

EDA Banking Credit Risk Analytics Case study

MANDAR DHUMAL

DS-37 PGDDS

IIIT-BANGALORE AND UPGRAD

Introduction

This is the banking credit risk analysis case study where in the bank wants to understand why are the major reason for the defaults in customer paying their Loans. Basically this study is to understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

Business Understanding

Credit risk refers to the uncertainty faced by an organization while lending money to an individual, business, or organization.

Credit risk analysis is defined as a detailed review, and inspection done by the lending organization regarding the borrower about their financial background, modes of earning, and the capability to repay the borrowed credit. This gives the lending enterprises a fair idea regarding the credit-paying capabilities of the borrower.

In simple terms, credit risk refers to the potential for loss due to the failure of a borrower to make a payment when it is due. The risk is mainly for the lender and it can include complete or partial loss of principal amount, loss of interest, and disruption of cash flow.

The purpose of credit risk analysis is to determine the creditworthiness of the borrower based on his financial background and repayment history and capacity. It means to determine the eligibility of the person to receive the amount of money he is requesting without causing any kind of complete or partial loss to the lending organization.

Business Objectives

- Bank wants to analyse the datasets to understand strong indicators that causes loan payment defaults
- This will ensure minimizing the Bank loses from the defaults by either denying loans, minimizing the loan amount, providing higher rate of interest to customers with such indicators
- Similarly they want to understand which customers should be targeted for giving loans

Datasets Understanding

There are three data sets

- Application Data,
- Previous Application Data, and
- Columns Description

The Application data set contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.

The Previous Application data set contains information about the client's previous loan data. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.

The Columns Description is data dictionary which describes the meaning of the variables.

Approach to the EDA

1. Understanding the Datasets especially columns
2. Understanding the Bank Credit Risk Analytics Market
3. Cleaning the datasets by dropping columns which are not required for the analysis, or data with more than 40% Null values
4. Identify the Outliers and treat them accordingly i.e if data column is continuous replace it with Mean or Mode and if the column is categorical, then replace it with mode. In this study, we have not deleted any datapoints
5. Checking the Data Imbalance
6. Merging the Application and Previous Application Data sets
7. Perform Univariate, Bivariate and Multivariate Analysis to understand the relations of the columns within the dataset
8. To present the Observations

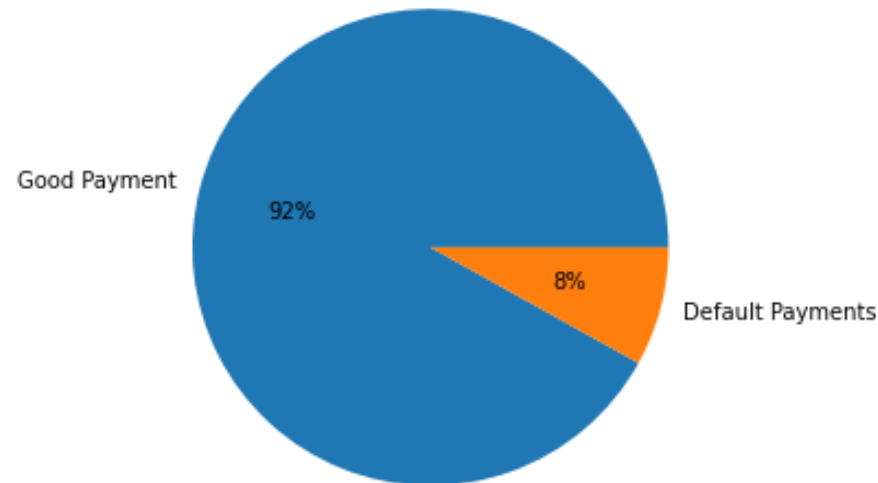
Data Cleaning, Null Values and Outliers

1. Both the datasets were cleaned by dropping columns with more than 40% Null values.
2. There were 49 columns in Application dataset and 11 columns in Previous Application dataset that were dropped as they had more than 40% of NaN values and this could have skewed the analysis
3. Columns with less than 40% Null values were either filled with median values in case of continuous variables and with mode with categorical variable
4. Outliers were shown wherever detected. As it was not required for the study, I have not deleted the outlier data points. I have imputed the median value to cap the outlier so that the analysis is not effected incorrectly. Also there are other ways to cap the value based on quantiles or by designing a model to come to a value for the outliers. But this has not been done for this study purpose

