Presentation on Credit EDA Case Study

Neha Kiran
Pavithra Sivanesan

Credit Risk Analysis:

Credit risk analysis is **a form of analysis performed to determine a borrower's ability to meet their debt obligations**. The purpose of credit analysis is to determine the creditworthiness of borrowers by quantifying the risk of loss that the lender is exposed to.

Problem Statement: The purpose of the Credit EDA Case Study is to understand how a bank approves and rejects a loan. With the help of EDA and basic understanding of risk analytics in business and financial services, we need to understand how data can be used to minimise the risk of losing money while lending to customers.

Steps

1.Data understanding and sourcing

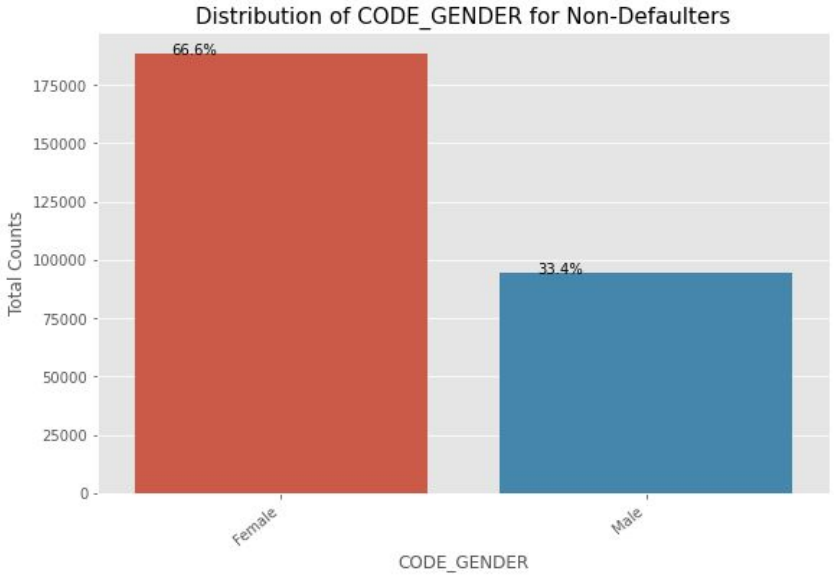
2.Check the missing values in the dataset and cleaning

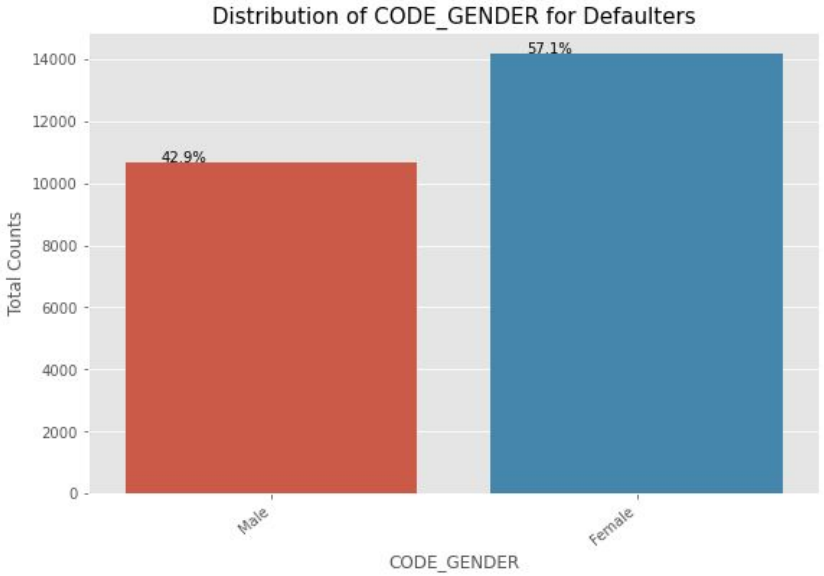
3.Merging the application data with previous data

4.Analysis by Univariate and Bivariate and Correlation

5.Risks and Recommendation

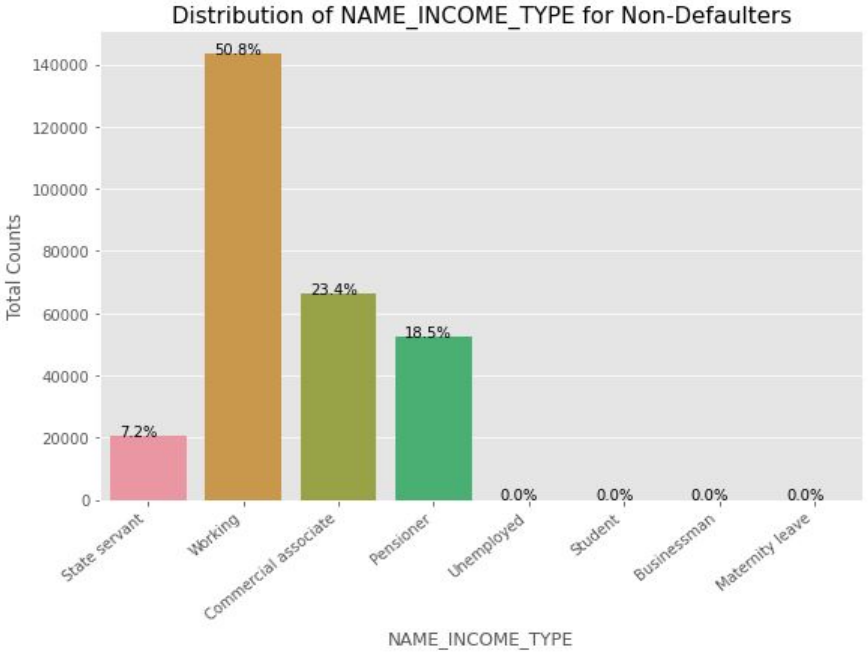
Univariate Analysis of Categorical Data in Application Data

