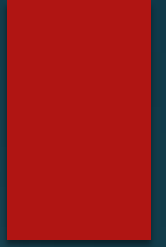




# Loan Providers' Risk Analysis Case Study

❖ Souvik Sarkar

# Business risks and Decision taking



Loan providers face two types of risks in order to rejecting/approving the loan applications

- ▶ If applicants likely to repay loan, then not providing the loan results in a loss of business
- ▶ If applicant is not likely to repay loan, then providing the loan also results in a loss of business.

Loan providers may receive applications from 2 types of applicant

- ▶ The Client with payment difficulties and he/she had late payment more than X days on at least one of the first Y instalments of the loan
- ▶ All other cases where no defaults happened in near past.

Based on the applicants loan payment history Loan providers can take following decisions

- ▶ Approved
- ▶ Cancelled
- ▶ Refused
- ▶ Unused Offer

# Data Imbalance

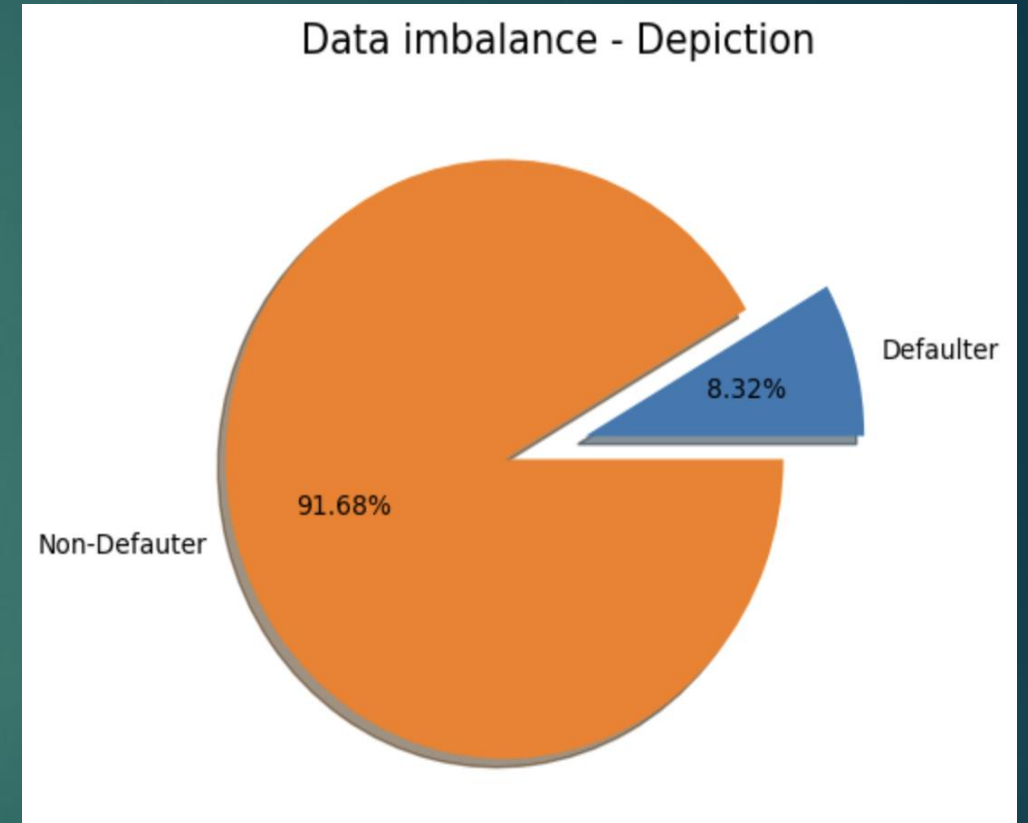
When data have been segregated into 2 types Defaulters(Target 1) and Non-Defaulters(Target 0), it reveals the data imbalance