Data Imbalance

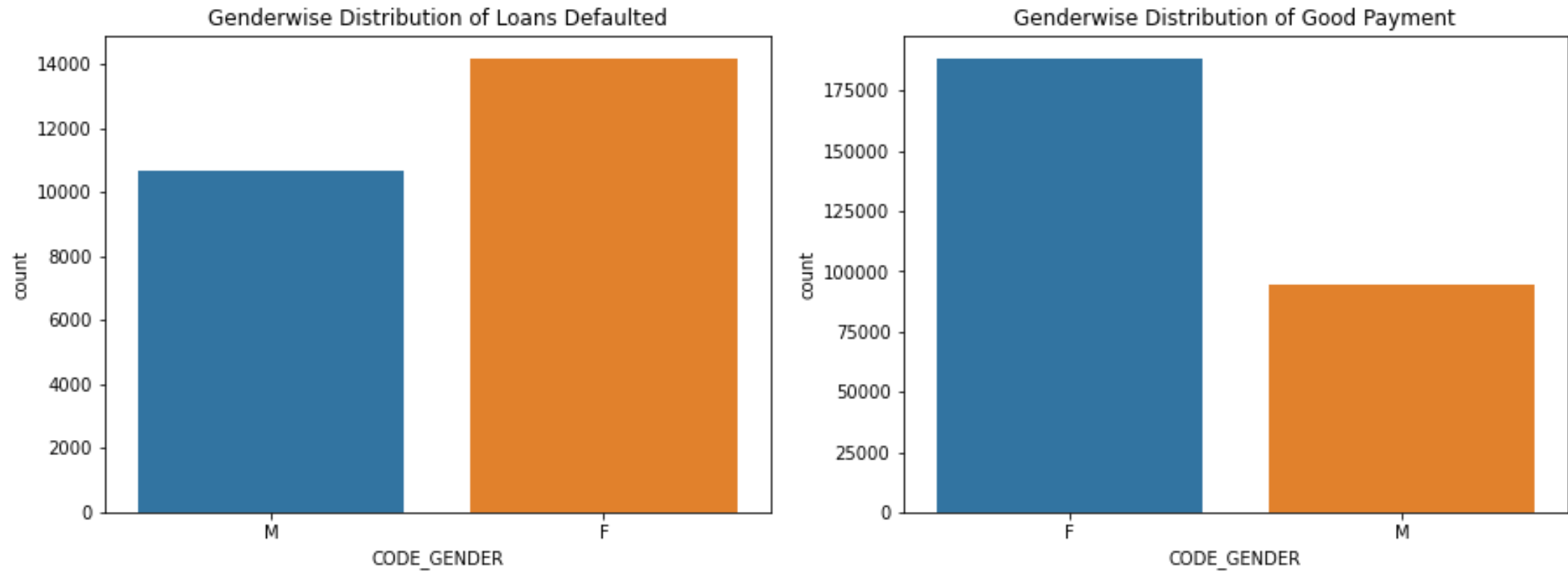
The Data imbalance was found for the Target column which is the main column of the study.

Pie chart of Good Payment Customer Vs Defaulters



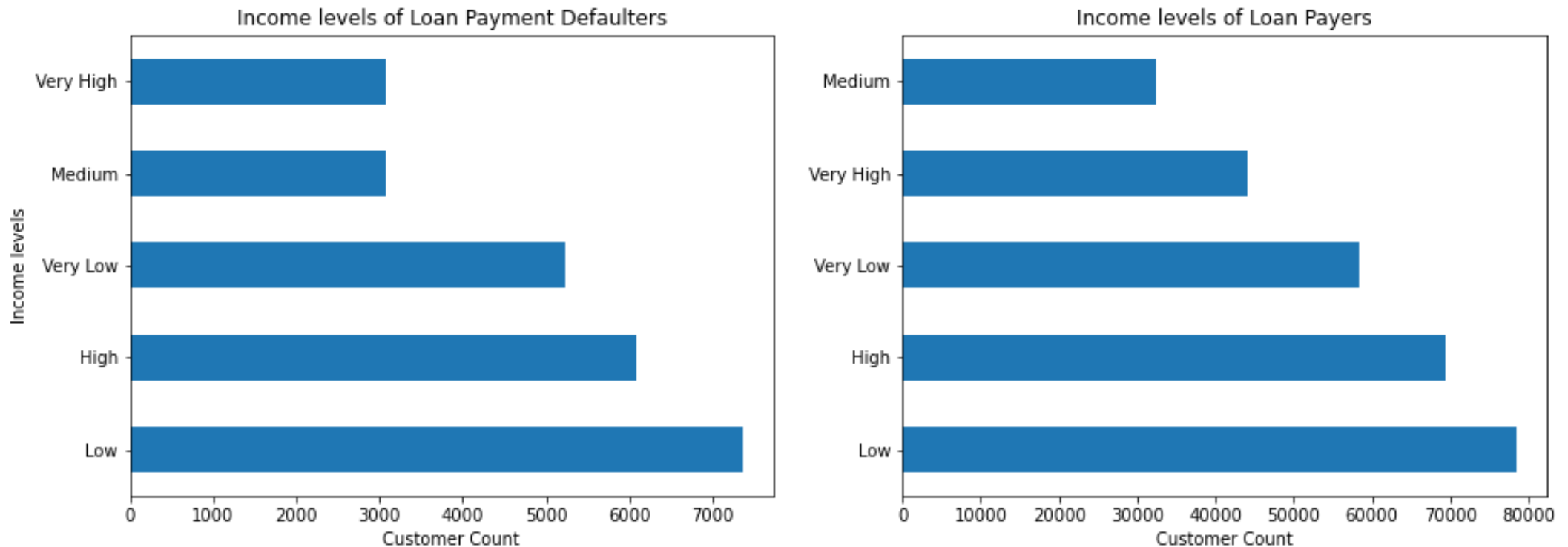
Only 8% of the Target customers have the payment related issue, rest have been 92% are regular in their dues

