



EDA – Bank Loan Applications

Problem Statement

- A Finance company wants to utilize its existing data on loan applications and derive insights from it to decide on whether to approve a loan for an applicant. They want to analyze the data with respect to two main features.
1. **The client with payment difficulties:** he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample
 2. **All other cases:** All other cases when the payment is paid on time
- They want to understand driving factors of the above 2 features and take loan approval decision accordingly.

Steps Involved in This Case Study

Understanding
Problem
Statement

Data
Understanding

Data Cleaning

Null Value
Imputation

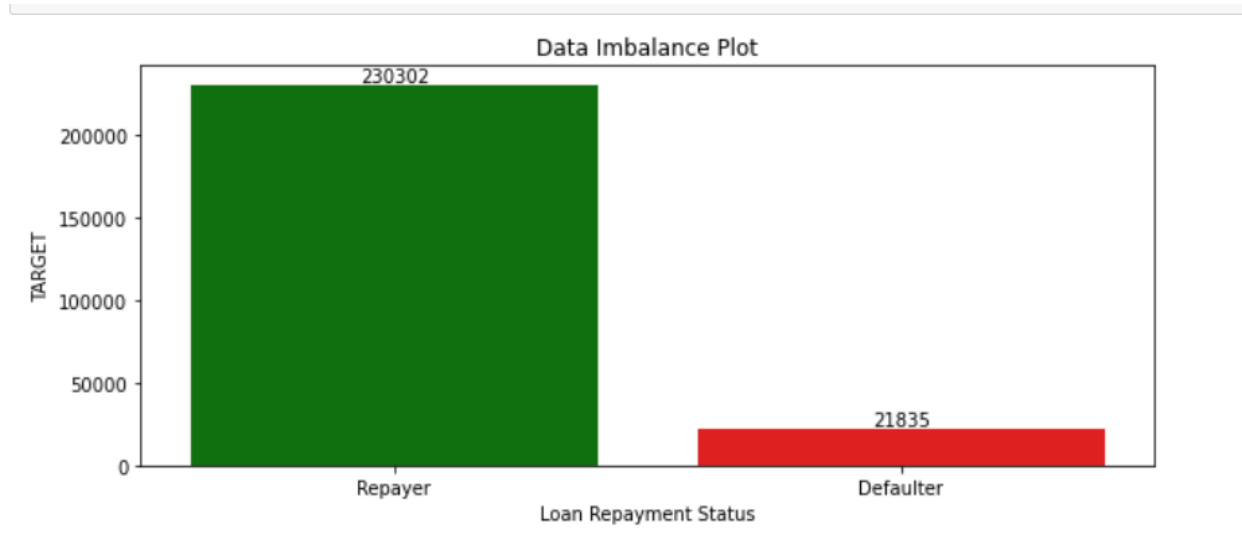
Outlier
Analysis

Univariate
Analysis

Bivariate /
Multivariate
Analysis

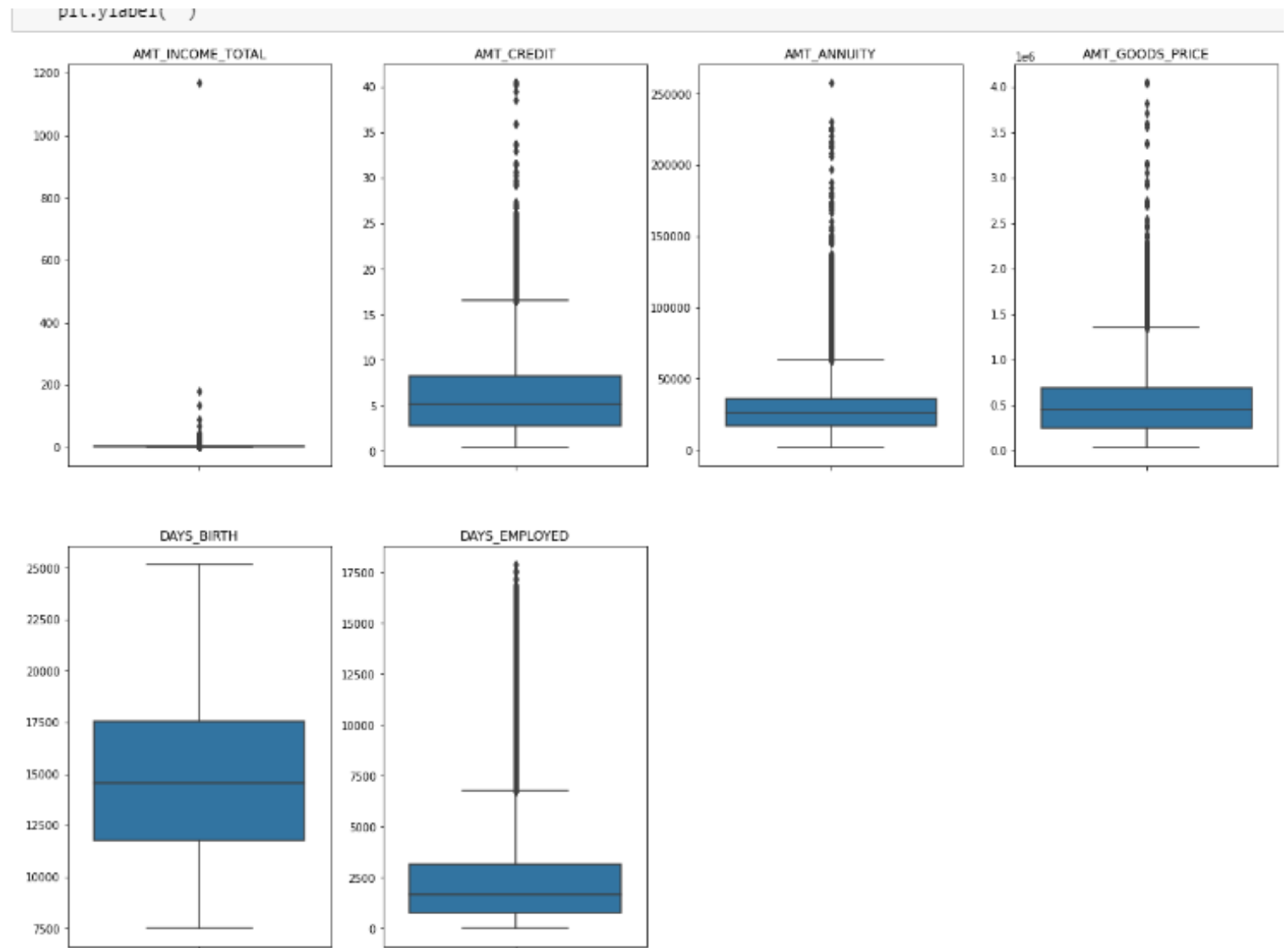
Data Imbalance

- **Target Variable**



