Presentation on Credit EDA Assignment

Understanding Risk Analytics of Bank & Financial Services

Project Submitted by-
Deepjyoti Chakraborty

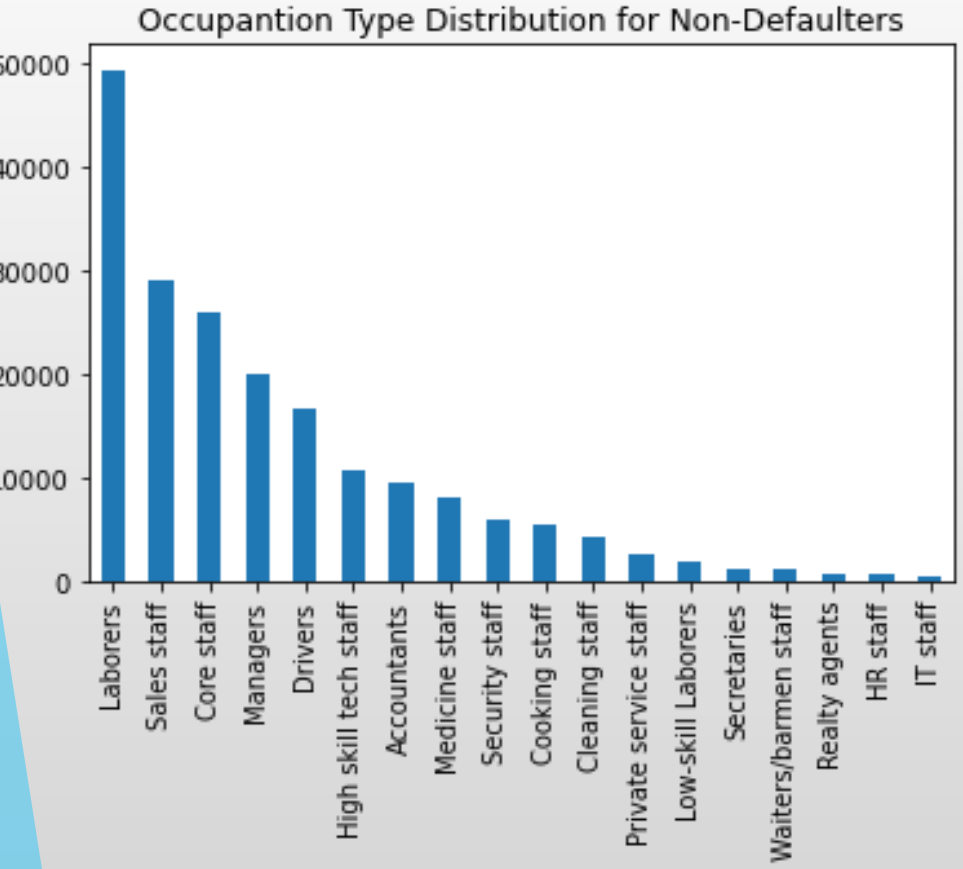
Business Objective

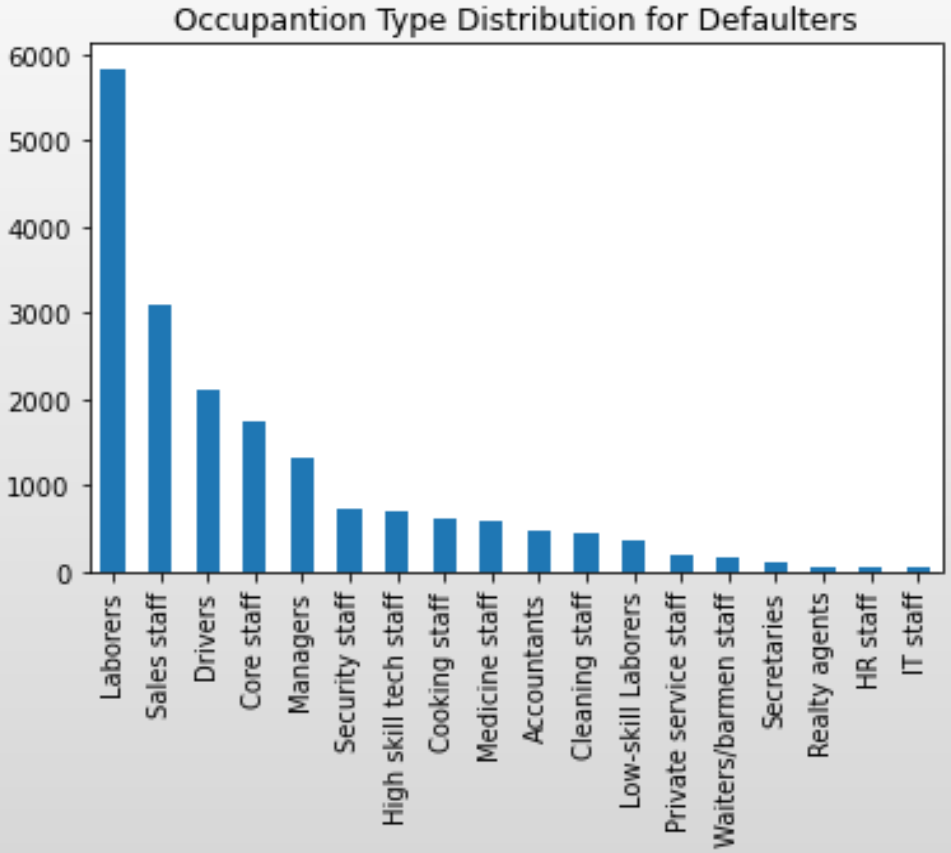
- This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

Performed Steps for Solution

- 1. Import Libraries and Datasets.
- 2. Observe the data set and do quality checks (Finding Missing values, Drop or replace values if needed, Check for outliers, bin variables wherever needed).
- 3. Check for data imbalance
- 4. Perform Univariate, Segmented Univariate, Bivariate analysis and also finding the correlations.
- 5. Merging application data with previous application data
- 6. . Perform data analysis by using Univariate, Segmented Univariate, Bivariate analysis and also finding the correlations.
- 7. Come to an conclusion

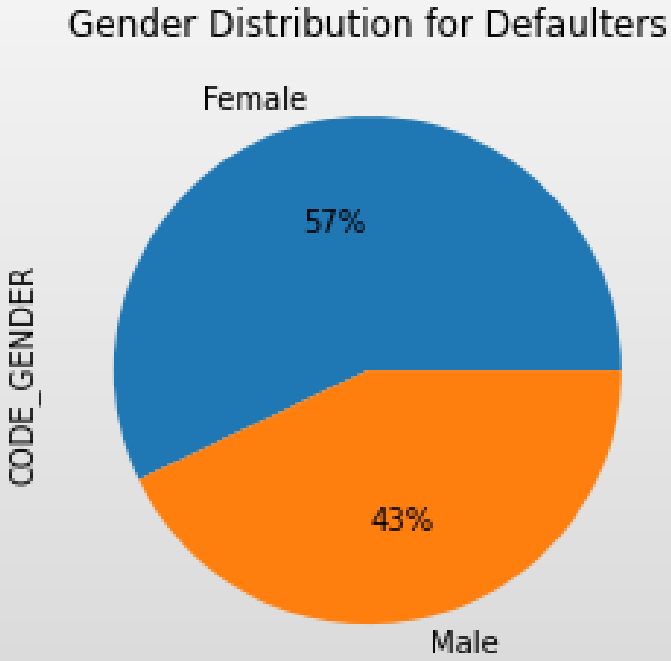
Univariate and Segmented Univariate Analysis for Application data

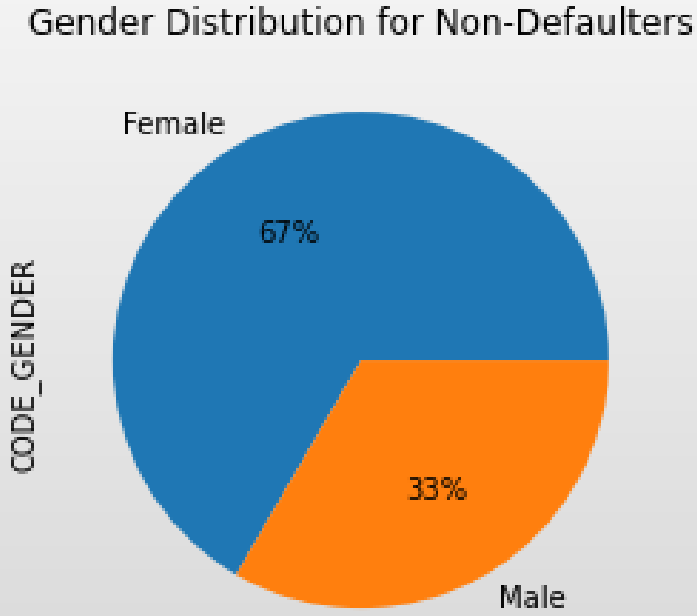




We can observe that laborer, Sales Staff and Core staff contribute major part of the application (for both Defaulters and Non-defaulters) and IT staff contributes the least (for both Defaulters and Non-defaulters).

