CREDIT EDA LOAN ANALYSIS:

BY

CHANDANA SARANGI

BUSINESS OBJECTIVE:

- This case study strives to ensure the highest possible turnout for the banks in case of processing of loans by assessing the risk to prevent the business loss.
- To improve the quality of Advances
- To maximize cross selling of products such as loans, Insurance and various deposits.

PROBLEM STATEMENT:

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

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

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

The data given below contains the information about the loan application at the time of applying for the loan. Two scenarios:

- The applicant with payment difficulties: Male/Female having late payment more than inordinate amount of days on atleast one of the first year installments of the loan.

- Other Cases when the payment is paid on time.

Using EDA we can analyse the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

EDA Analysis steps:

1. Understanding the data

- Reading
- Checking Datatype
- Data Imbalance

2. Data Cleaning and Manipulation

- Missing Values
- Imputation
- Dropping

3. Data Analysis

- Univariate Analysis
- Bivariate Analysis
- Correlations

4. Conclusions

Datasets:

1. **Application_data.csv :** contains all the information of the client at the time of application

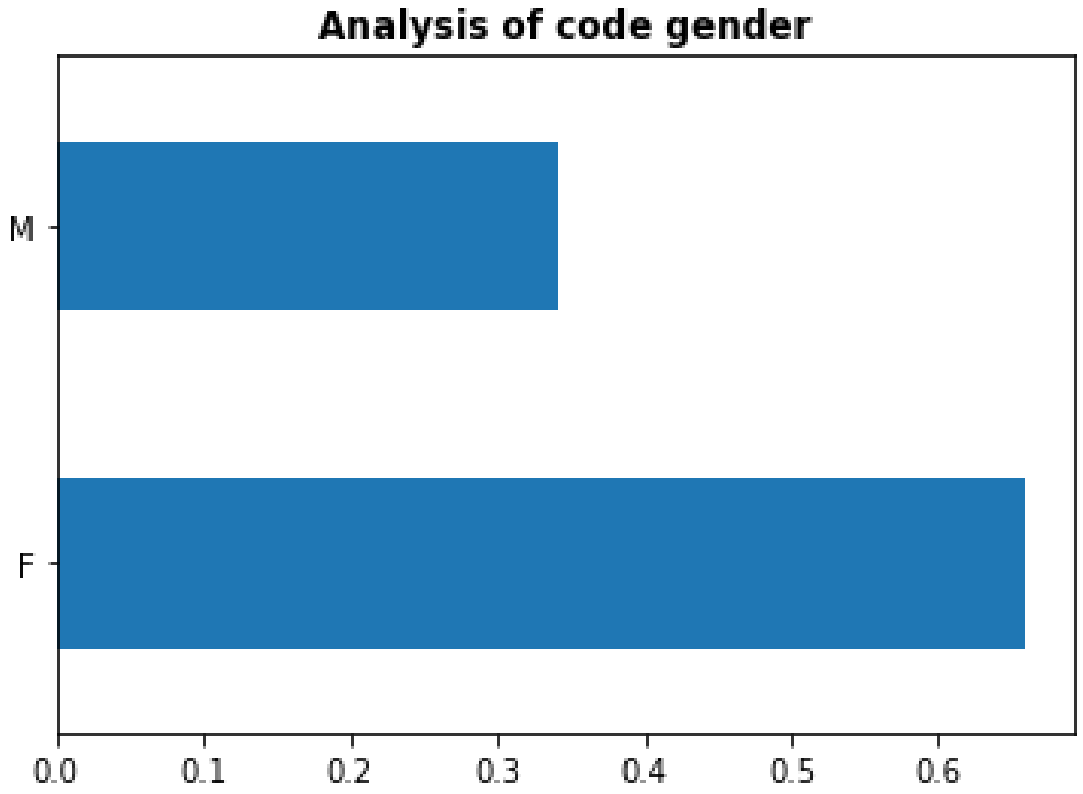
2. **Previous_application.csv:** contains information about the clients previous loan. It contains data whether the application had been Approved, Canceled, Refused or Unused Offer.

3.**Coulmns description.csv:** is the data dictionary which describes the meaning of variables.