Analysis - 1



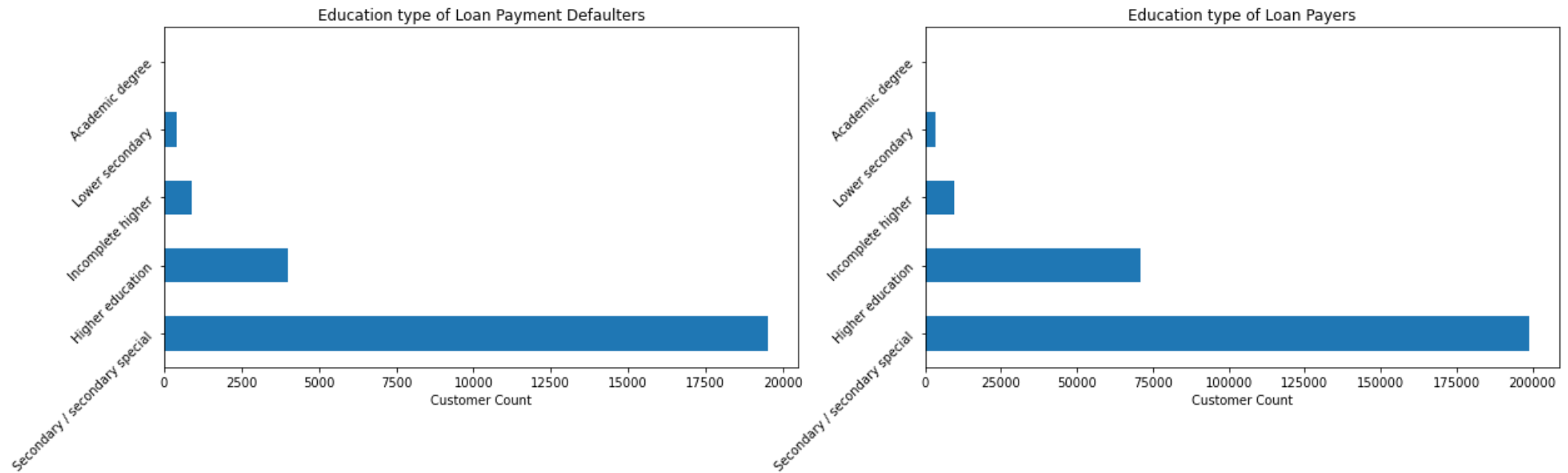
It can be observed that Female gender is more prone to default the payments but this scenario can be because the loans are given to more number of females as compared to males