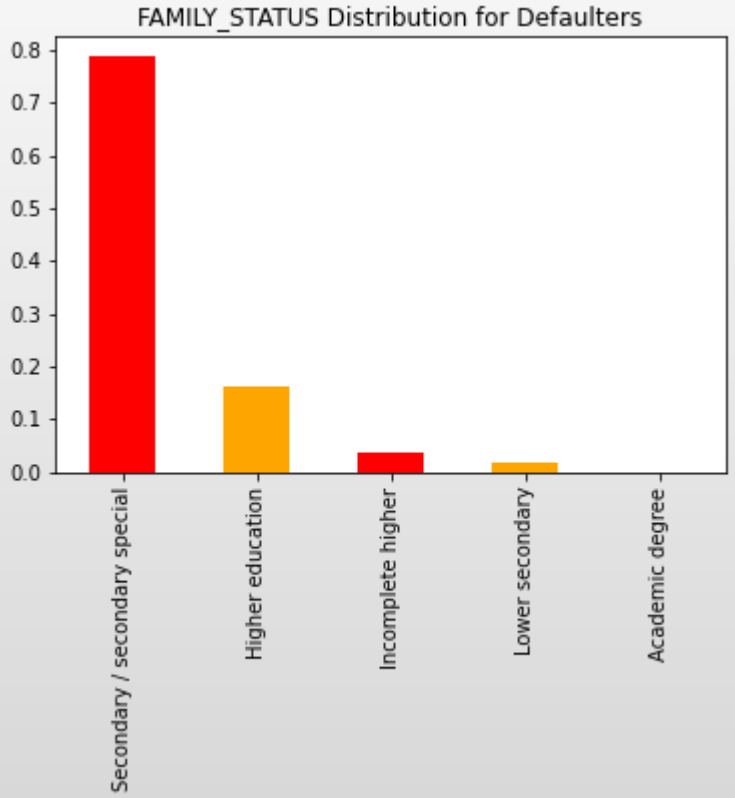
Univariate and Segmented Univariate Analysis for Application data

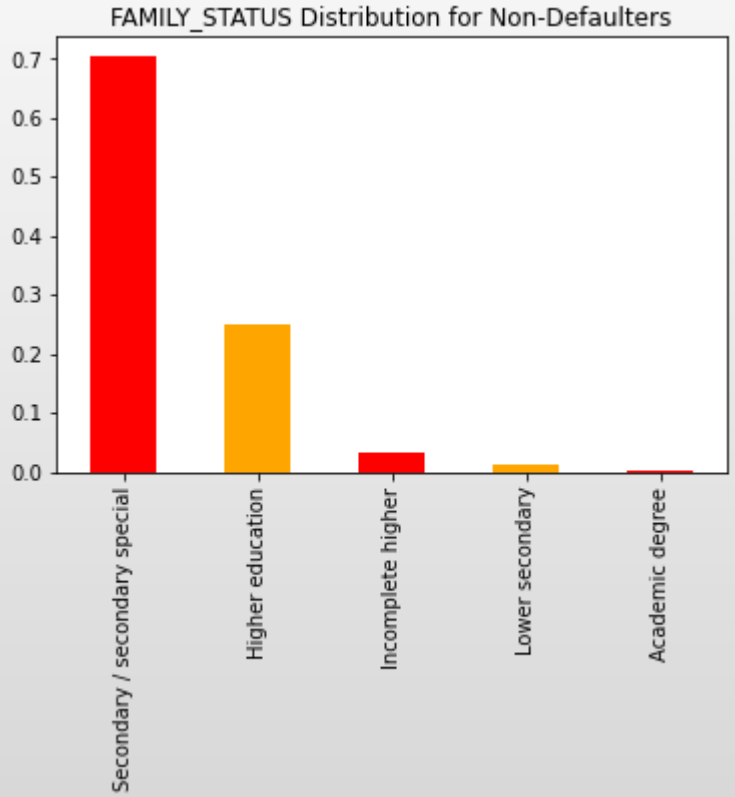




We can see that more Number of Female apllies for more so, for both Defaulters and Non-Defaultes female should contribute more. %The plots also shows that there are more number of female in both the groups(for Defaulters 57%, for Non-defaulters 67%)as well

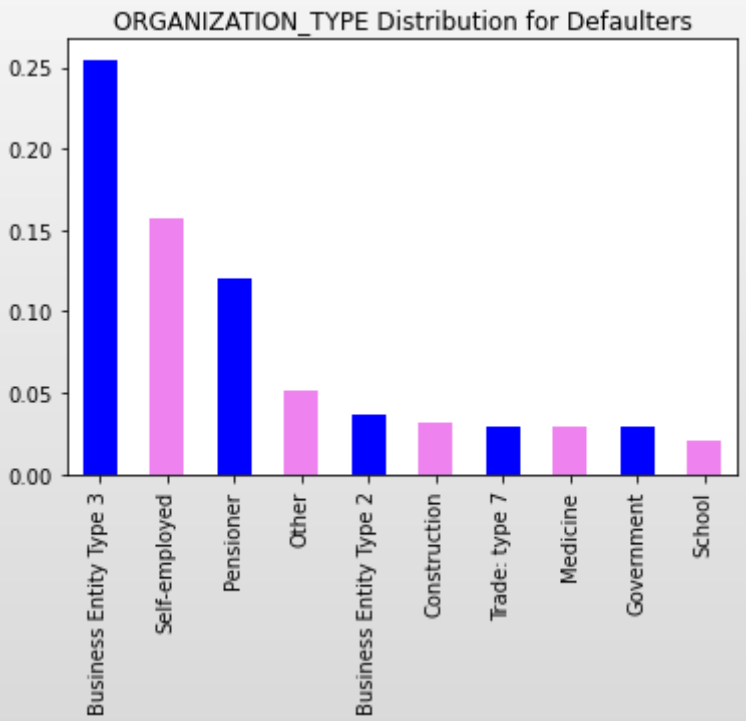
Univariate and Segmented Univariate Analysis for Application data

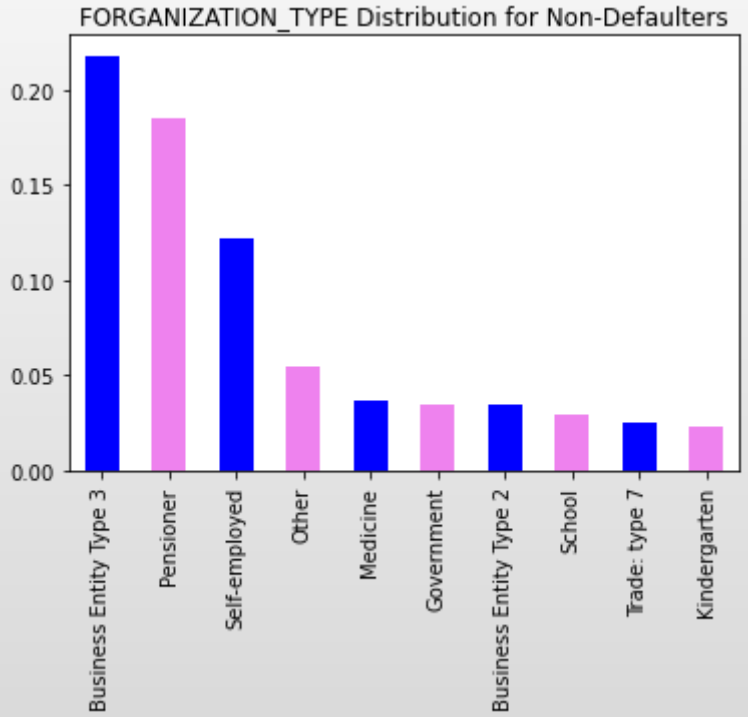




We can see all the education category people tend to default expect the **Higher Educated** people