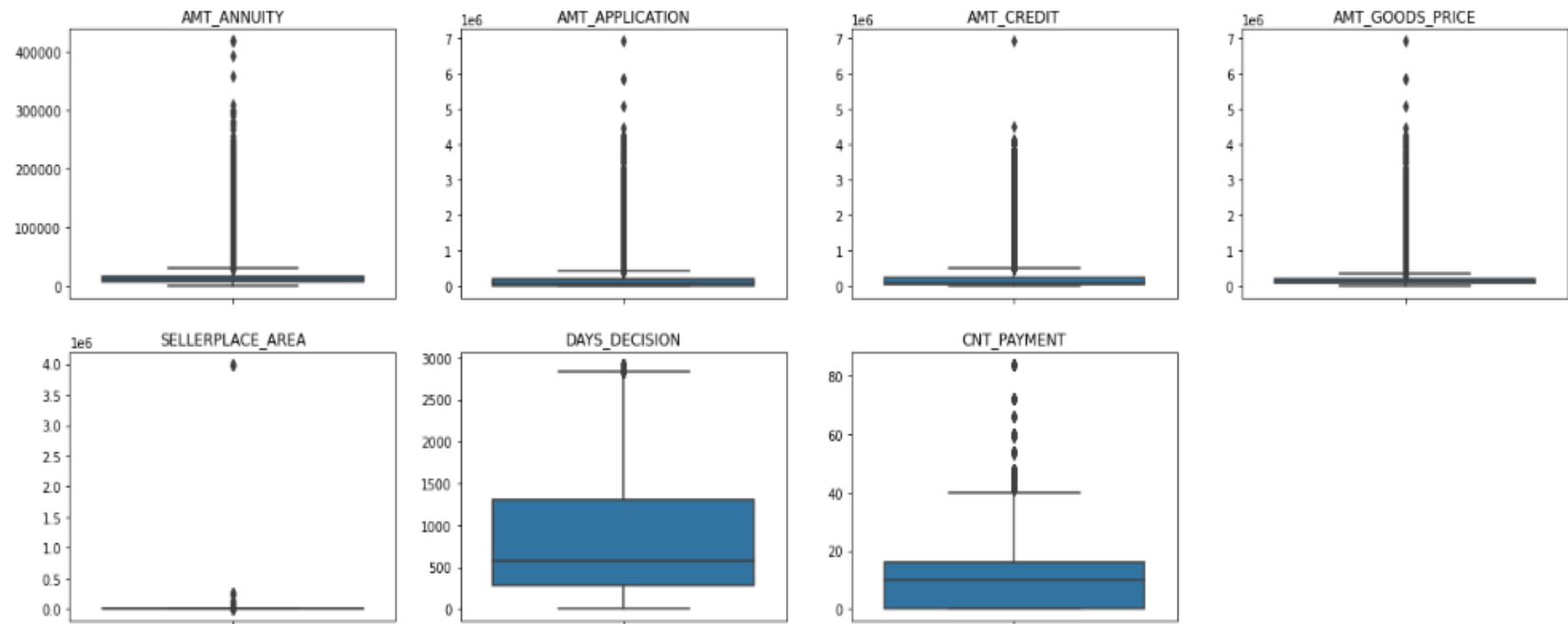
- Ratios of imbalance for Repayer and Defaulter in Percentage is: 91.3 and 8.6
- Ratios of imbalance for Repayer Vs Defaulter is: 10.55 :1 (approx.)

Outliers – Application Data



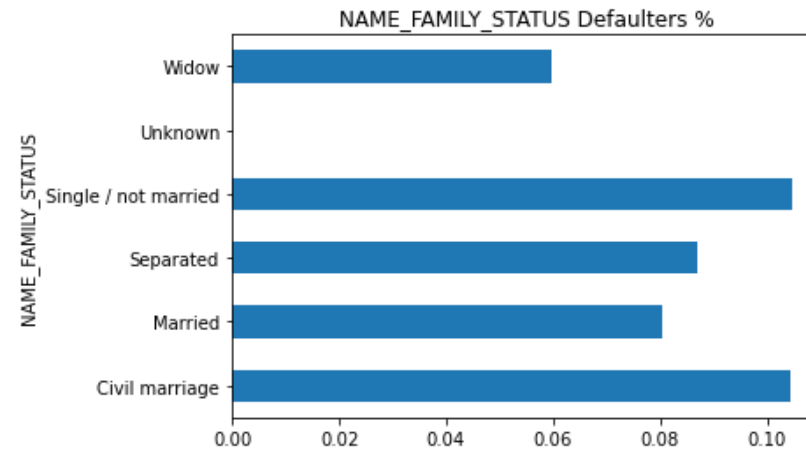
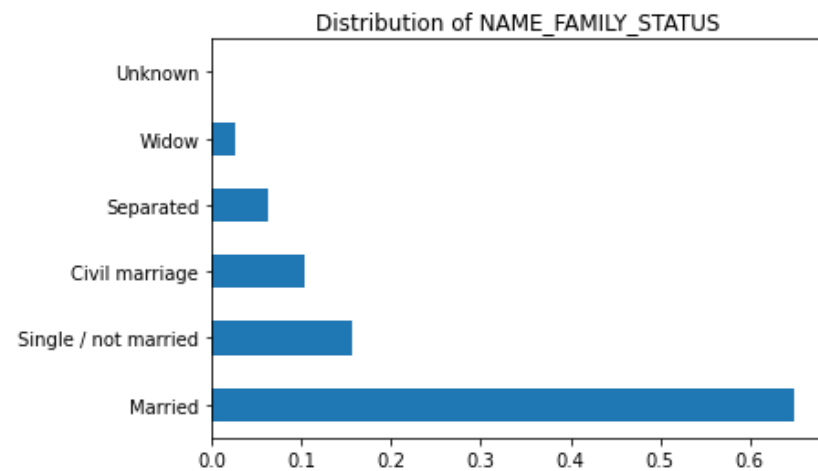
- AMT_CREDIT, AMT_ANNUITY, AMT_GOODS_PRICE, DAYS_EMPLOYED, have some number of outliers.
- AMT_INCOME_TOTAL has huge number of outliers which indicate that few of the loan applicants have higher income when compared to the others.

Outliers – Previous Application Data



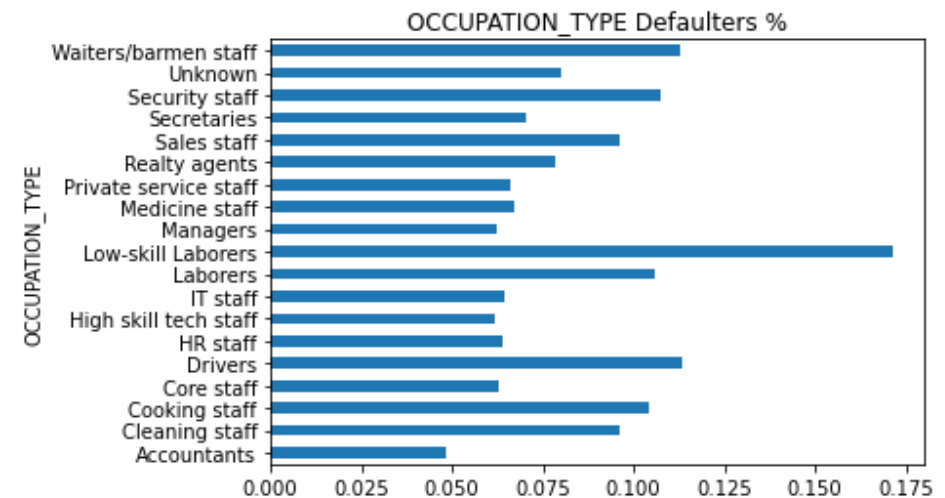
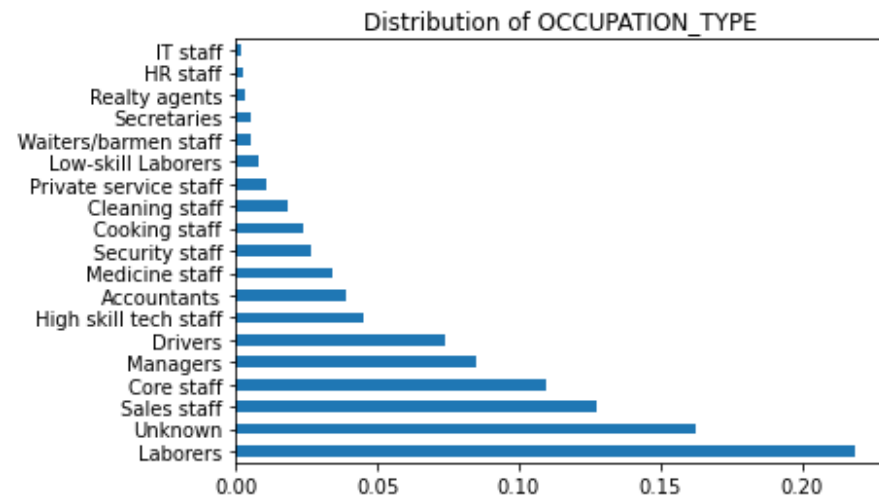
- AMT_ANNUIITY, AMT_APPLICATION, AMT_CREDIT, AMT_GOODS_PRICE, SELLERPLACE_AREA have higher number of outliers.
- CNT_PAYMENT has few outlier values.
- DAYS_DECISION has little number of outliers indicating that these previous applications decisions were taken long back.