UNIVARIATE ANALYSIS

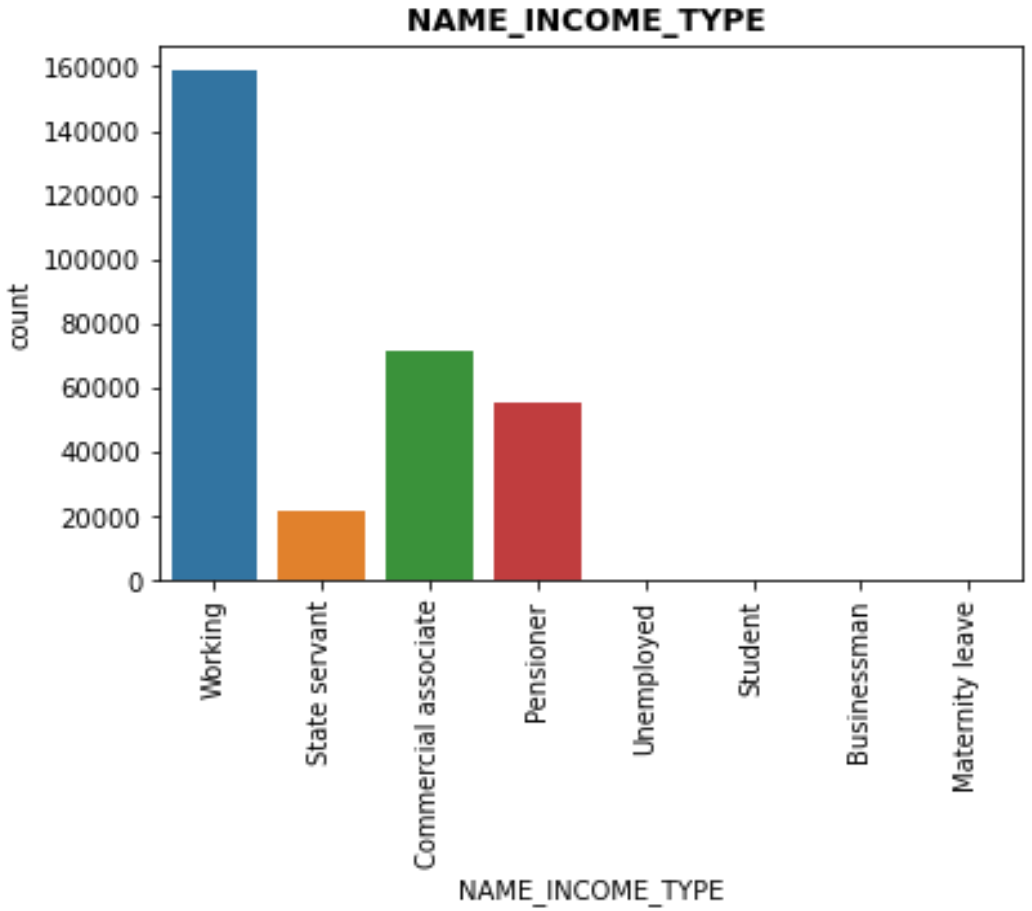
Insights:

Female have taken more loans in comparison to male counterparts.



UNIVARIATE ANALYSIS

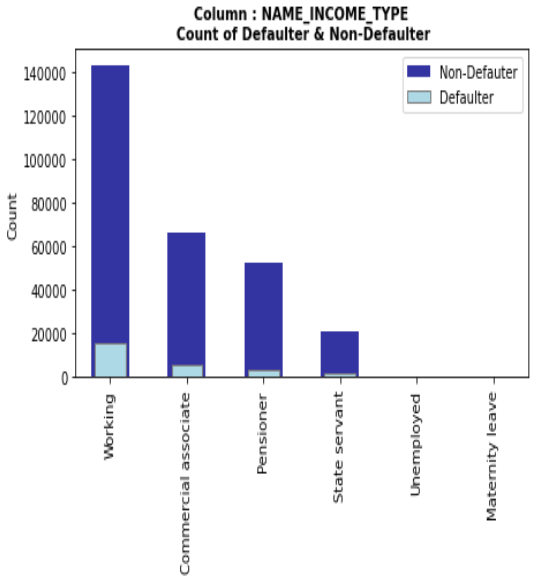
- Working, Commercial associate, Pensioner have taken most loans.
- Unemployed, students, Businessman, Maternity leave have taken fewer loans.