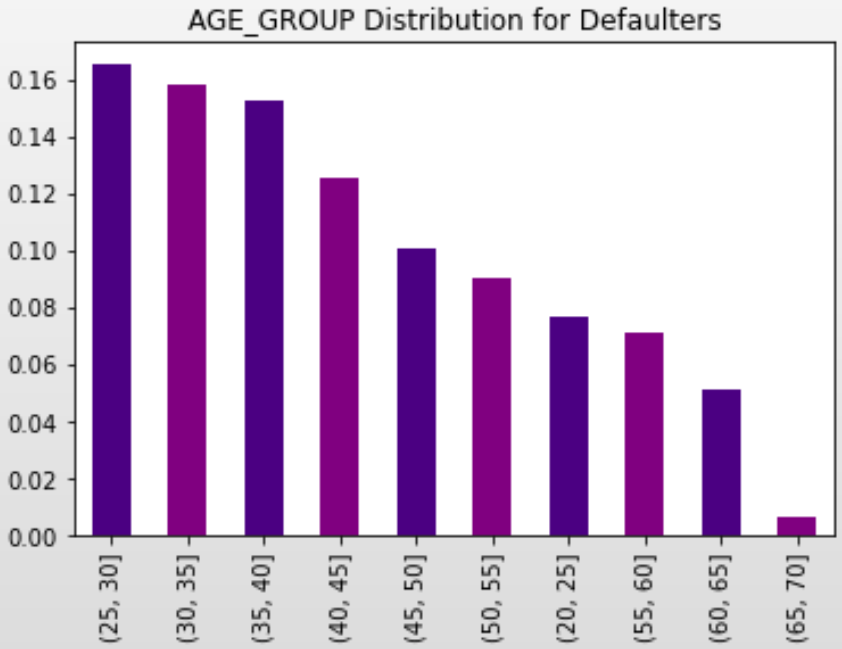
Univariate and Segmented Univariate Analysis for Application data

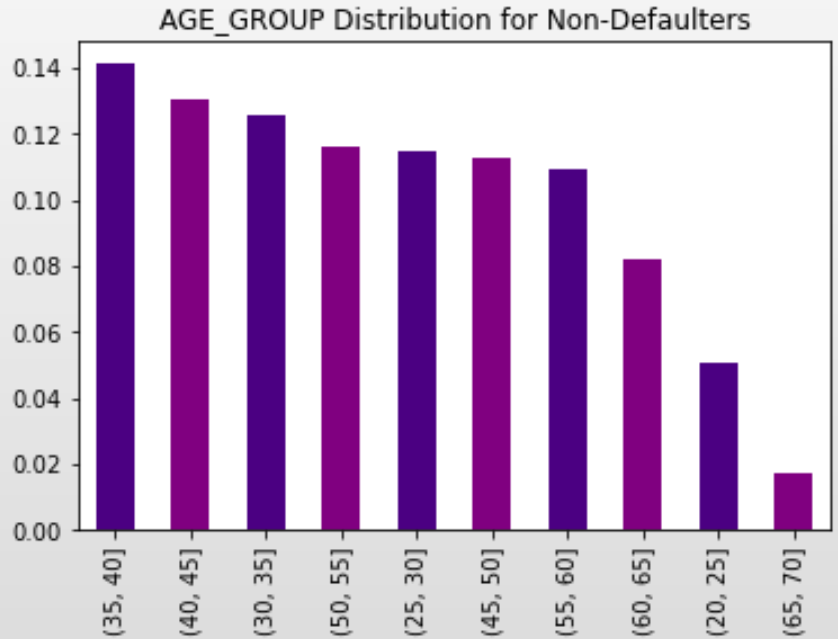




- We can observe that Self Employed people have a higher chances to default with respect to their loan application.
- Pensioner or Government Employees have a good chance to repay the loan

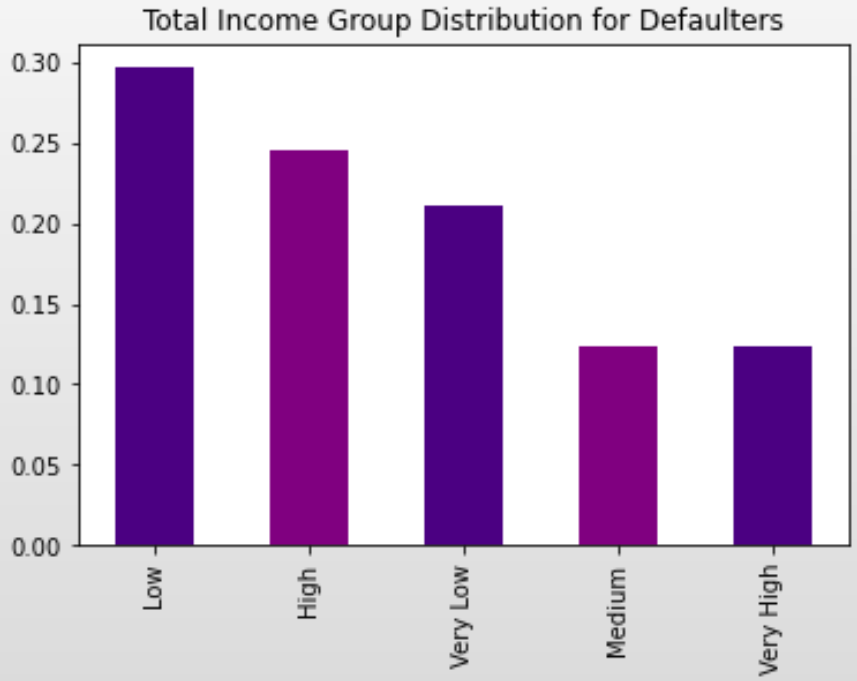
Univariate and Segmented Univariate Analysis for Application data

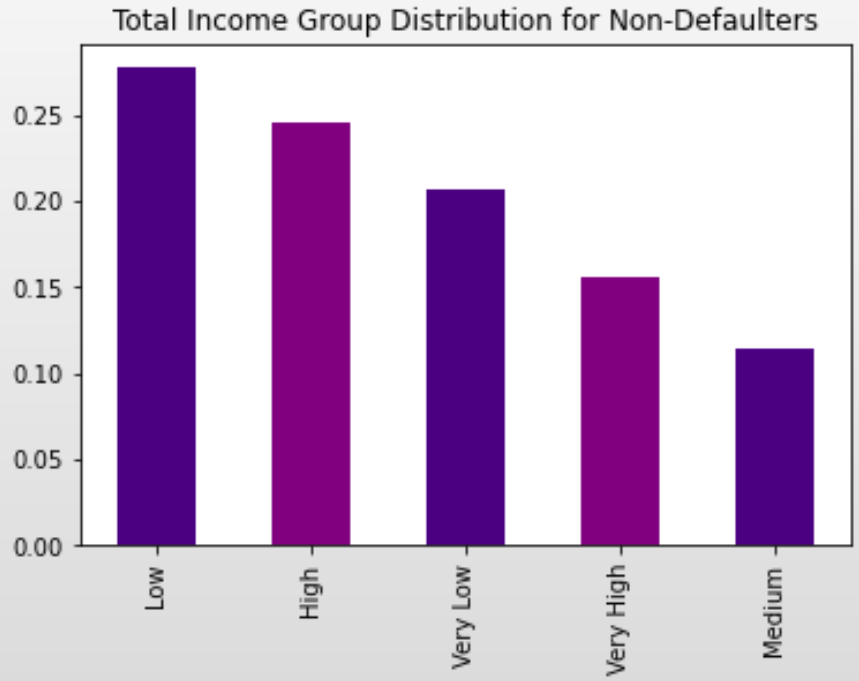




- We can say that people from 35 to 40 years old apply for loans much more than others
- 25-30 years age group tends to default most, they are very risky to loan.