Univariate Analysis - NAME_FAMILY_STATUS



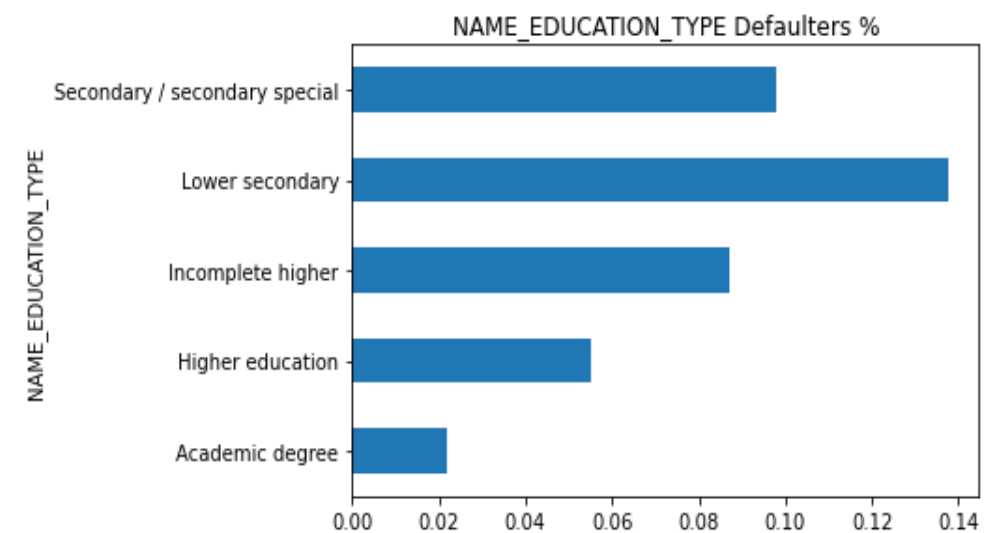
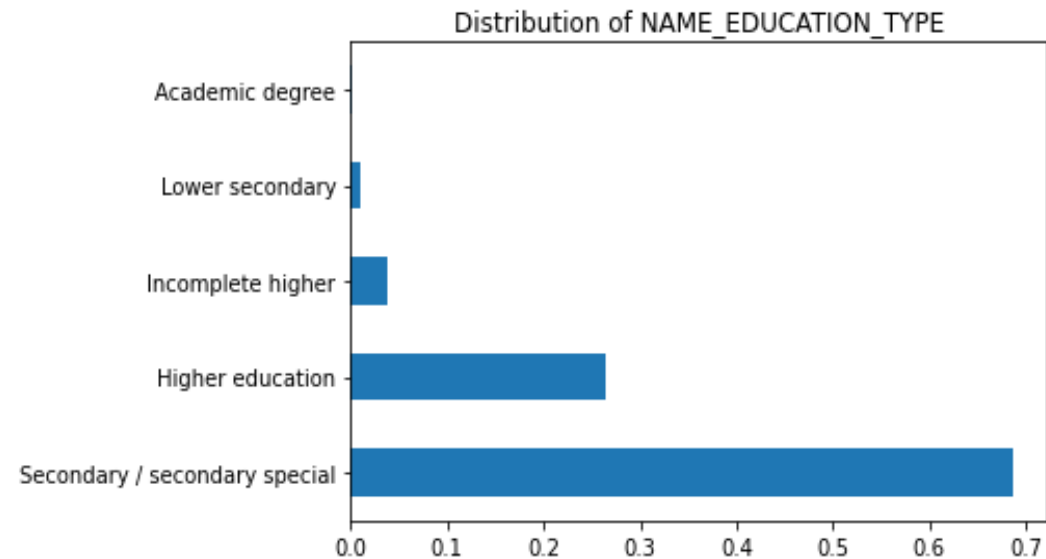
- Single / not married or Civil marriage applicants have higher default rate

OCCUPATION_TYPE



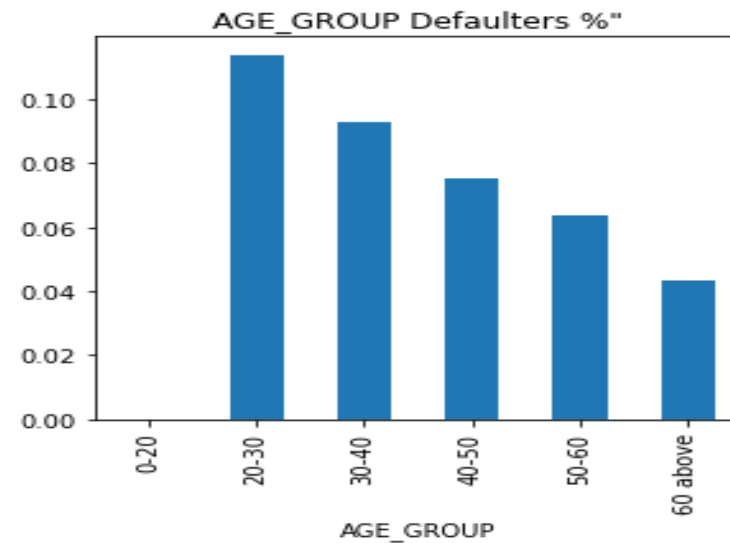
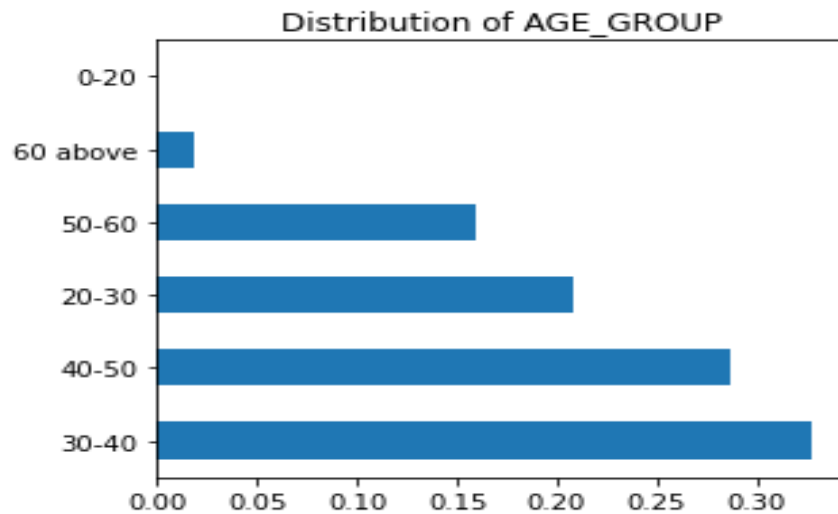
- Low skill Laborers applicants have higher default rate
- Accountants have very low default rate

