.

REGION RATING







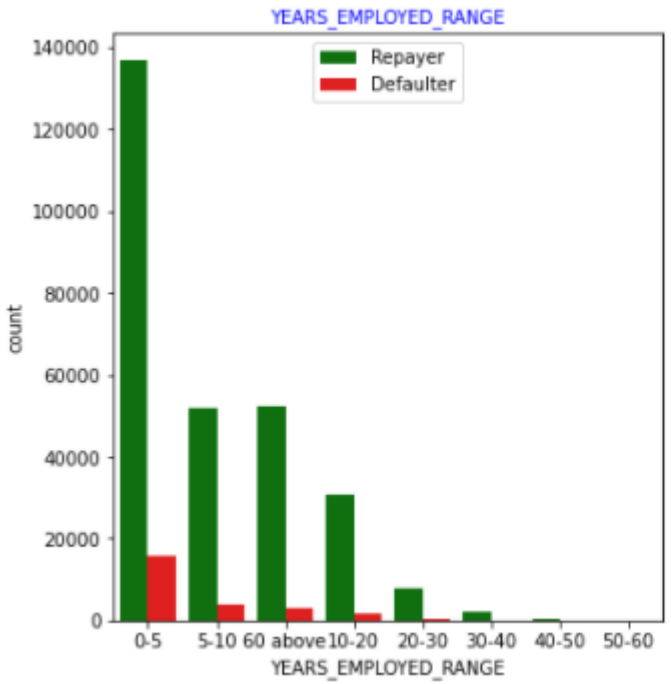
- Most of the applicants are living in Region_Rating 2 place.

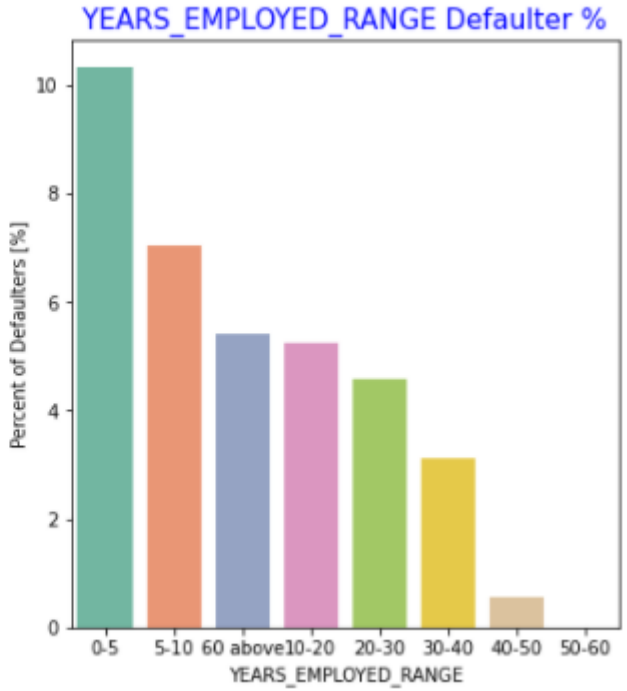
- Region Rating 3 has the highest default rate (11%) , fol1owed by 2(around 8%) and 1(around 5%)
- Applicant living in Region_Rating 1 has the lowest probability of defaulting, thus safer for approving loans

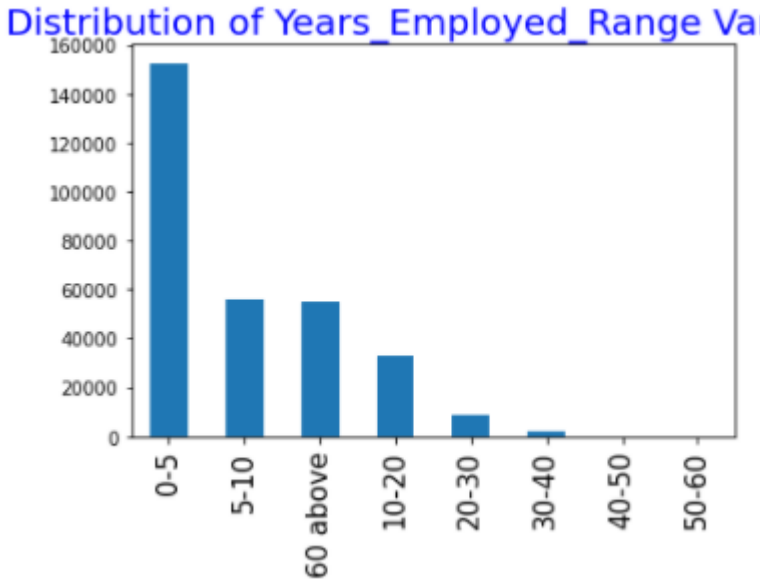
Inference:-

Applicants who live in areas with Region Rating 1 are safe borrowers

YEARS EMPLOYED







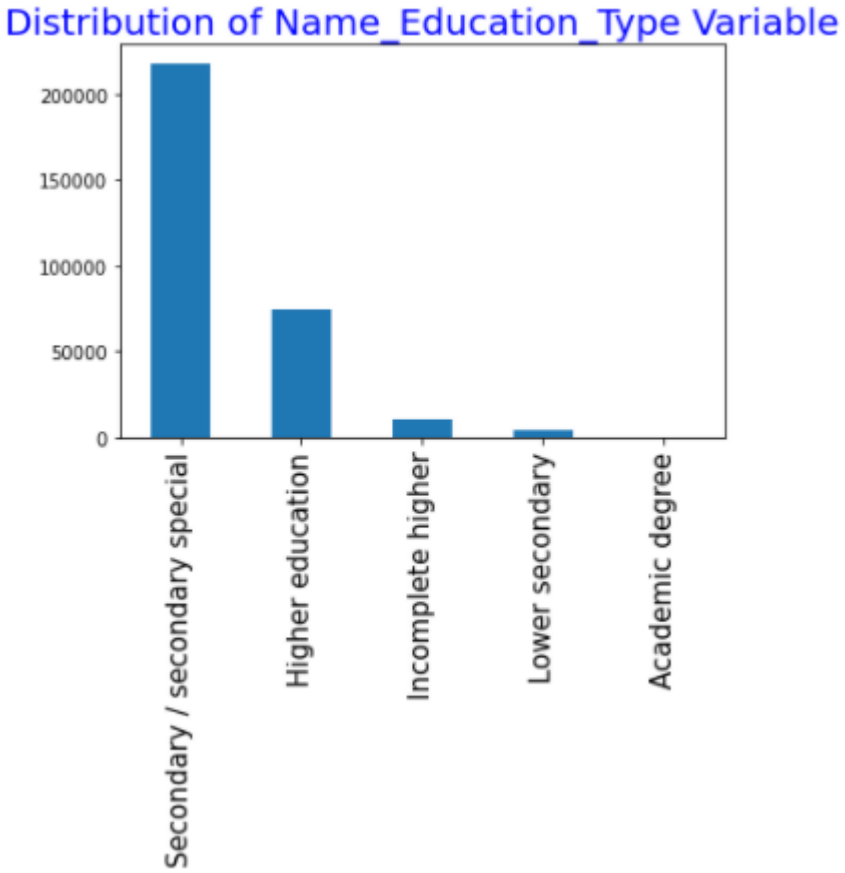
Majority of the applicants have been employed in between 0-5 years