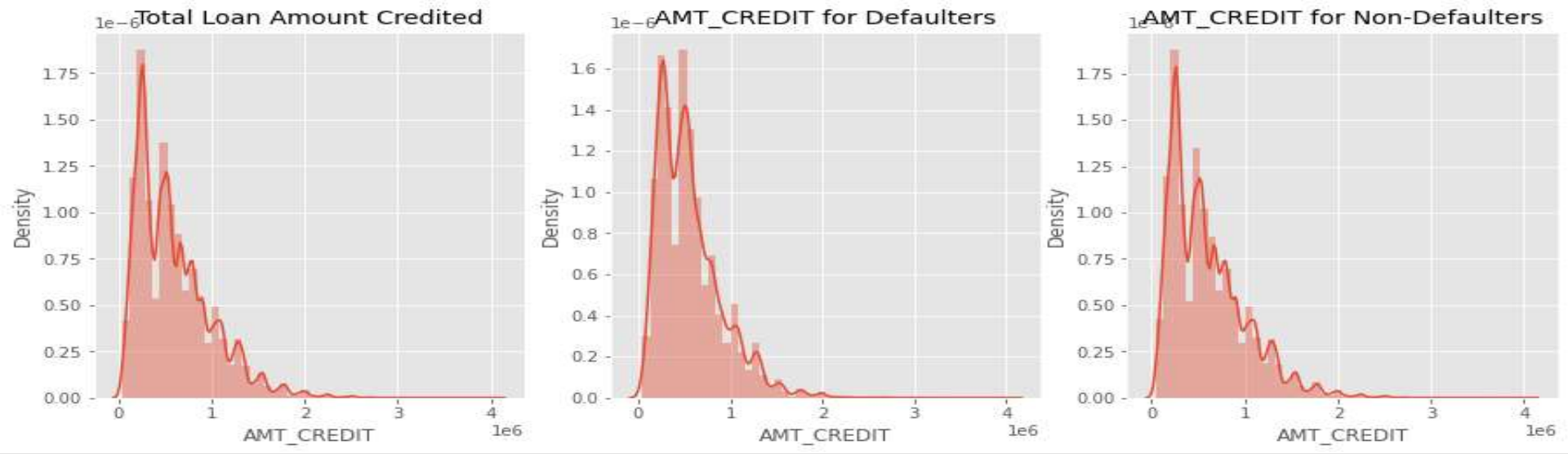
Univariate and Segmented Univariate Analysis for Application data

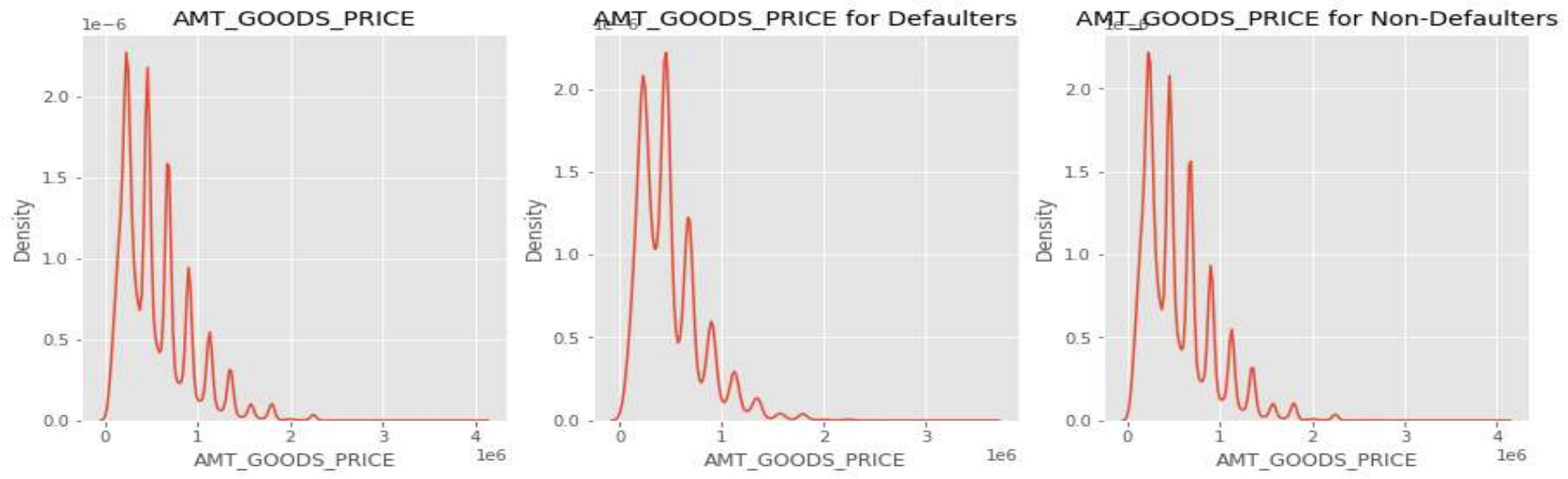




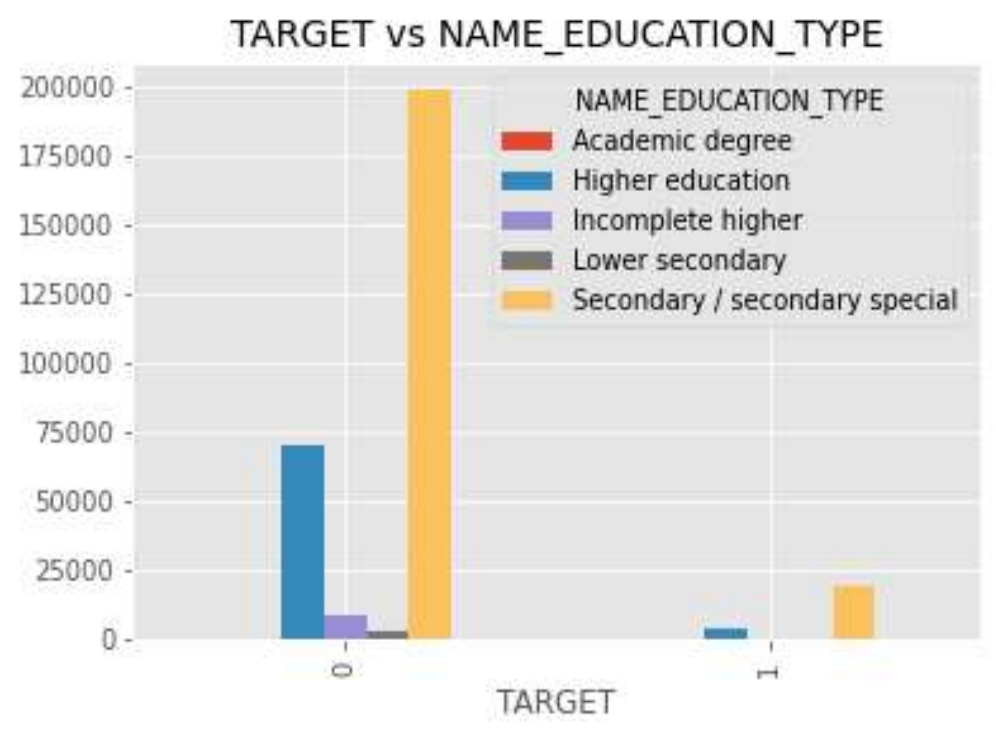
- Low income group people tend to take more loans as well as default more.
- Very High Income category people have very less defaulting percentage (Very safe for granting Loans).

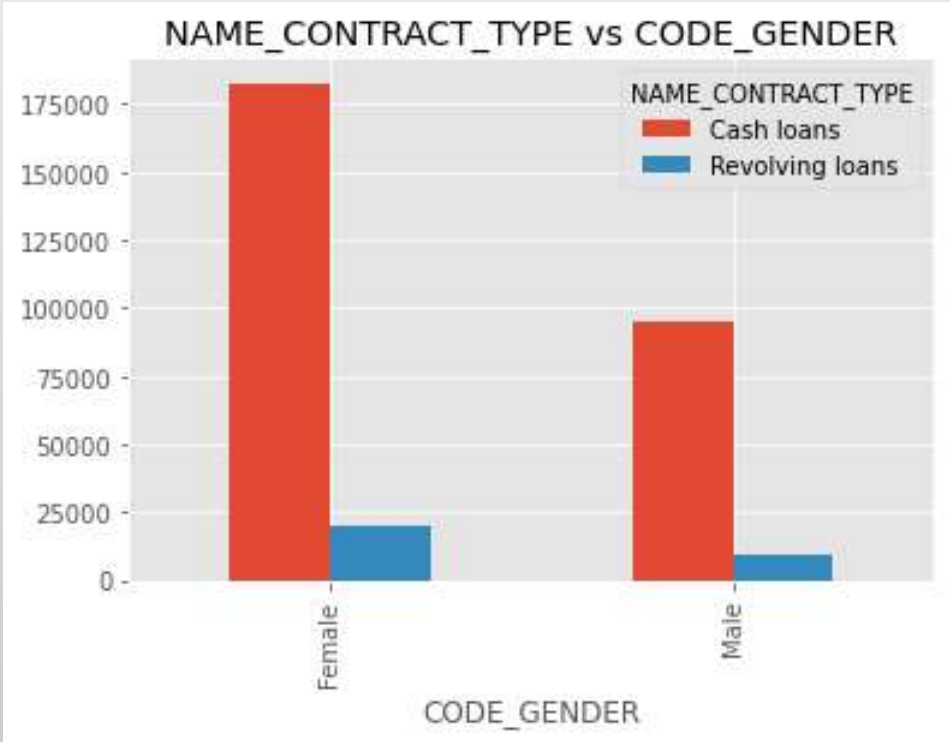
Univariate and Segmented Univariate Analysis for Application data





