

Credit EDA Case Study

By - Nilesh Rathor

Business Objective

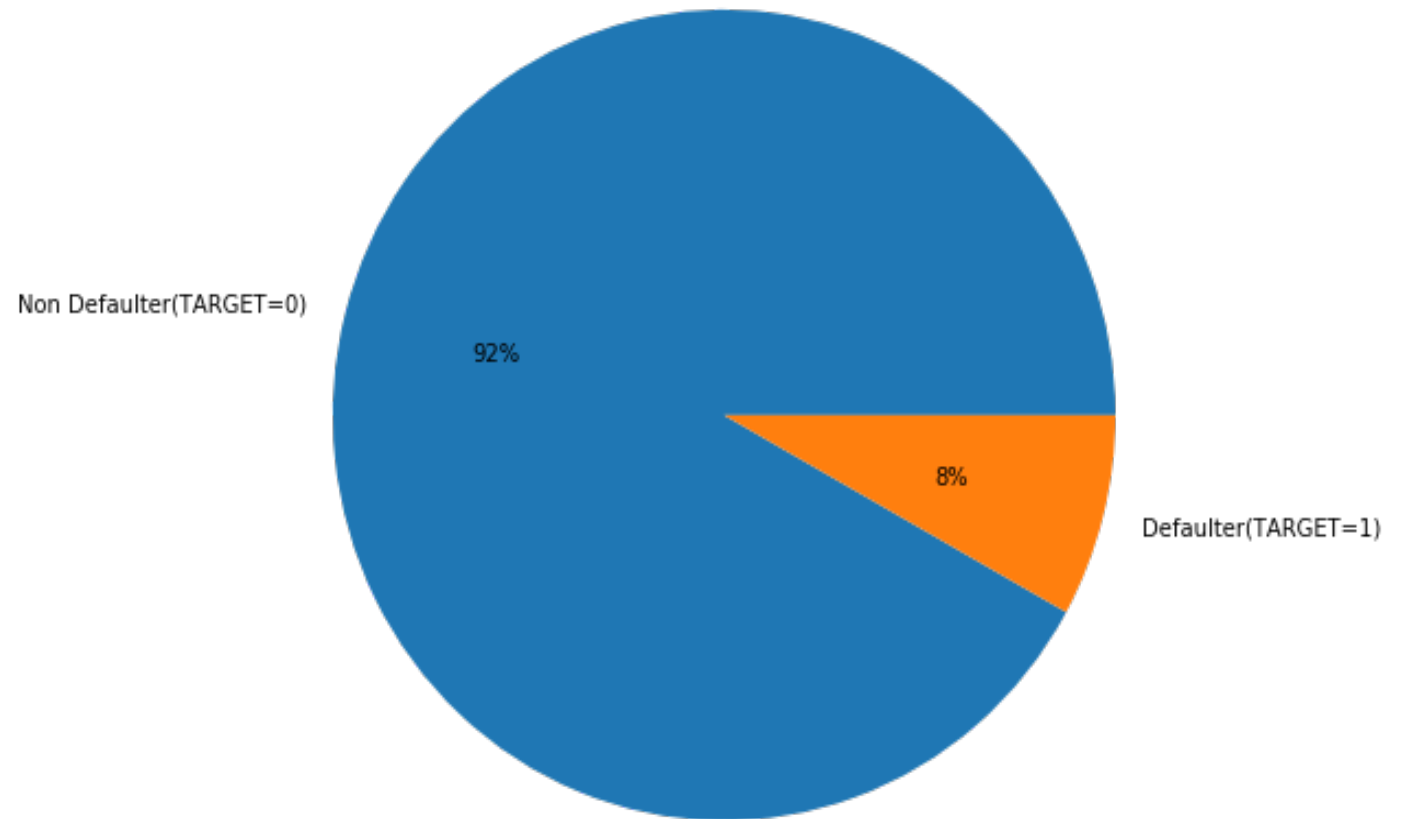
The aim of this case study is to identify the applicants who are capable of repaying the loans and in order to do so the patterns are identified which indicate if a person is having difficulty in paying their installments.

The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Data Imbalance of Target

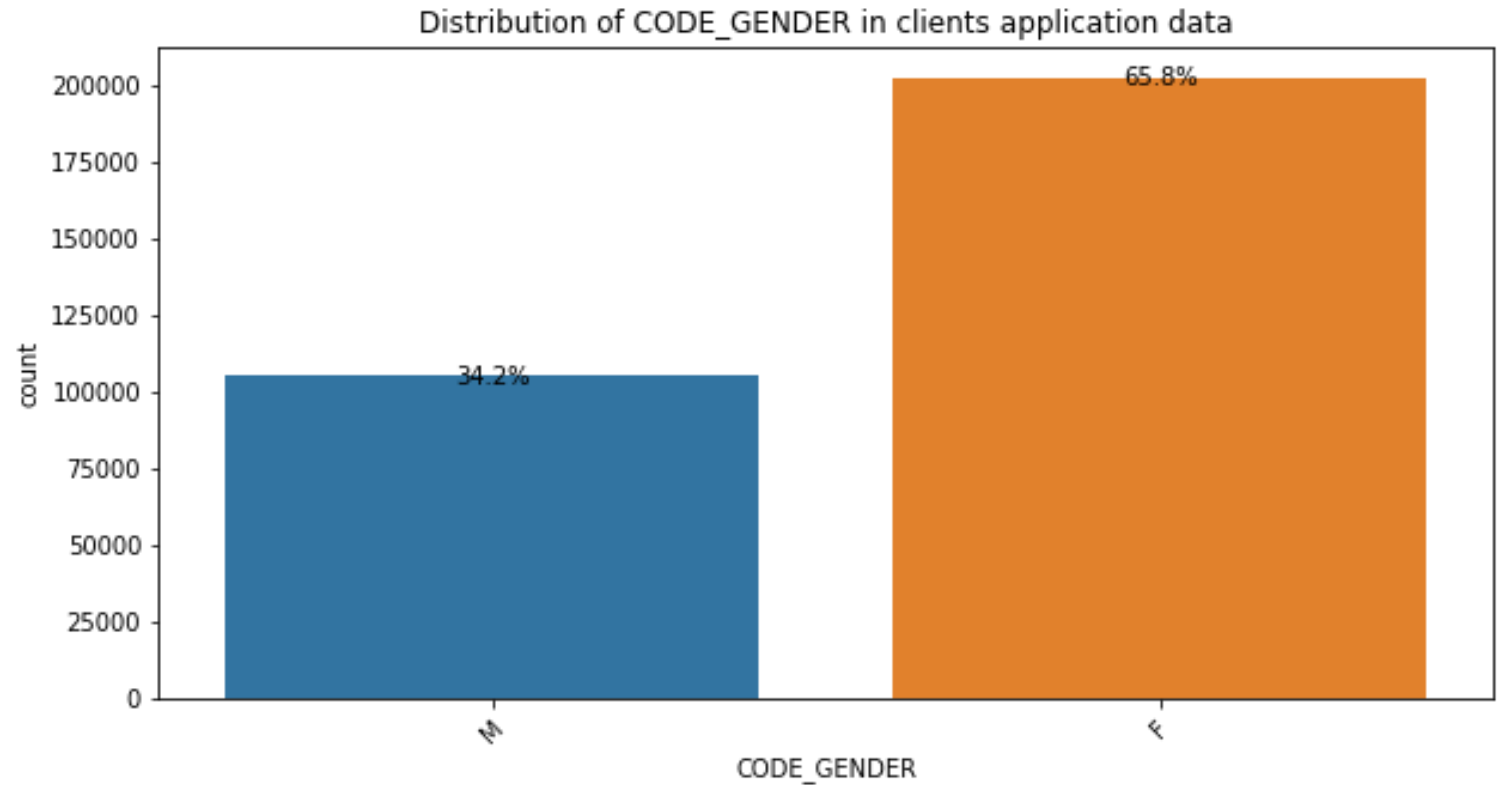
- There is an imbalance between defaulters and not defaulters. The non defaulters are 12 times the defaulters.