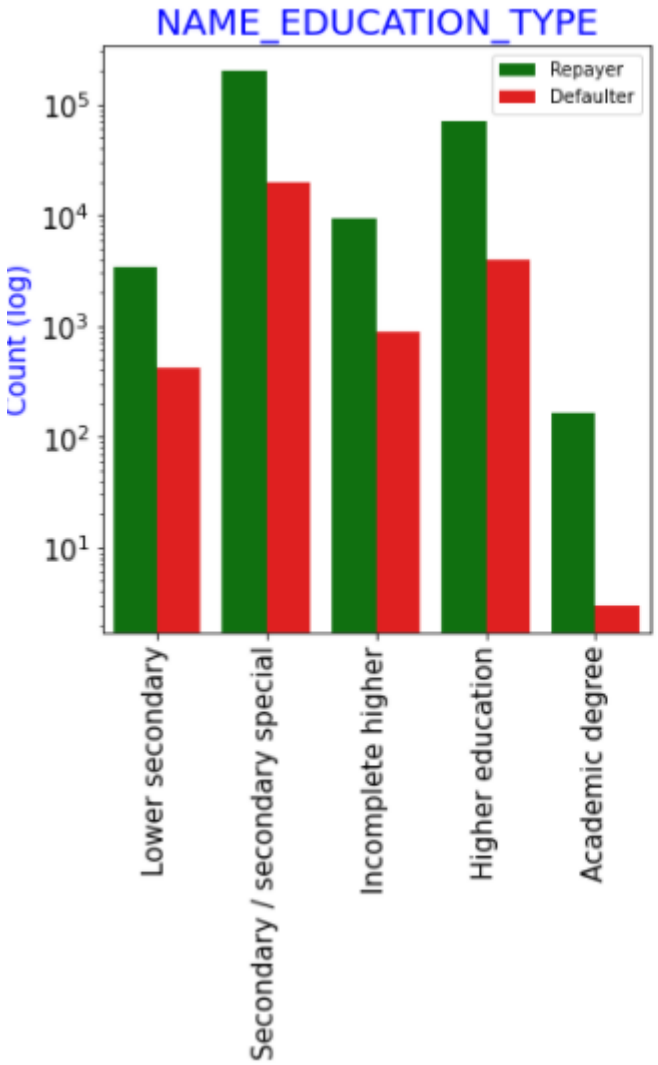
- Majority of the applicants have been employed in between 0-5 years. The defaulting rating of this group is also the highest which is 10%
- With increase of employment year, defaulting rate is gradually decreasing with people having 40+ year experience having less than 1% default rate

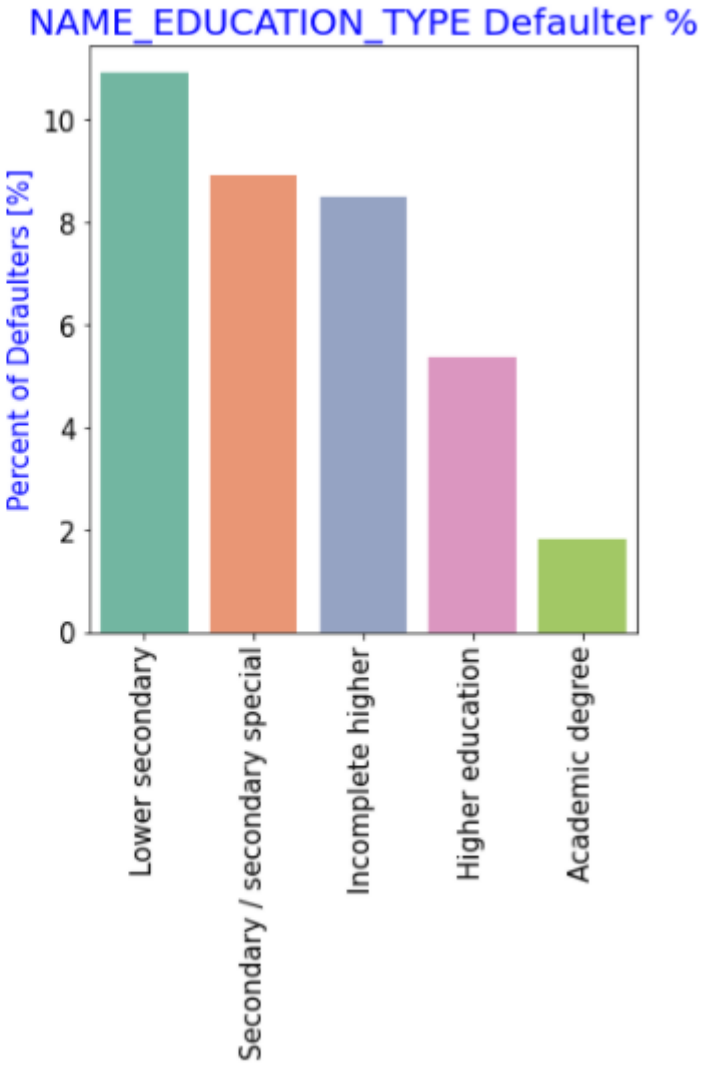
INFERENCE:-

Applicants with 40+ year experience having less than 1% default rate

EDUCATION TYPE







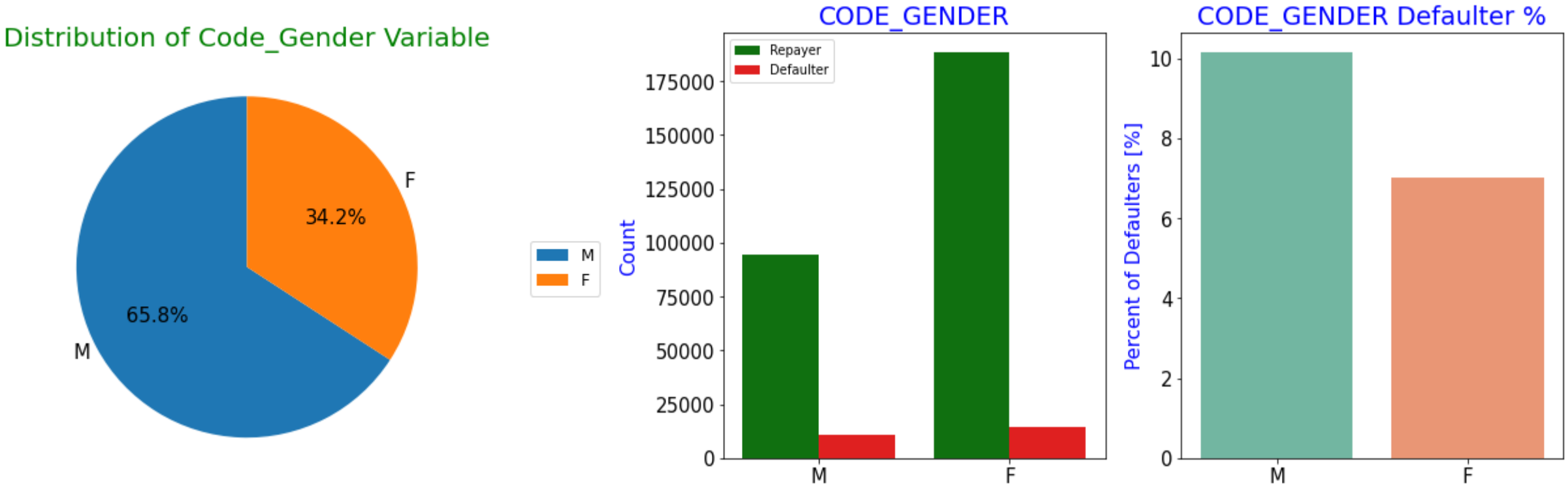
Majority of the clients have Secondary / secondary special education, followed by clients with Higher education. Only a very small number having an academic degree

The Lower secondary category, although rare, have the largest rate of defaulters (11%). The people with Academic degree have the lowest defaulting rate(around 2%).