NAME_EDUCATION_TYPE



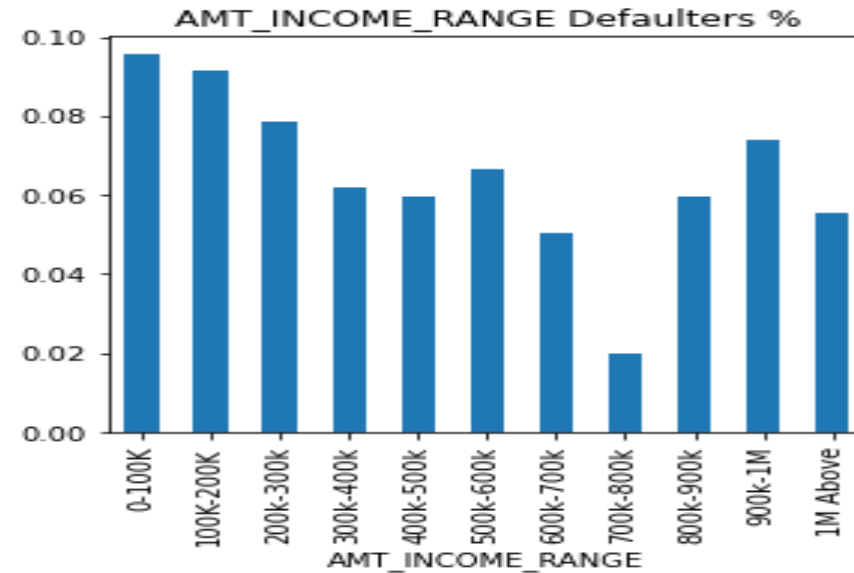
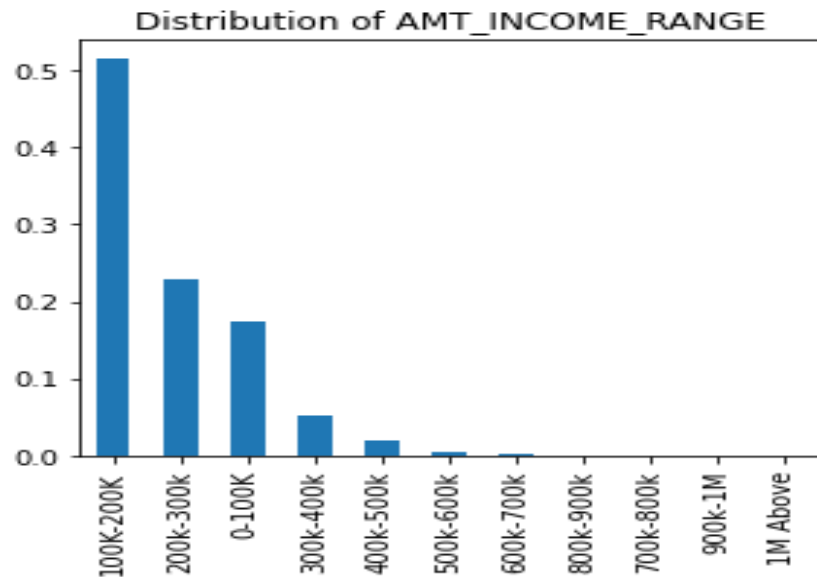
- Lower secondary applicants have higher default rate
- Academic degree applicants have lower default rate

AGE_GROUP



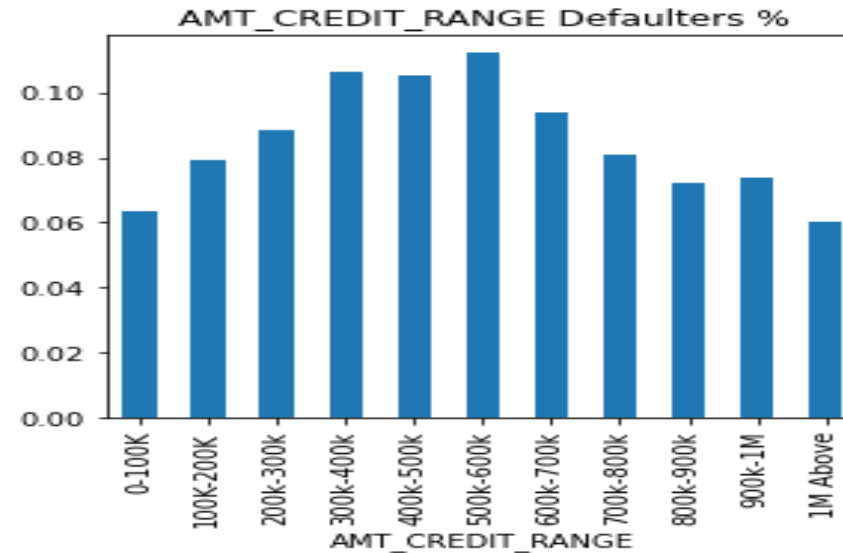
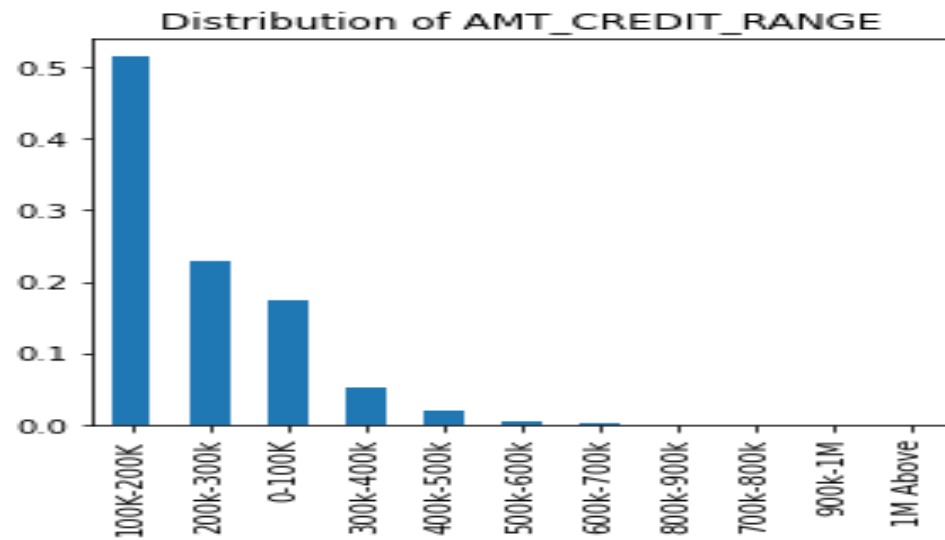
- 20-30 Age group applicants have higher default rate
- Above 60 Age group applicants have low default rate