TARGET Imbalance - Defaulter vs Non Defaulter



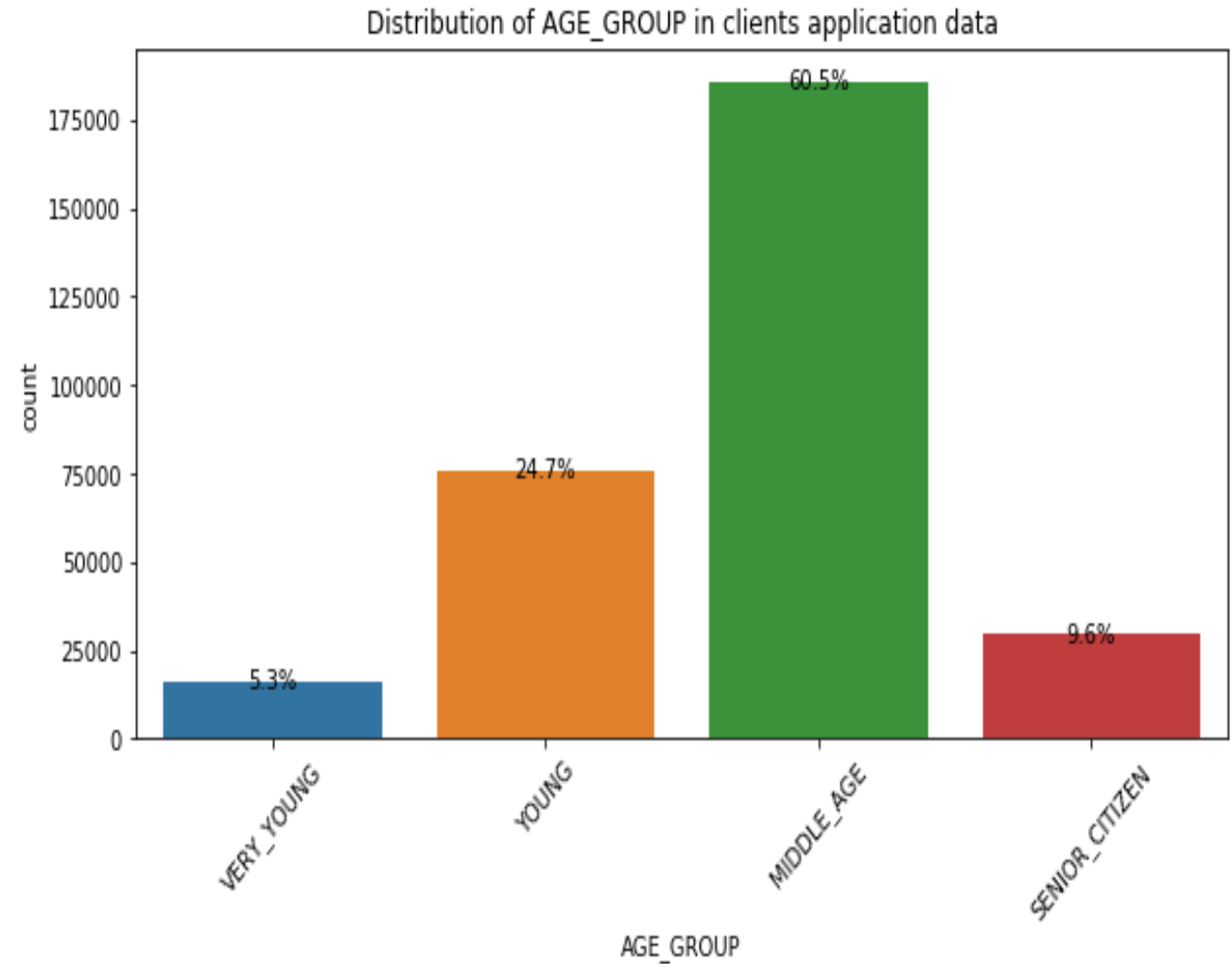
Gender Distribution in Loan Application

- The above graph shows that the Females applied for loans more than Males, with Female percent as 65.8% and 34.2% for Males
- Females applies for loans almost twice as males



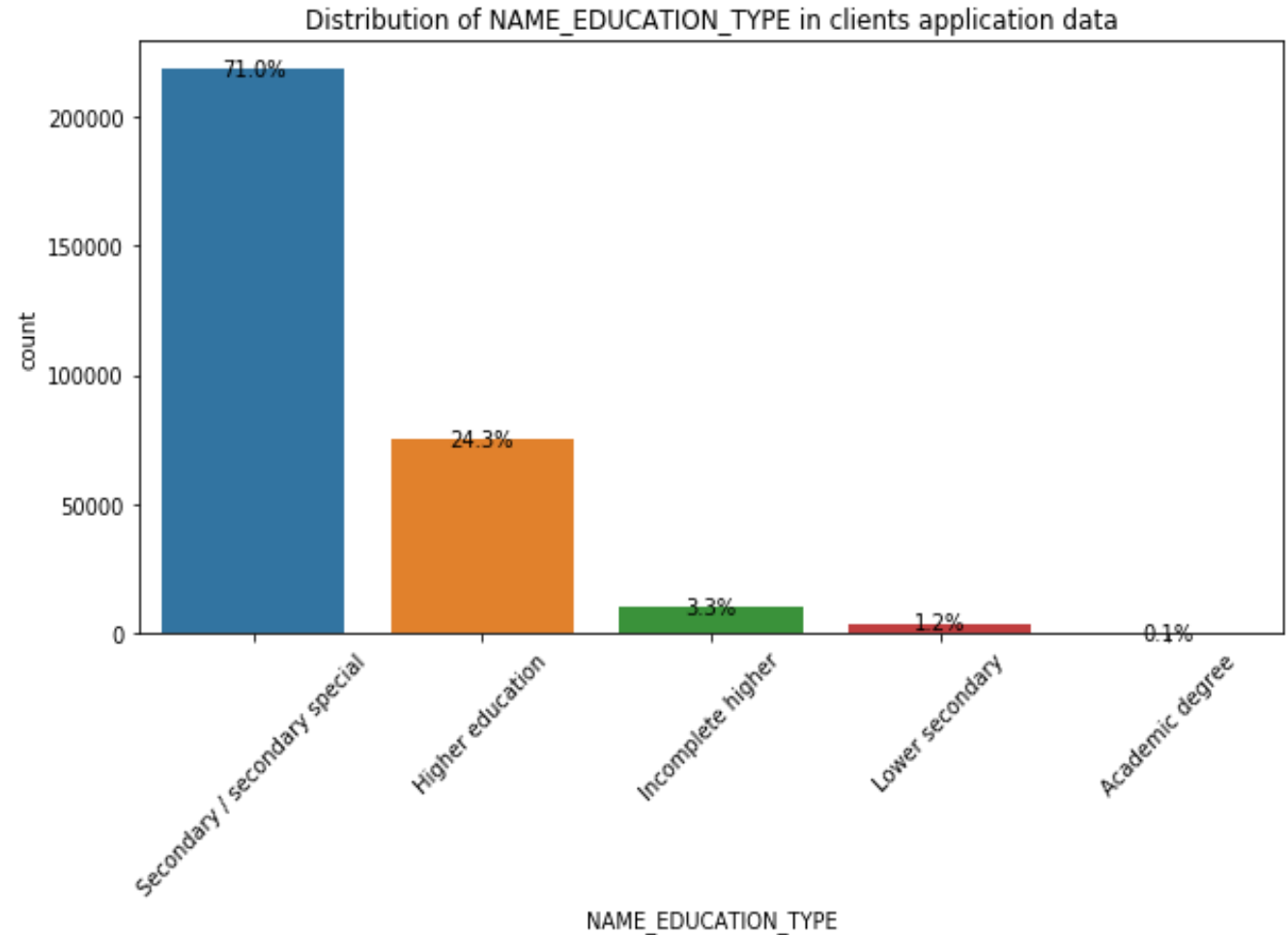
Age Group Distribution for Loan Application

- The middle age group applications have the highest percentage of 60.5



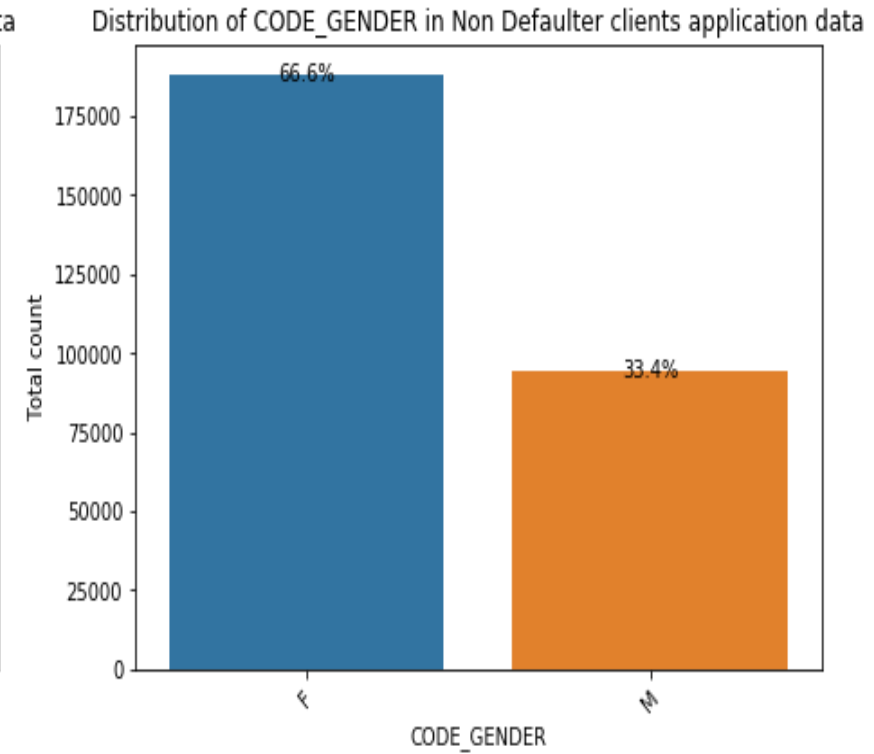
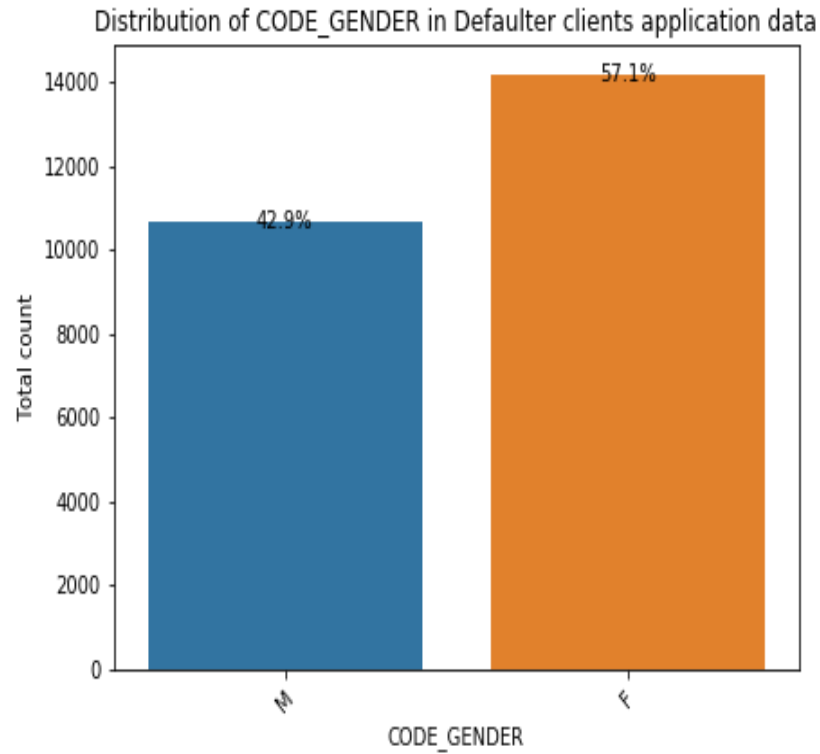
Education Type Distribution for Loan Application