Inference:-

Applicants who are Academic degree holder have lower default rate

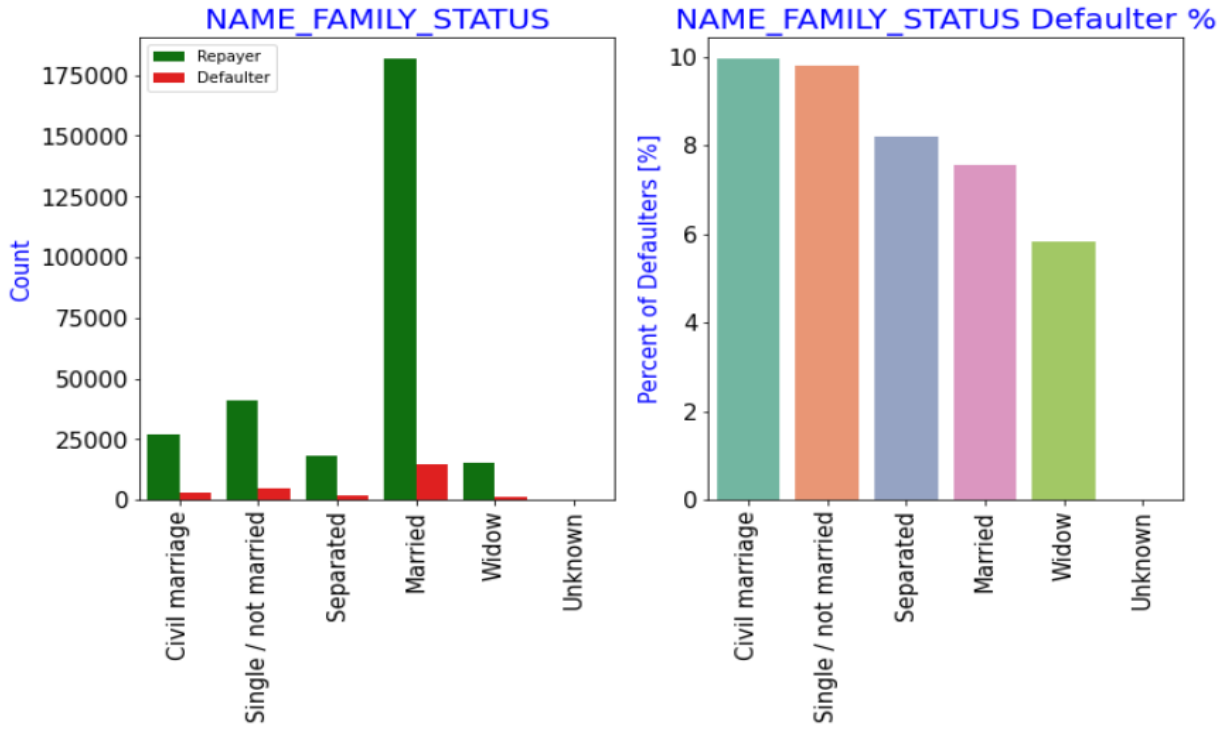
GENDER CODE



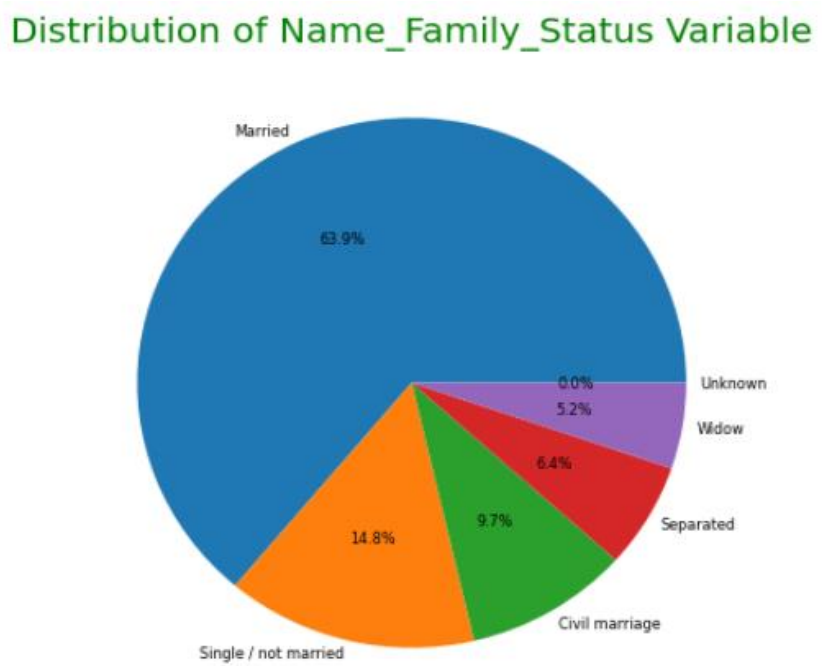
Inferences:

The number of female clients is almost double the number of male clients. Based on the percentage of defaulted credits, males have a higher chance of not returning their loans (approx 10%), comparing with women (~7%)

FAMILY/MARITIAL STATUS



In terms of percentage of not repayment of loan, Civil marriage has the highest percent of not repayment (10%), with Widow the lowest (exception being Unknown).