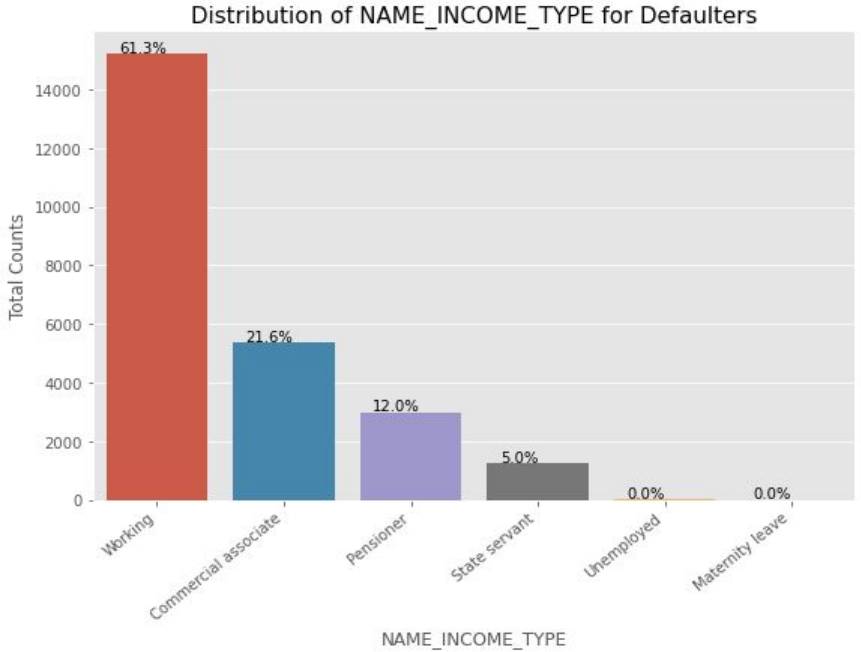




It can be observed that Female contribute **67%** to the non-defaulters while **57%** to the defaulters. We see more female applying for loans than males and hence the more number of female defaulters as well.