- The clients with secondary education applies highest with 71% and least being applications with academic degree



Target Segmented Univariate Analysis – Gender Distribution

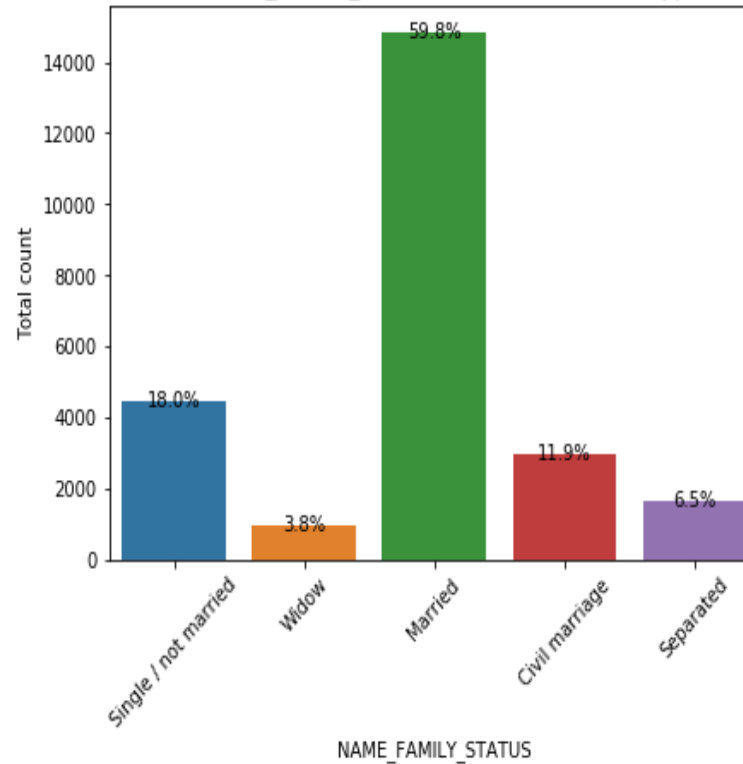
- The female clients applies for loans more than males, and higher values in both segments, the values being 57.1% and 66.6% in Defaulters and Non Defaulters Data respectively
- It can be inferred that females defaults less than males



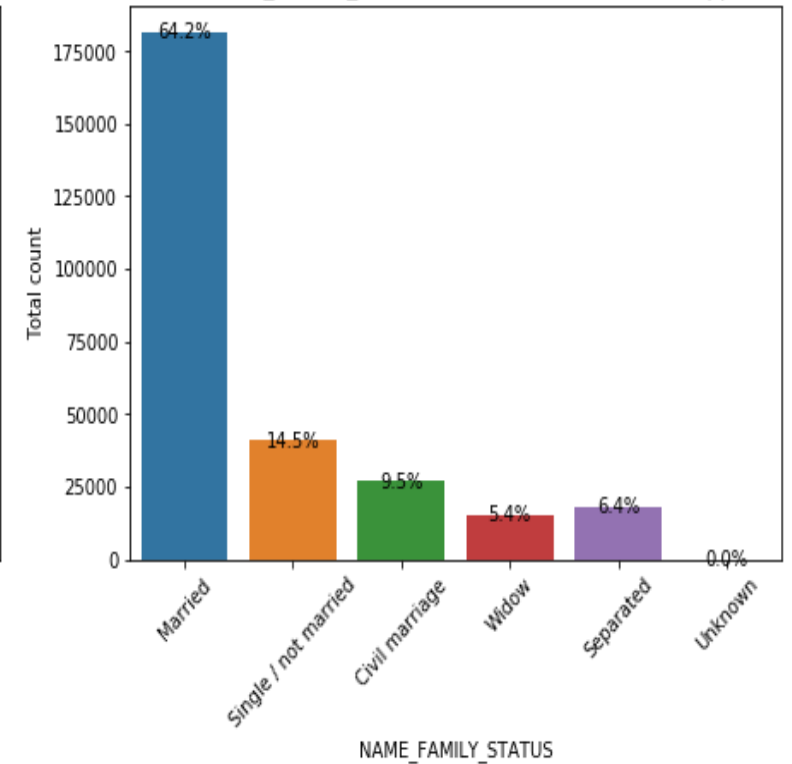
Target Segmented Univariate Analysis – Family Status

- Married people applies the most for loans
- only single people have higher defaulter percentage than their non defaulters percentage, with 18% defaulter rate and 14.5% non defaulter rate
- Providing loans to single people have higher risks of defaulters