Bi-variate Analysis for Application data





Bi-variate Analysis for Application data

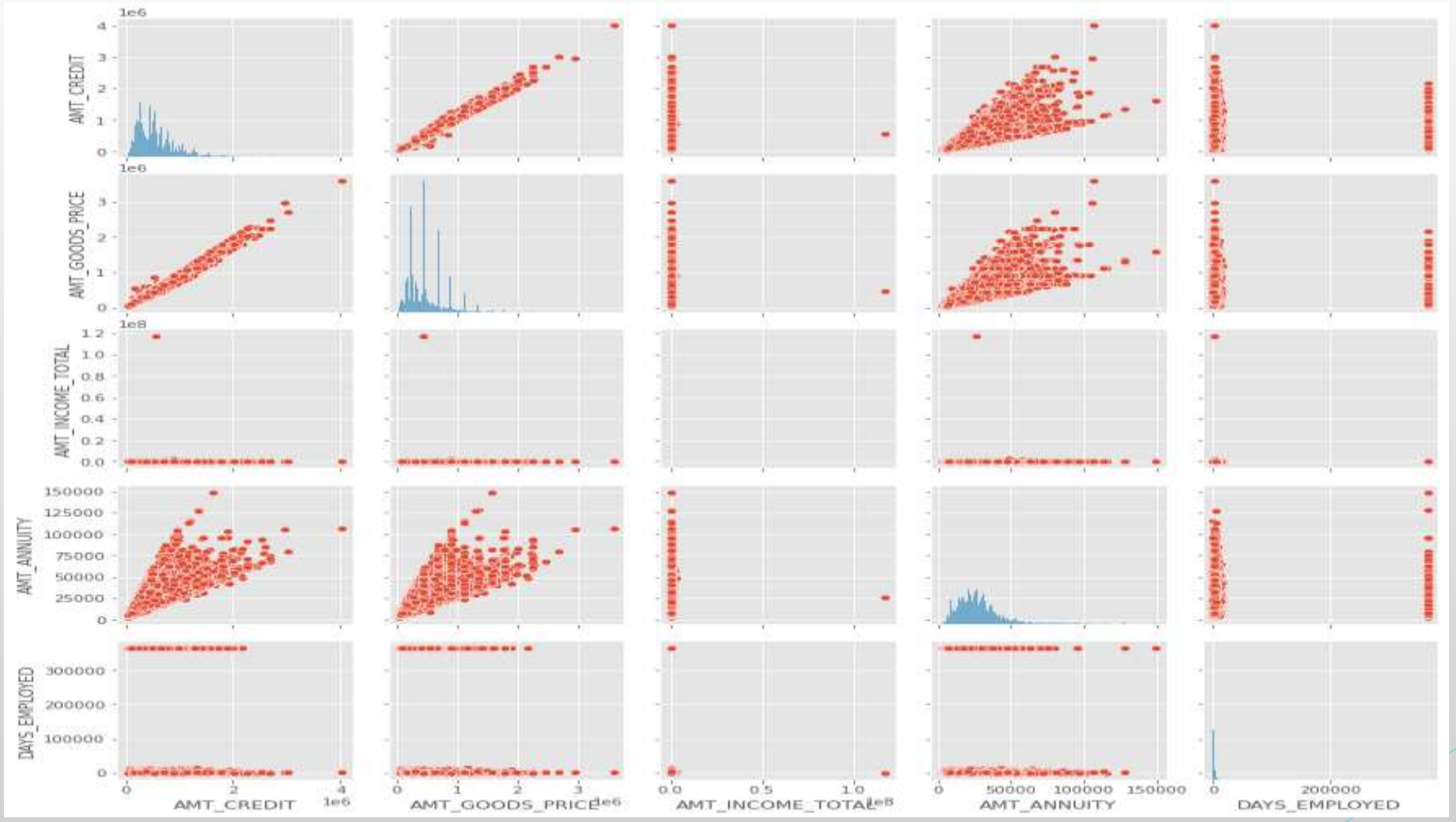
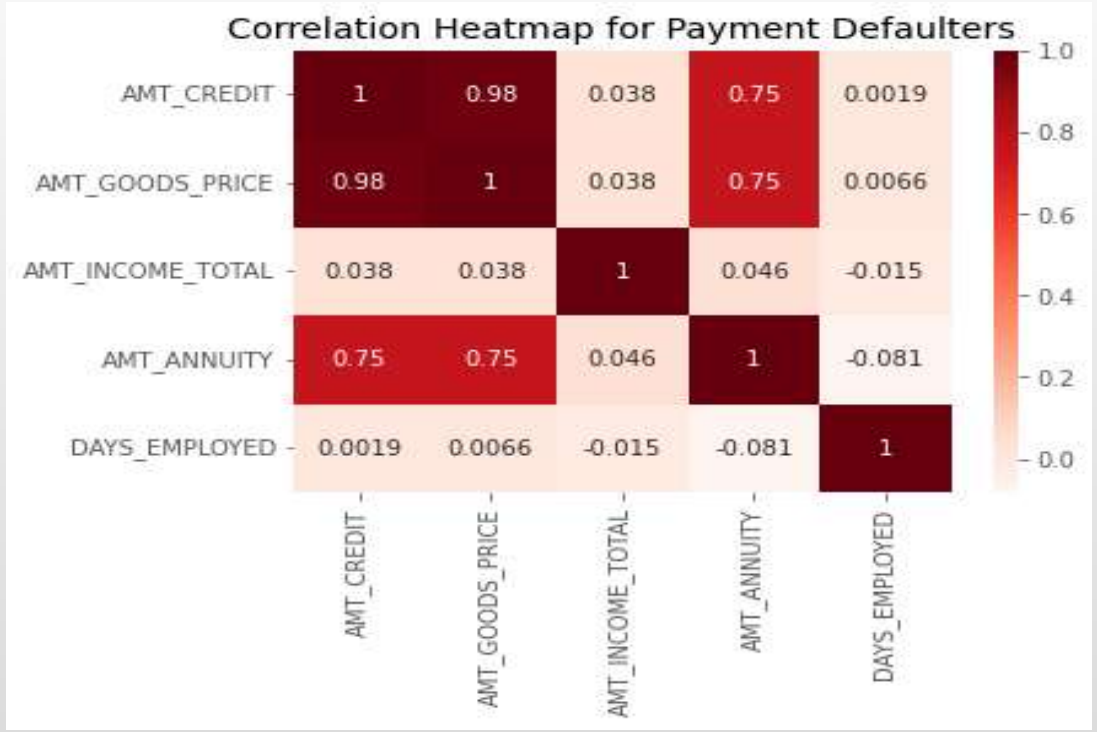
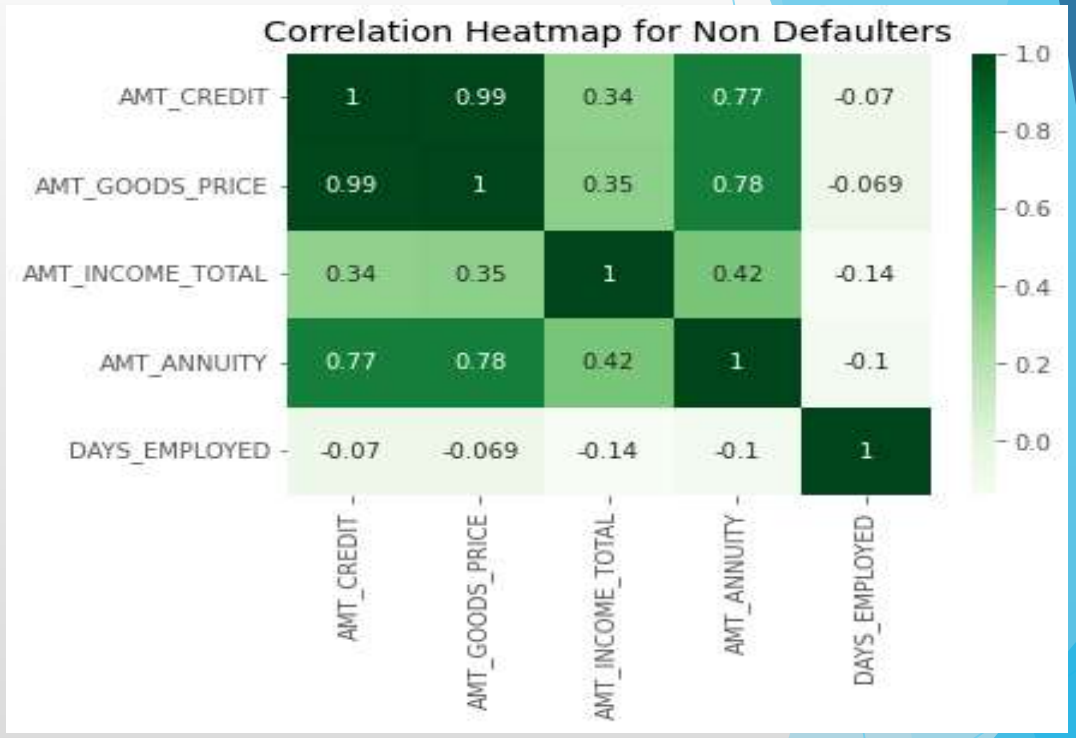