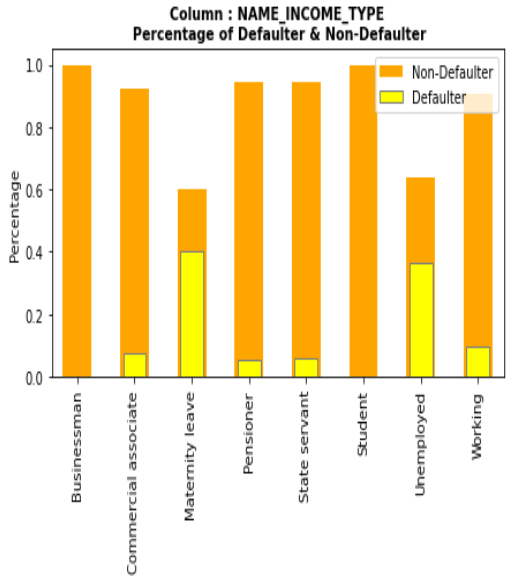


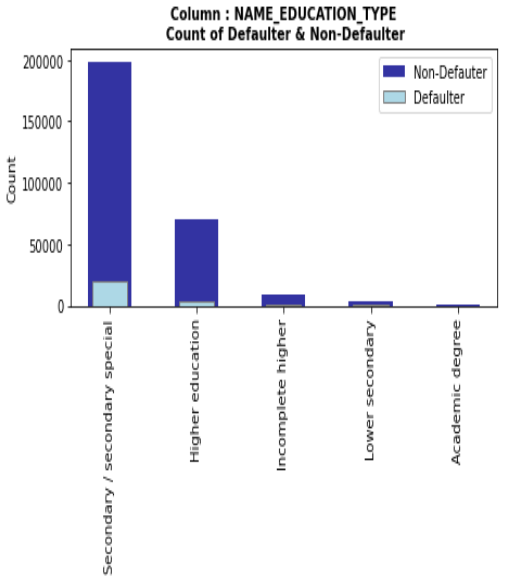
Observations on Univariate analysis:

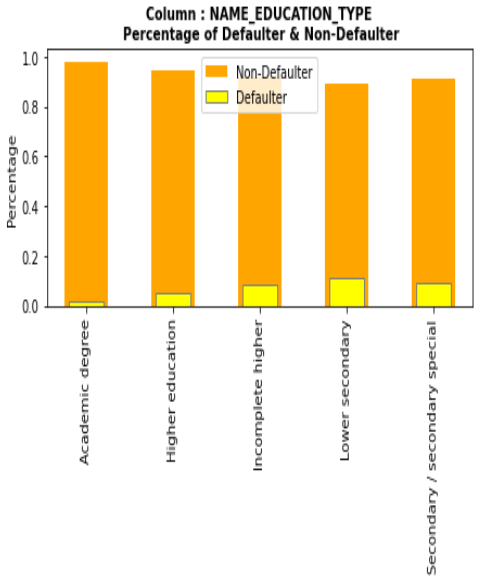
- Analysis on CODE GENDER entails that when Women borrow more loans than men.
- Similarly, in case of Education type, people with secondary/special education are most likely to apply for the loan
- People tend to take out more cash loans than revolving loans.
- It is seen that People who own a house or apartment have applied for loans
- as compared to others, married people have taken more loans.
- Also it has been observed that people who do not own a car take out more loans.