Analysis - 2



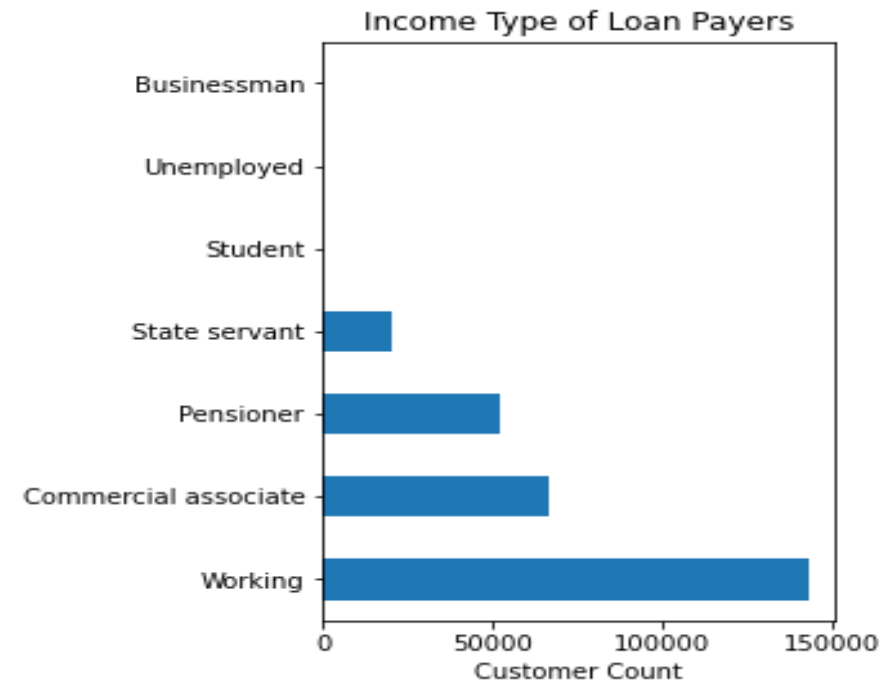
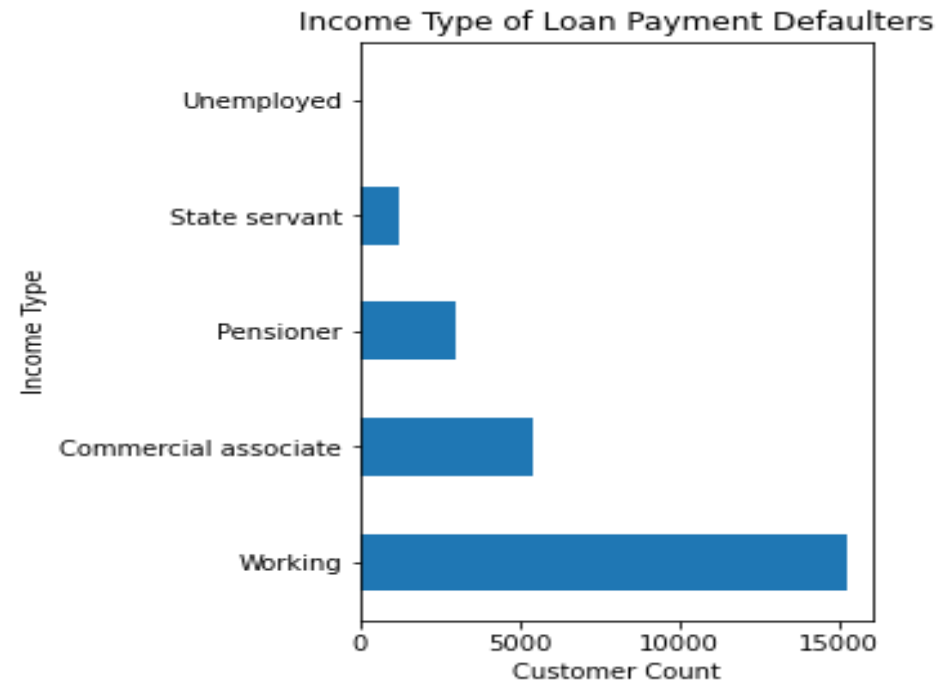
We can observe that Very Low, Low and High level of Income Customers are major defaulters. It is quite understandable that Low levels of Income can be a reason for defaults, but we need to dig deeper to understand the reason behind payment defaults by High Income Customers

Analysis - 3



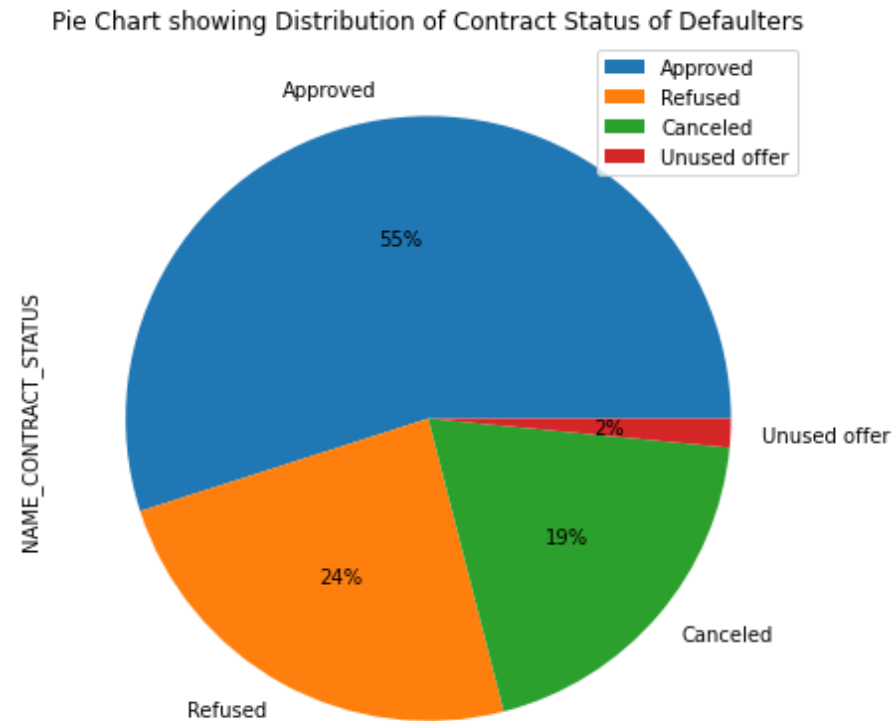
We can infer that maximum defaulters are with secondary level education