AMT_INCOME_RANGE



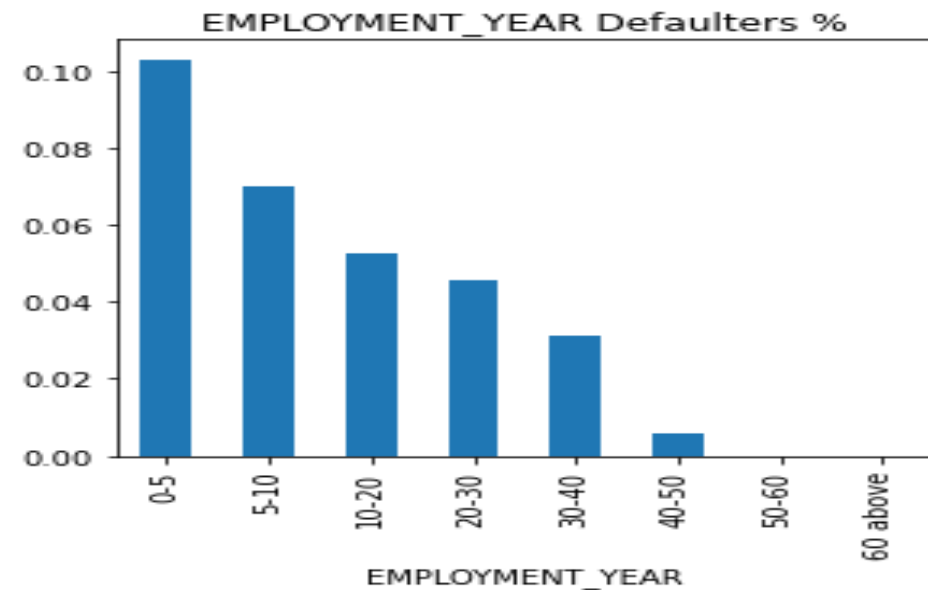
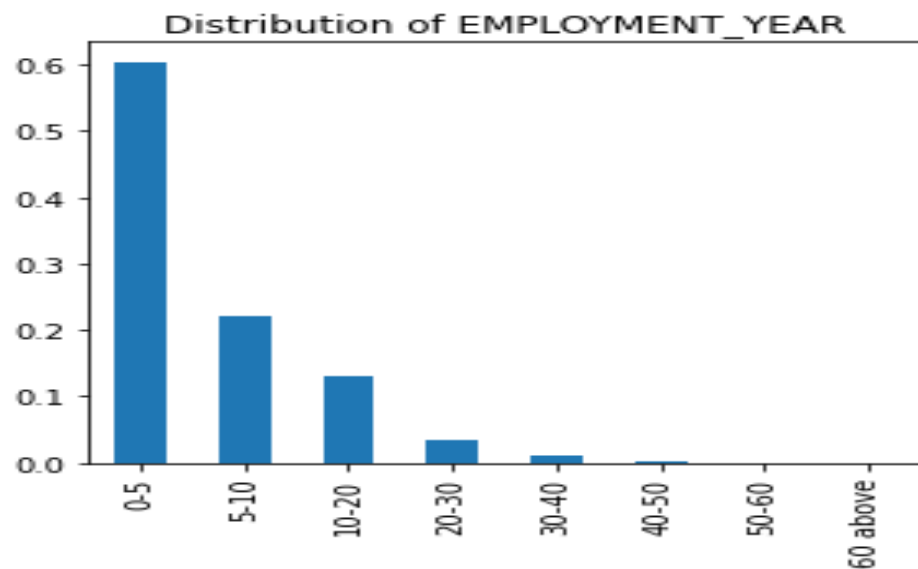
- 0-100k income range applicants have higher default rate
- 700k-800k income range applicants have low default rate

AMT_CREDIT_RANGE



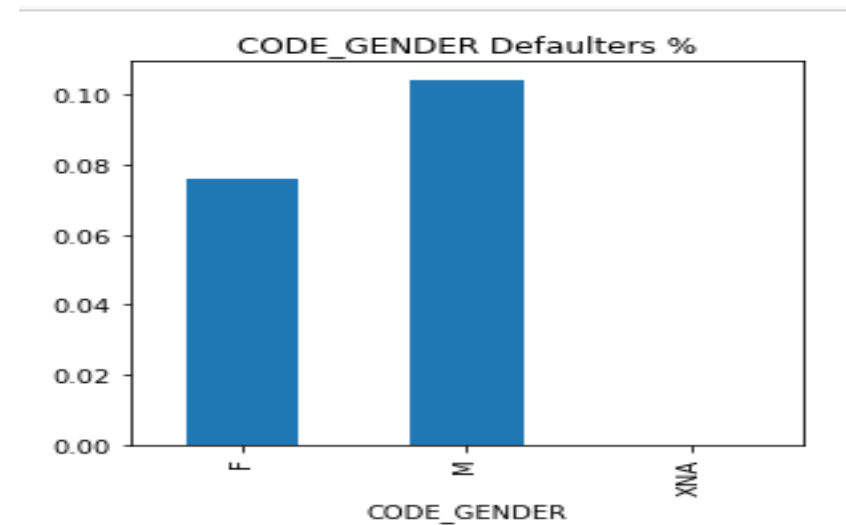
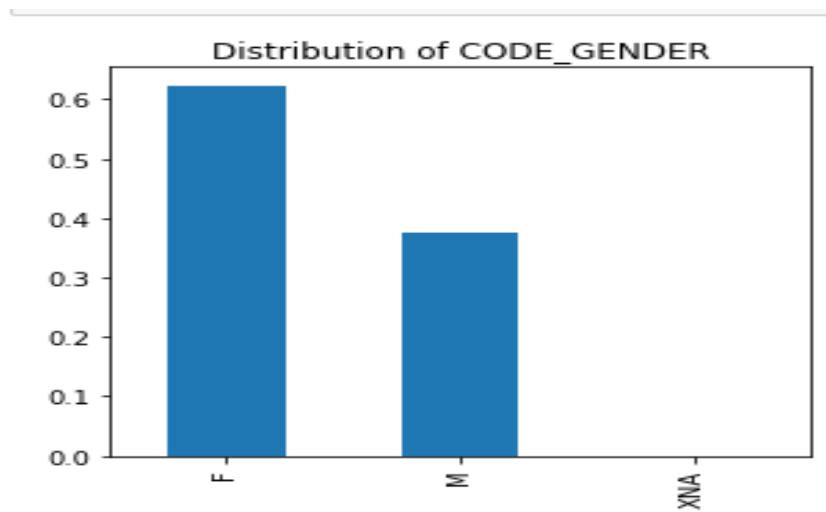
- 500k-600k Credit range applicants have higher default rate
- Above 1M Credit range applicants have low default rate

