

EDA Case Study

Credit Risk Analysis

Amit Mishra

Problem Statement

There are two different types of risk associated with any loan application

- 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. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

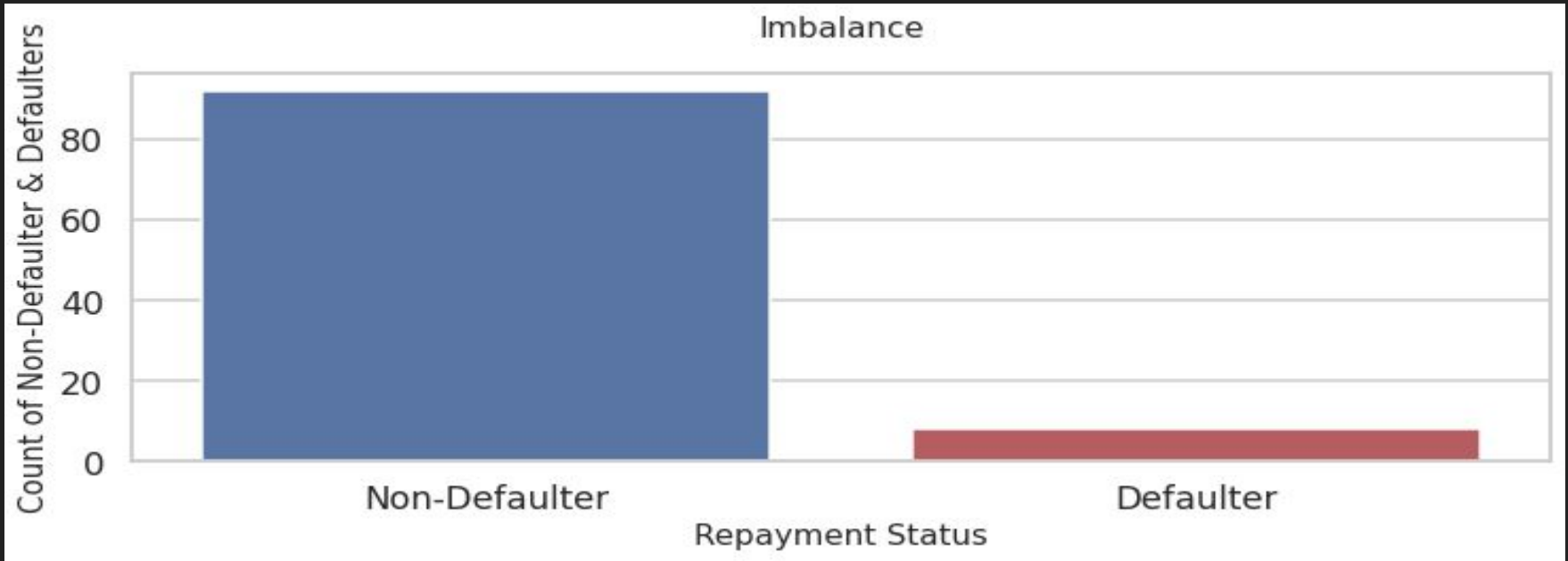
Tasks performed

- Understanding the Data
 - Carefully analysed all the columns and their relevance, to make meaningful insights
- Identifying and handling missing values
 - Used techniques like Mean and Median for continuous numerical columns
 - Used Mode for categorical column
 - Created common methods to reuse for the data cleanup activity
 - Dropped columns with high missing percentage
- Identifying and handling outliers
 - Used techniques like Boxplot to identify outliers in continuous numerical column

Tasks performed

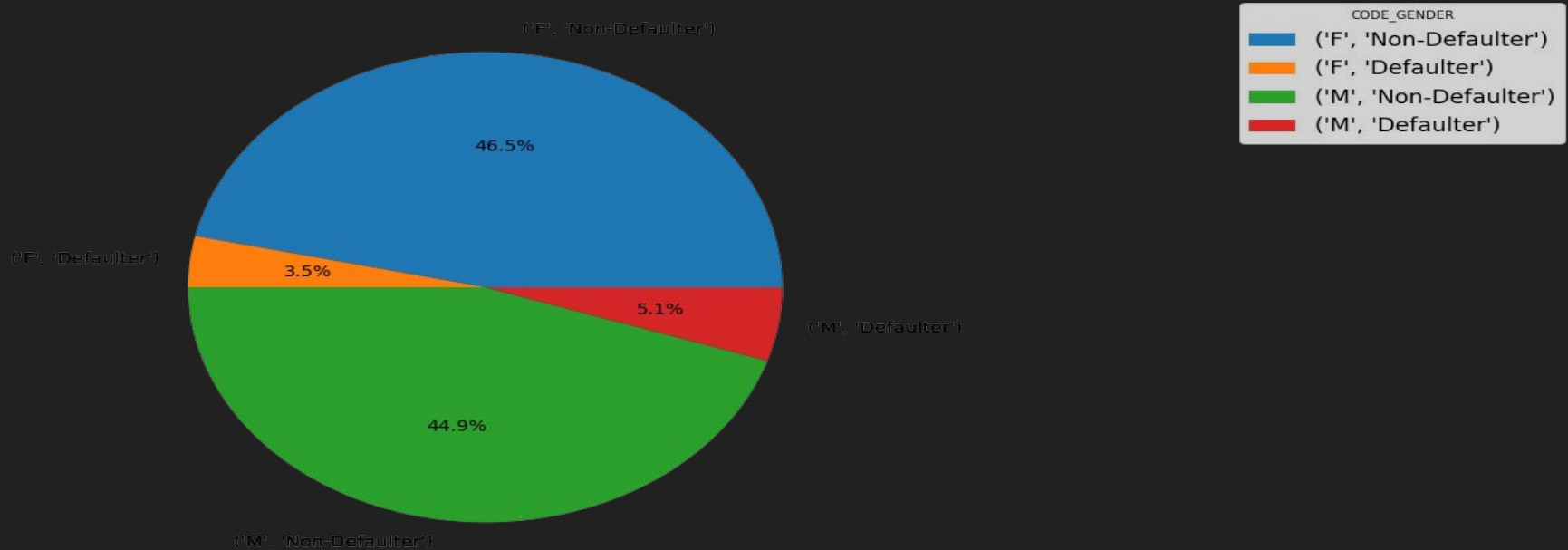
- Identifying the correlation of data with the target variable
 - Used python library methods for identifying correlation between numerical variables
- Making use of plotting techniques for better visualization
 - Used barplot / countplot for understanding relationship between numerical variables
 - Used pie chart for displaying in percentages
- Performed analysis
 - Analysed the output of plots to determine relationship between them