Analysis - 4



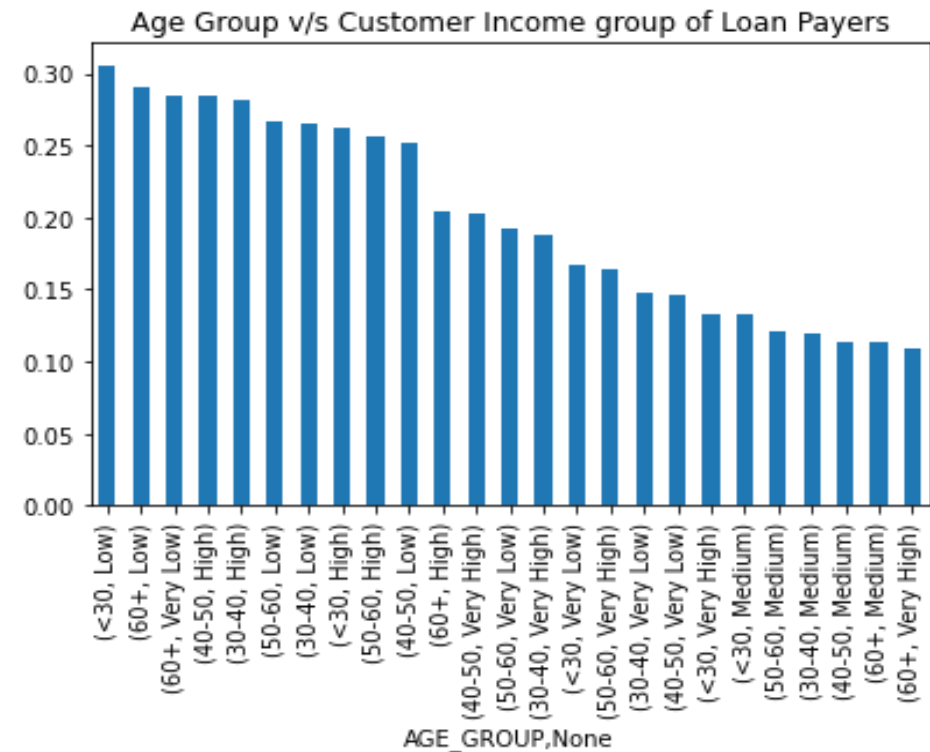
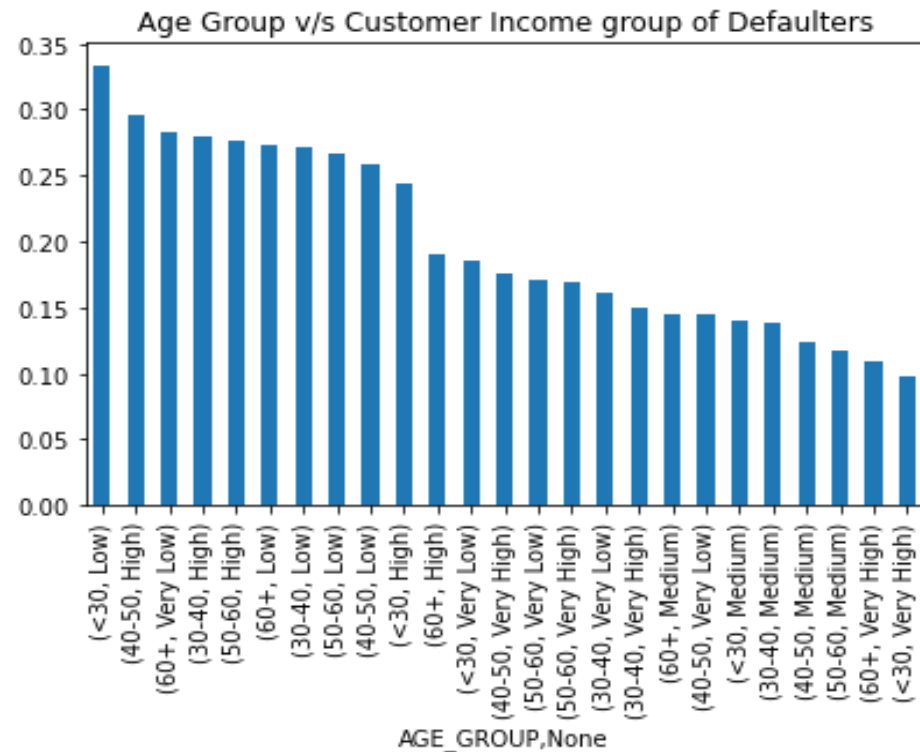
It is a strange result that Working customers are mostly the defaulters. We need to further dig into this aspect to understand why a stable income customer is defaulter as against unemployed or Pensioners, who have low income/no income and yet are less in defaults