Categorical Univariate Analysis



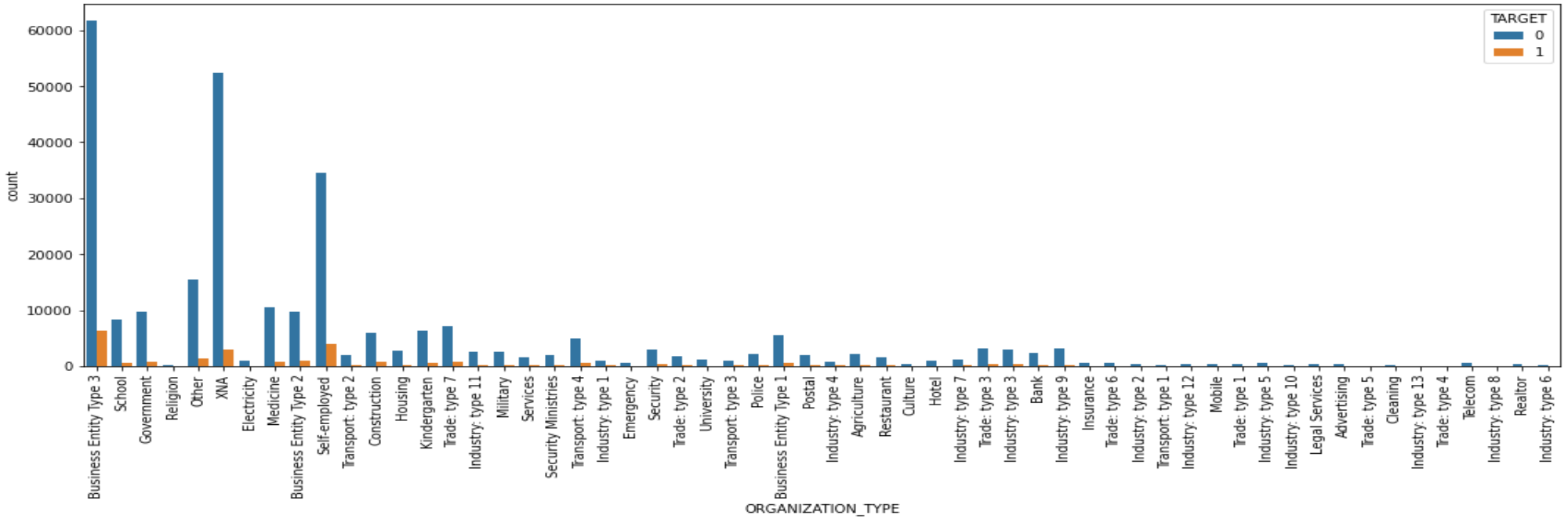






- The percentage of defaulter In case higher education is less as compared to others.
- The percentage of defaulters in case of people who are on Maternity Leave is higher than others.