Imbalance percentage



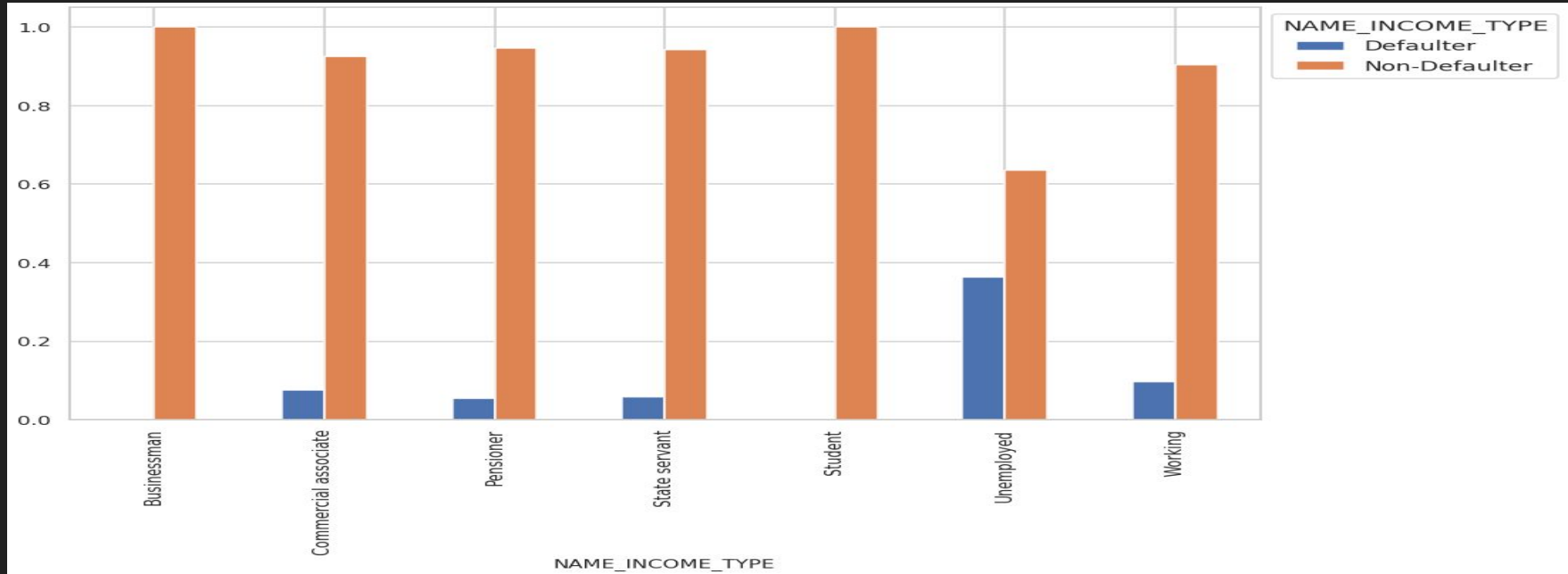
- More than 80% are Non-Defaulter
- Less than 20% are Defaulters

Gender Imbalance percentage



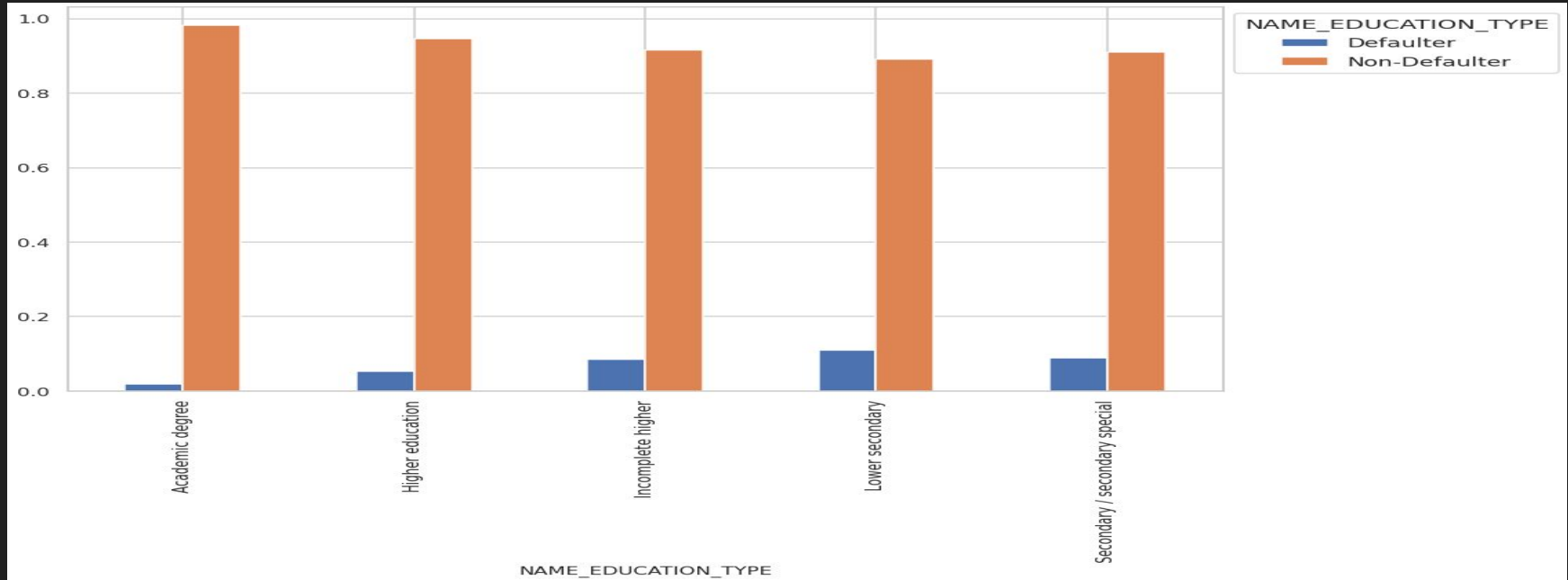
- Males are more likely to Default compared to females

Income Type



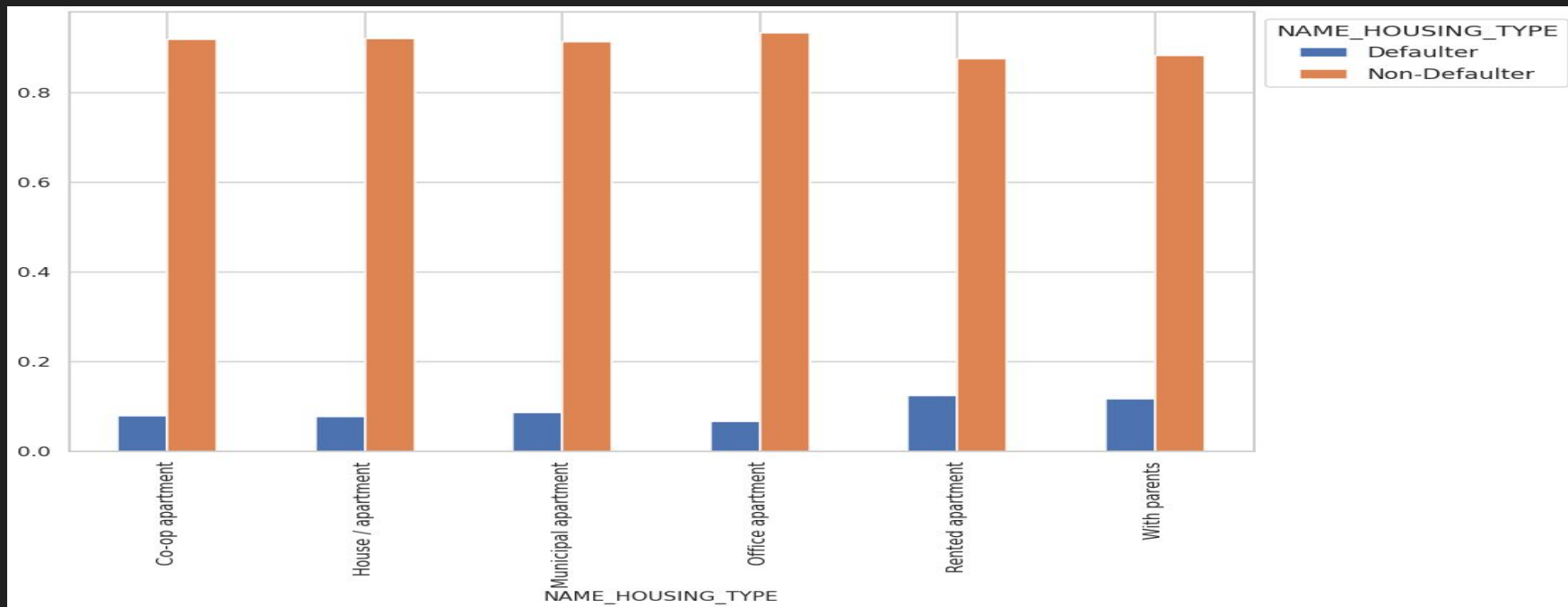
- Business man are the least likely to Default
- Unemployed are most likely to Default