Analysis - 5



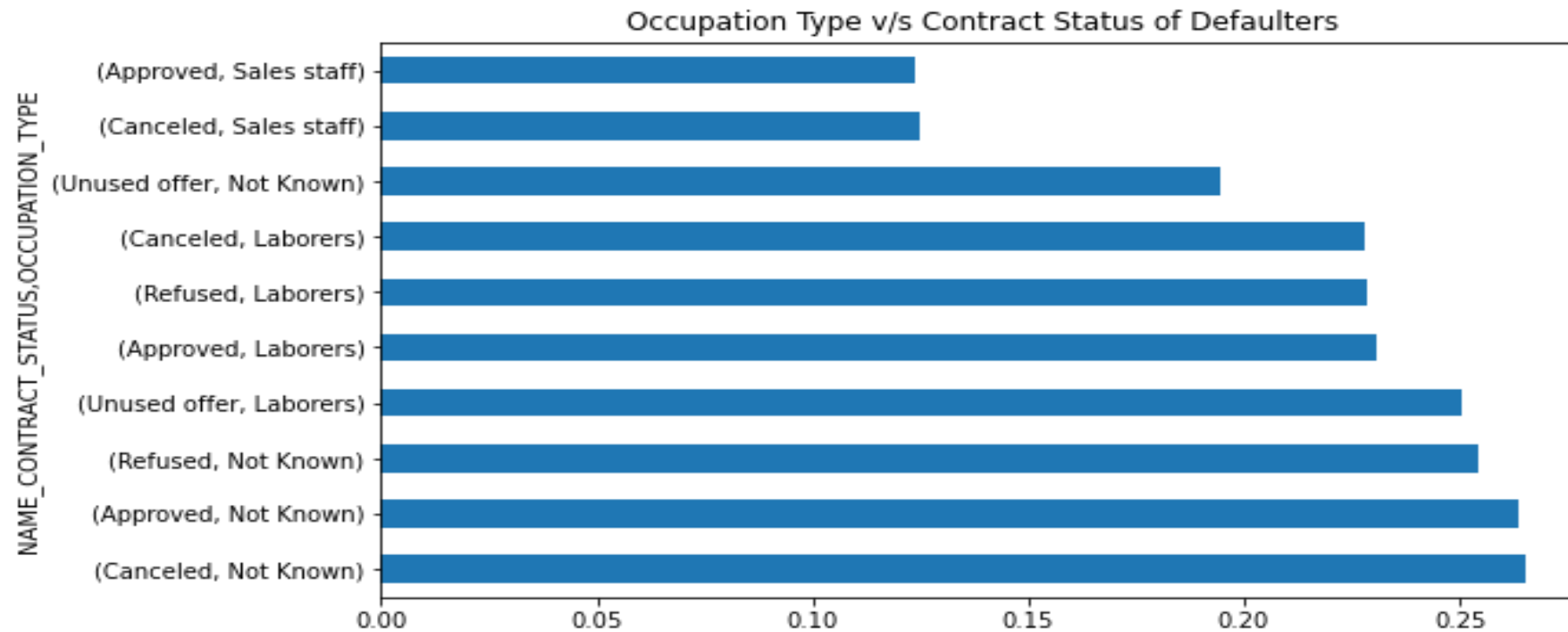
It can be seen that around 24% defaulters were refused the loan in the previous application. This should have been considered while giving the current loan

Analysis - 6



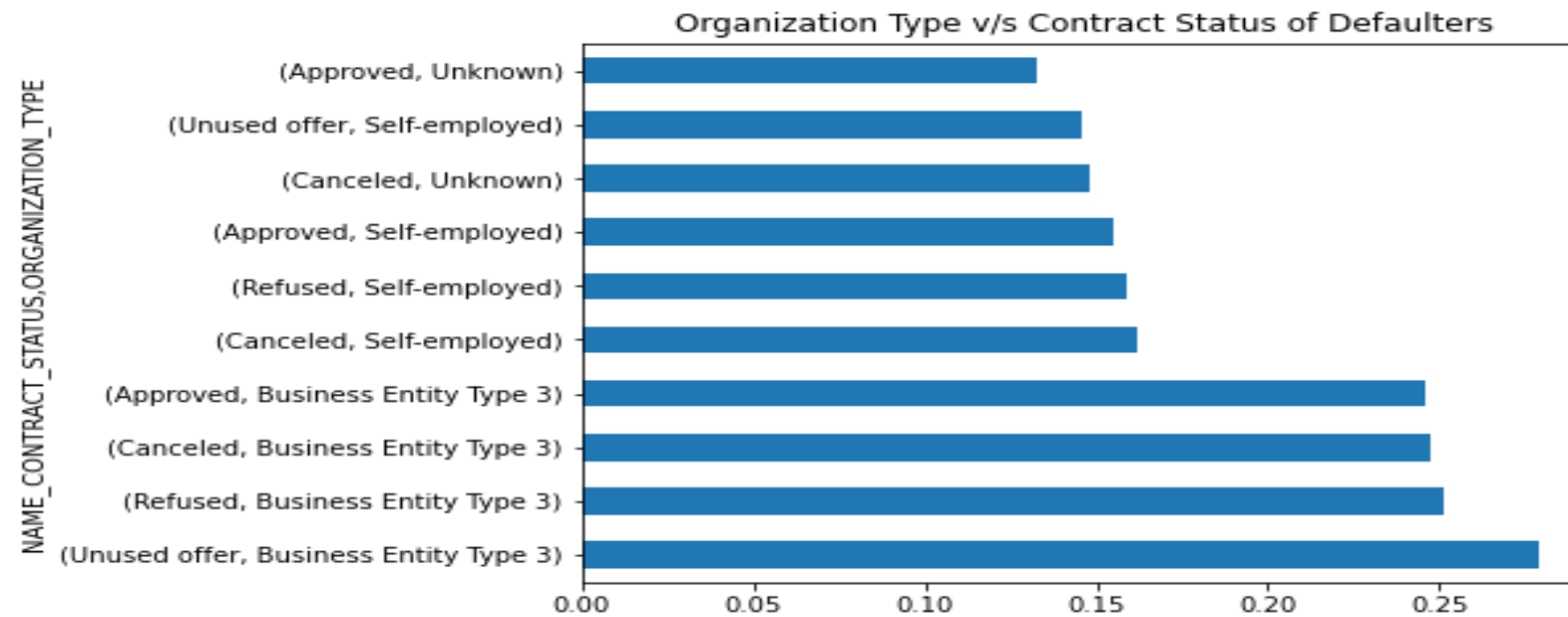
It can be seen that majority of the defaulters are from low income