Distribution of NAME_FAMILY_STATUS in Defaulter clients application data

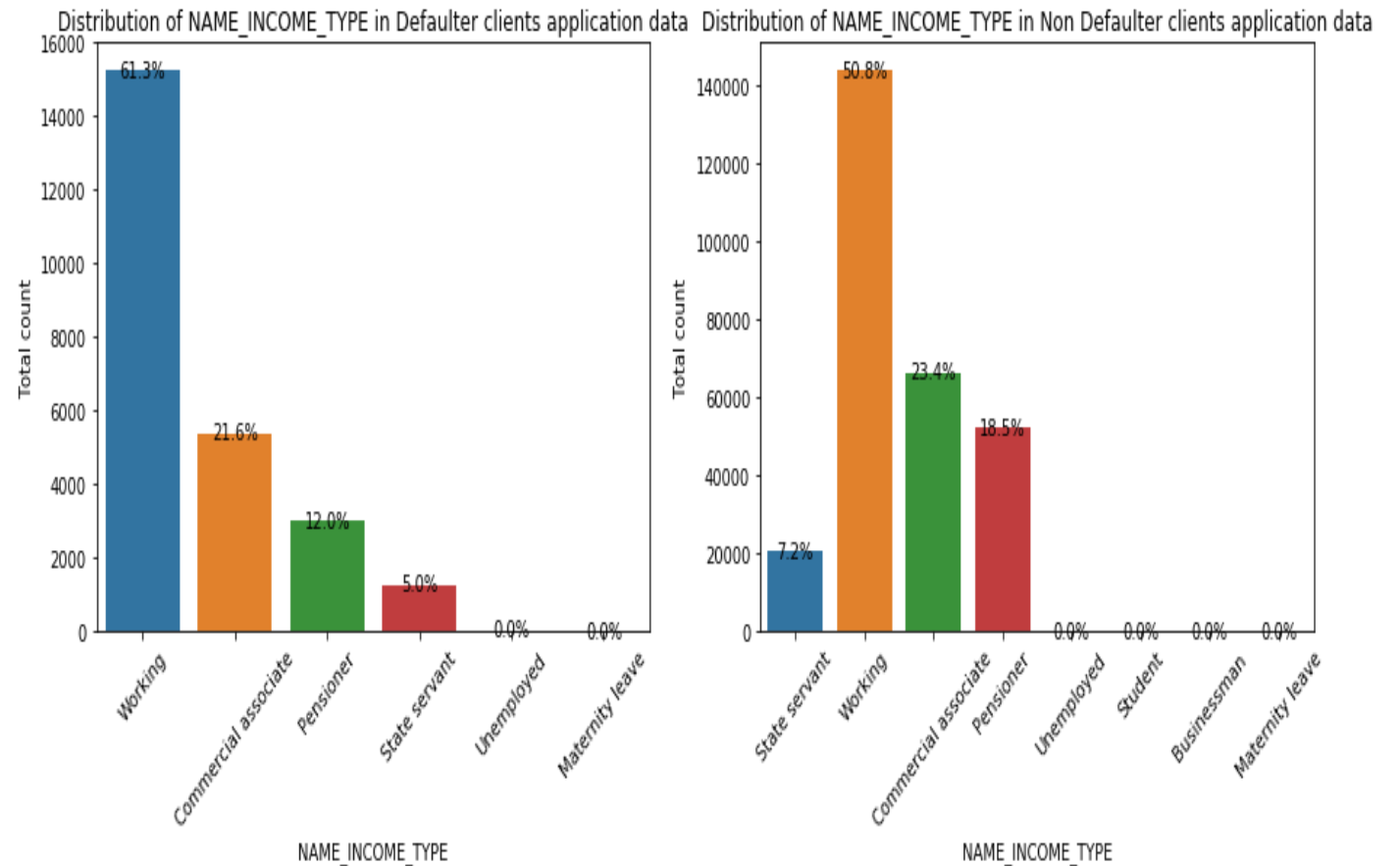


Distribution of NAME_FAMILY_STATUS in Non Defaulter clients application data



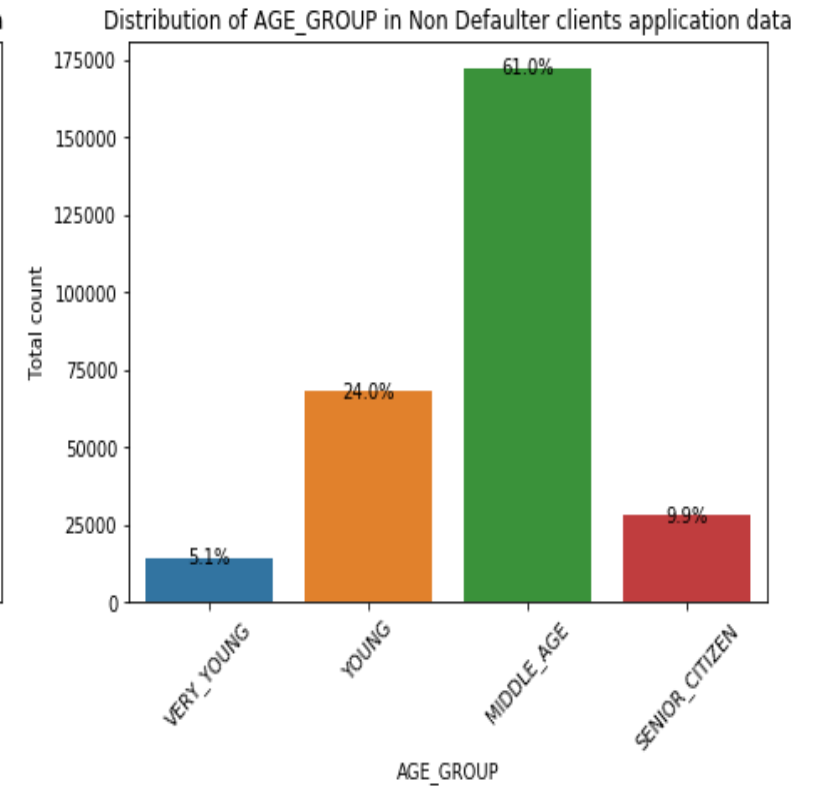
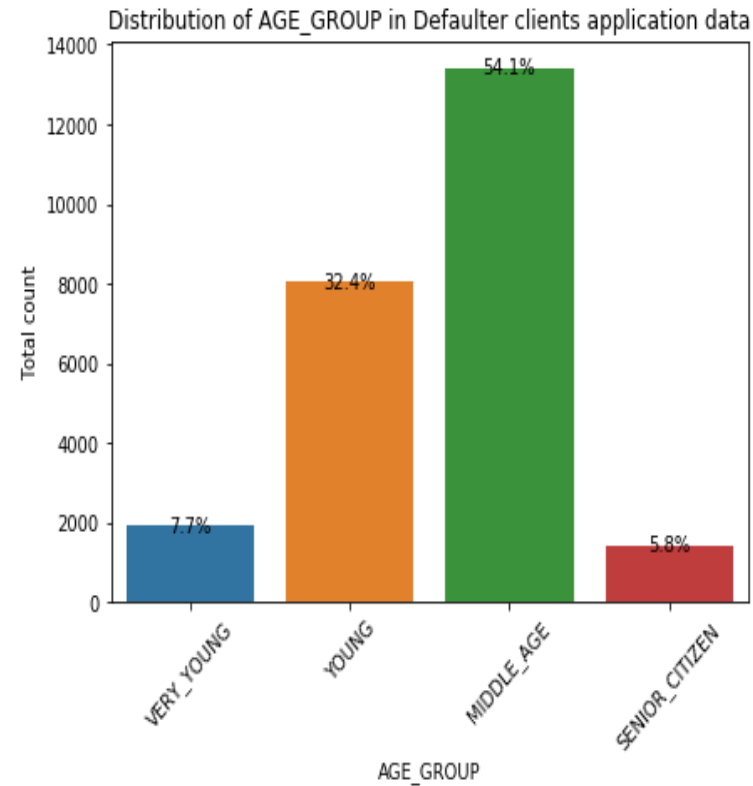
Target Segmented Univariate Analysis – Income Type

- The plot provides insights that the students and Businessmen don't default



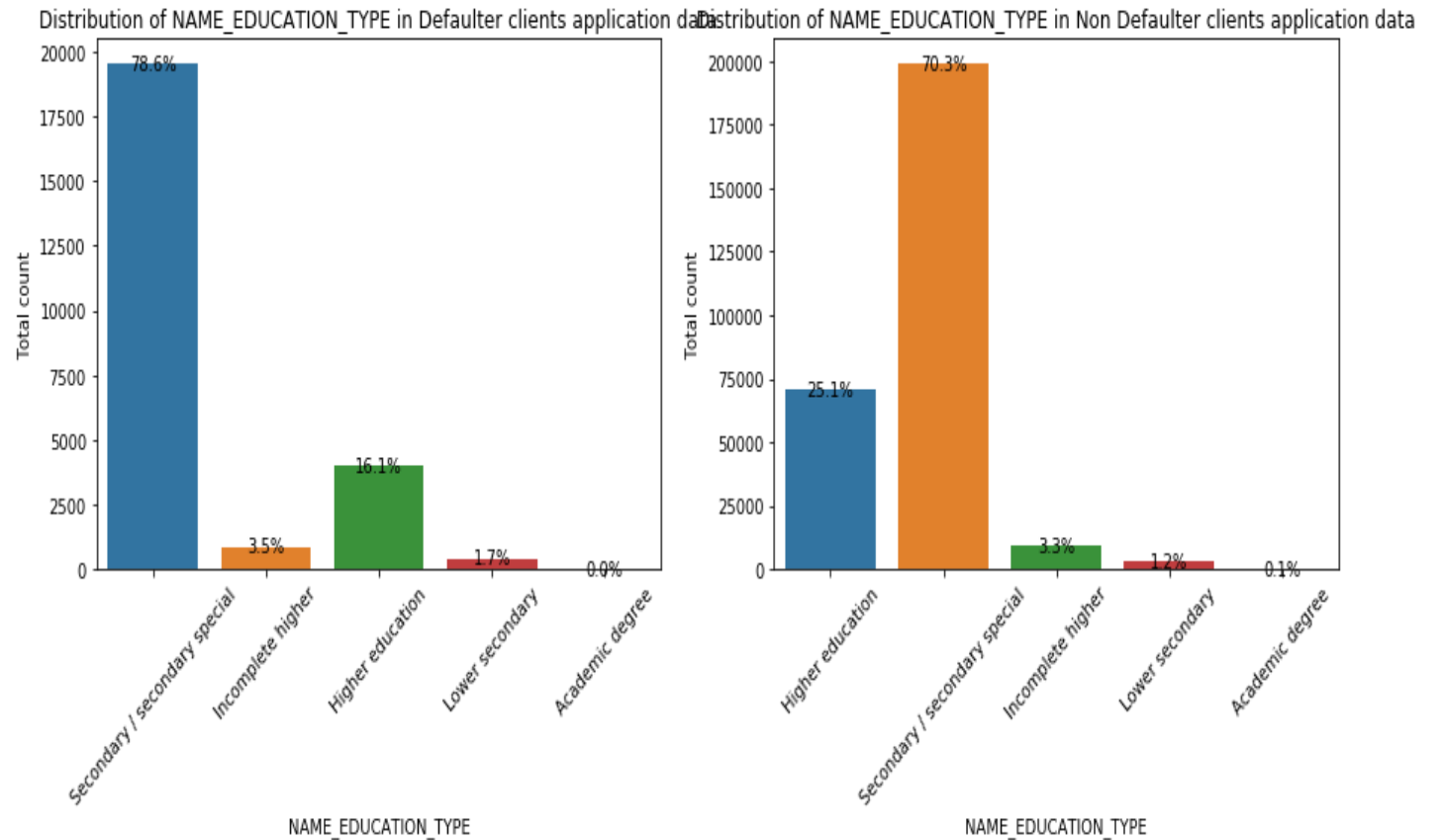
Target Segmented Univariate Analysis – Age Group

- It can be inferred that very young and young age group defaults are more than their non default rate hence more risk associated with these two age groups