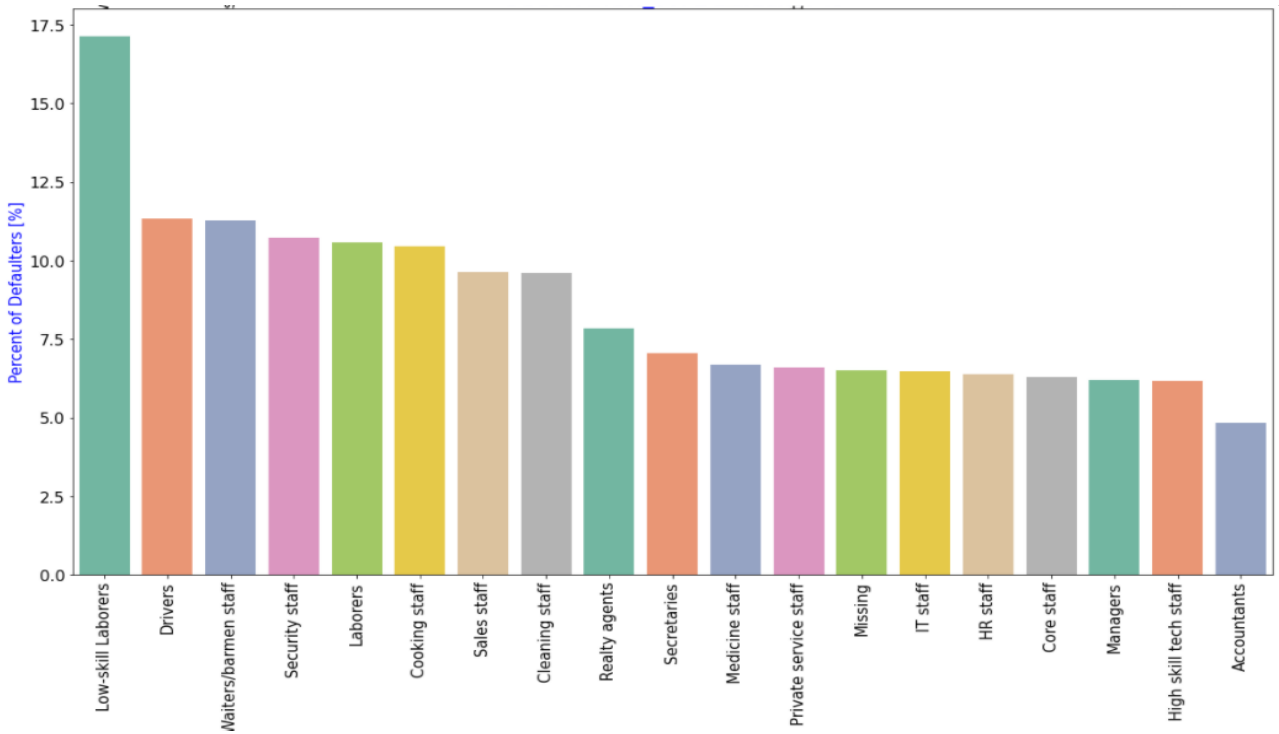


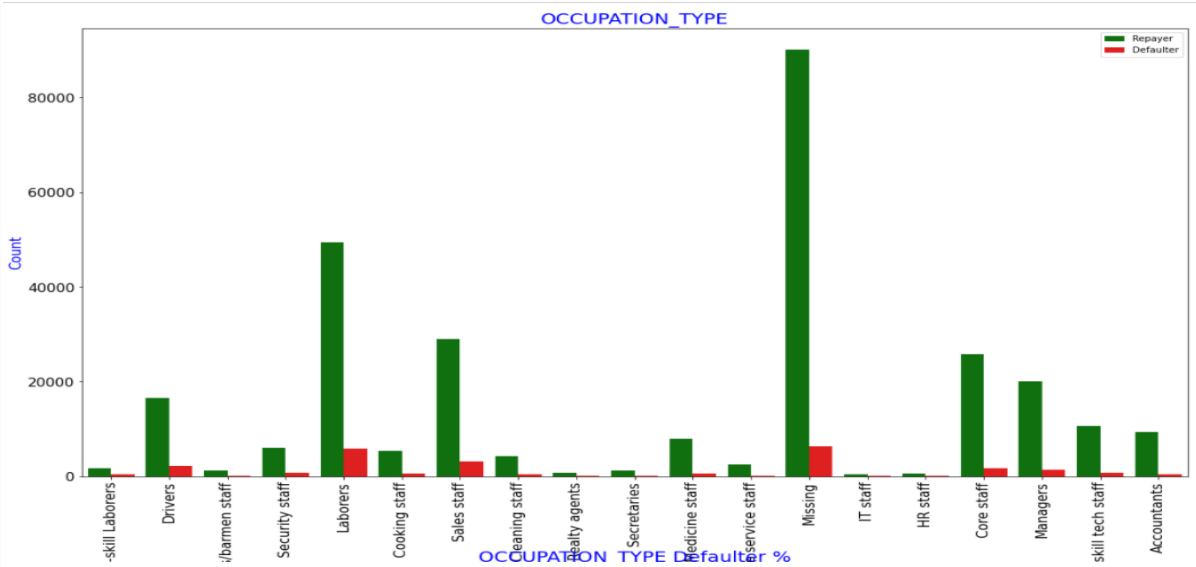
Most of the people who have taken loan are married, followed by Single/not married and civil marriage

Inference

Applicants in civil marriage or who are single have higher default rate

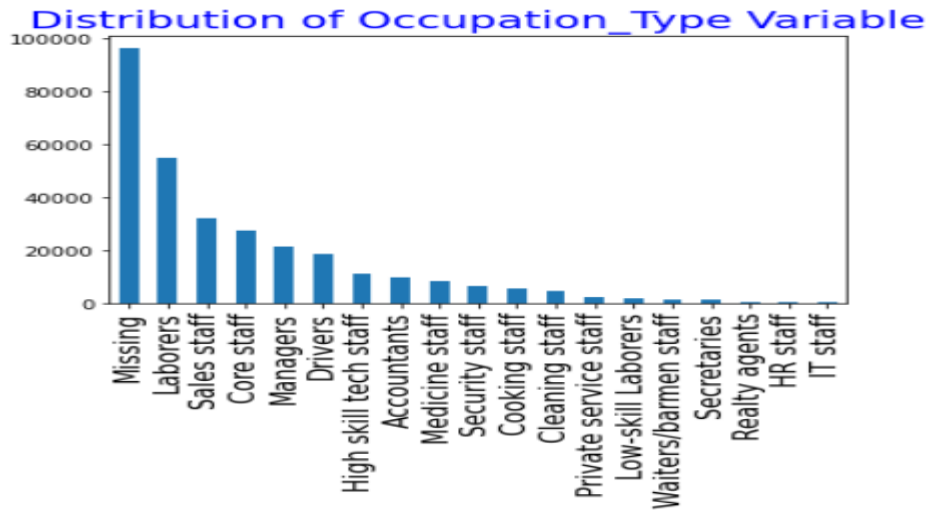
OCCUPATION TYPE





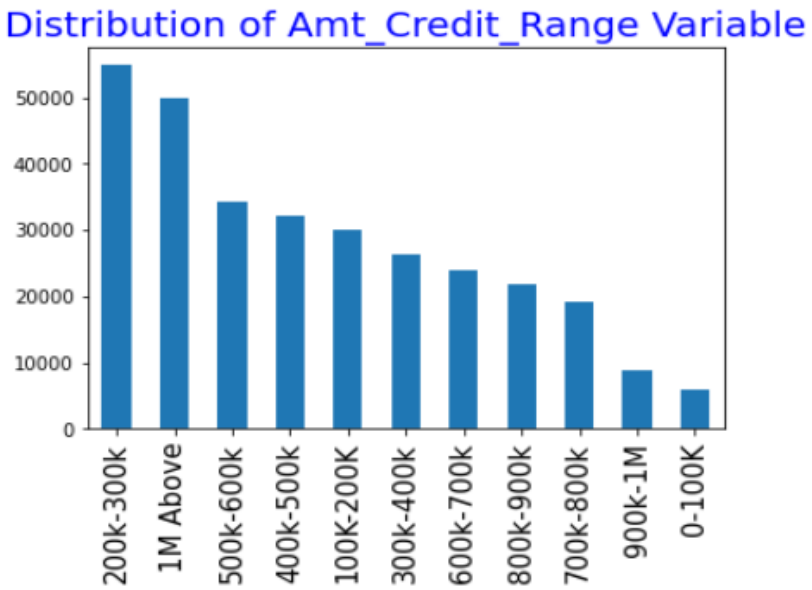
The category with highest percent of not repaid loans are Low-skill Laborers (above 17%), followed by Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff.

Inference:- *Applicants who are Low-skill Laborers, Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff as the default rate is huge.*