Analysis - 7



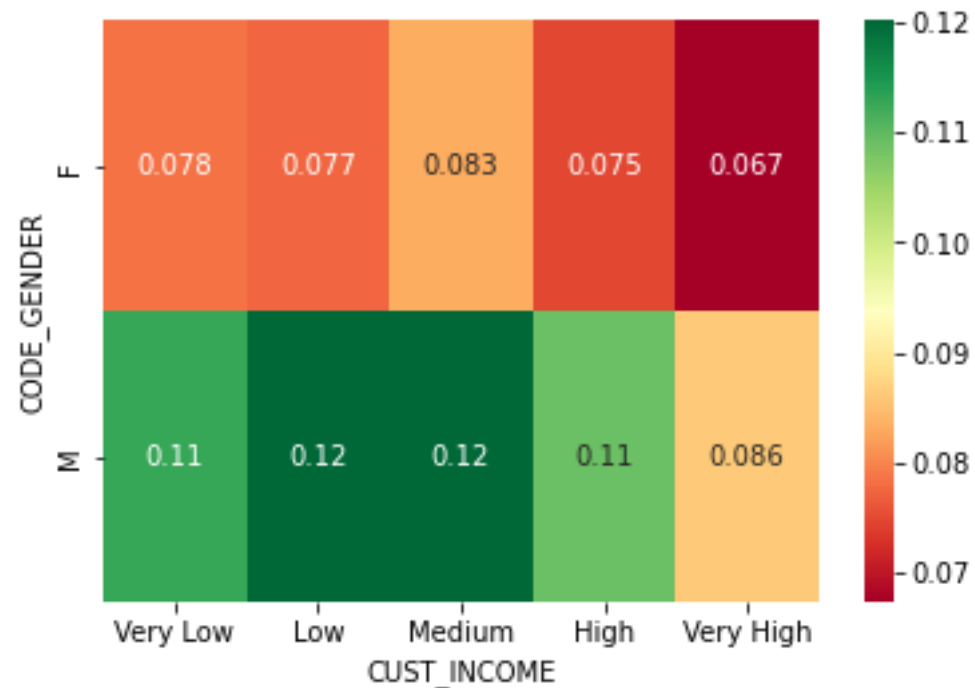
It can be observed that majority of the defaulters are Laborers who majorly have low income levels.

Analysis - 8



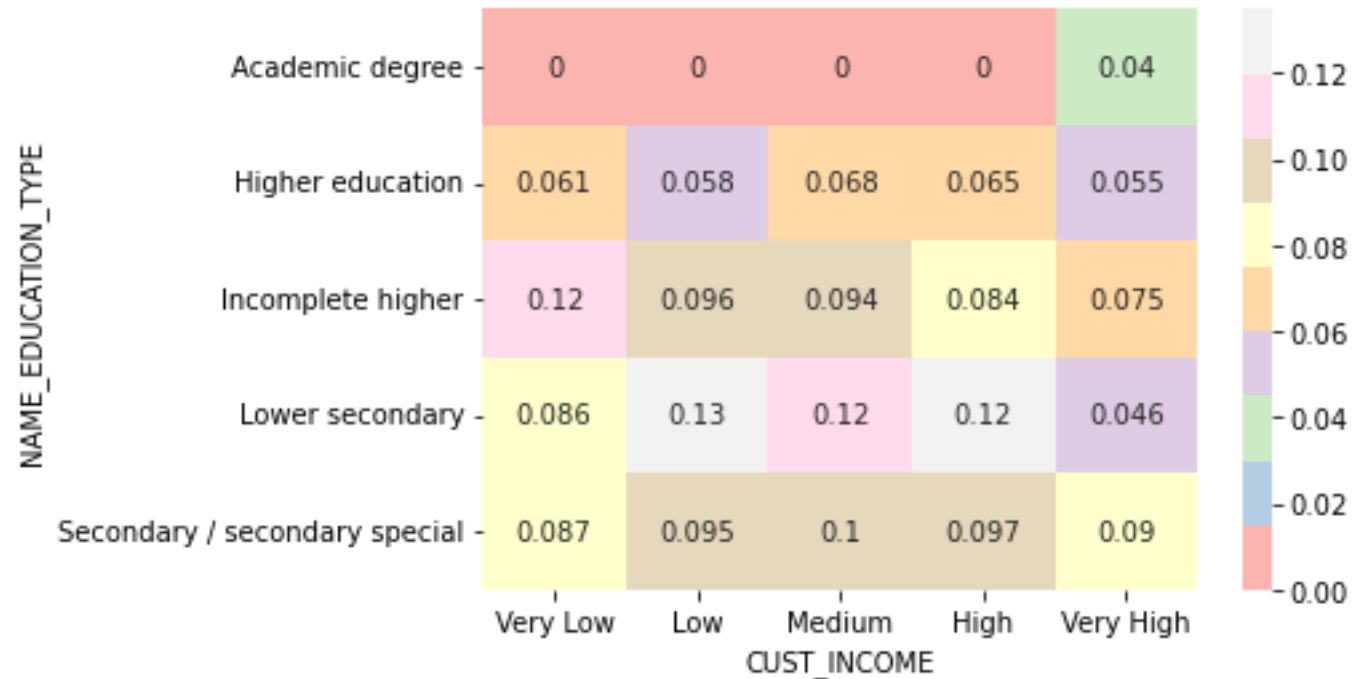
Most of the defaulters are the ones with Business Entity Type - 3 and Self-Employed