We can see data imbalance ratio is almost 11:1

```
TARGET
0    91.675101
1     8.324899
Name: proportion, dtype: float64

91.675101/8.324899

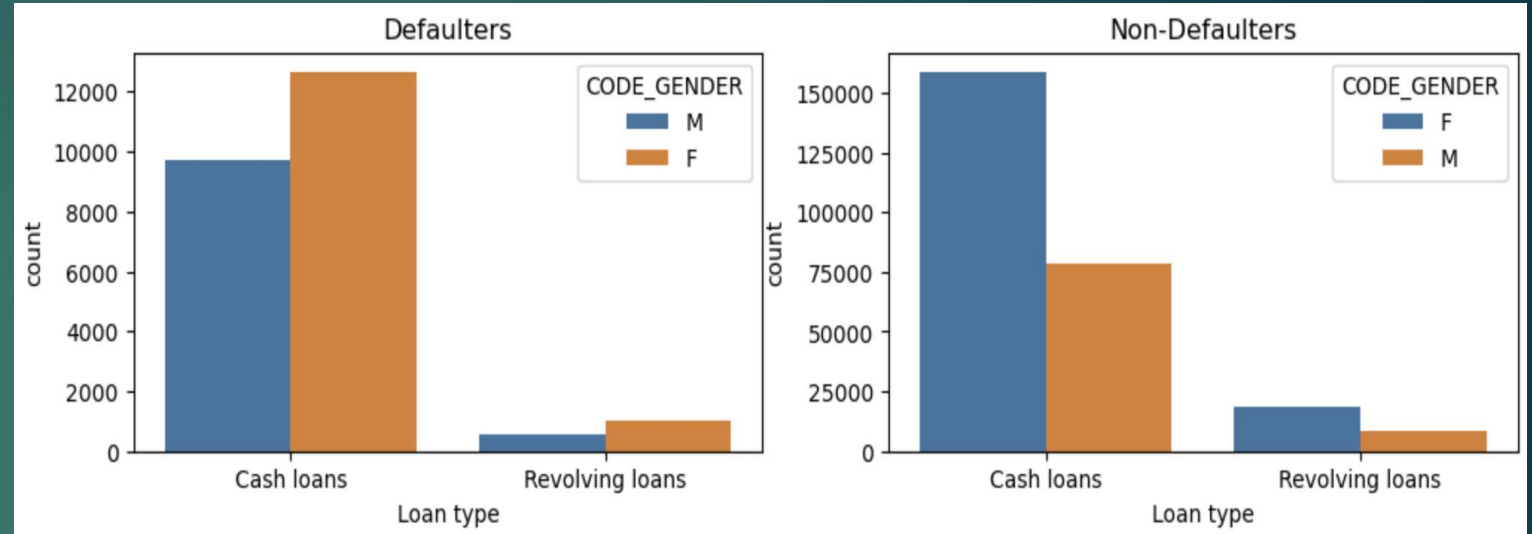
11.012157745096967
```