Categorical Univariate Analysis

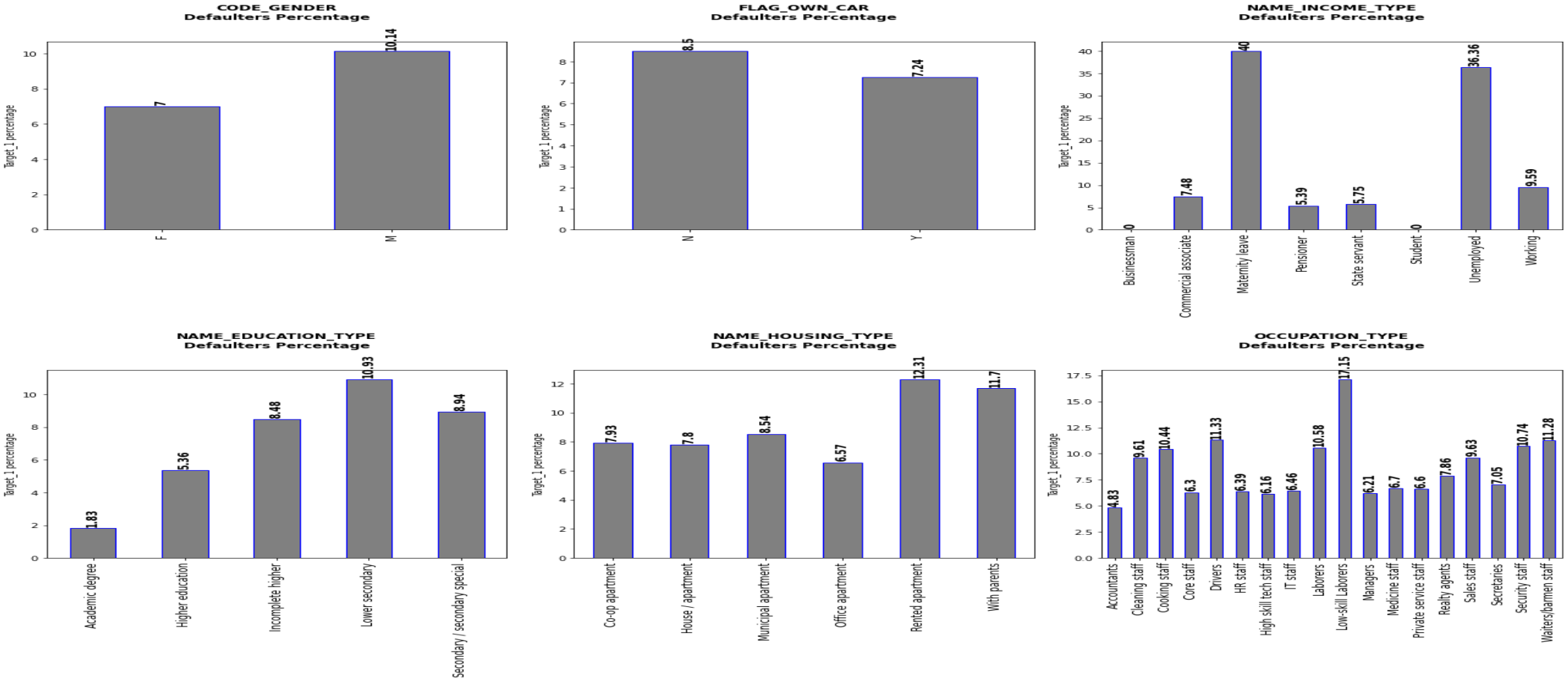


- Organisation type such as Business Type 3, XNA and Self employed have got more approved loans as they are less likely to default.

Observation on Numerical univariate analysis:

- DAYS_BIRTH - People aged 10000 days (27 years) to 15000 days (41 years) are more likely to take out loans and have a higher rate of default.
- DAYS_EMPLOYED - People who have recently been employed prior to applying for a loan are more likely to default.
- DAYS_ID_PUBLISH -People whose IDs were published between 4000-5000 days take out more loans and have a higher default rate.
- DAYS_LAST_PHONE_CHANGE - People who have changed their phone number within the last 200 days borrow more and have a higher default rate.
- AMT_INCOME_TOTAL - Low-income individuals are more likely to default.
- AMT_ANNUITY - Annuity with a low payout has a large number of loans and are more likely to default.
- AMT_GOODS_PRICE - People Having Low Price of goods default the most.
- AMT_CREDIT - The people who take less amount of loans are more likely to default.

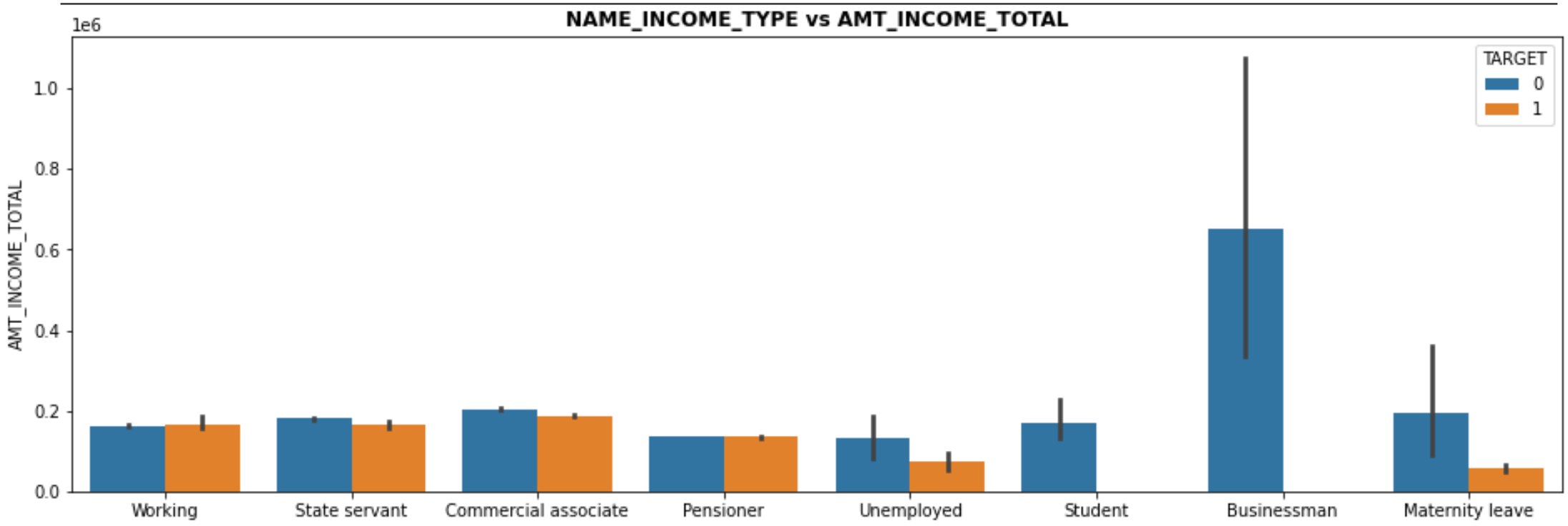
Bivariate Analysis:



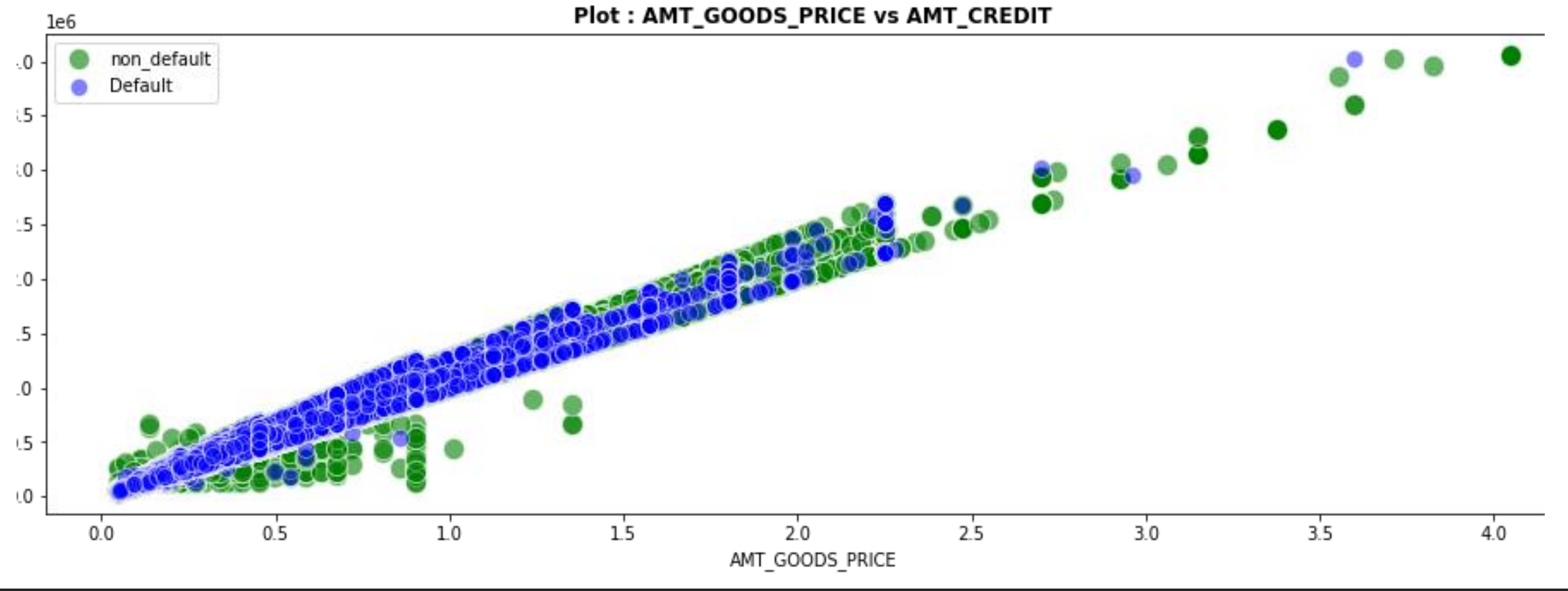
Observation on Bivariate Analysis:

- Males are having highest Default rate than Females.
- People who do not own car are having very high default rate than people who own car.
- People from the categories like Unemployed, Maternity leave & Working are having highest default rate in comparison to others(i.e. commercial associate, pensioner & state servant).
- Less educated people are having high default rate than highly eduacted people.
- People who are leaving in rented apartments and who leaves with their parents are being noticed as the highest defaluters in comparison to others.
- it is being observed that Low-skill labourers, waiters/barmen staffs,Drivers, labourers are the highest defaulters.