Education Type



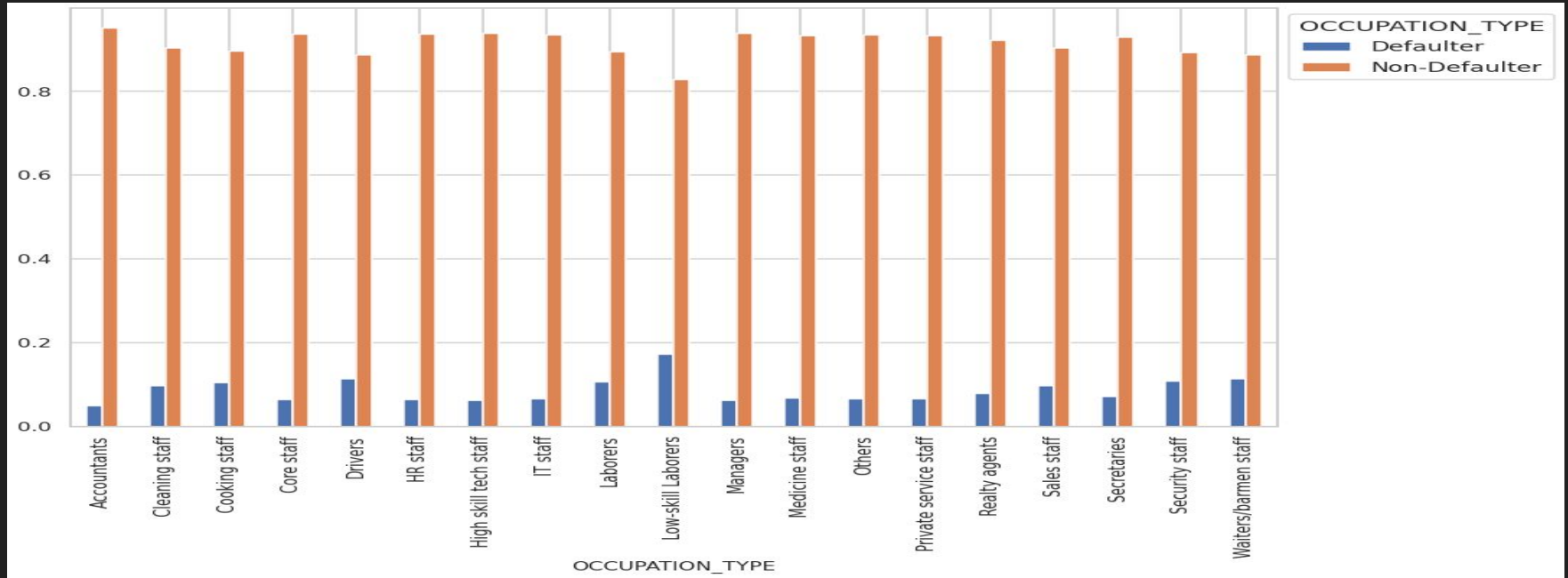
- People with Lower secondary degree is more likely to default

Housing Type



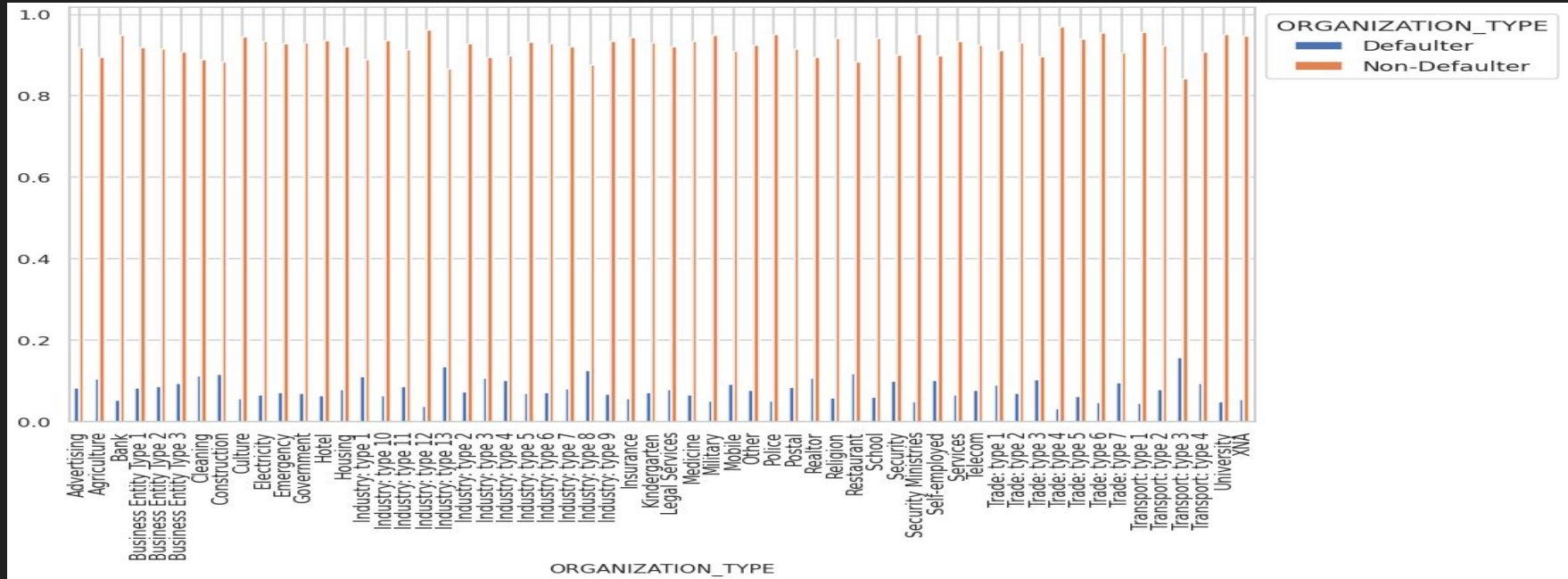
- People with Rented apartment are more likely to default