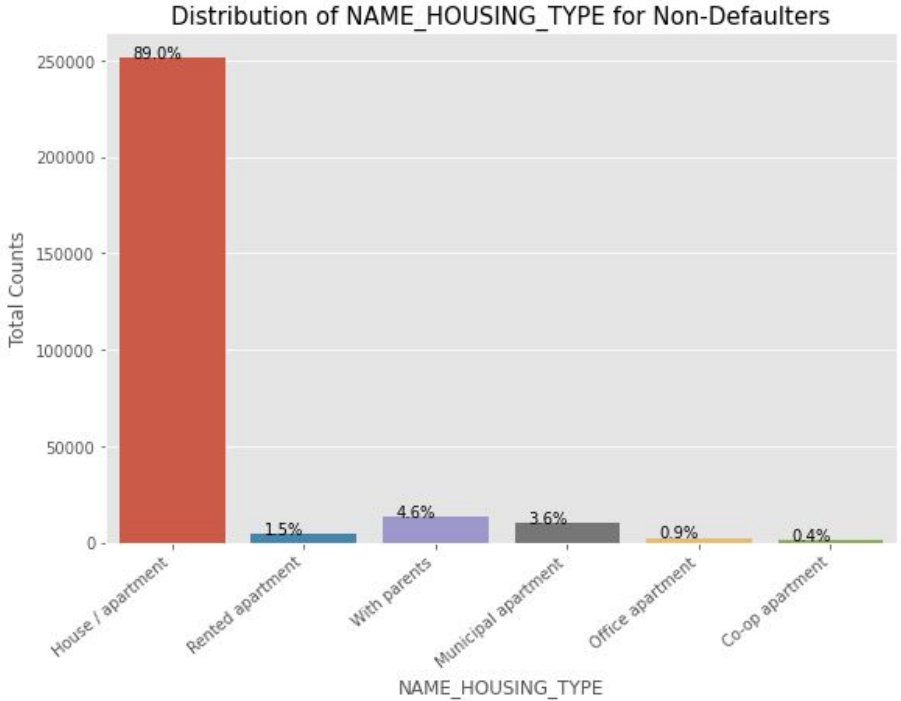
Name_Income Type Variables

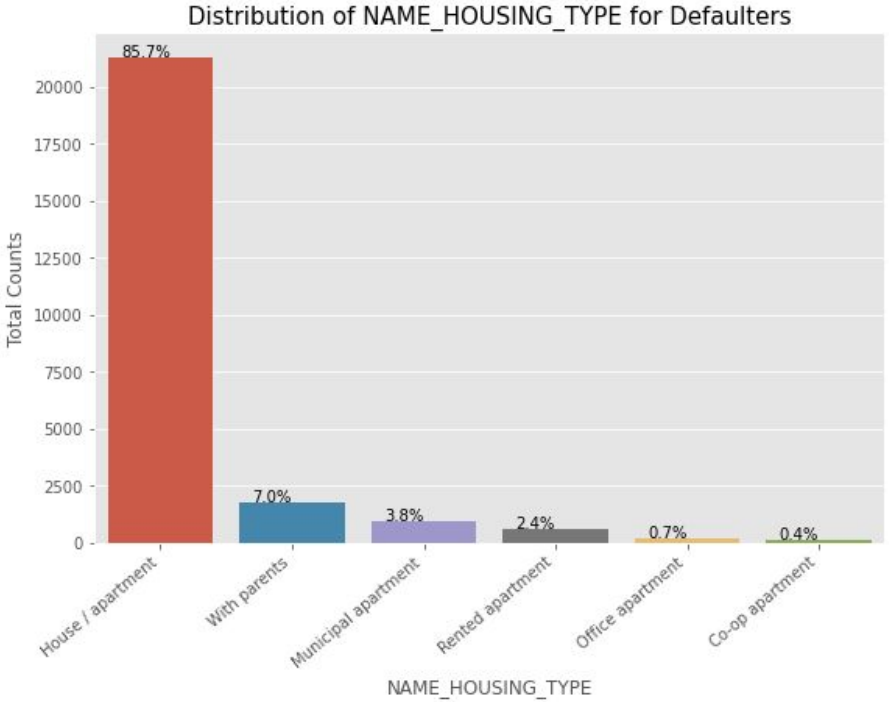




It can be observed that the working class people are the ones taking maximum loans and they contribute to 51% of Non- Defaulters and 61% of Defaulters.

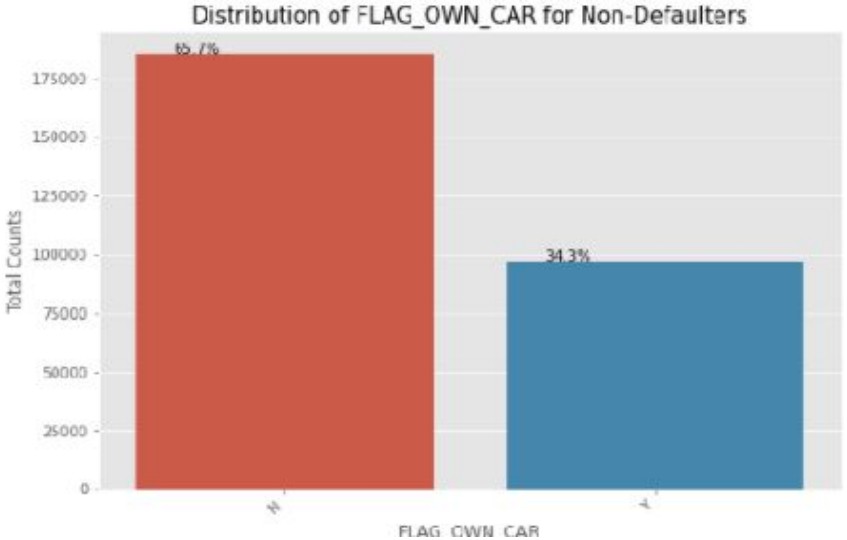
Housing type variable

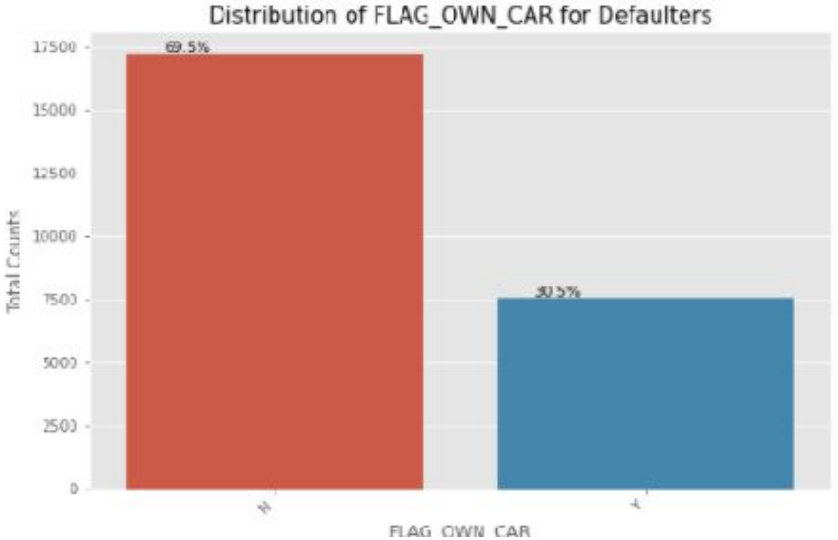




It is clear from the graph that people who have House/Appartment, tend to apply for more loans.
Whereas People living with parents tend to default more often when compared with others.