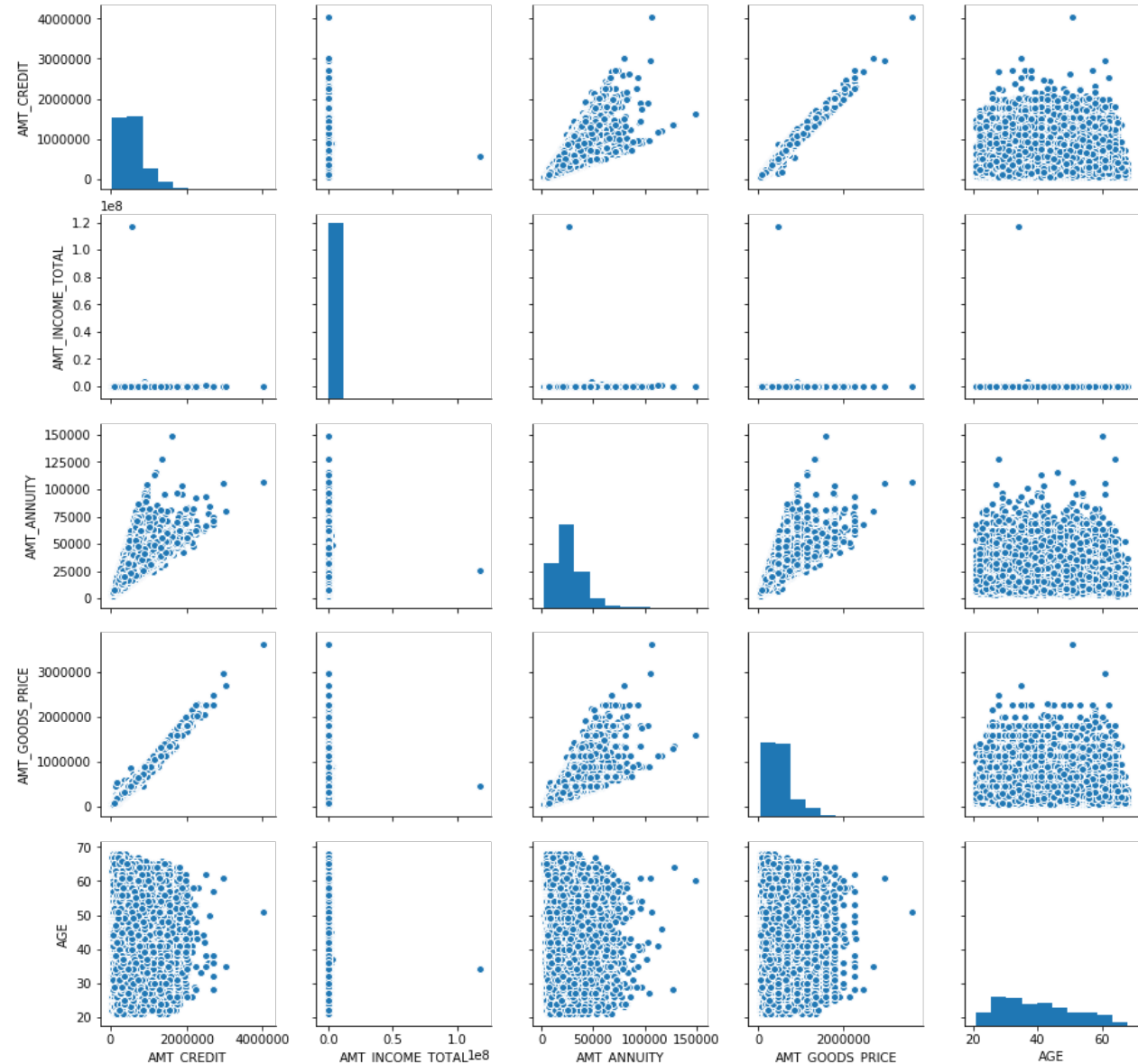
Target Segmented Univariate Analysis – Education Type

- Secondary educated clients have higher defaulter percentage than their non default rates
- Higher Risks associated with this group



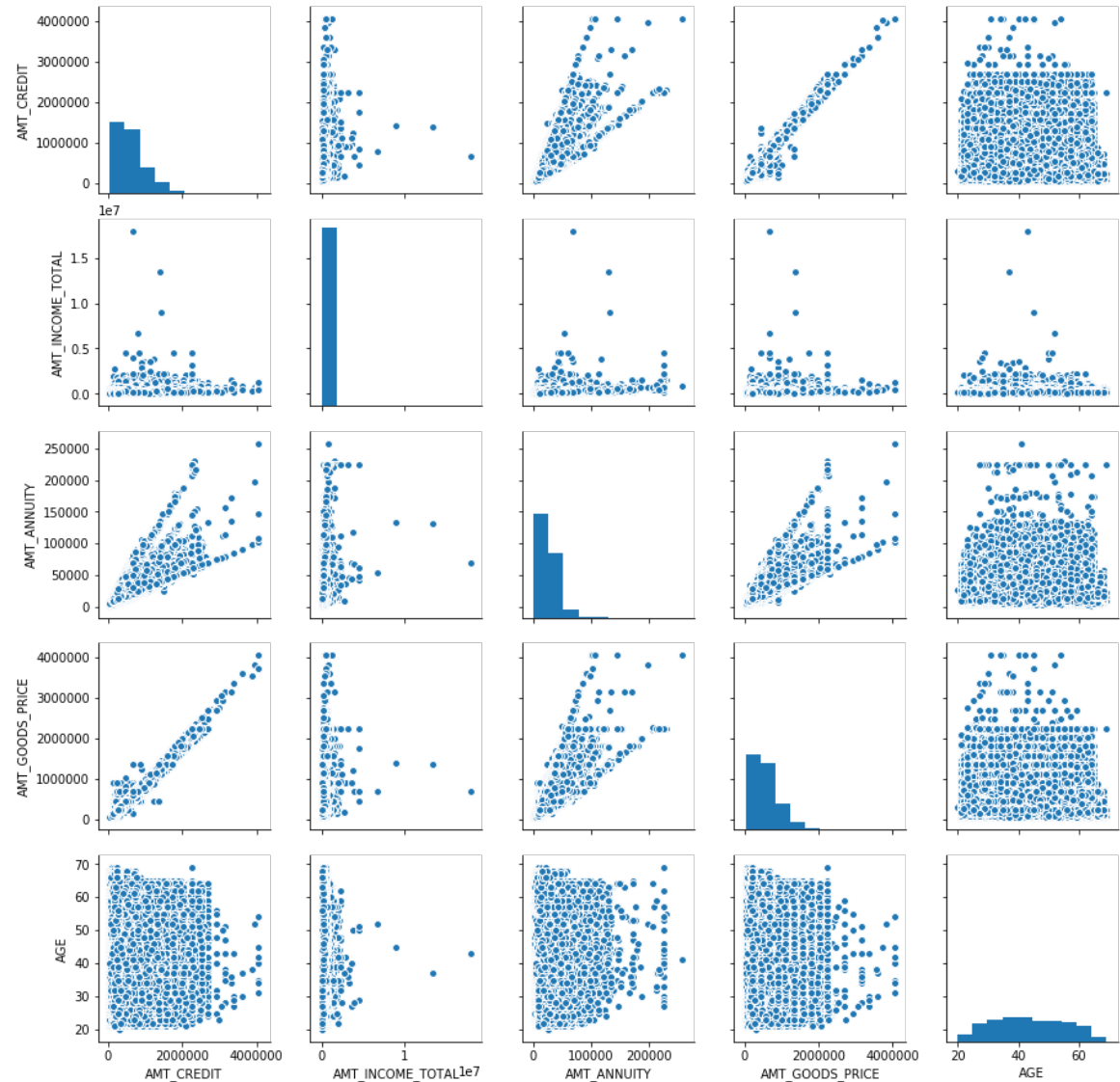
Bivariate Analysis – Numeric Vs Numeric – Defaulters data

- There is a high correlation between amount goods price and amount credit for non defaulter data
- There are irregularities present between income and credit amount
- There are irregularities present between income and credit amount



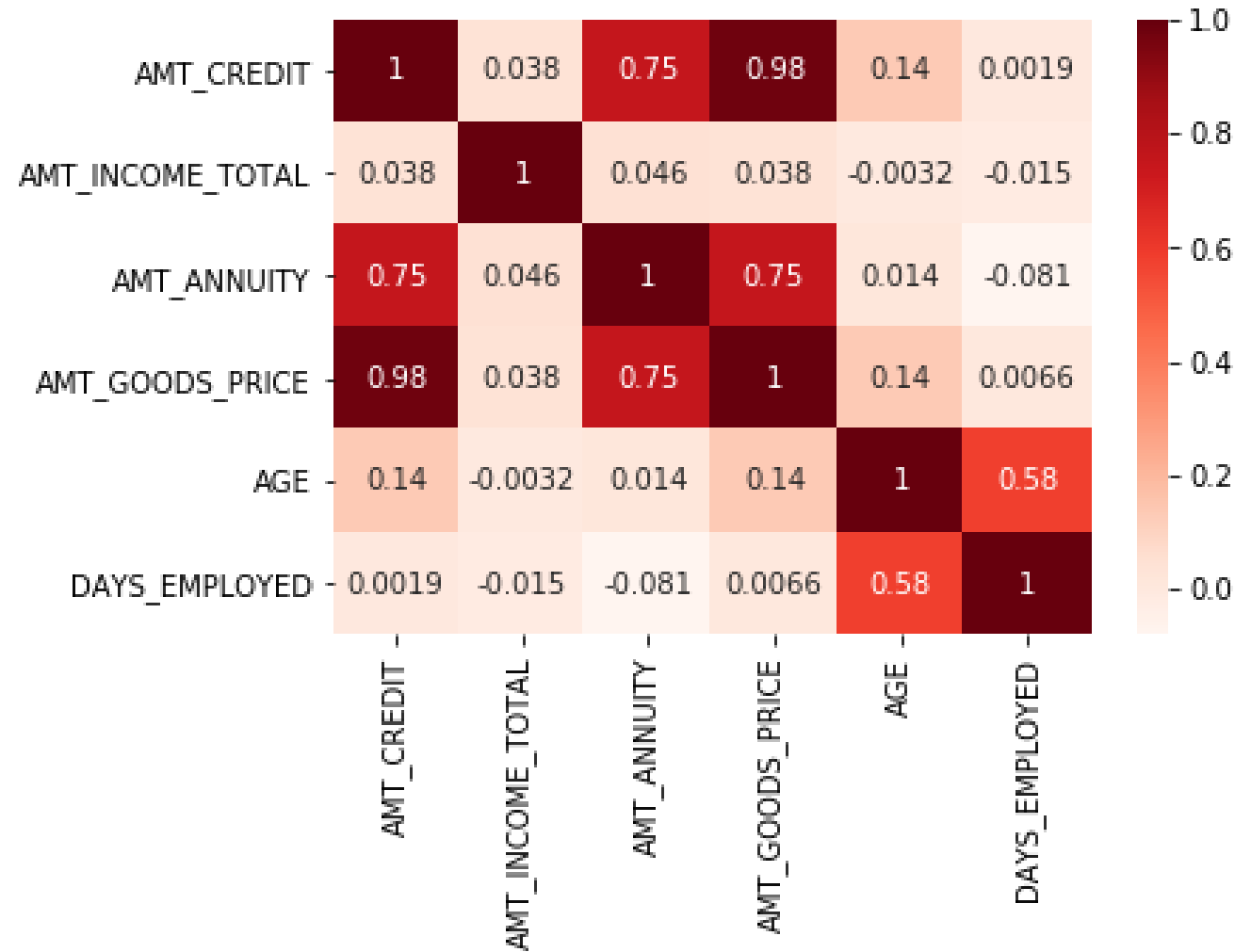
Bivariate Analysis – Numeric Vs Numeric – Non Defaulters data