Occupation Type



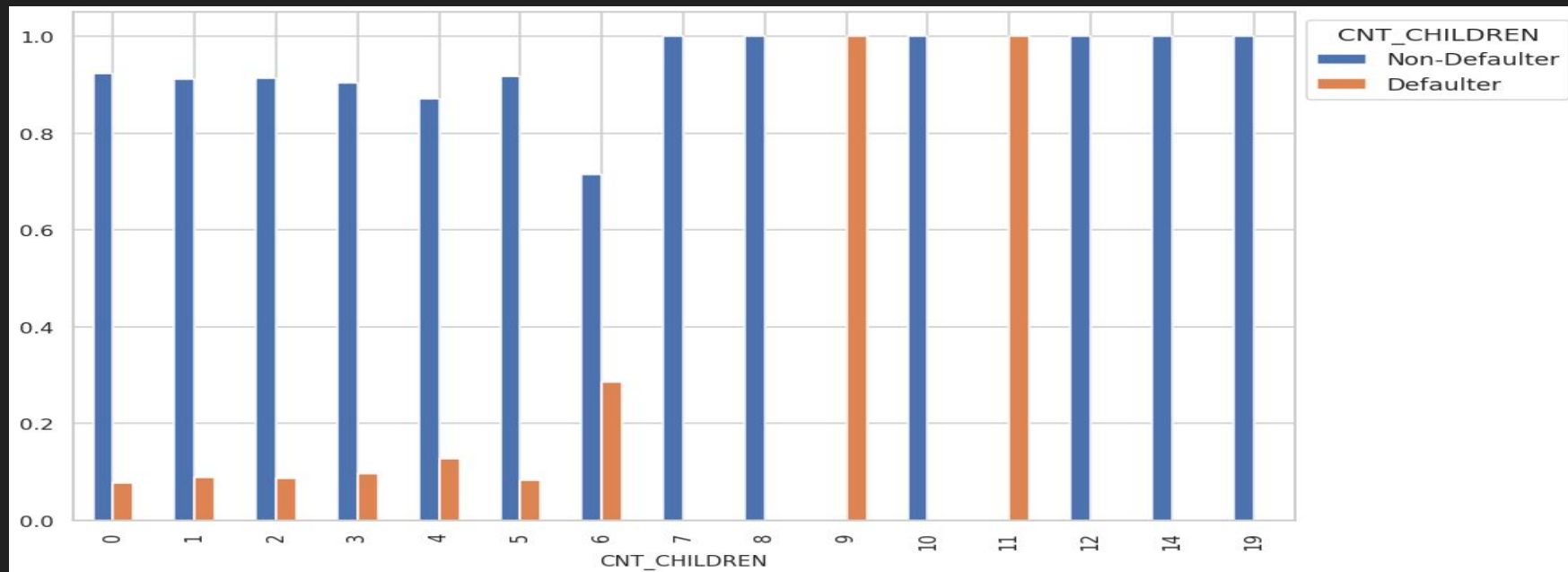
- Low skill labour have higher chances of Defaulting

Organization Type



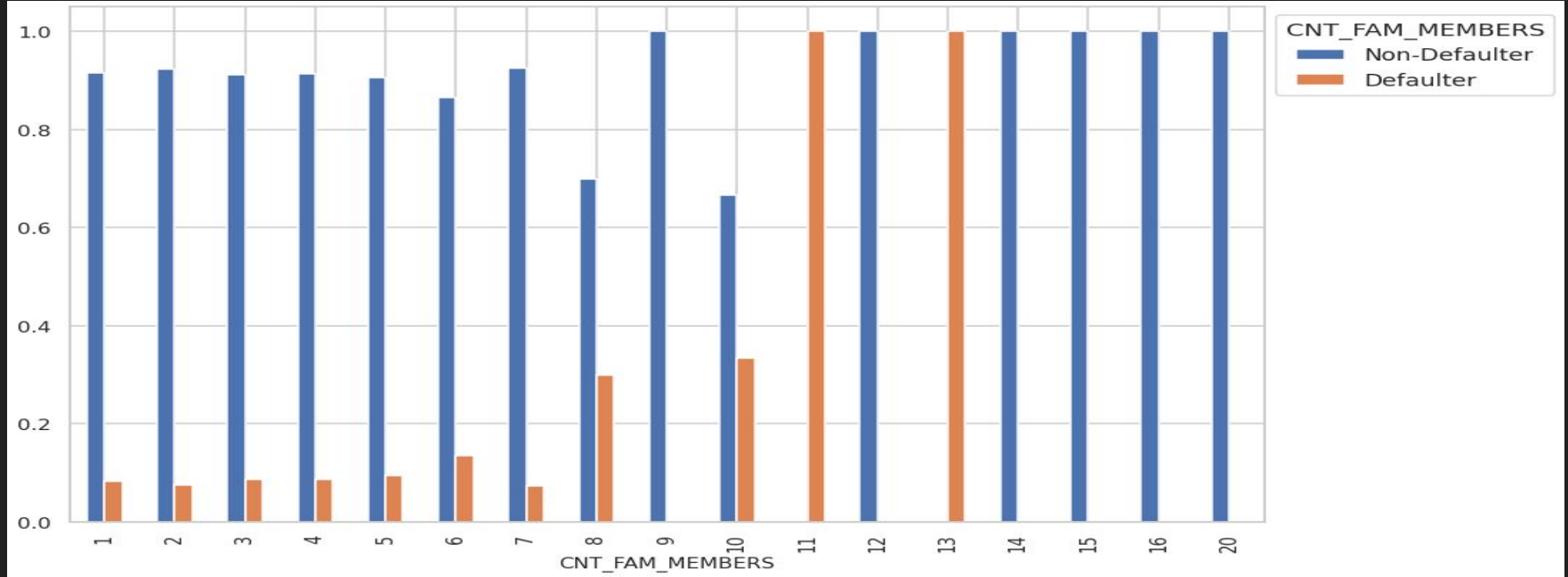
- Mostly Transport type 3 Organizations has the highest percent of people having payment issues