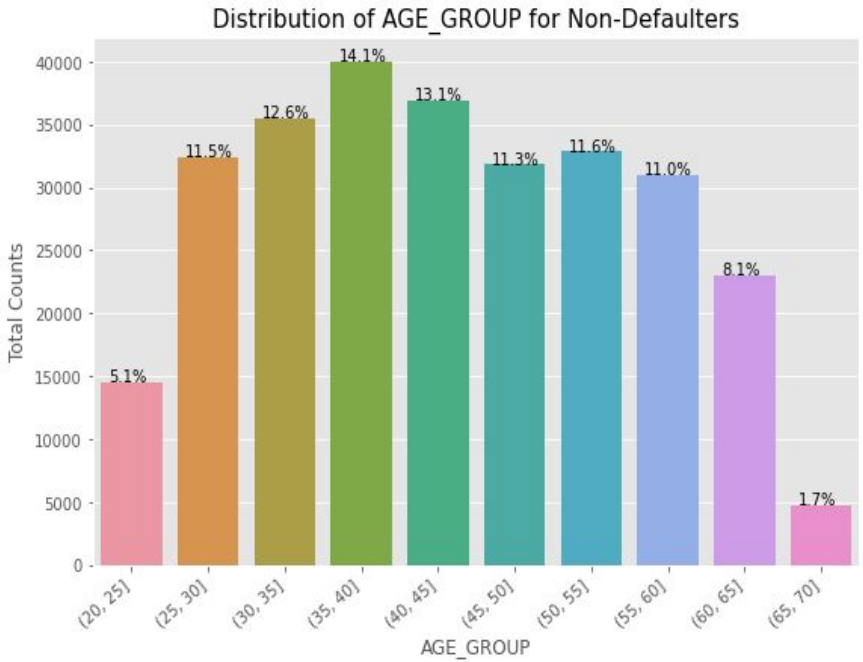
Own car type Variable

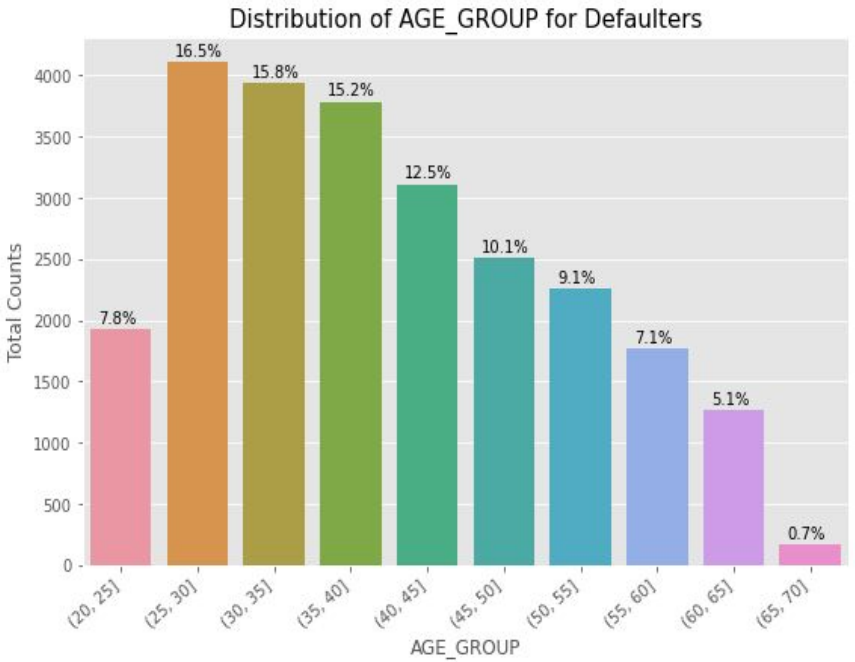




It can be observed that people with cars contribute 65.7% to the non-defaulters while 69.5% to the defaulters. We can conclude that While people who have car default more often, the reason could be there are simply more people without cars
Looking at the percentages in both the charts, we can conclude that the rate of default of people having car is low compared to people who don't.

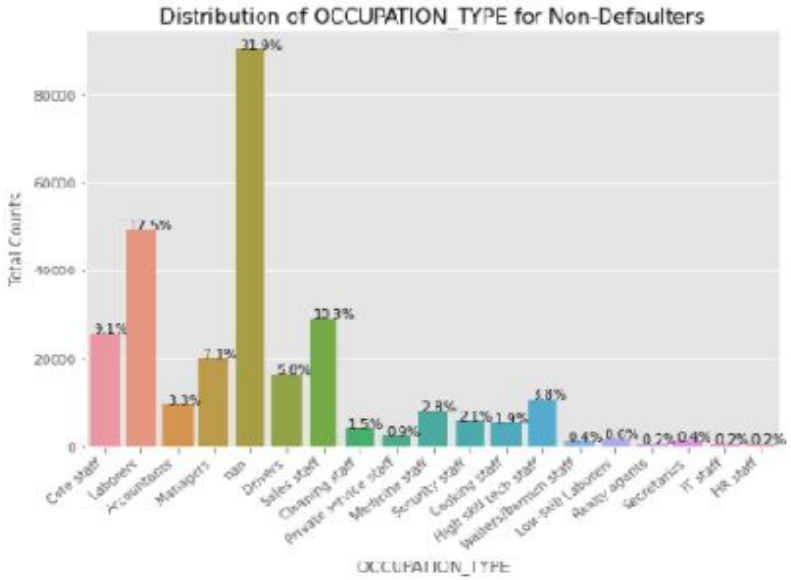
Age Group Data

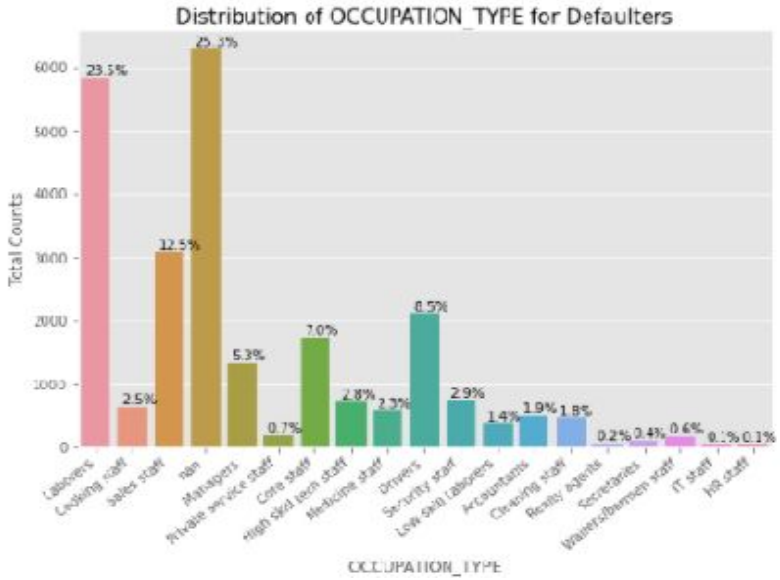




We see that (25,30] age group tend to default more often. So they are the riskiest people to loan to.