Bivariate Analysis:

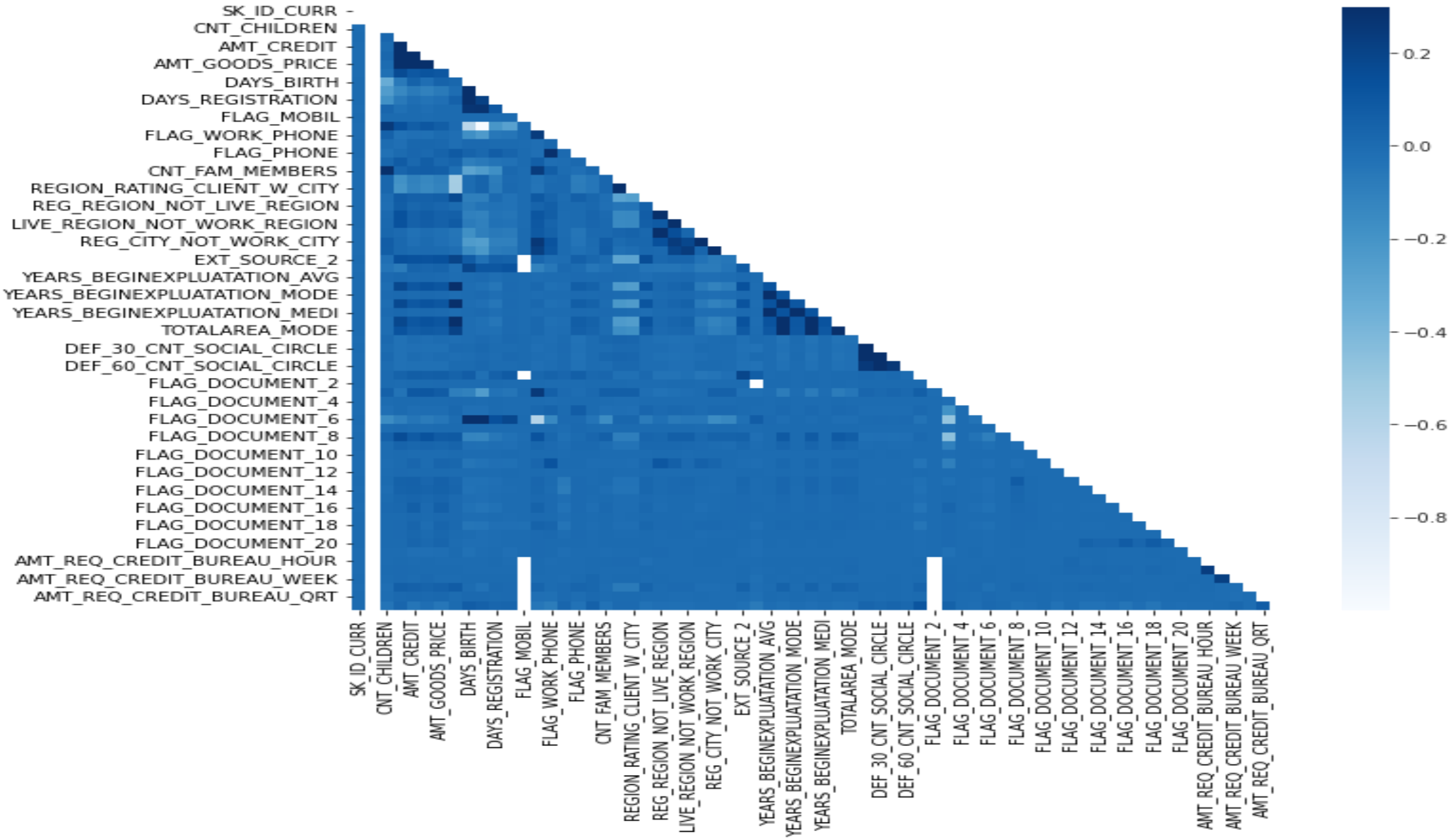


- People with higher income such as businessman are having 0% default rate. Likewise people with low income range less than or equal to 2 lakh are falling either in default or non-default category.