Children Count



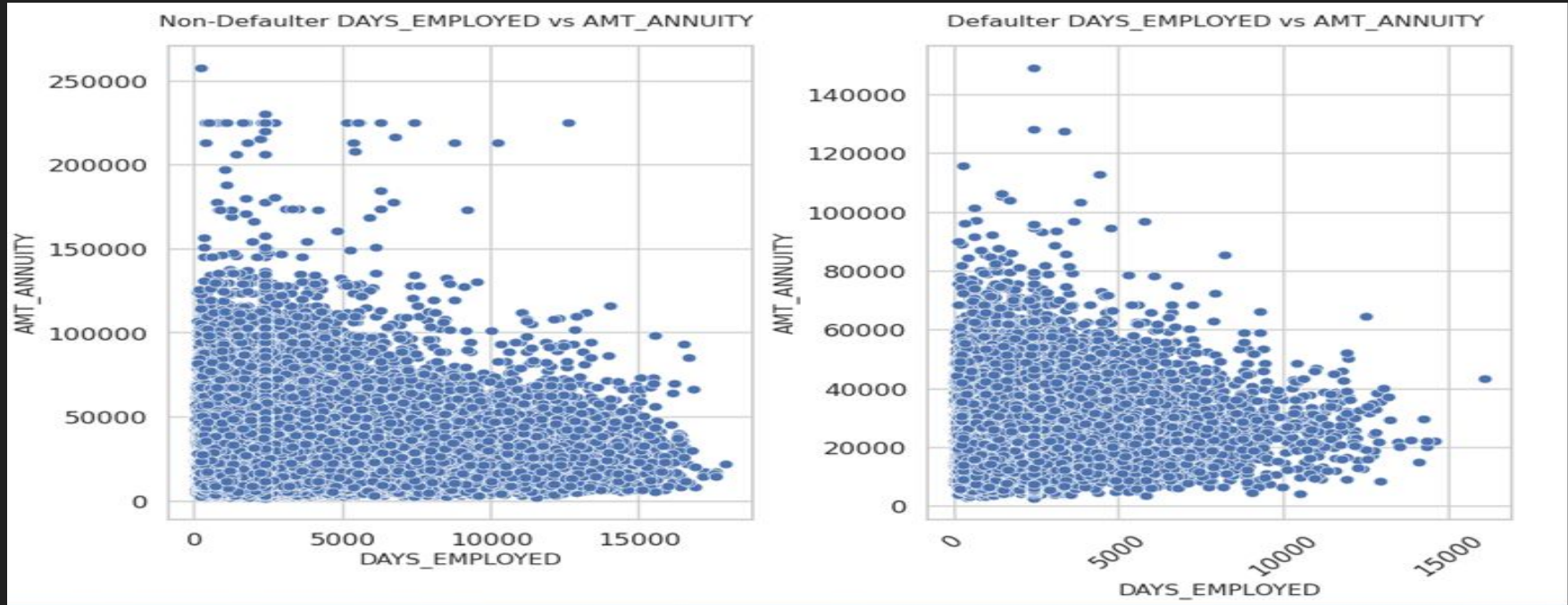
- We see that as the children count increases Default rate increases

Family Count



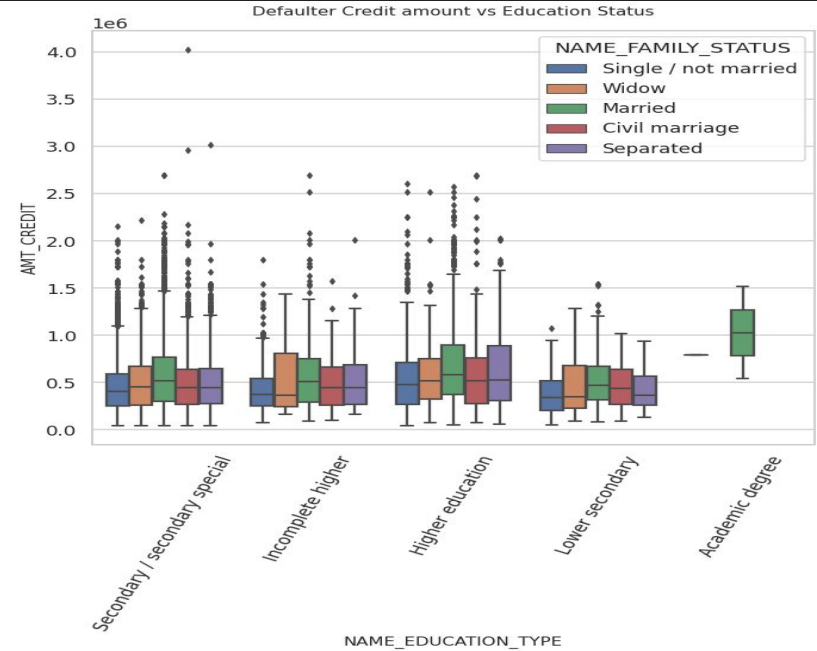
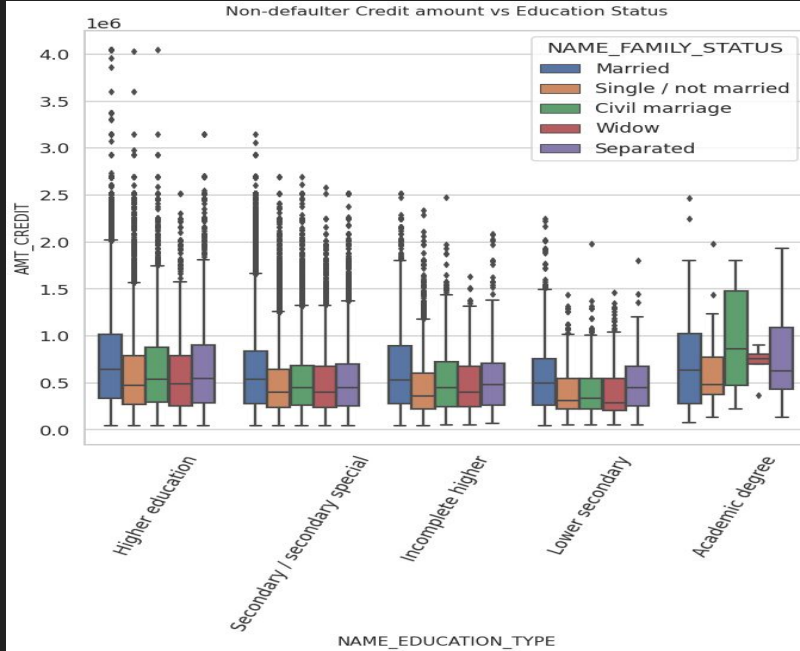
- We see that as the family count increases Default rate increases