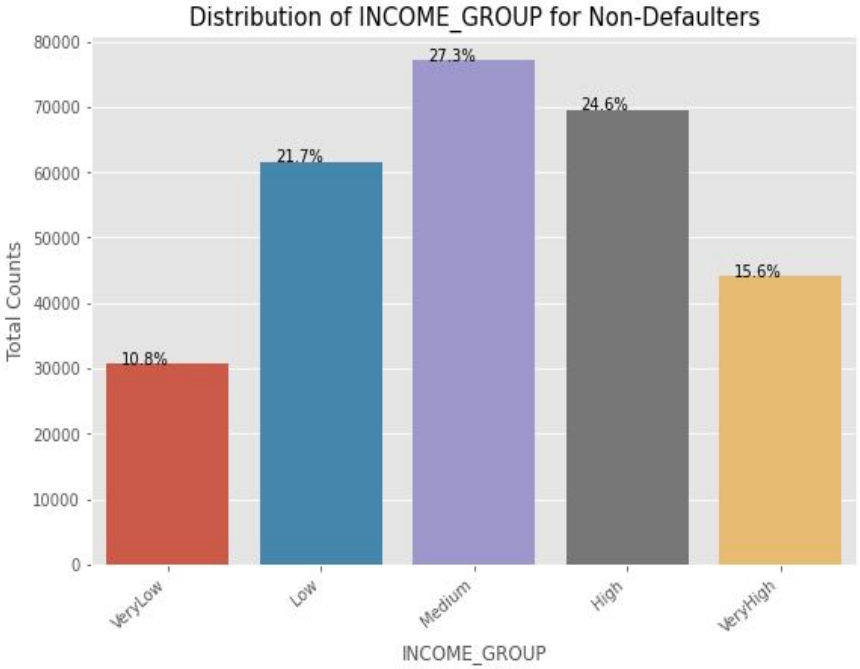
Occupation type Data

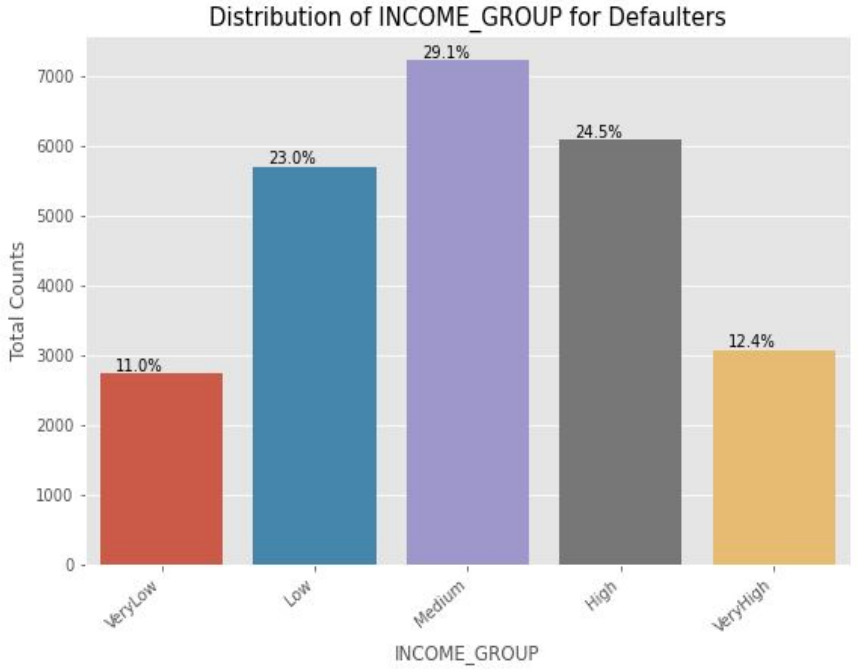




It is clear from the graph that in both defaulter and non-defaulter the labourers percentage. Comparing to defaulter the non defaulter laborer percentage are little bit high