EMPLOYMENT_YEAR



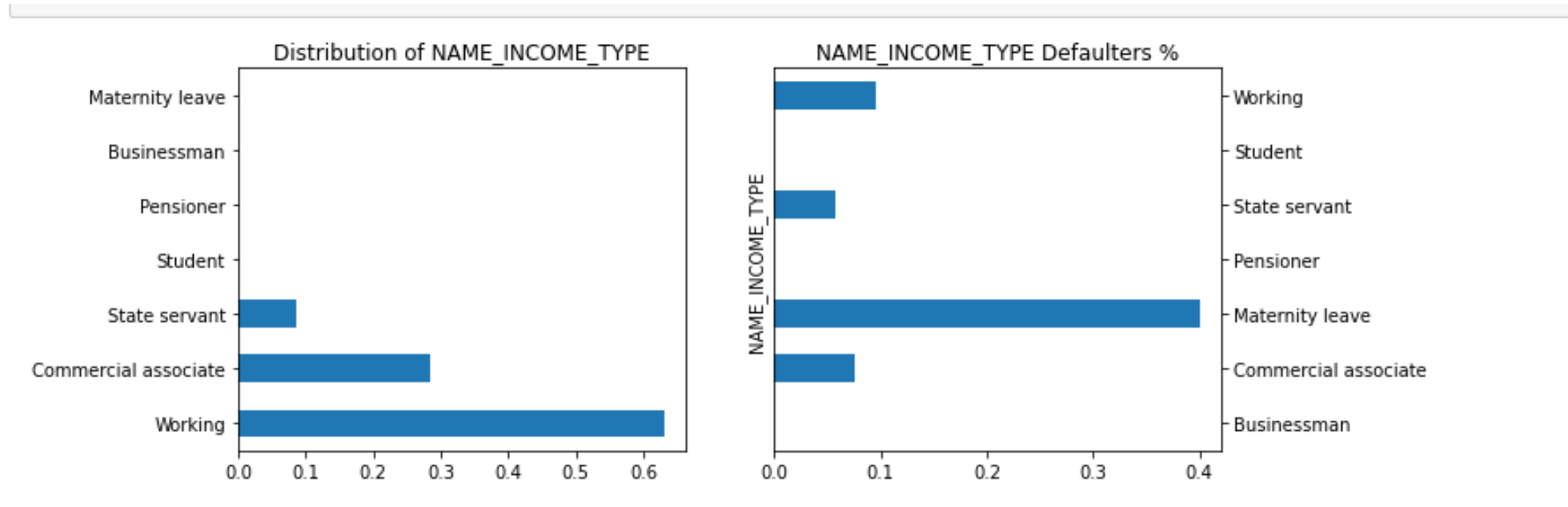
- We can observe that applicants with lower working years have higher default rate

CODE_GENDER



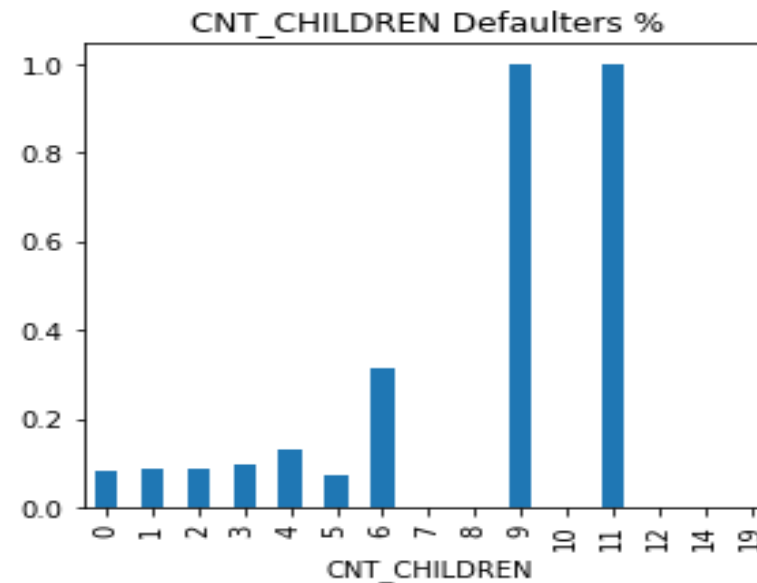
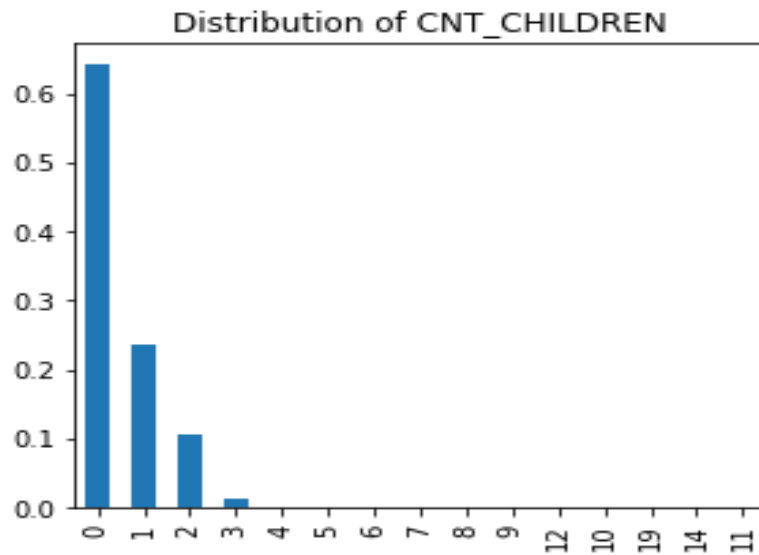
- We can observe that bank has more female applicants than male but more male defaulters than female.

NAME_INCOME_TYPE



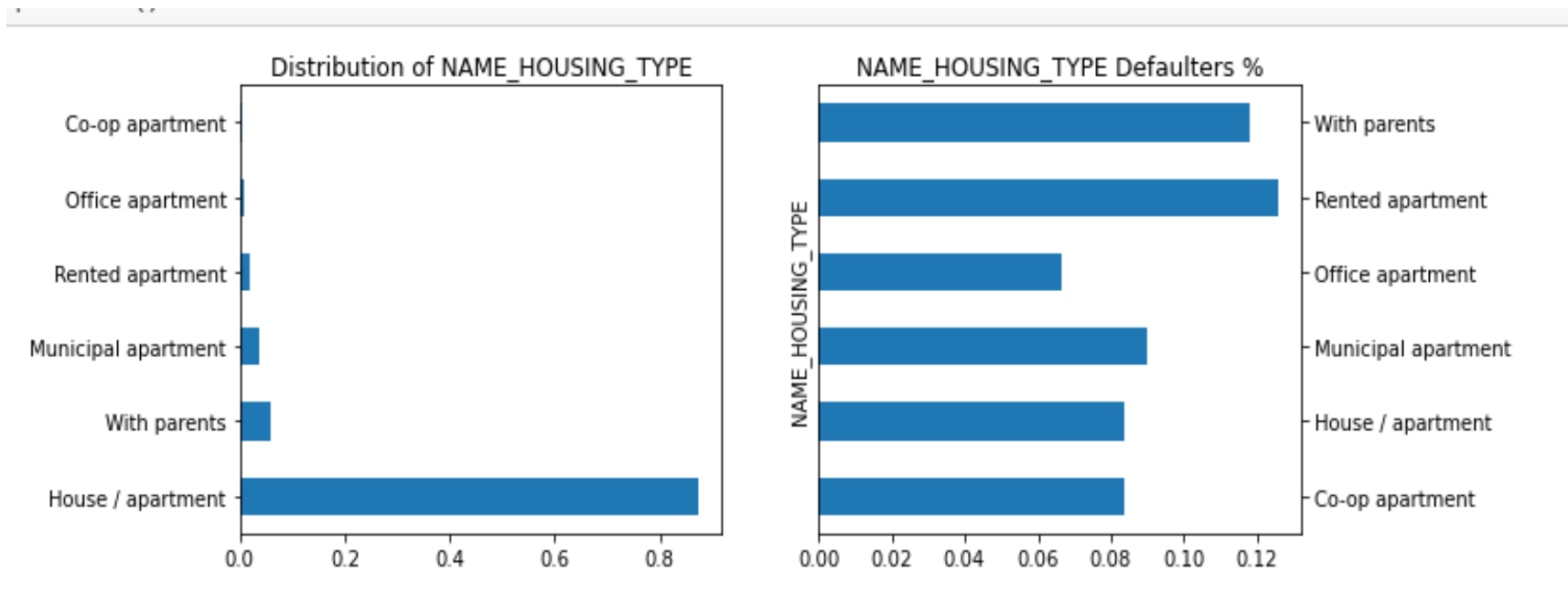
- We can observe that Maternity leave income type applicants are very low in number and very high in default rate