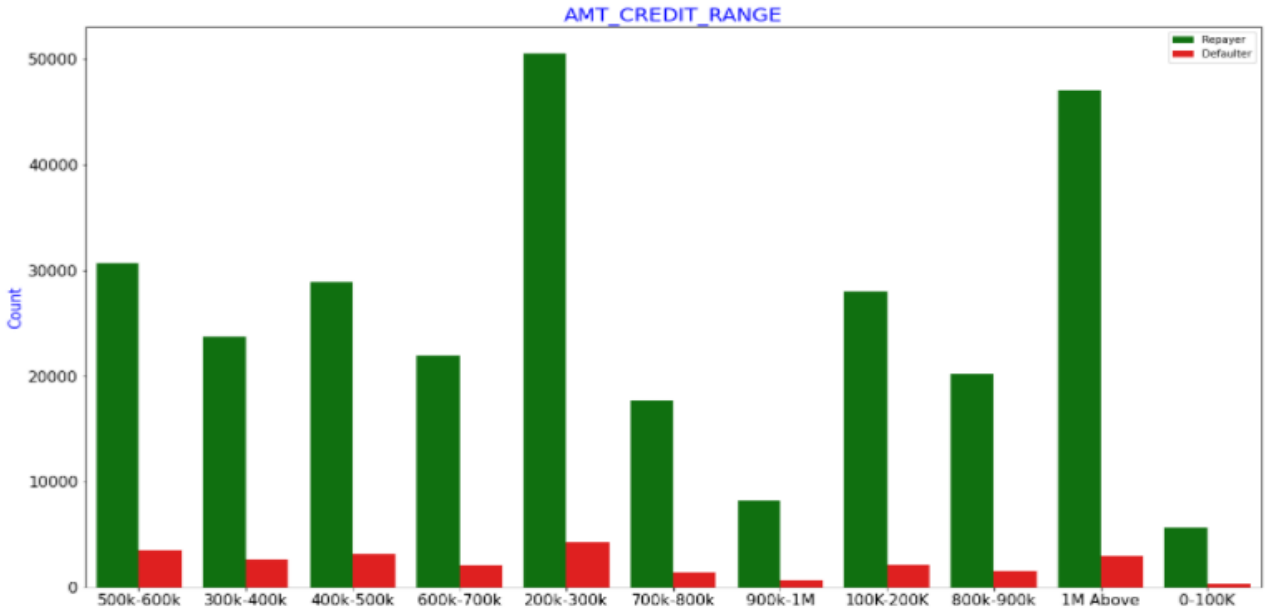


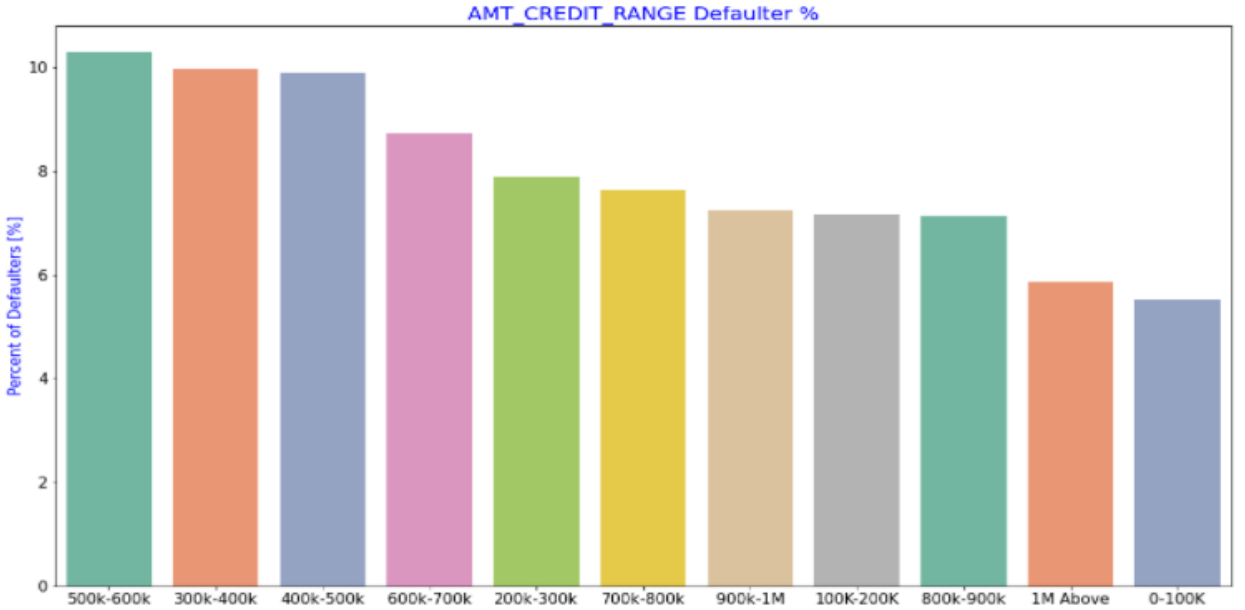
Most of the loans are taken by people whose Occupation is "Missing" in the dataset followed by Laborers, Sales staff. IT staff take the lowest amount of loans.

CREDIT AMOUNT



Majority of the Loan amount is between 200-300K



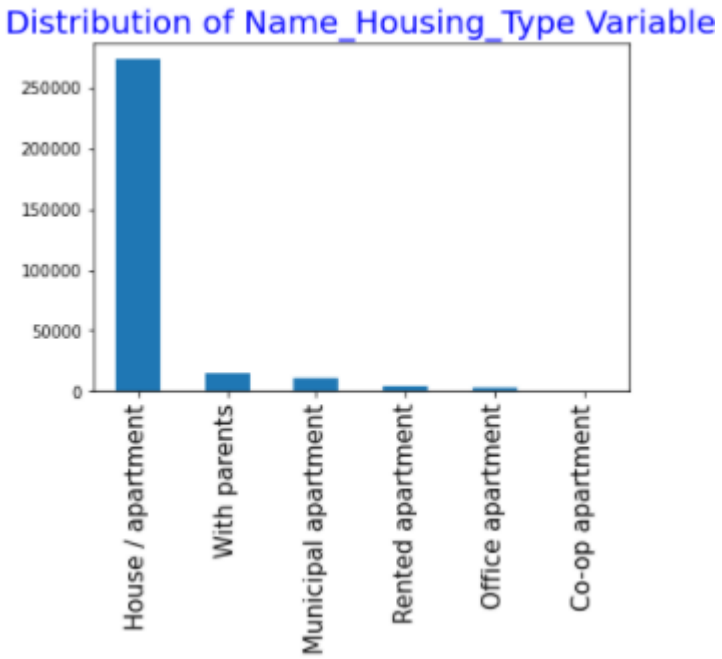


Inference:-

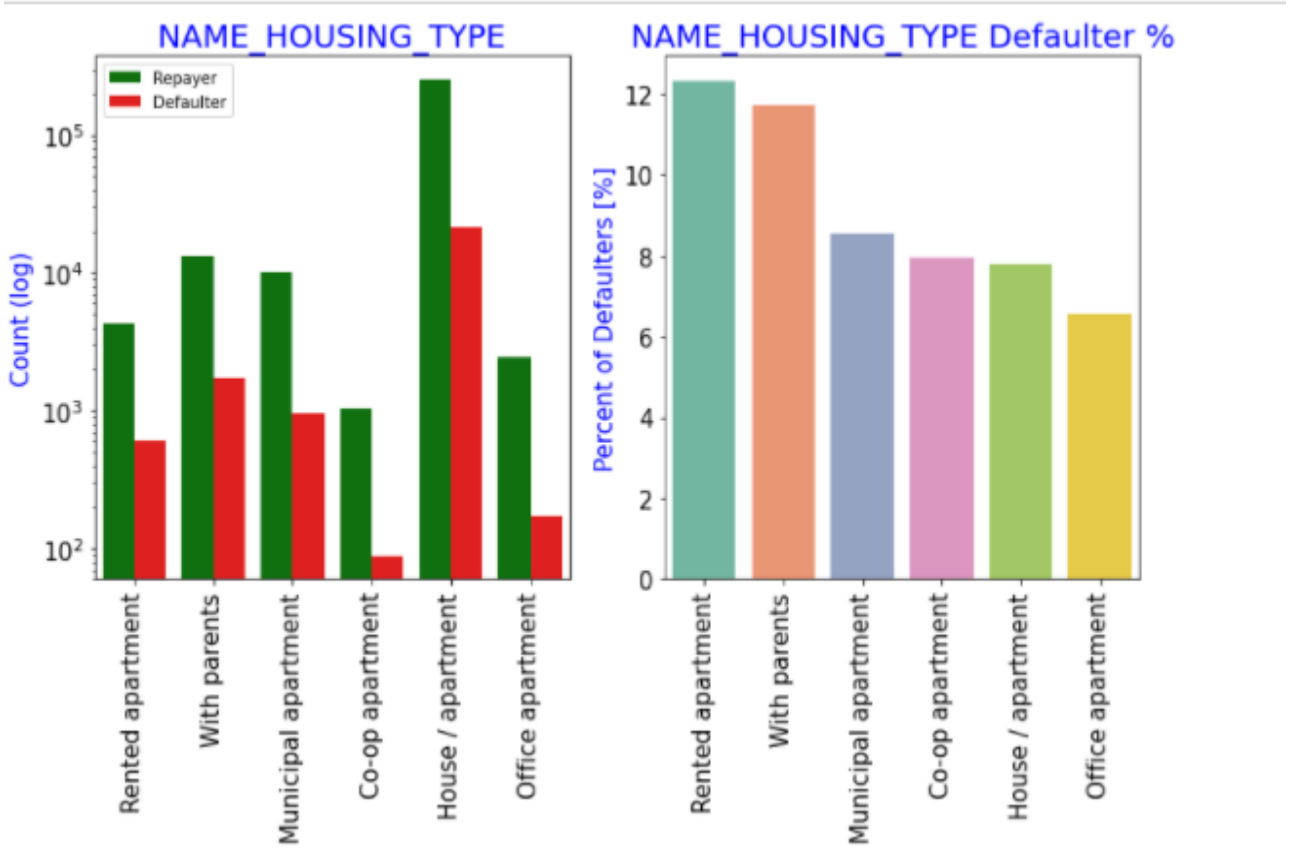
Applicants with Income more than 700,000 are less likely to default