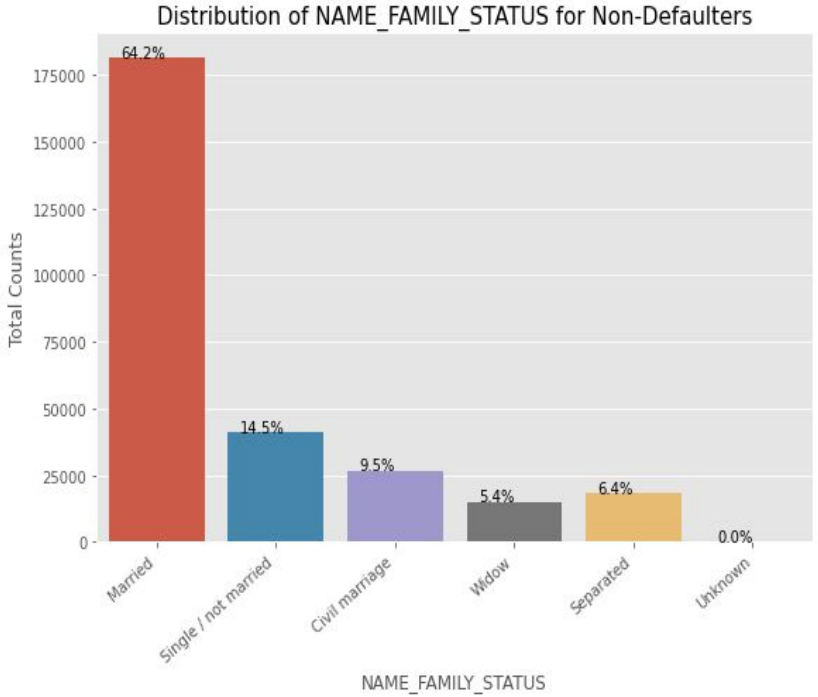
Income group Data

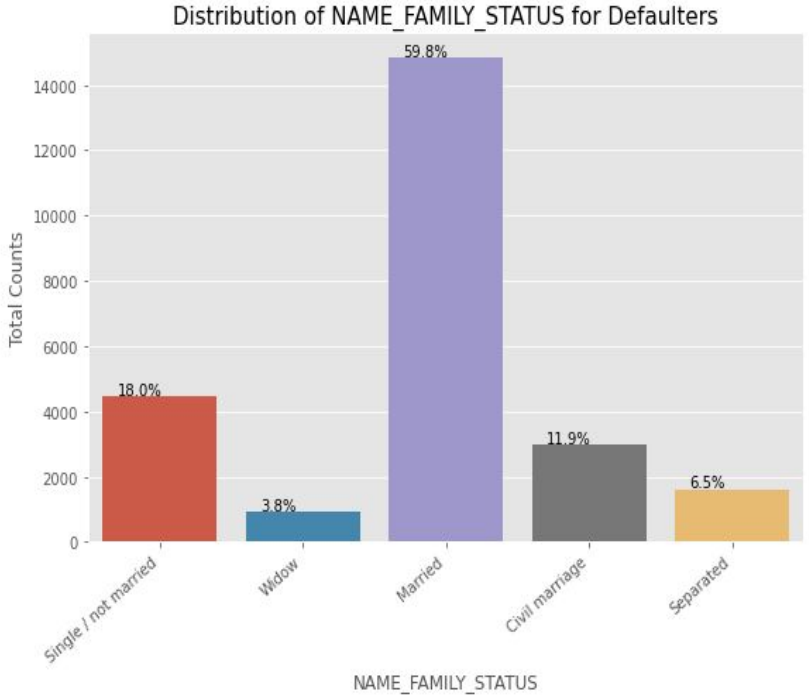




The Medium Income group people tend to take loans more frequently and they contribute the most in the defaulters.

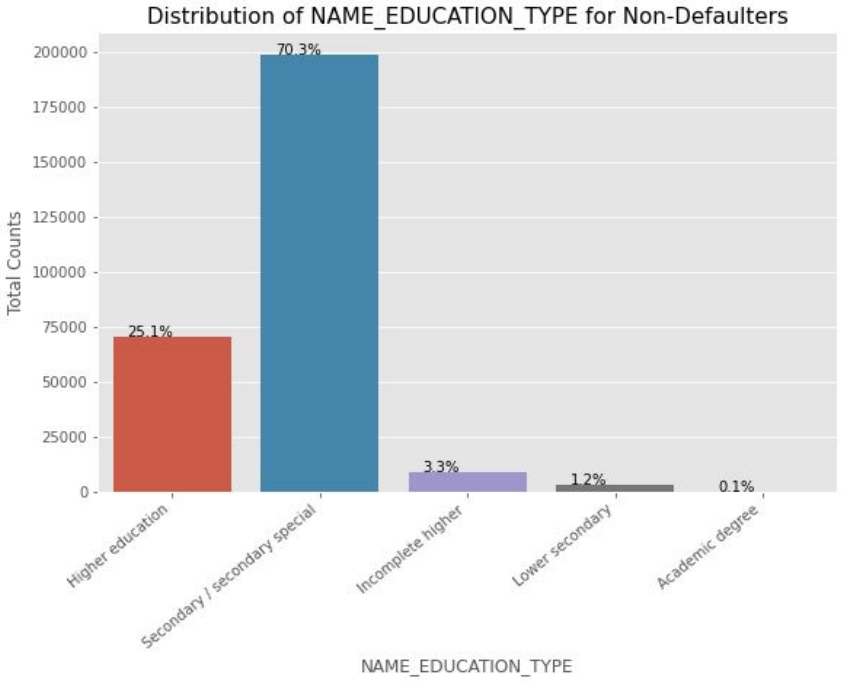
Family Status Data

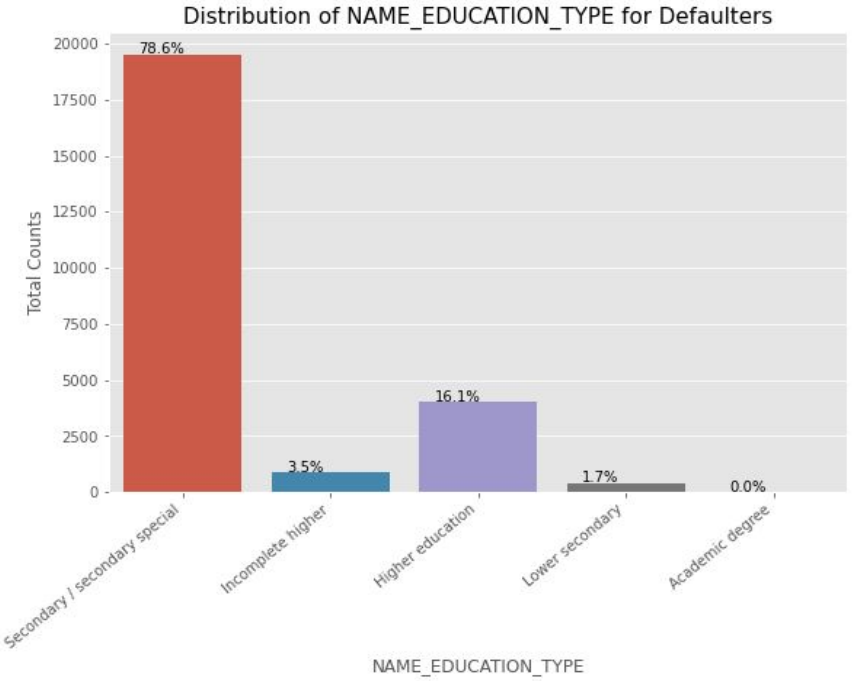




Married people tend to apply for more loans compared to others.
But from the graph we see that Single/non Married people contribute 14.5% to Non Defaulters and 18% to the defaulters. So there is more risk associated with them.