# Univariate Analysis of Categorical columns

Two types of loans

- ❖ Cash Loans
- ❖ Revolving loans

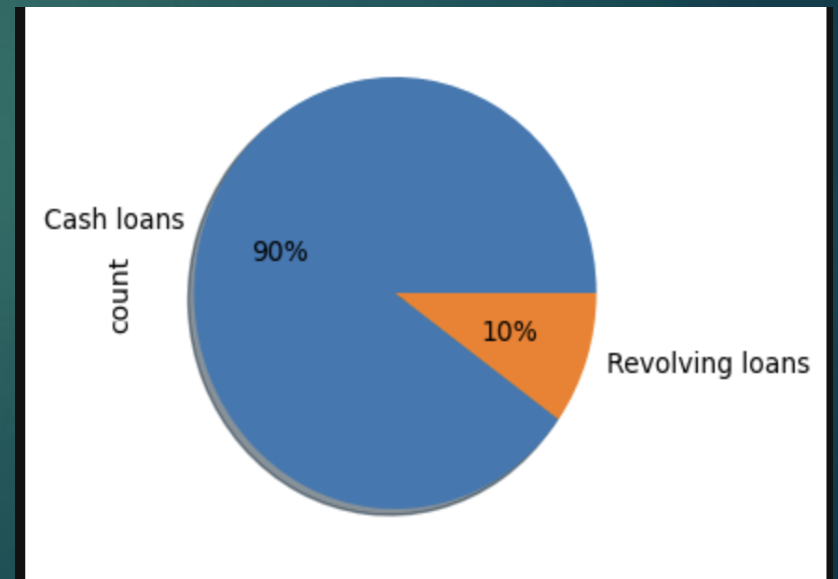


Analysis :

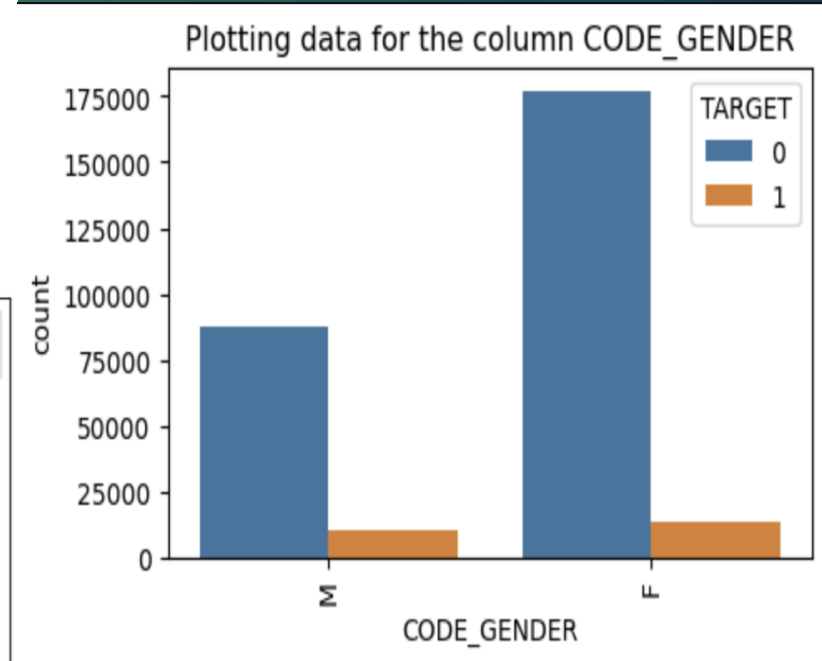
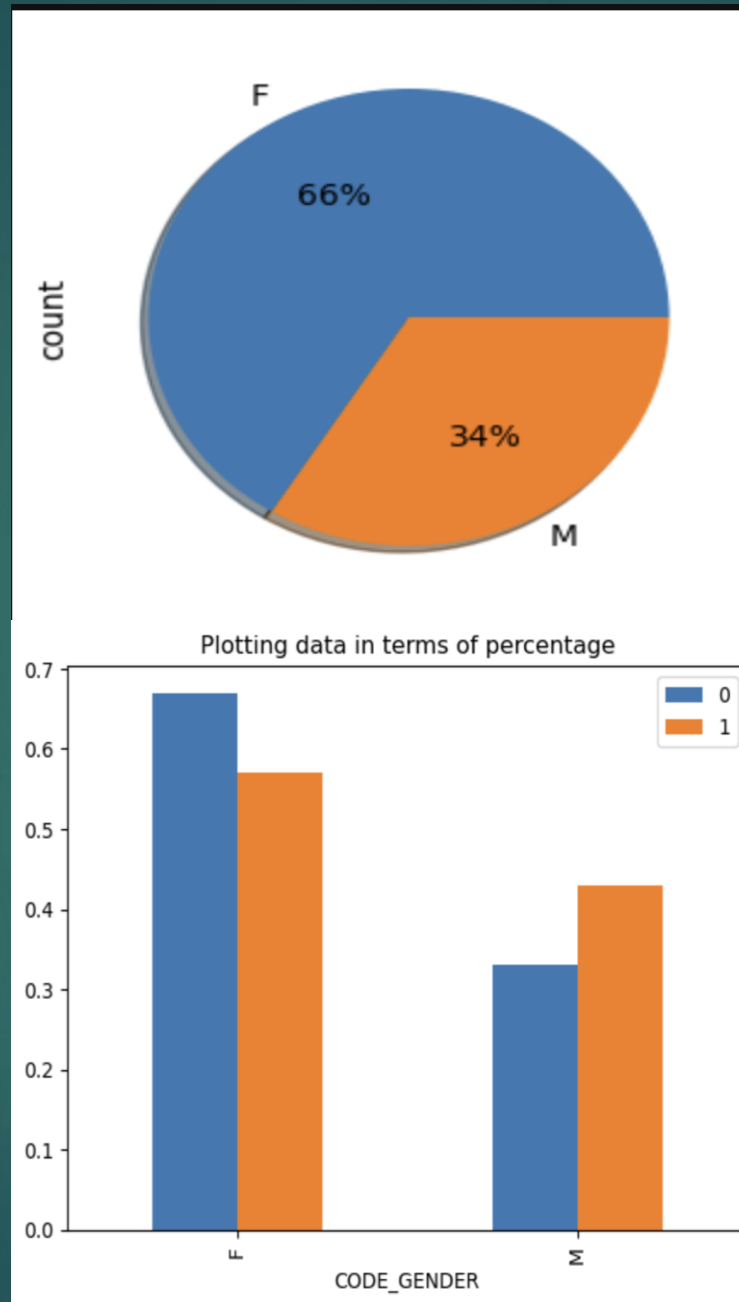
Cash loan clients has more payment