- There is a high correlation between amount goods price and amount credit for non defaulter data
- There are irregularities present between income and credit amount
- There are irregularities present between income and credit amount



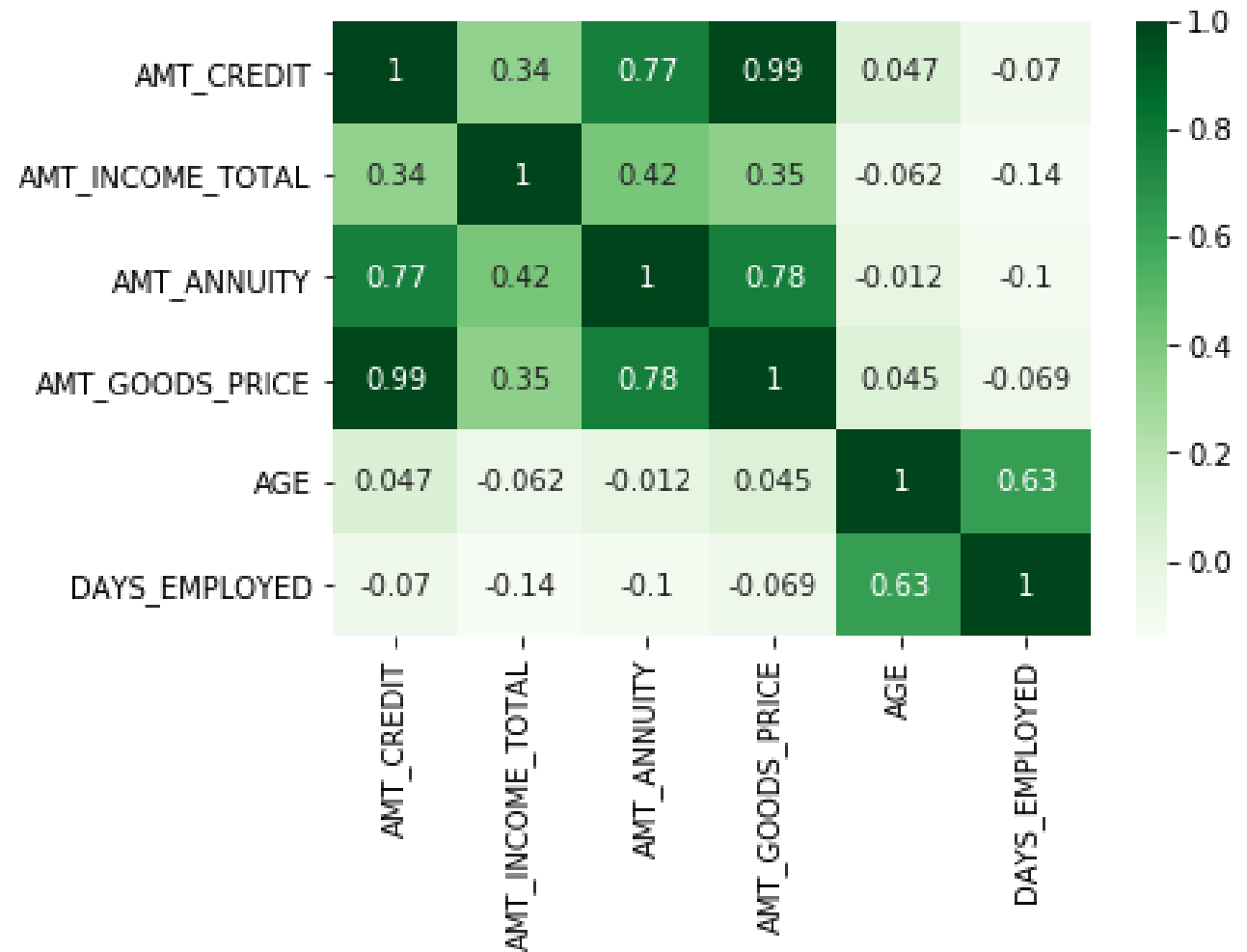
Correlation for Defaulters Data

- Plotting heatmaps for correlation in continuous numeric variable



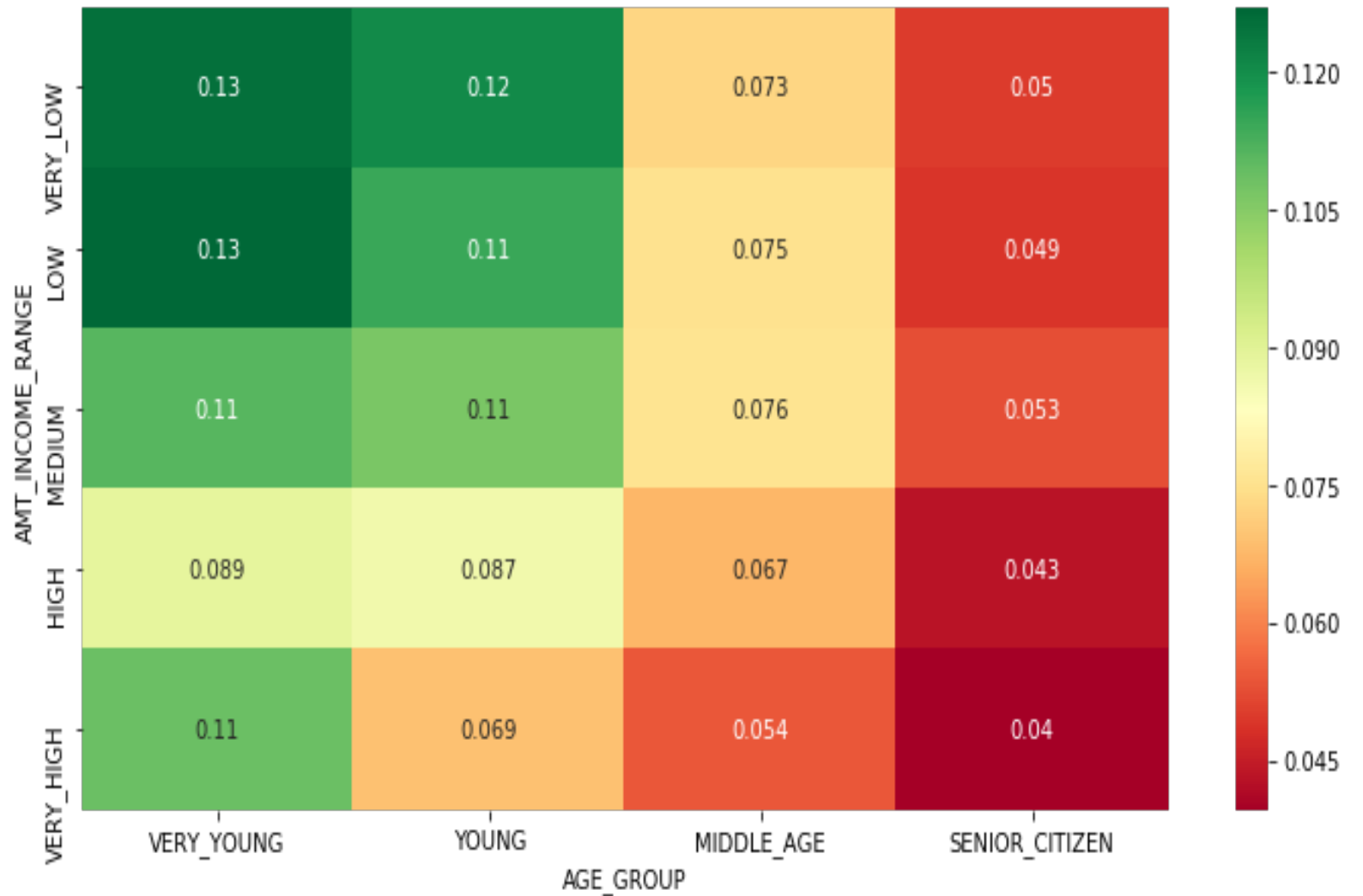
Correlation for Non Defaulters Data

- Plotting heatmaps for correlation in continuous numeric variable



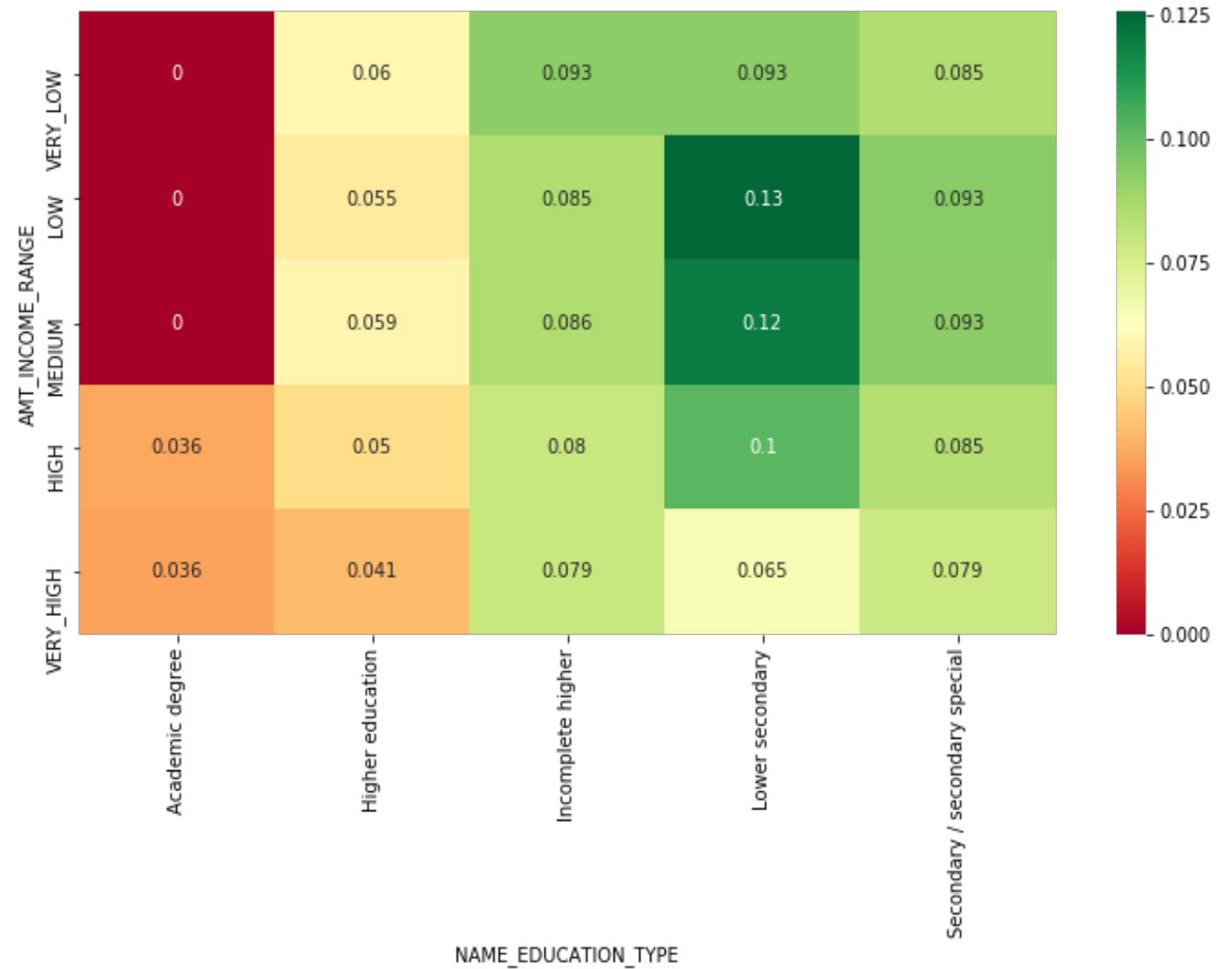
Multivariate Analysis

- The variables used are age group and income range
- It can be inferred that clients in the very young age group with very low and low income range have higher chances of defaults.