CNT_CHILDREN



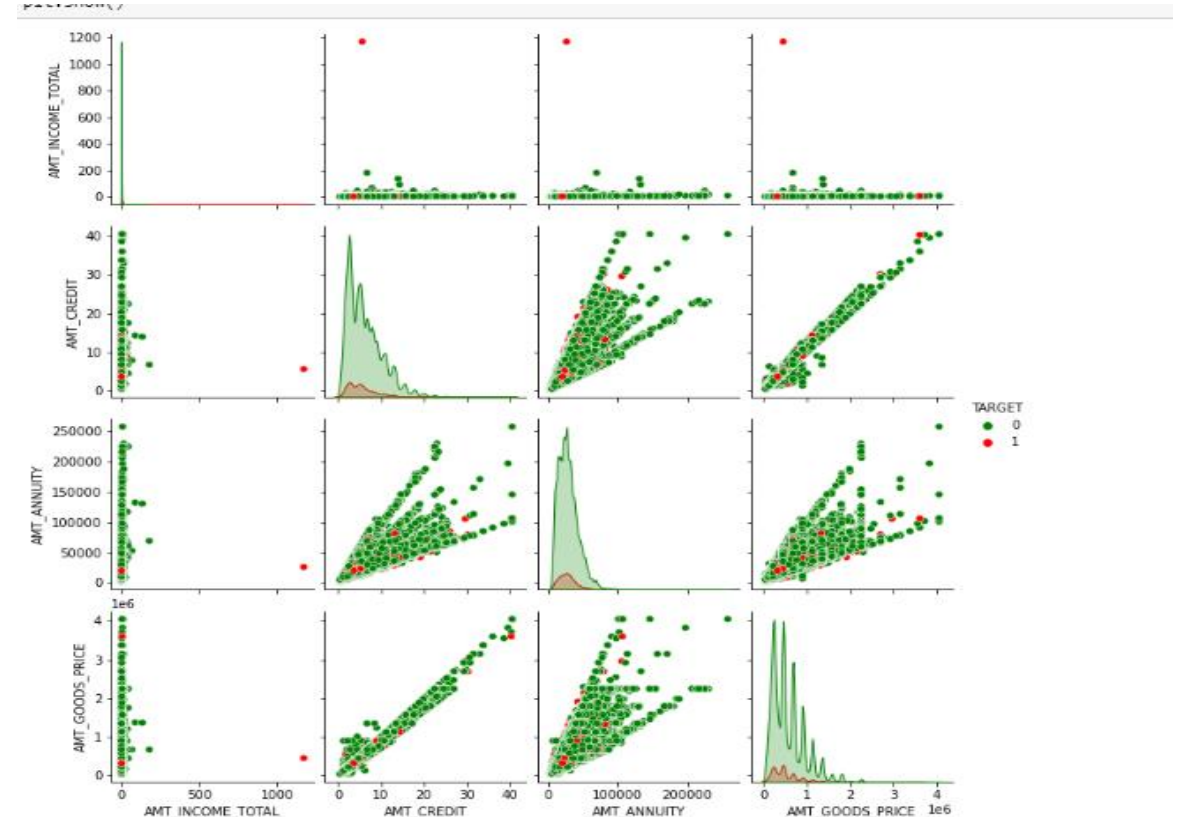
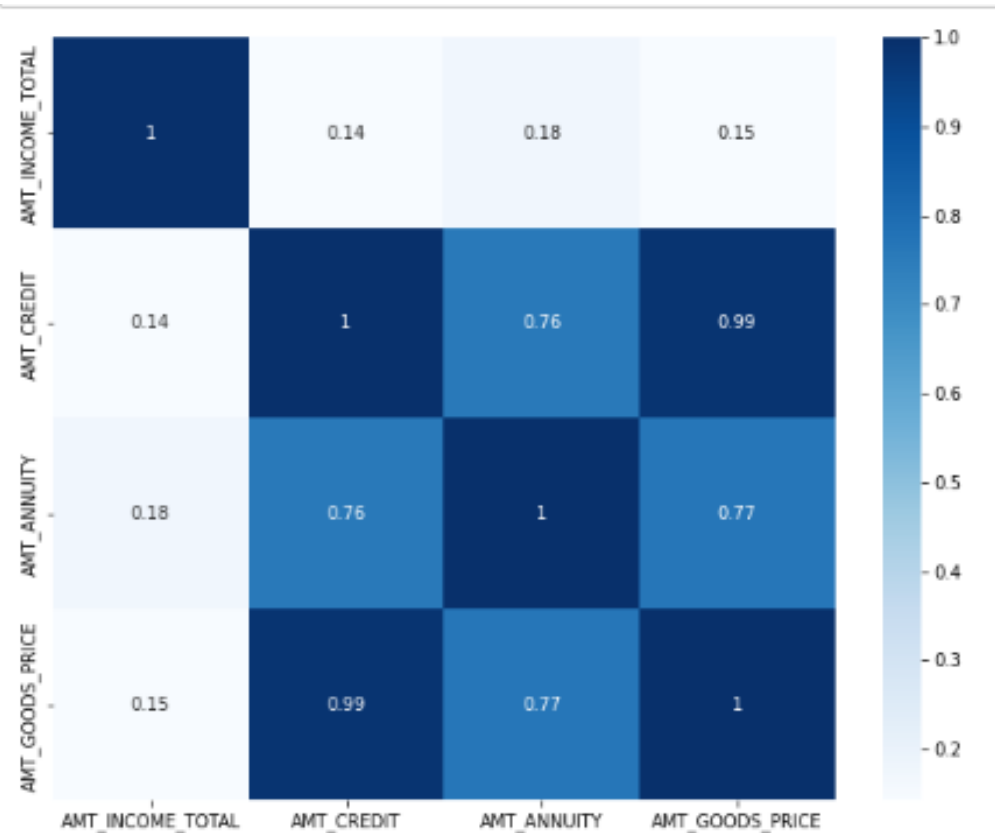
- We can observe that applicants with 0 children are very high in number and very low in default rate whereas applicants with 9,11 are very low in number and very high in default rate

NAME_HOUSING_TYPE



- We can observe that applicants with 0 children are very high in number and very low in default rate whereas applicants with 9,11 are very low in number and very high in default rate