Annuity vs Days employed



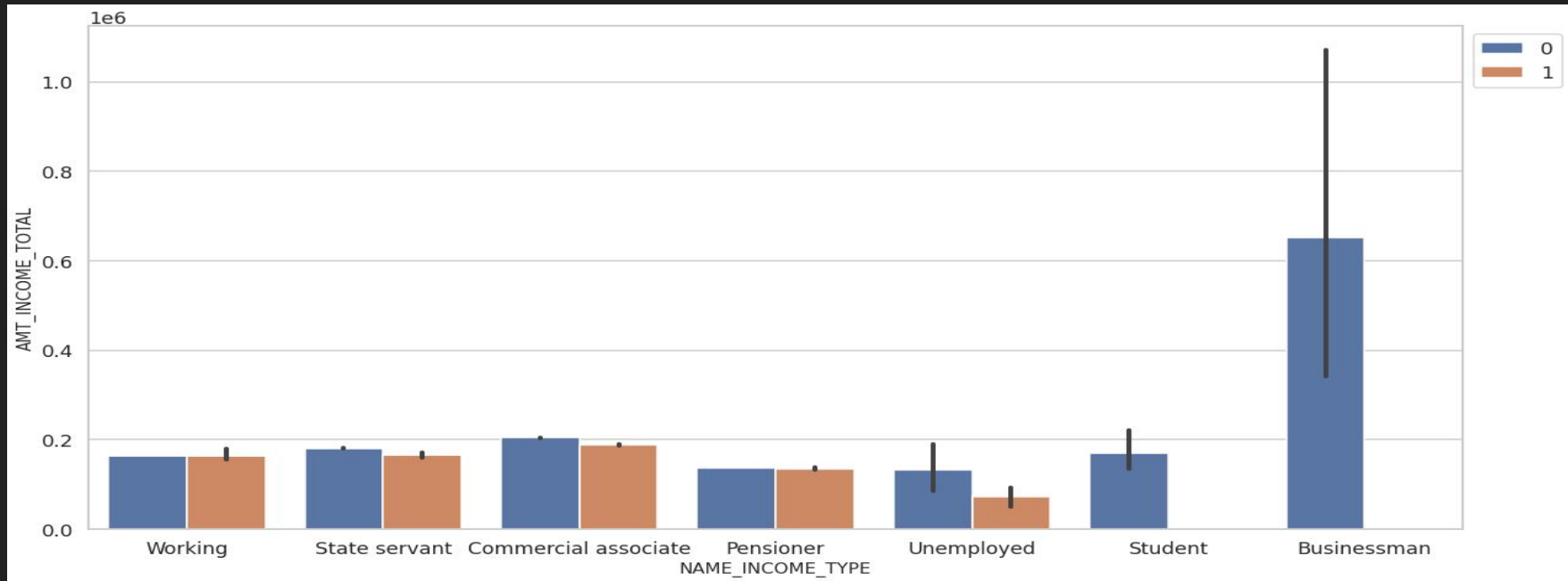
- Non defaulters are like to be employed longer
- Non defaulters have annuity much higher than the defaulters

Credit vs Education vs Family Status



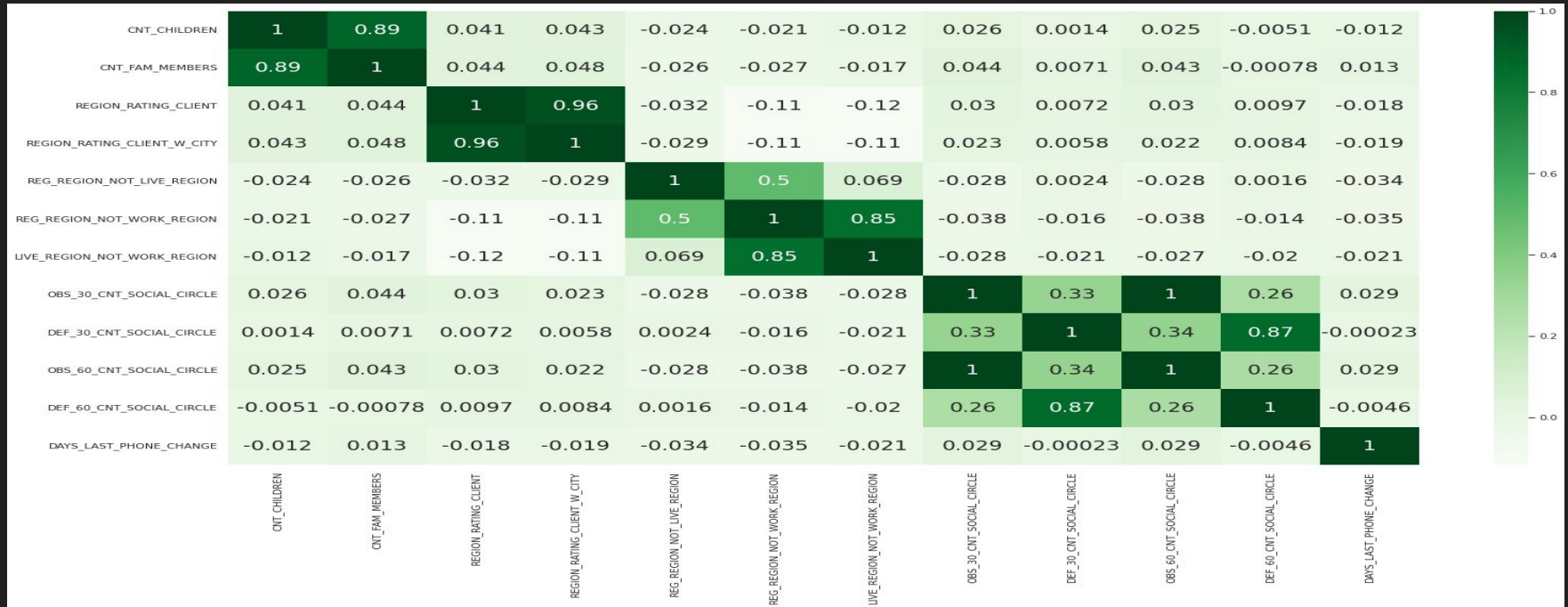
- Married customer with higher education are having the highest credit limits
- Customer with higher education are less likely to default
- Married customer with Academic degree are more likely to default

Income vs Income type



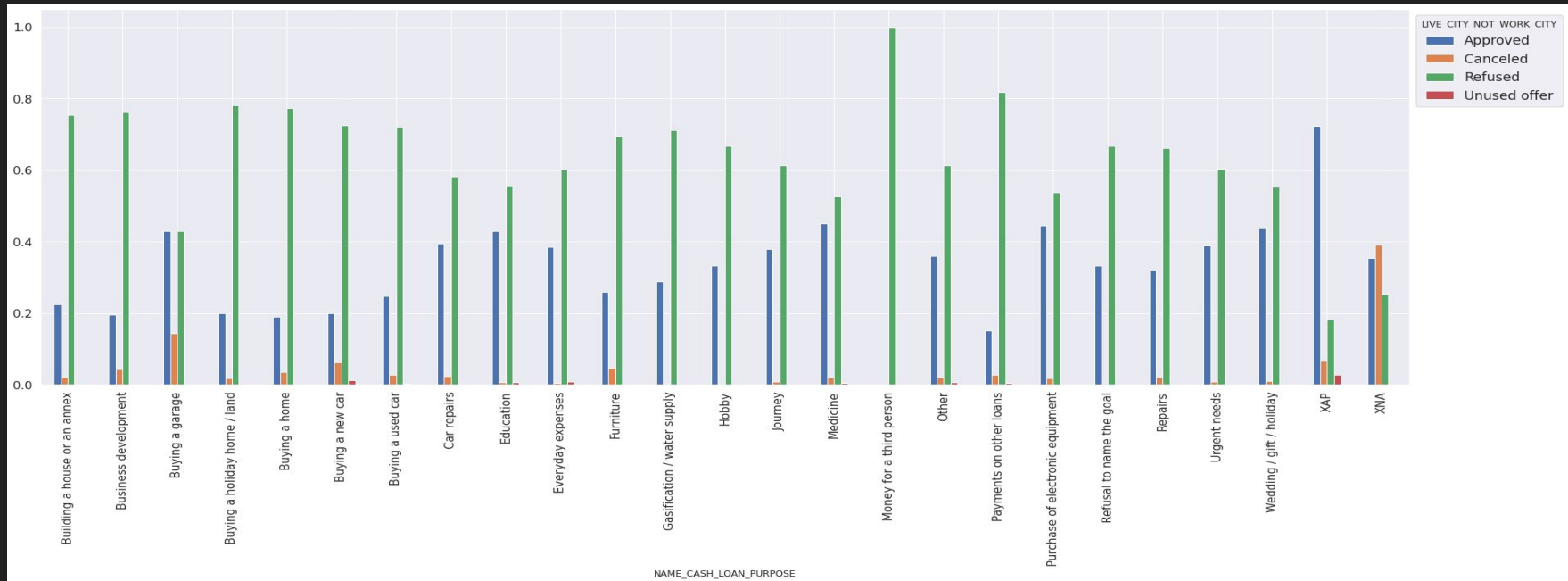
- Businessman has the highest income also the least default rate