- Senior citizen with very high incomes are least likely to default



Multivariate Analysis

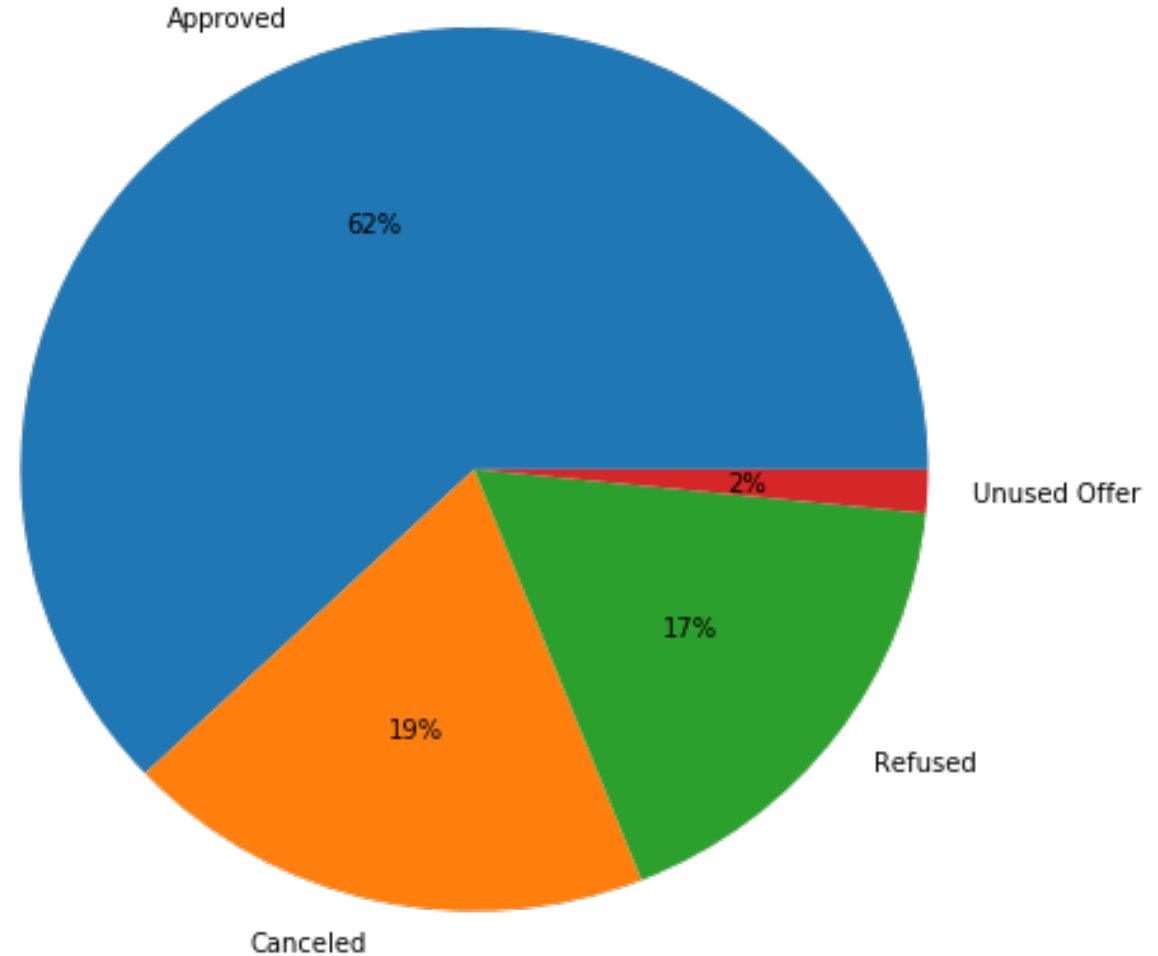
- The variables used are education type and income range
- It can be inferred that clients with lower secondary education and low income range are likely to default



Previous Application Data – Contract Status Distribution

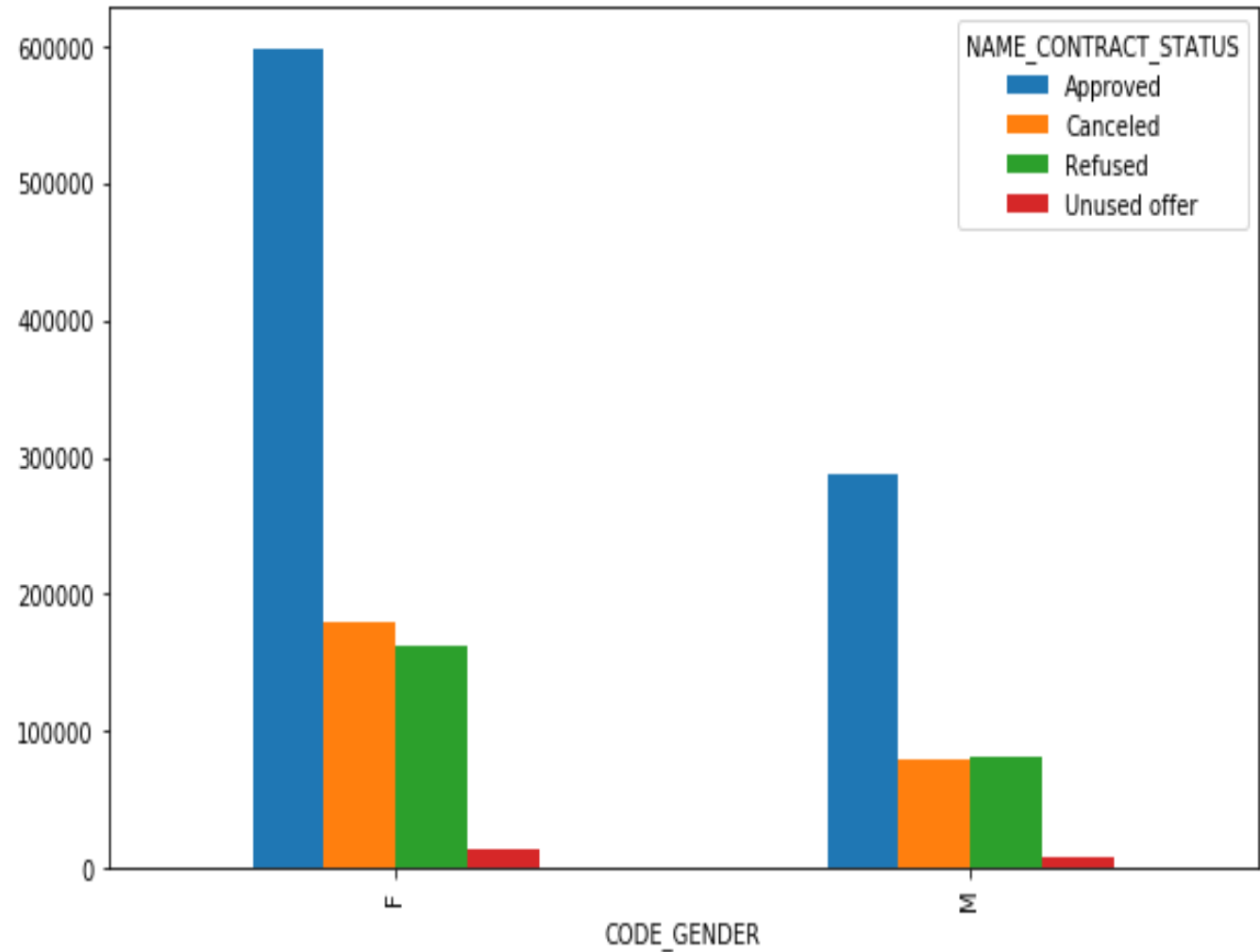
- Approved: The Company has approved loan Application
- Cancelled: The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.
- Refused: The company had rejected the loan (because the client does not meet their requirements etc.).
- Unused offer: Loan has been cancelled by the client but on different stages of the process