Fig : Pairplot for Loan Defaulters

Bi-variate Analysis for Application data

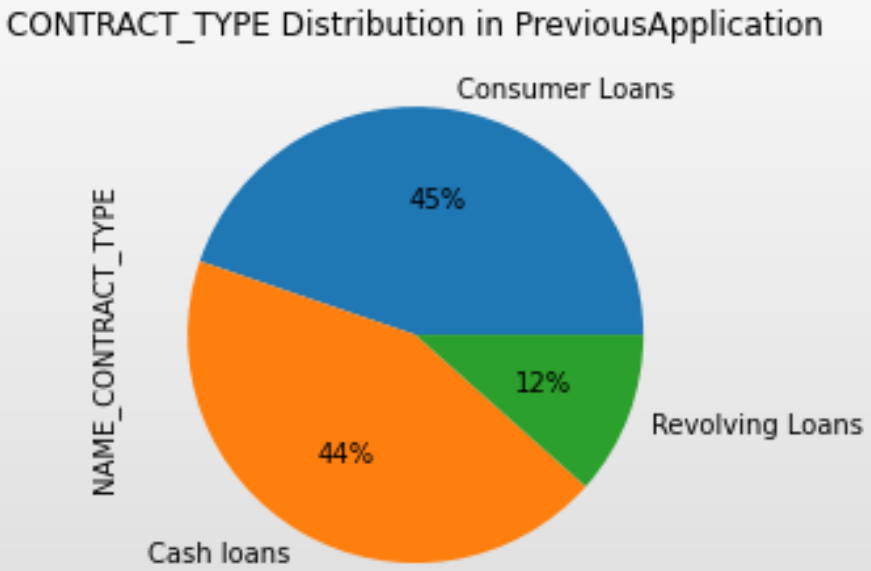


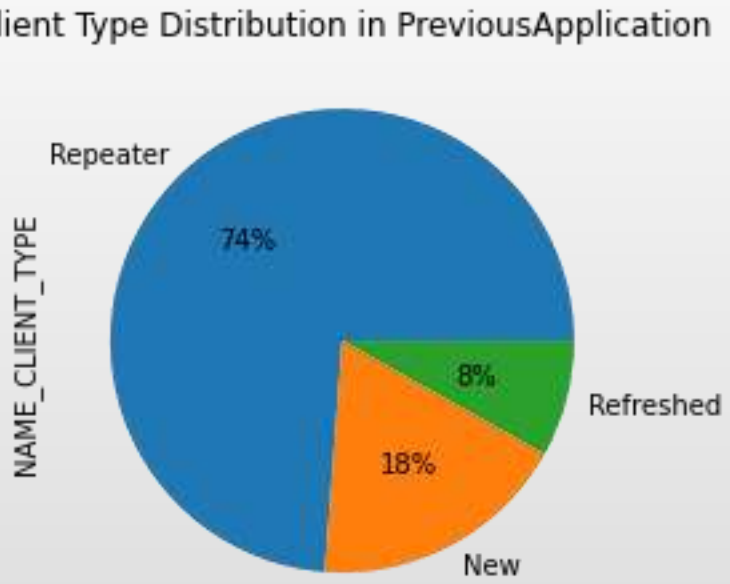


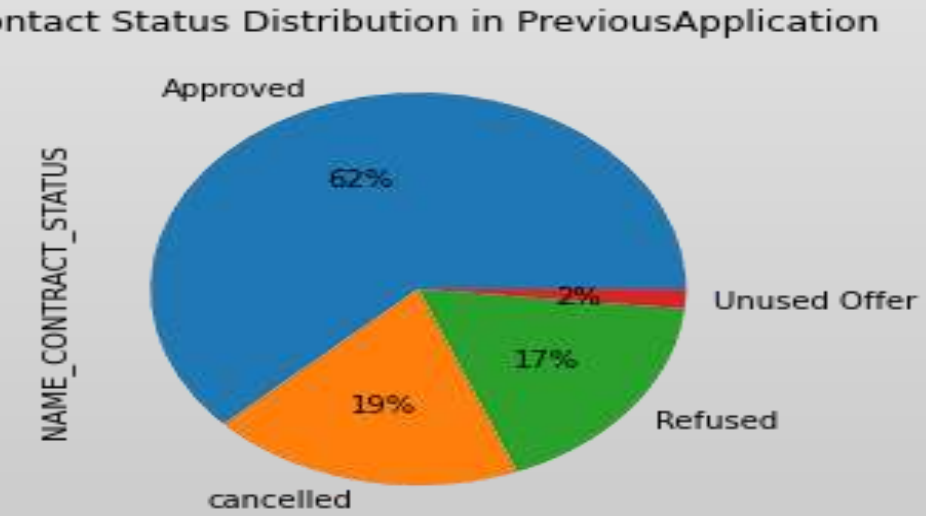
Insights:

- There is a high correlation between Credit amount and Goods Price
- There is also some correlation between Annual Income and Goods price for Defaulters.
- There is also some correlation between Annual Income and Credited amount for Defaulters.