Heatmap



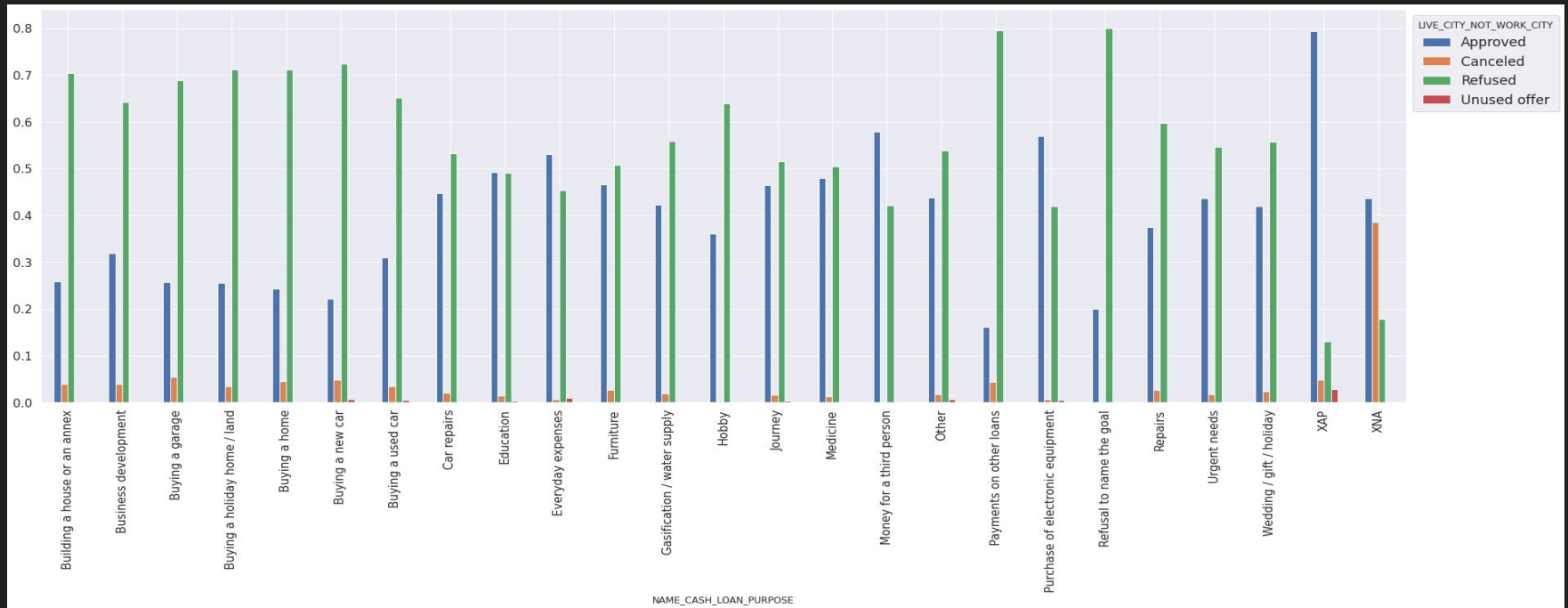
- REGION_RATING_CLIENT AND REGION_RATING_CLIENT_W_CITY are highly correlated
- OBS_30_CNT_SOCIAL_CIRCLE and OBS_60_CNT_SOCIAL_CIRCLE are highly correlated
- DEF_30_CNT_SOCIAL_CIRCLE and DEF_60_CNT_SOCIAL_CIRCLE are also correlated

Defaulter - Loan vs Status



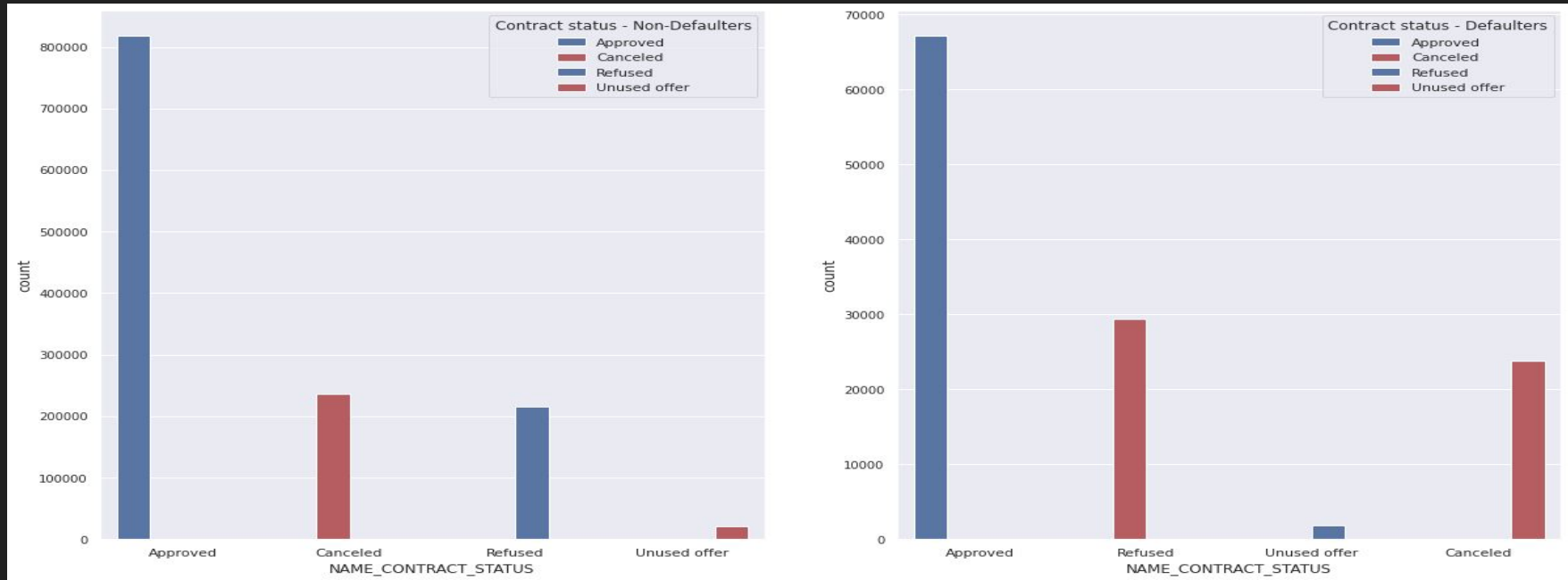
- Defaulter has mostly refused the contract where purpose was
 - Money for third person
 - Refusal to name the goal