More than 80% of the loan provided are for amount less than 900,000 People who get loan for 300-600k tend to default more than others.

HOUSING TYPE



Majority of people live in House/apartment

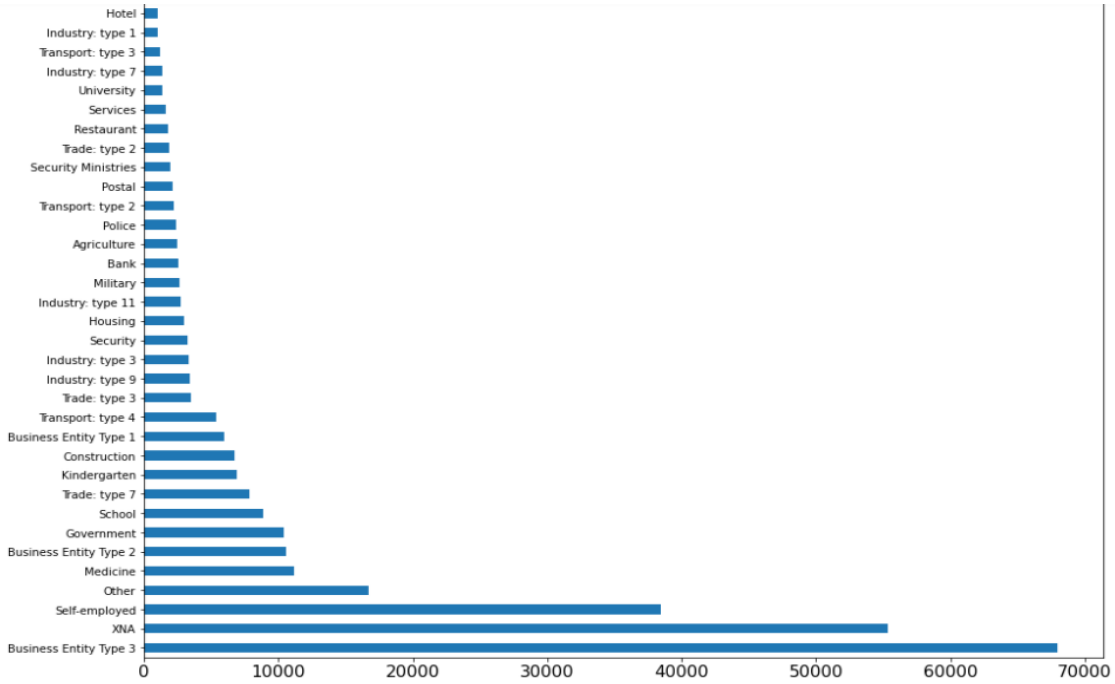


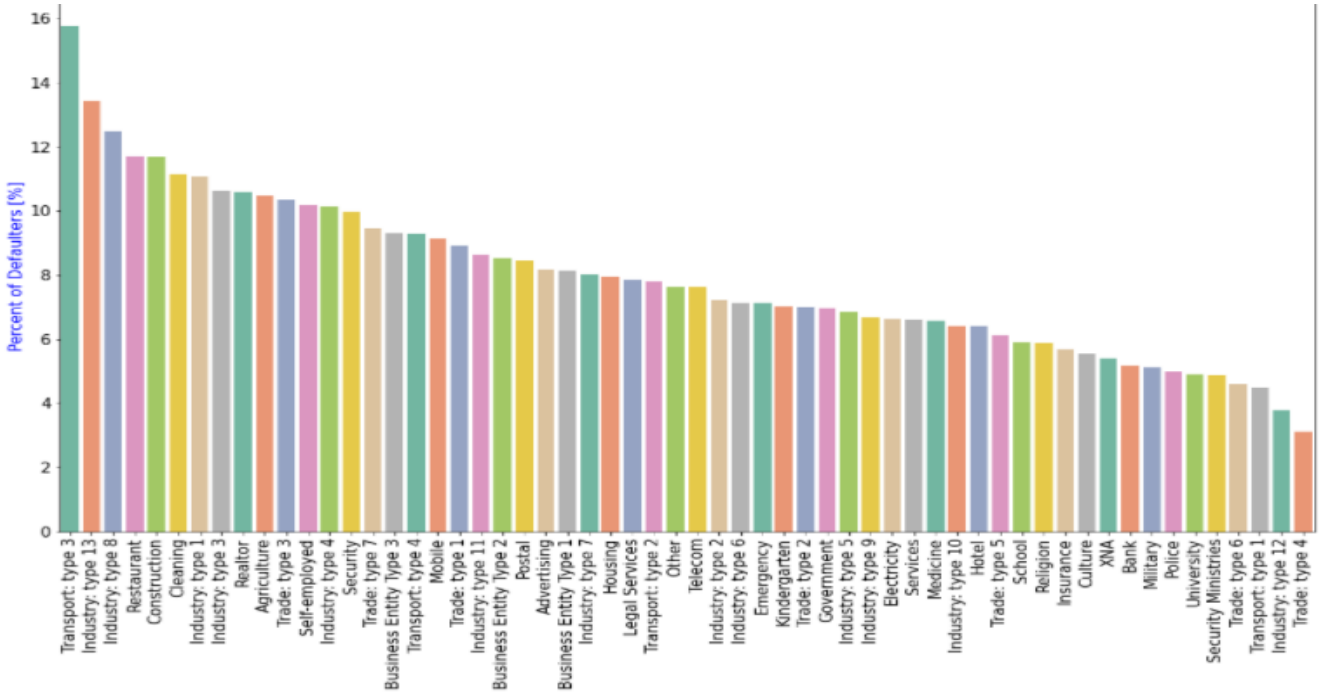
People living in office apartments have lowest default rate
- People living with parents (around 11.5%) and living in rented apartments(> 12%) have higher probability of defaulting

Inference:

High number of loan applications are from the category of people who live in Rented apartments & living with parents and hence offering the loan would mitigate the loss if any of those default

ORGANIZATION_TYPE





Inferences:

- Most of the applications for loan are from people working in Business Entity Type 3 organization

- Organizations with highest percent of loans not repaid are Transport: type 3 (around 16%), Industry: type 13 (13.5%), Industry: type 8 (around