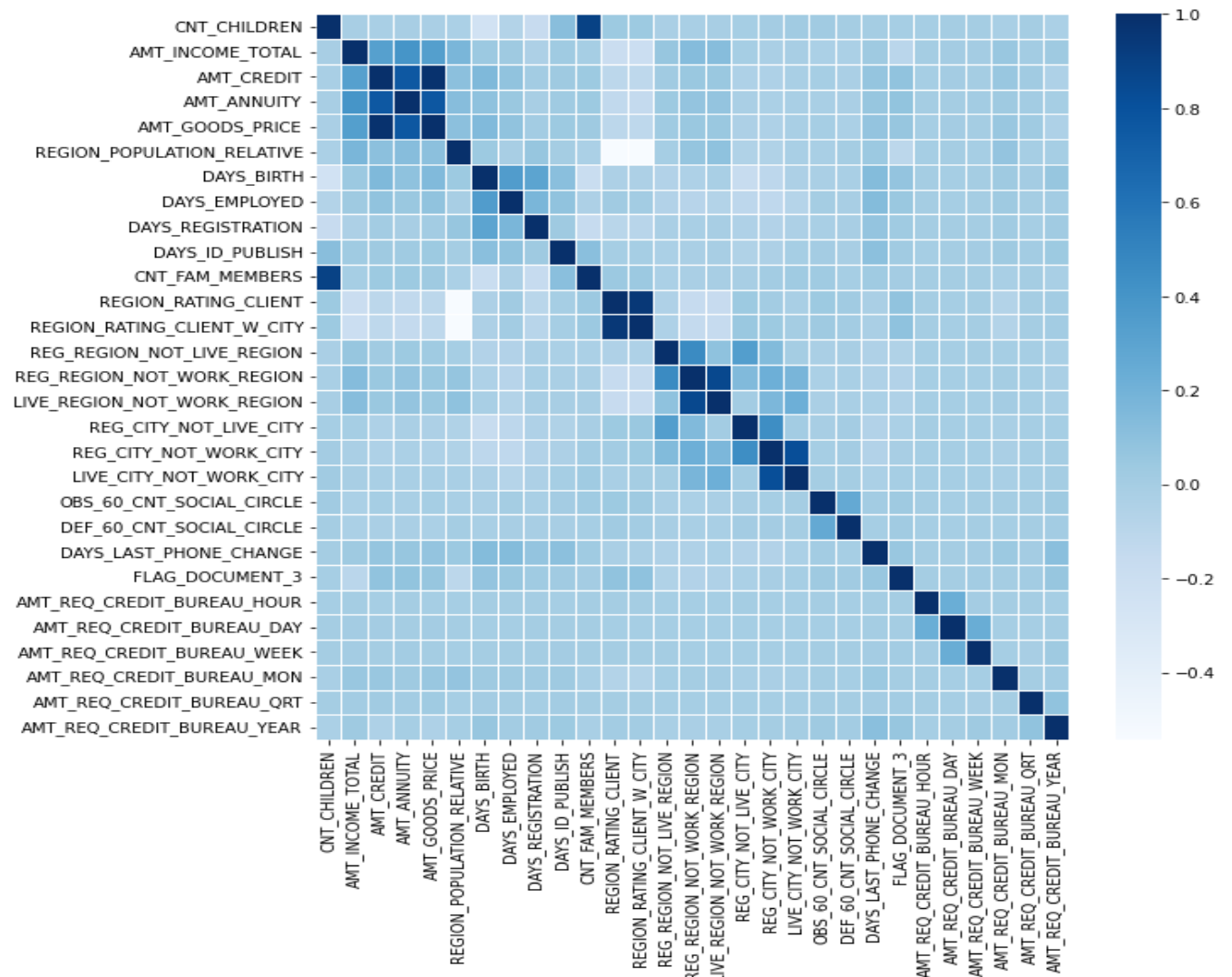
Multivariate & Bivariate Analysis – Amount variables



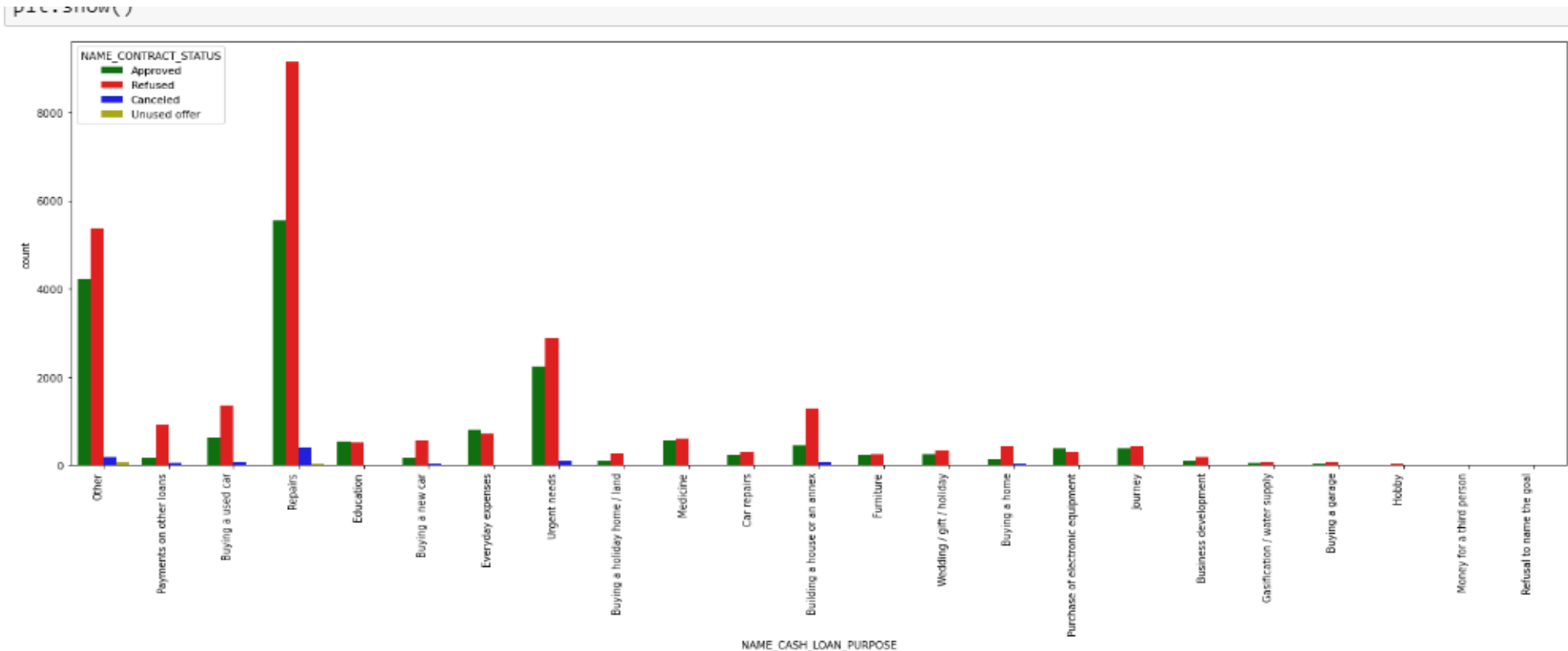
- We can see high correlation between AMT_CREDIT and AMT_GOODS_PRICE -Applicants owning goods of high value can take loans of higher amounts.

Multivariate & Bivariate Analysis – Numeric columns

*Re-payer correlation factors-
Credit amount is highly correlated with
loan annuity
amount of goods price
total income*



NAME_CASH_LOAN_PURPOSE(re-payers)



Applicants who has taken loan repairs tend to repay more.

Summary:

Driving factors for an applicant to be a re-payer

Applicants who are Accountants are less likely to default.

Applicants who have Academic degree are less likely to default.

Applicants with 40+ years of work experience are less likely to default.

Applicants with above 60+ age years are less likely to default.

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

Student have no defaults.

Applicants with Trade Type 4 and 5 and Industry type 8 are less likely to default.

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

Driving factors for an applicant to be a defaulter

Applicants who are skill Laborers are more likely to default.

Applicants who have Lower secondary education are more likely to default

Applicants who have higher count of children(≥ 9) have higher default rate and their applications can be rejected.

Avoid approving young applicant loans who are in age group of 20-40 as they are more likely to default.

Male applicants are more likely to default

Applicants who have less than 5 years of employment are more likely to default.

Applicants who get loan for 300-600 k are more likely to default than others.

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



Thank you