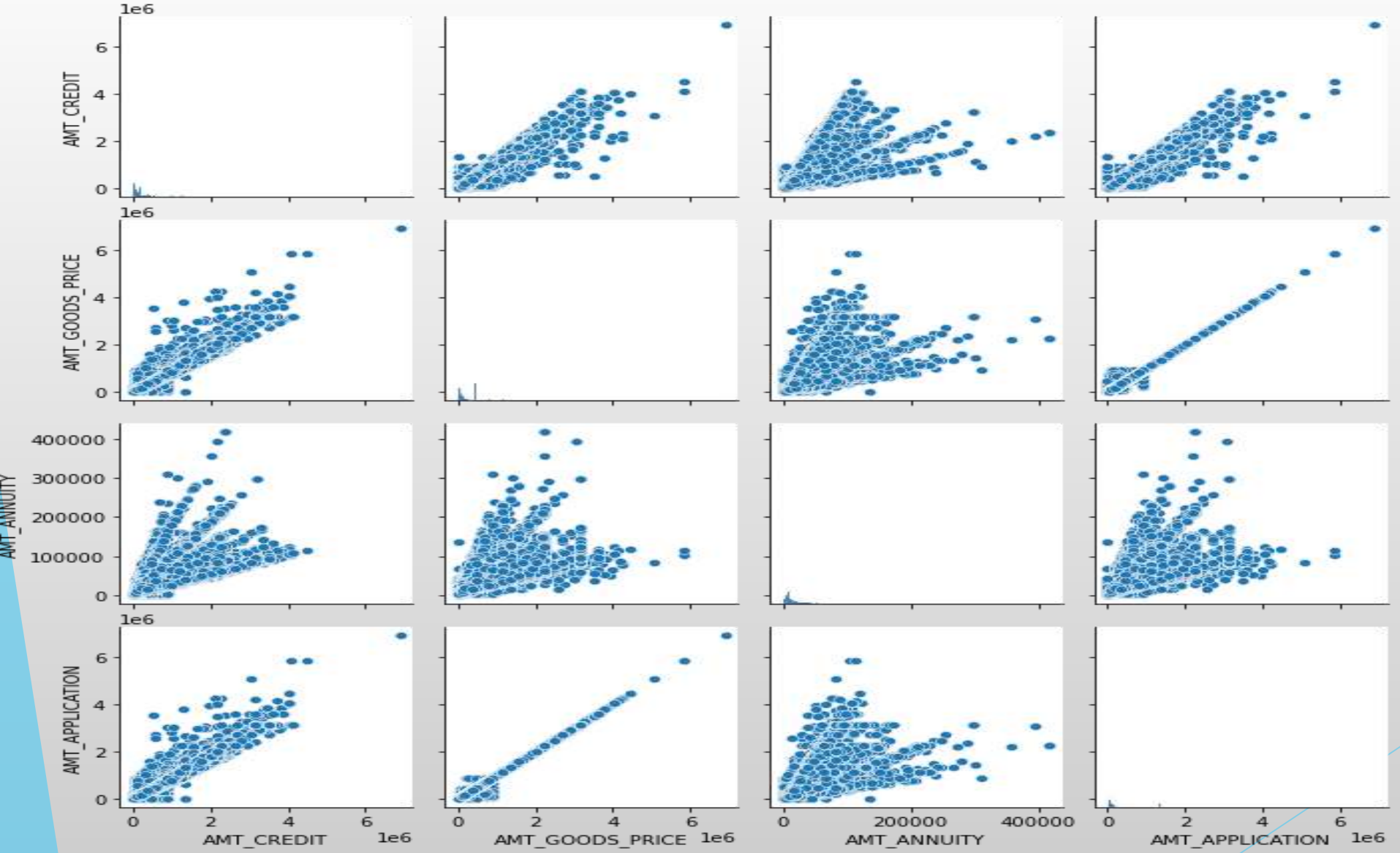
Analysis for Previous Application data



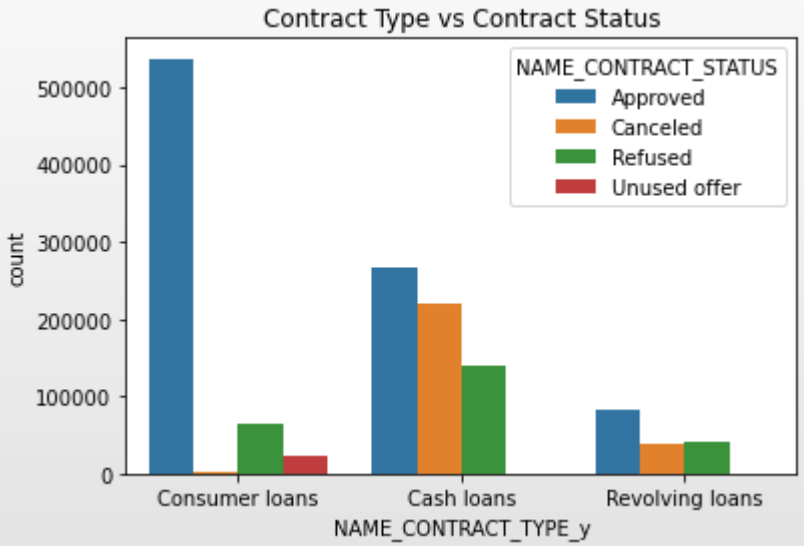


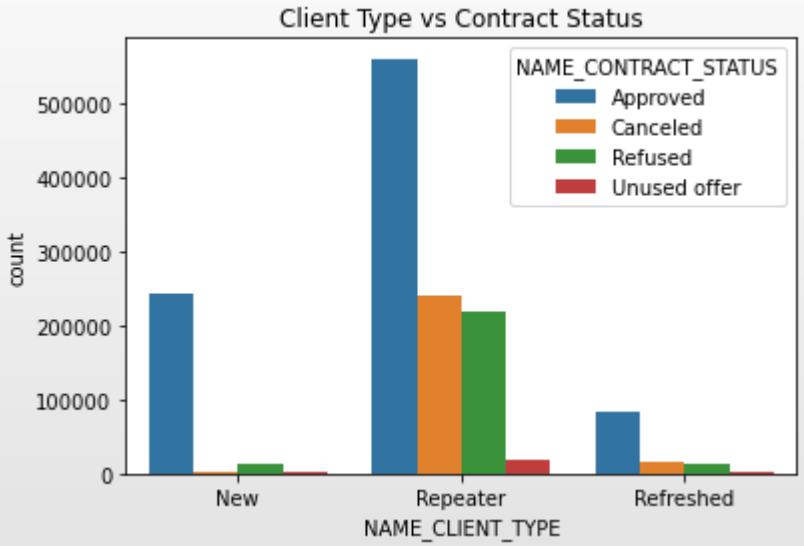


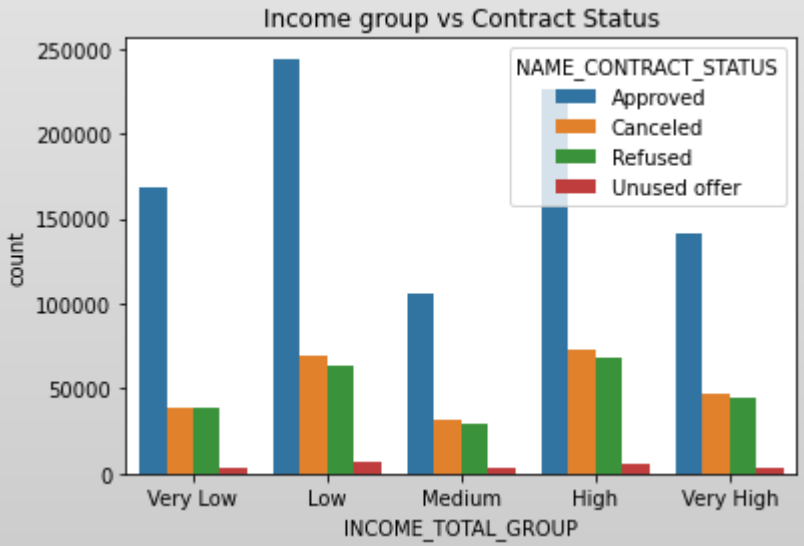
Bi-variate Analysis for Previous Application data

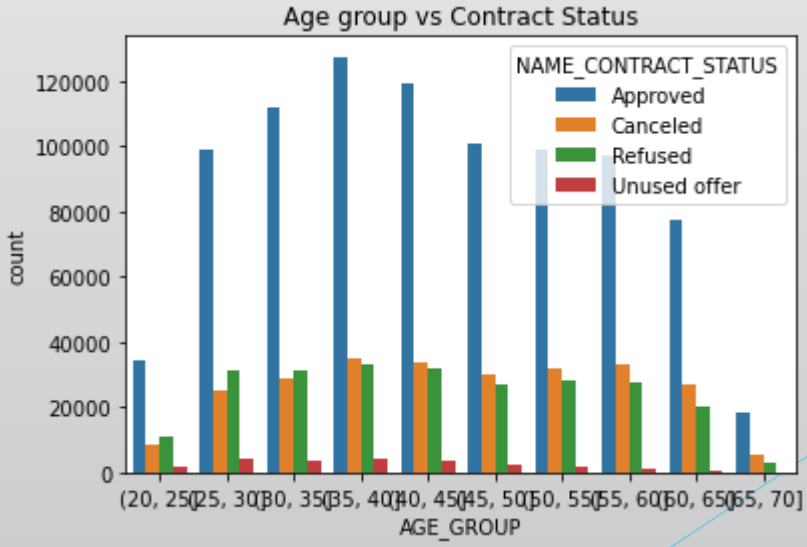


Analysis for Final Data (Merged Dataset)