Distribution of contract status of previous application



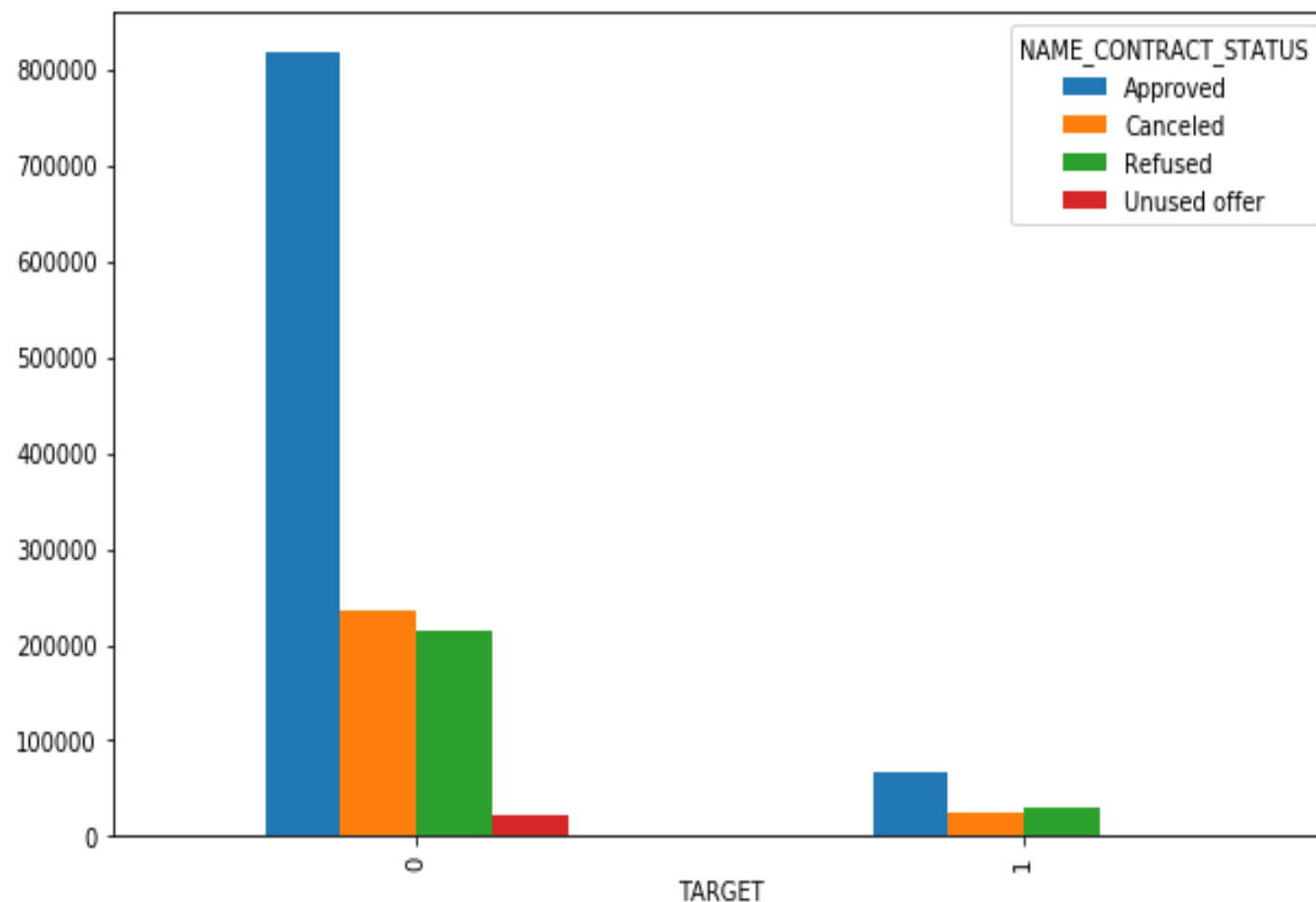
Merged Application and Previous Application Data Analysis

- Gender Distribution
- It can be inferred that female clients applications have higher approval rates for loans and the reasons associated can be less defaulters than males and higher number of applications



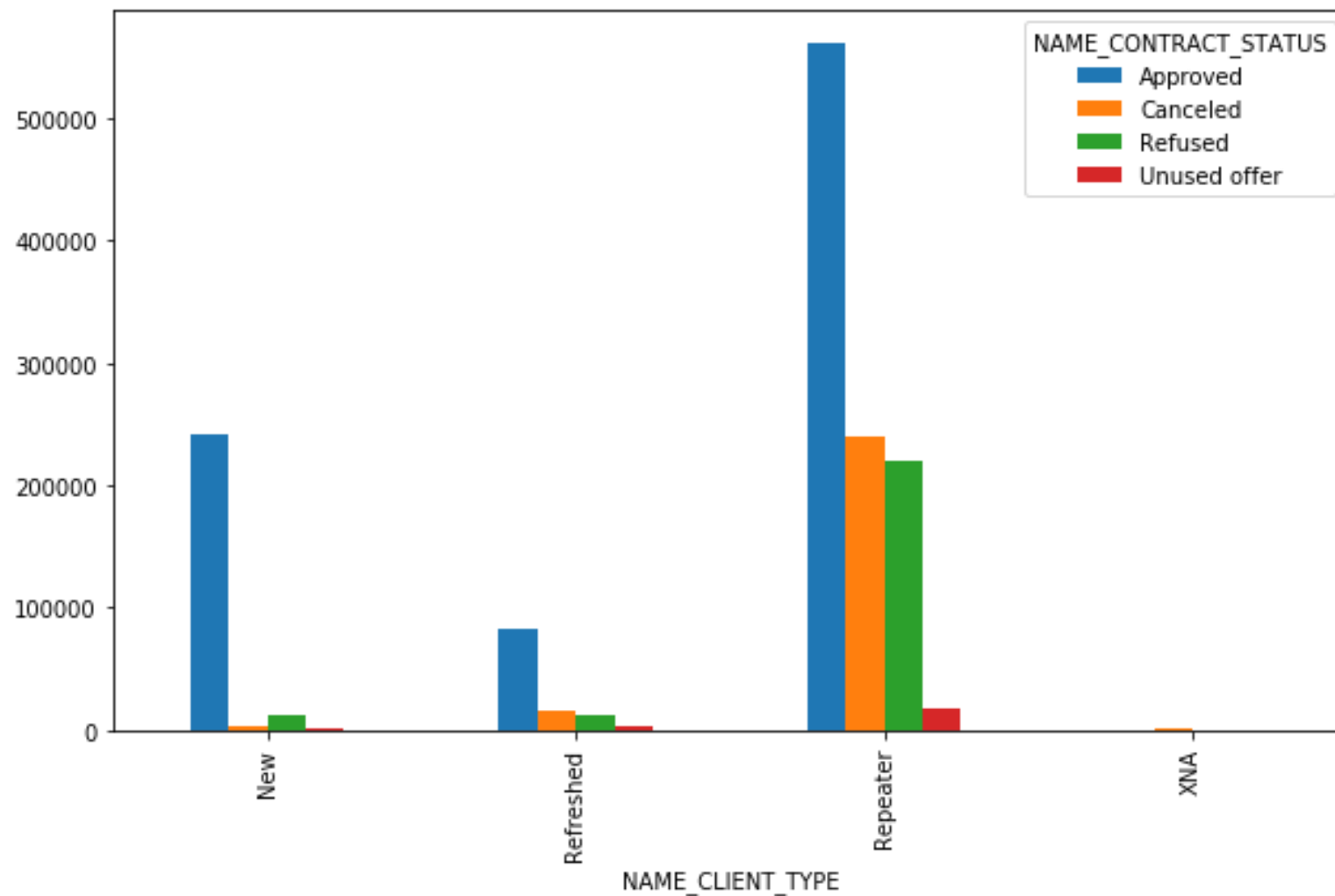
Merged Application and Previous Application Data Analysis

- Target Distribution – Defaulter and Non Defaulter
- It can be inferred that clients with previous approved application status have lower defaulter rates



Merged Application and Previous Application Data Analysis

- Client Type Distribution
- It can be inferred that clients old clients have higher approval rates



Thank You