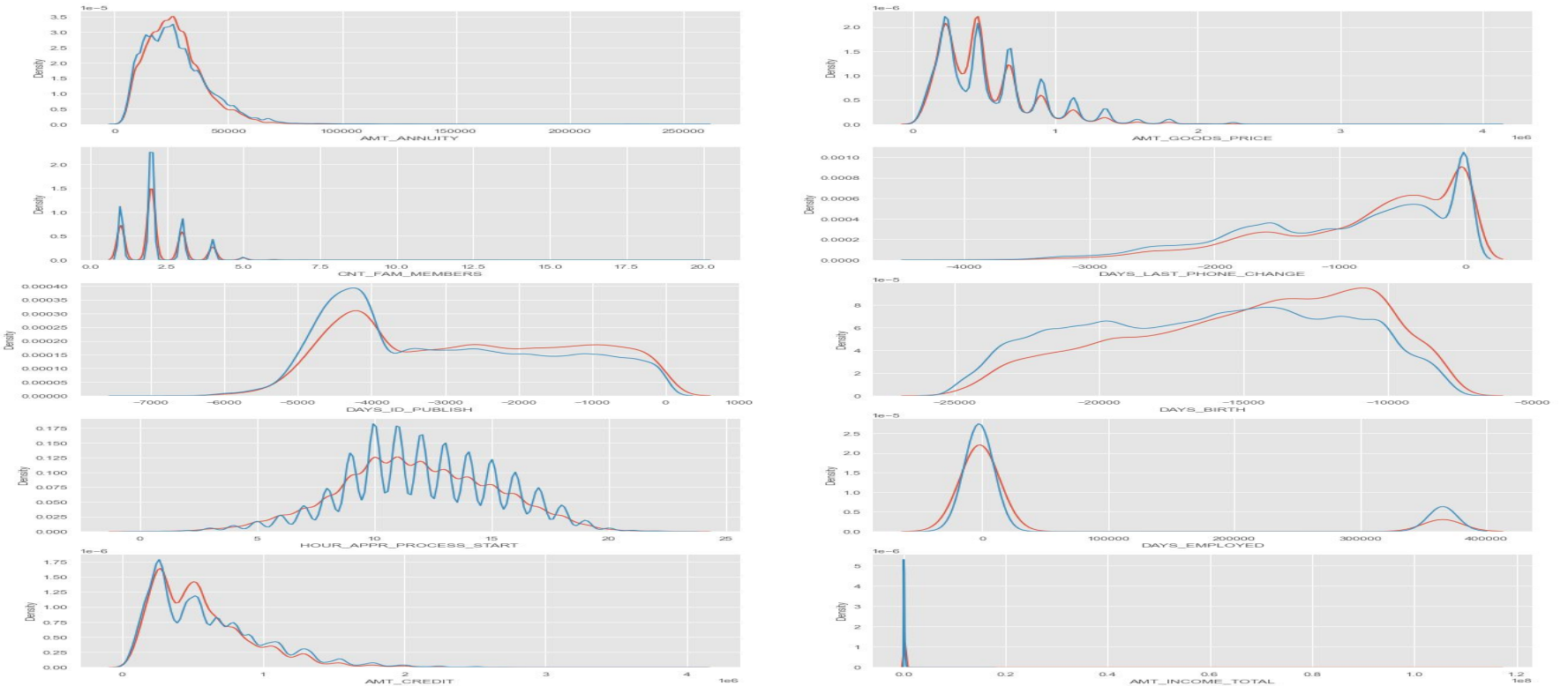
Education Type Data



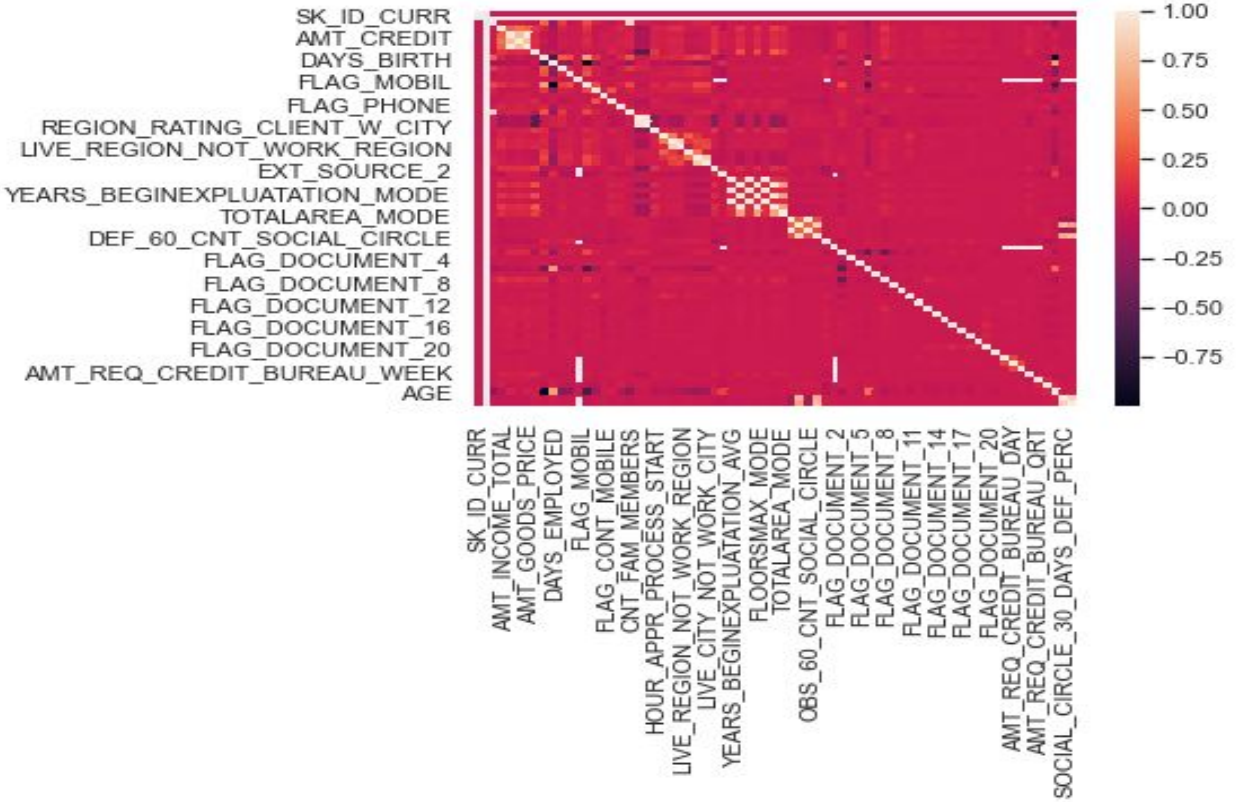


Almost all of the Education categories are equally likely to default except for the higher educated ones who are less likely to default and secondary educated people are more likely to default.

Univariate Analysis on Numerical Data

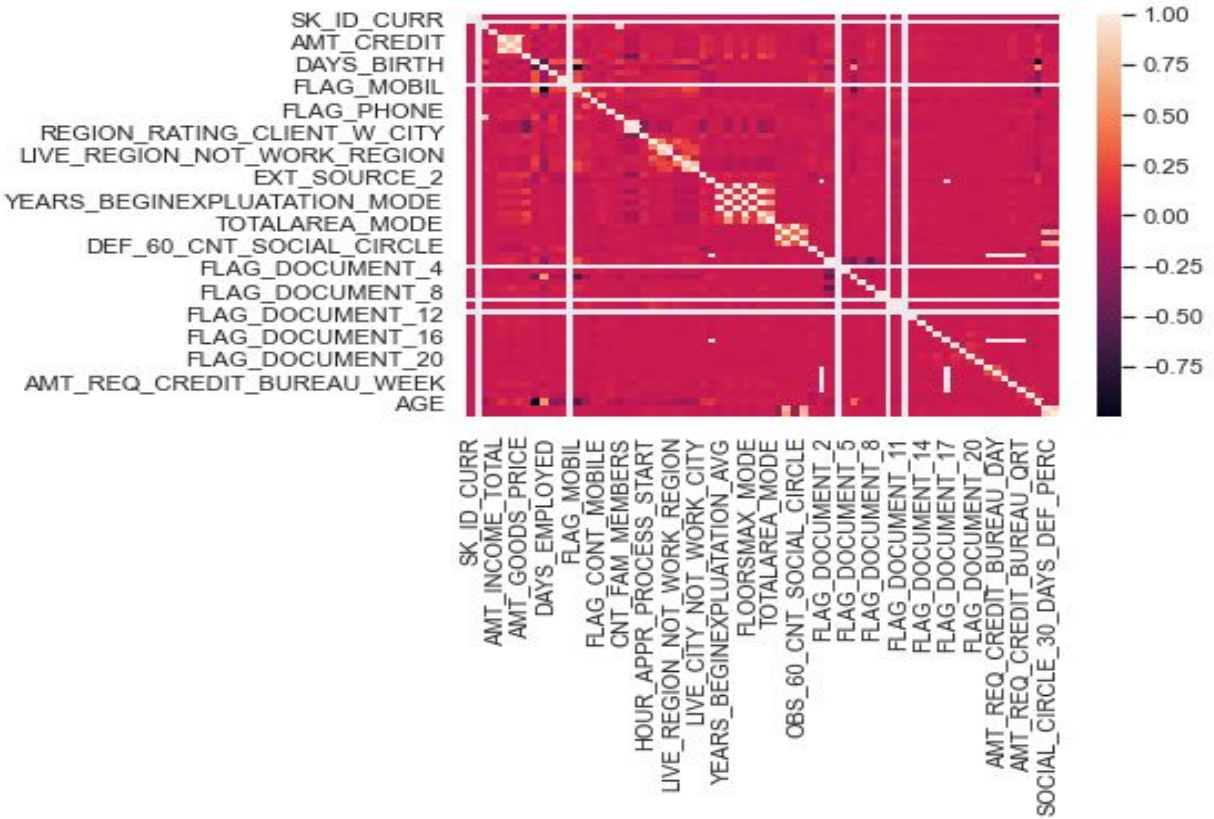


Top 10 Correlation



Target0

Top 10 Correalation



Target1

Group more likely non-defaulter(Recommandation)

1. Client working as a State servant.
2. People who having Pension.
3. People who has own car.
4. People who are core staff.
5. People who are in the age between 40-65.
6. People with high income.
7. People who are having higher education level.
8. Client whose previous loan was approved.

Group more Likely to Default(Risky group)

1. People between the age of 25 to 30 are more likely to default.
2. People with lower income range.
3. People with Lower Education are likely to default.
4. People whose loans have been refused earlier.
5. Civil Marriage and Single/not married people tend to default more.
6. People falling in the Labourer Occupation type.