Analysis - 9



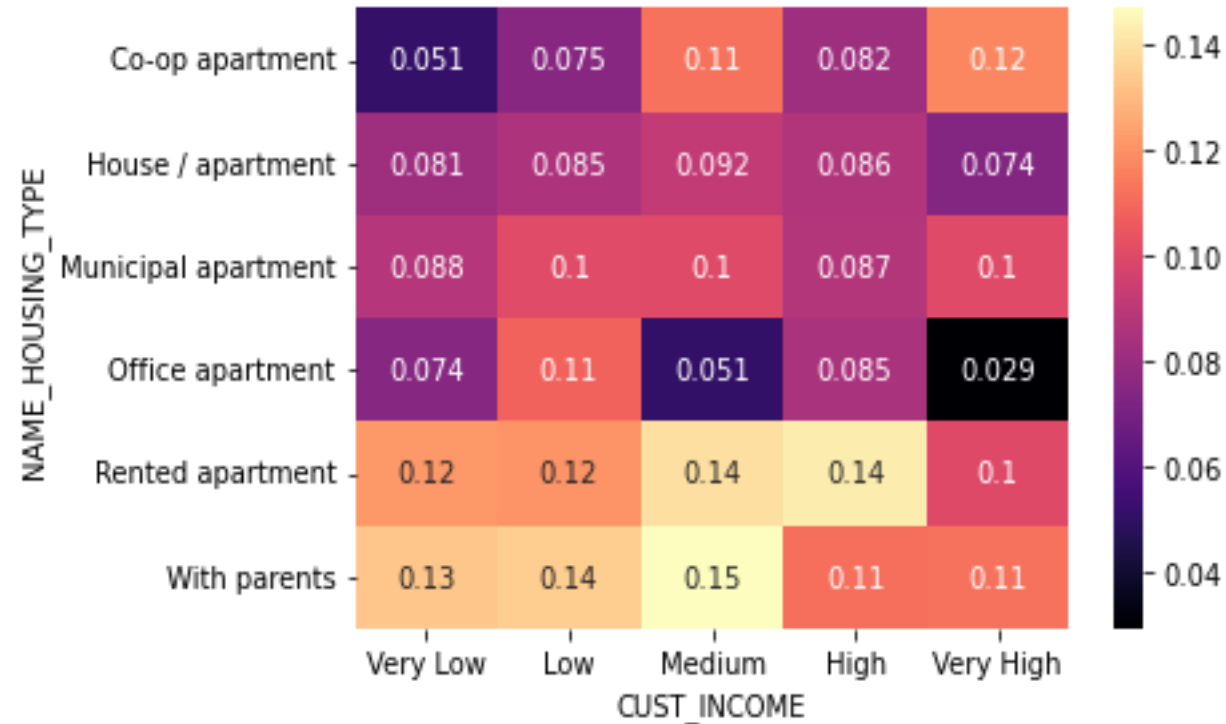
We can see that there is strong correlation between Females and Medium Income group, Males and Very High Income group among Defaulters

Analysis - 10



Lower Education levels and Lower Salary levels show strong correlation while defaulting loans

Analysis - 11



We can observe that there is strong correlation with all income levels and customers Owning House have defaulted while customers staying with parents or on rent have very weak correlation with defaults

Observations

1. Majorly customers defaulted the cash loans.
2. Females were mostly the defaulters but also they were major customers who paid the Loans on time
3. The Secondary level education customers were the defaulters
4. Working class customers were major defaulters but also major customers paying the loan
5. As the no. of family members and children increased customers were more inclined to pay loans on time
6. Low Salary was the main cause for the defaults
7. People who were self employed or who worked for Business Entity Type 3 were major defaulters
8. People who worked as Laborers were most defaulters

Thank You