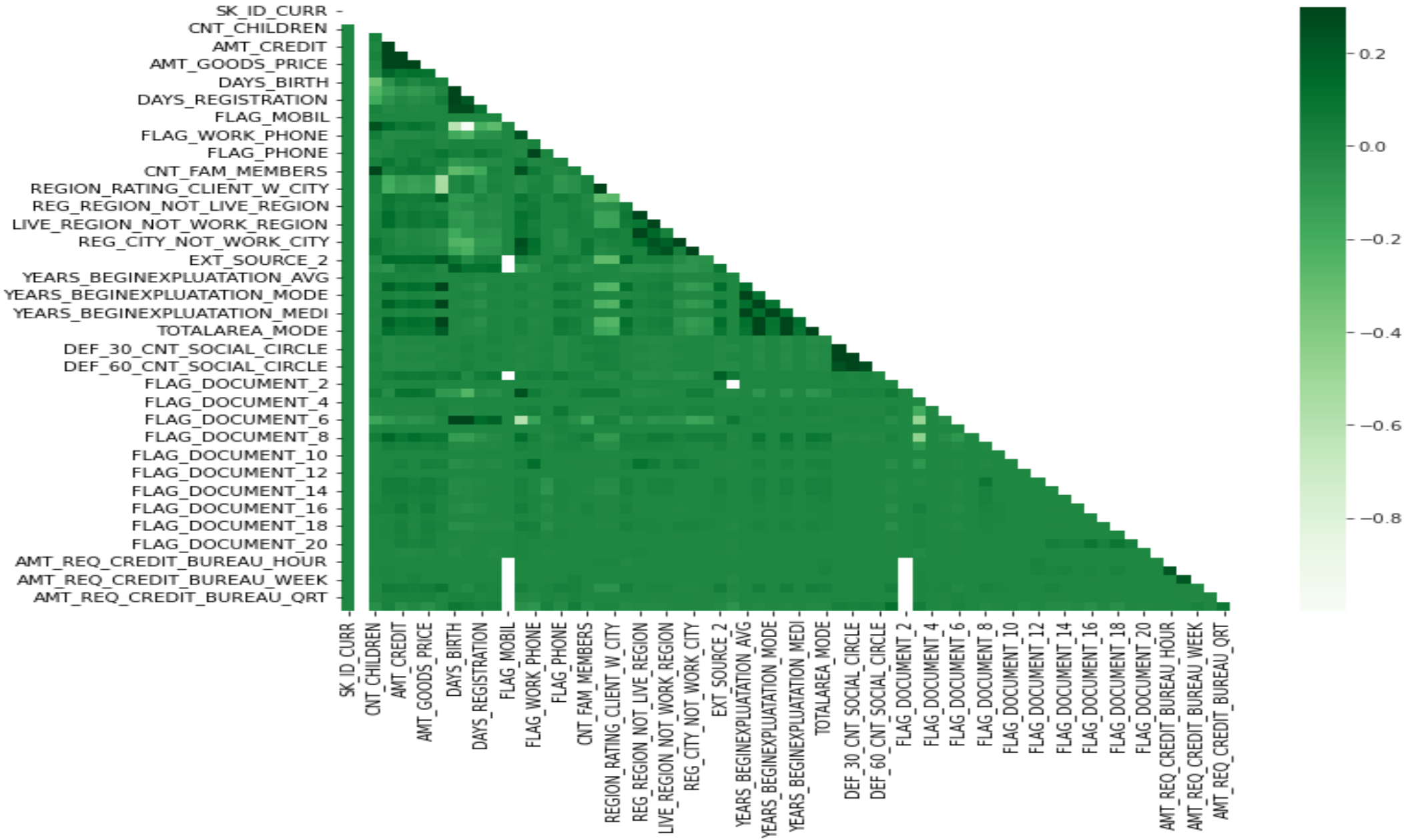


Numerical Bivariate Analysis:

- people with low AMT_CREDIT and AMT_GOODS_PRICE have higher default rate
- people with high AMT_CREDIT and AMT_GOODS_PRICE have lower default rate.



Correlation of Defaulters



Correlation of Non-defaulters:

Conclusions:

- From the analysis which we have done for the two data sets, it is being observed that there are certain criteria which needs to be fulfilled for getting the loan approved by the bank.
- Certian parameters to determine whether an applicant will be a Repayer are: the default rate on consumer loans and revolving loans, CODE_GENDER Females default less and age group .
- Conclusive factors to determine the defaulters are: the percentage of default rate in education sector, income range and previous loans which were approved or rejected etc.

THANK YOU

Submitted by
Chandana Sarangi