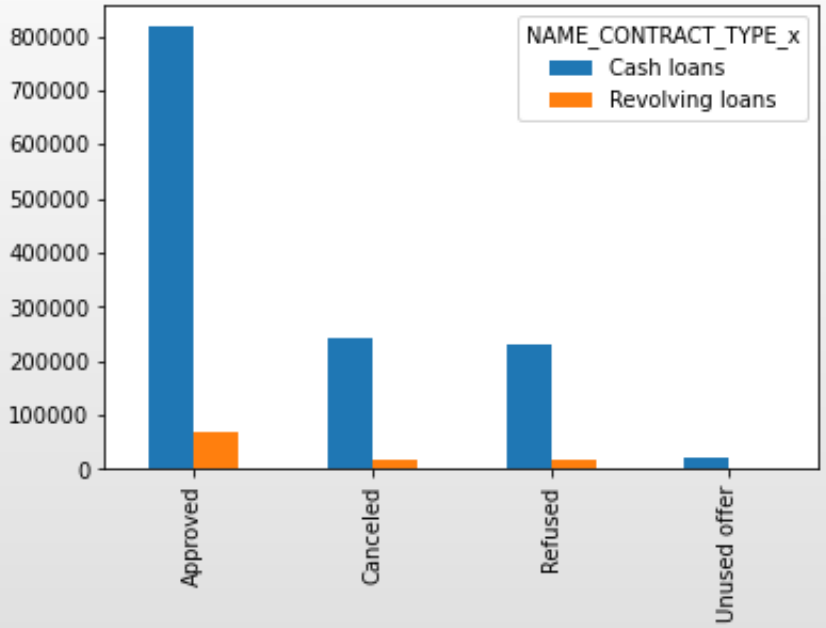


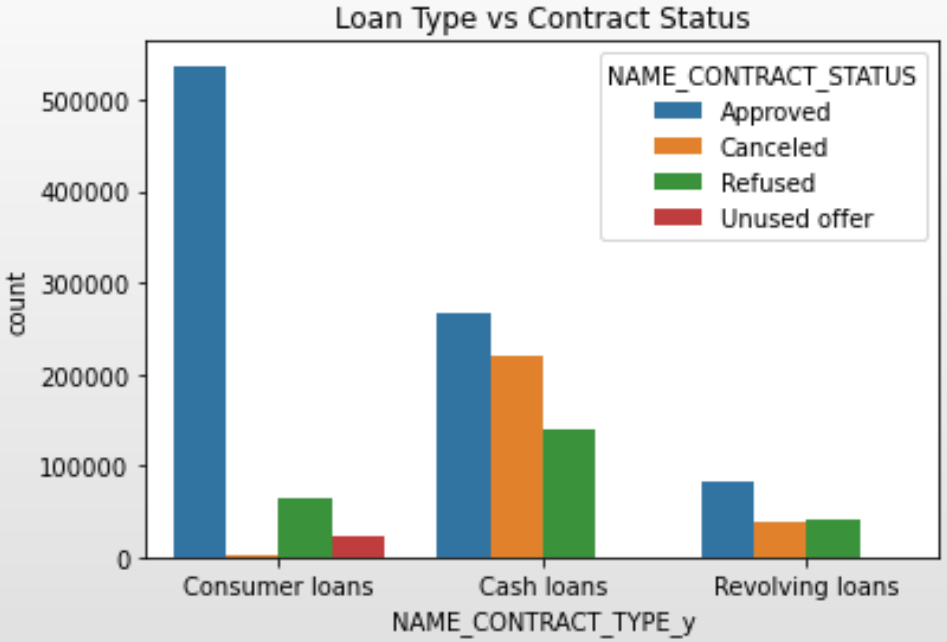


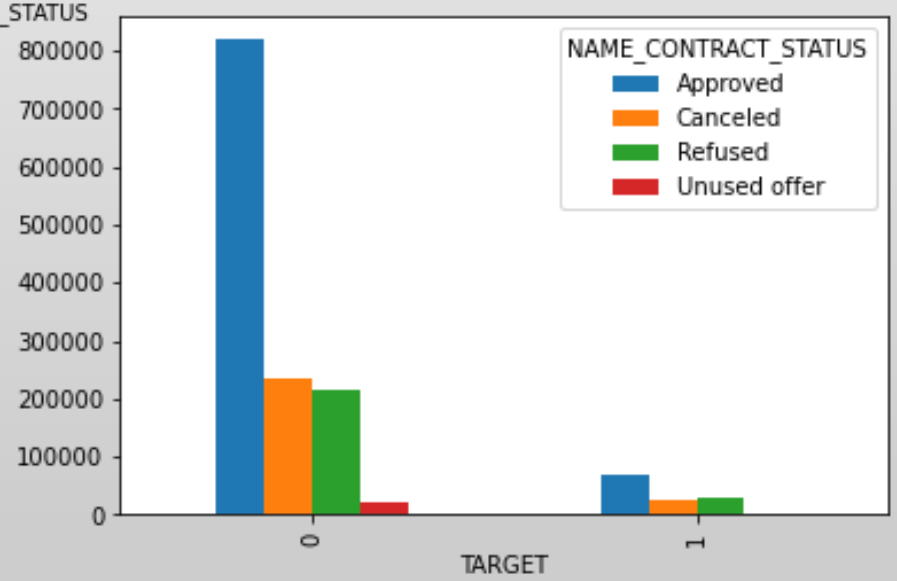




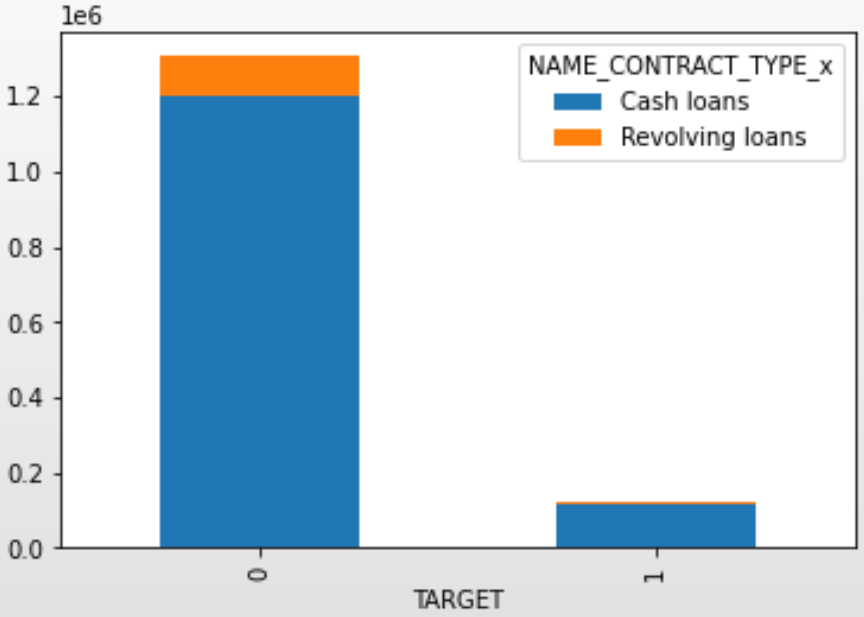
Analysis for Final Data (Merged Dataset)

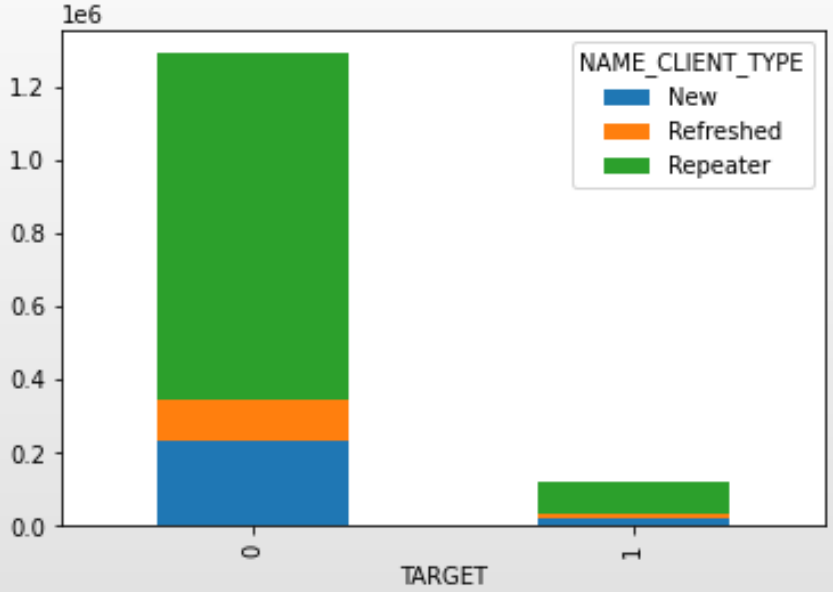






Analysis for Final Data (Merged Dataset)





- **Insights:**
- Consumer loans has higher percentage of approval, where Cash loans has higher Canceled/ refused percentage
- New Applicants have better chances for loan approval
- 35-40 years people tend to get more approvals. for 60+ years people the approval percentage is less
- Low income category people have higher number of refused/ canceled loans but Medium Income category people has highest percentage of refused/ canceled loans
- Defaulters have very lower chances to get a loan.
- People with Revolving loans have less chance to default
- large Segment of Repeater Loan applicants don't default and also New Applicant's have a high contribution among non-defaulters

Observation 1:
Recommended Group for Loan Approval

- Client with higher education
- State Government Worker
- Old People
- Client who has no default record
- People with 35+ years age

Observation 2:
Risky Group for Loan Approval (Chances to Default)

- Previously refused clients
- Clients who have already defaulted
- Single Male
- Self Employed people
- Clients with lower than Secondary Education

Assumptions:

- We have considered 'XNA' ,'XAP' as null values for our help in analysis
- Target Variable for New Application Dataset = “TARGET”
- Focused Variable for Previous Application Dataset = “NAME_CONTRACT_STATUS”

Conclusion

- In this case study we have understood and used various techniques of Exploratory Data Analysis and gathered a lot of insight regarding the Risk Analysis of Financial Sectors.

Thank You