Non-Defaulter - Loan vs Status



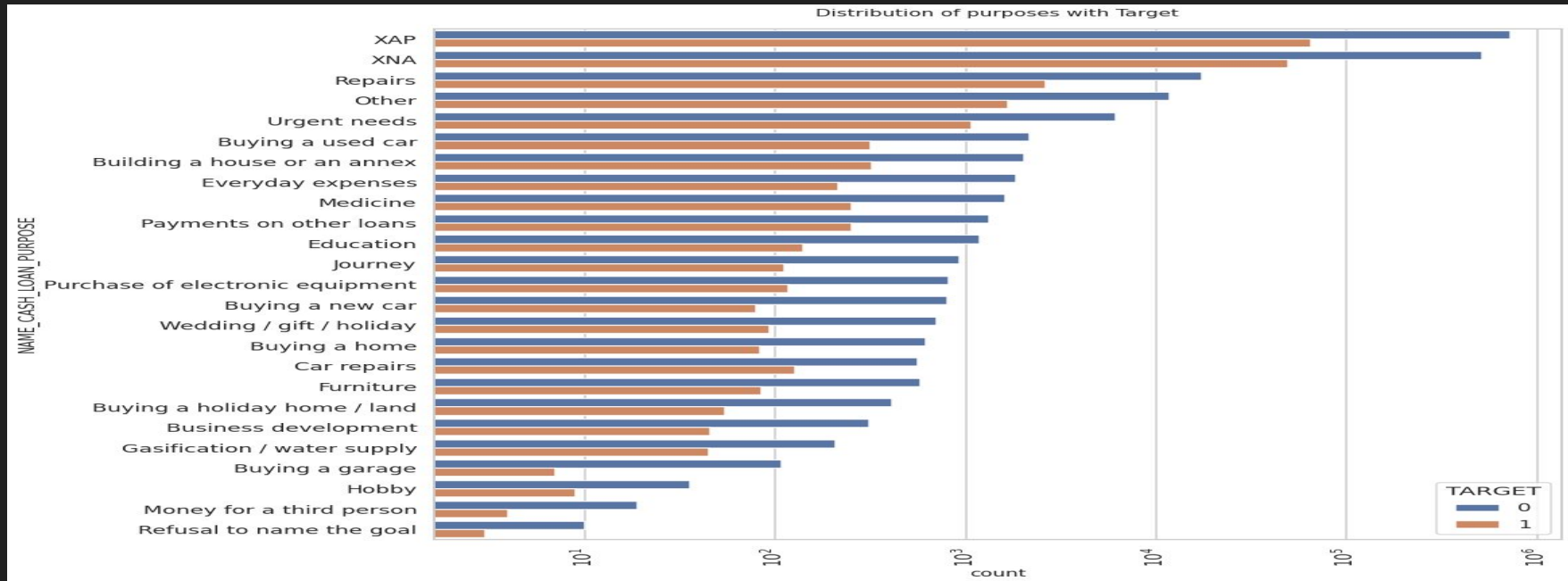
- Majority of the contract was approved for Money for third person and refused for payment on other goals and also where customer refused to name the goal

Contract Status - Imbalance



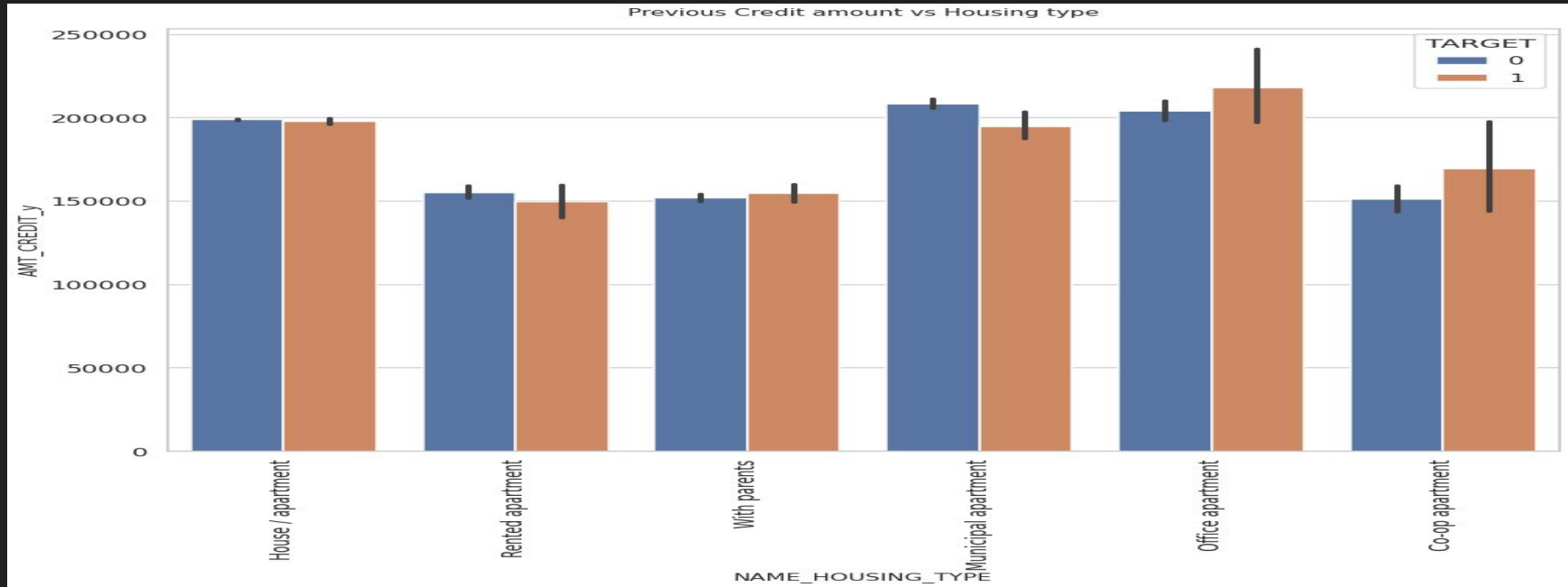
- Non-Defaulters have the highest approval count, where as Defaulters have the highest Refusal count

Distribution of Loan purpose



- Most of the Defaulter are those who takes loan for repairs
- Other categories with payment difficulties are Other, Urgent needs and buying a used car

Credit Amount vs Housing Type



- Customers with previous loan and Housing type Office apartment / Co-operative apartment have difficulties with payment

Analysis Verdict

- Companies should focus on customers like Businessman and people who stayed employed longer
 - They should also focus more on people with higher education
 - For Non-Defaulters contracts were refused for customer who wish not to name the purpose of credit, this could lead loss of business
-
- Companies should avoid unemployed customers, and people with lower secondary degree
 - Should be careful when dealing with customer with rented apartment and customers with higher children count
 - Customer with previous loan and Housing type Office apartment / Co-operative apartment may have difficulties with payment