- Self employed people have relative high defaulting rate (10%), and thus should be thoughly scrutized before being approved for loan or provide

- For a very high number of applications, Organization type information is unavailable(XNA)

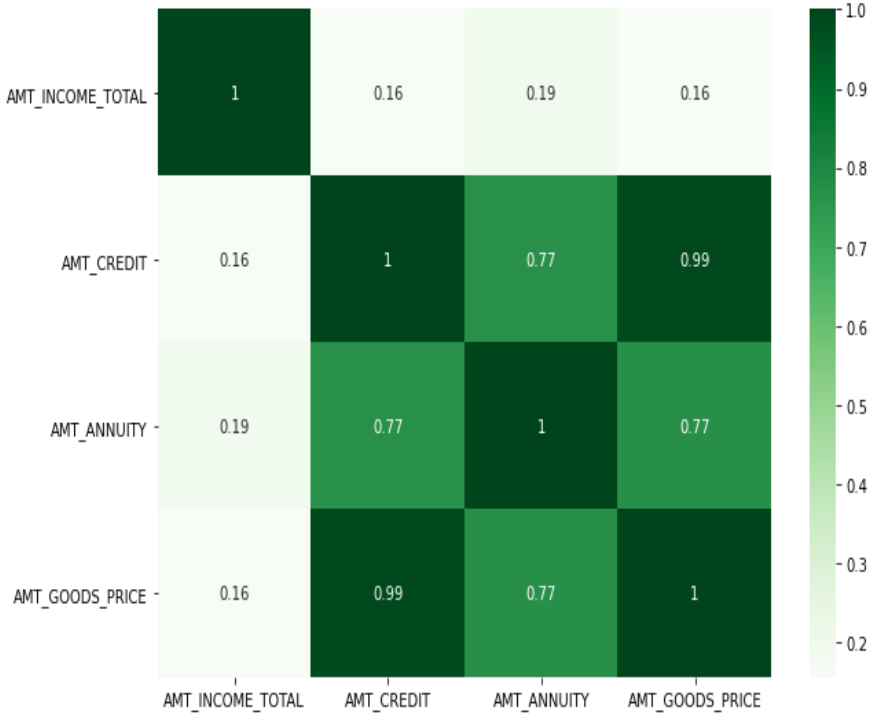
- It can be seen that following category of organization type has lesser defaulters thus safer for providing loans:

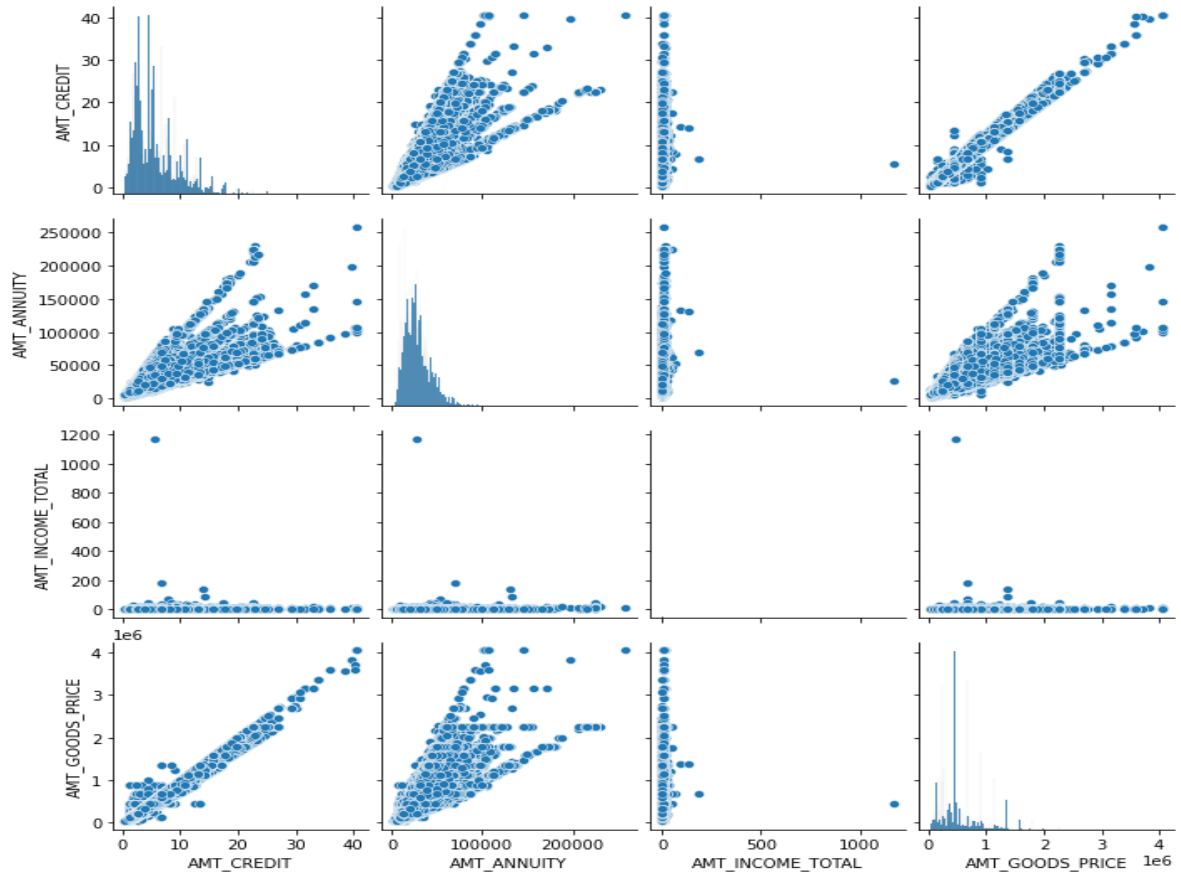
- Trade Type 4

- Industry type 122%).

Bivariate & Multivariate

AMT_CREDIT, AMT_ANNUITY, AMT_INCOME_TOTAL , AMT_GOODS_PRICE





Inferences:

Very high correlation between AMT_CREDIT and AMT_GOODS_PRICE - Applicants owning goods of high value can take loans of higher amounts.

Multivariate (Numeric Columns)

Inferences:

Correlating factors amongst re-payers:

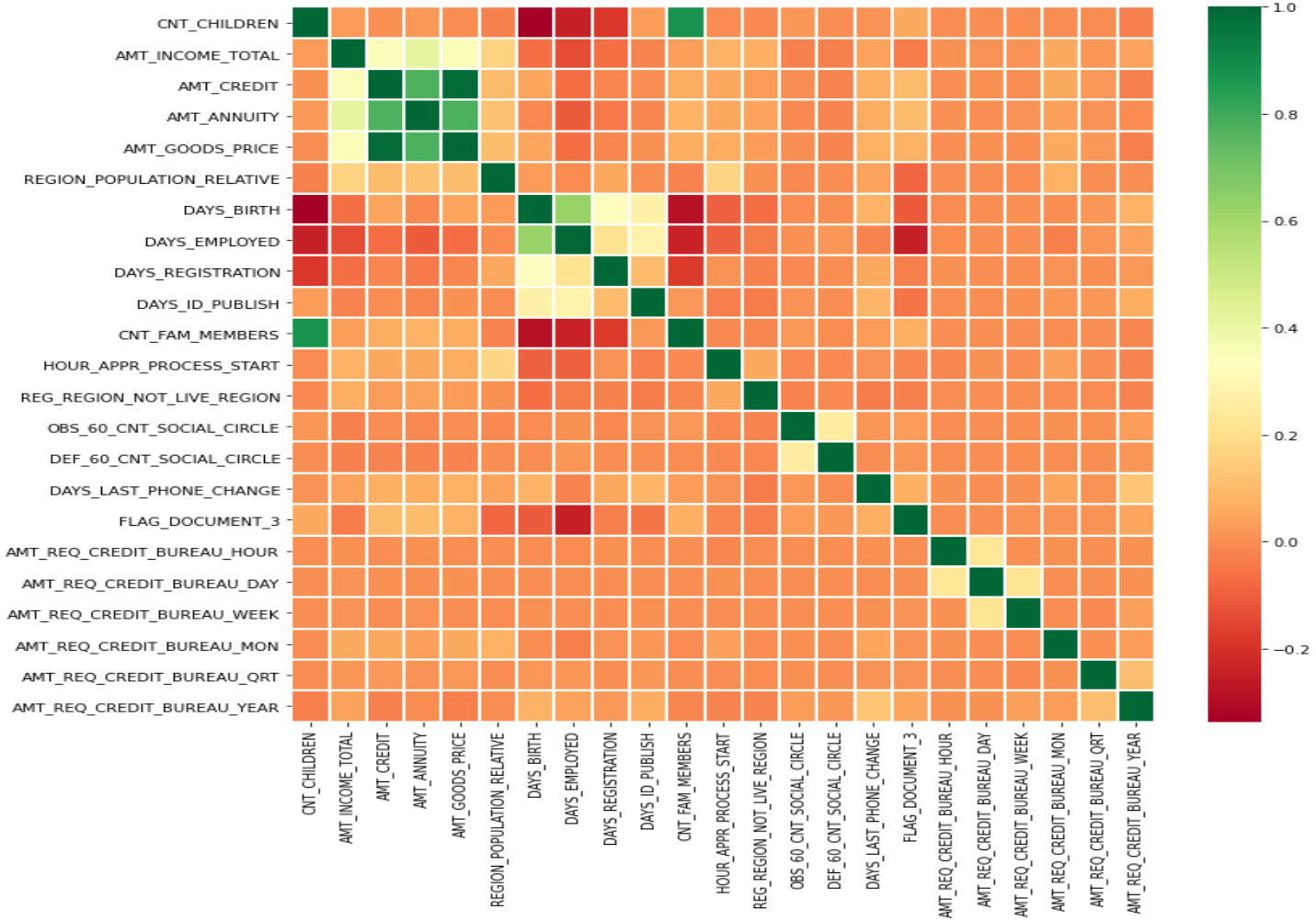
Credit amount is highly correlated with

amount of goods price

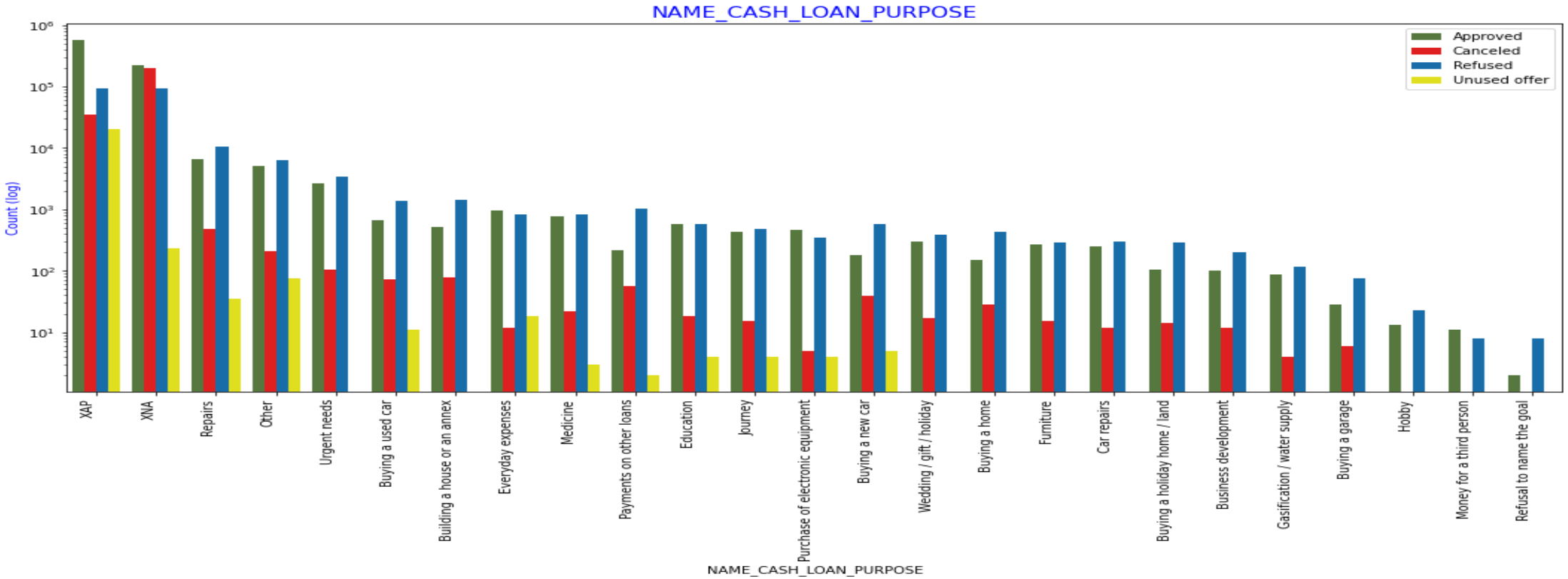
loan annuity

total income

We can also see that re-payers have high correlation in number of days employed.



NAME_CASH_LOAN_PURPOSE



Case Study

Summary

Decisive Factors for an applicant to be safe borrowers

Applicants with Income more than 700,000 are less likely to default

Applicants with 40+ year experience having less than 1% default rate

Applicants with zero to two children tend to repay the loans.

Applicants above age of 50 have low probability of defaulting.

Academic degree has less defaults.

Applicants with Trade Type 4 and 5 and Industry type 8 have defaulted less than 3%.

Decisive Factors for an applicant to be safe borrowers

Loans bought for Hobby, Buying garage are being repaid mostly.

Student have no defaults.

Applicants who live in areas with Region Rating 1 are safe borrowers.

Decisive Factors for an applicant to be a potential Defaulter

When the credit amount goes beyond 3M, there is an increase in defaulters.

Male applicants have relatively higher default rate

Applicants who have children equal to or more than 9 default 100% and hence their applications can to be rejected.

Applicants who have higher family members (>=11) have higher default rate and their applications can be rejected.

Avoid young applicants who are in age group of 20-40 as they have higher probability of defaulting

Applicants who have less than 5 years of employment have high default rate.

Decisive Factors for an applicant to be a potential Defaulter

Applicants with Lower Secondary education , ,incomplete education have higher default rate.

Applicants who are either at Maternity leave or Unemployed have higher default rate.

Applicants who live in areas with Region Rating as 3 has highest defaults.

Applicants in civil marriage or who are single have higher default rate

Low-skill Laborers, drivers and Waiters/barmen staff, Security staff have huge default rate.

Industry type 3 , type 13 and type 8 have high defaulting rate

Decisive Factors for an applicant to be a potential Defaulter

Applicants who get loan for 300-600 k tend to default more than others and hence having higher interest specifically for this credit range would be ideal.

Applicants with family members between 8 to 10 have a very high default rate and hence higher interest should be imposed on their loans.

Since 90% of the applications have Income total less than 300,000 and they have high probability of defaulting, they could be offered loan with higher interest compared to other income category.

Loan for house Repairs seems to have highest default rate. A very high number applications have been rejected by bank .

Applicants who have 4 to 8 children have a very high default rate and hence higher interest should be imposed on their loans.

People living in rented apartments & living with parents and hence offering the loan would mitigate the loss if any of those default

Thankyou !