Defaulters than revolving loans.

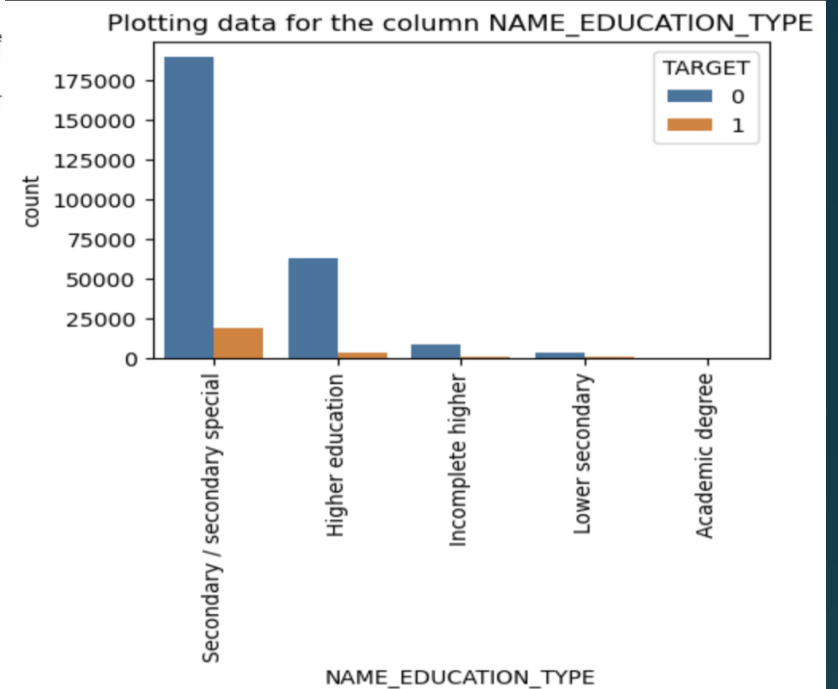
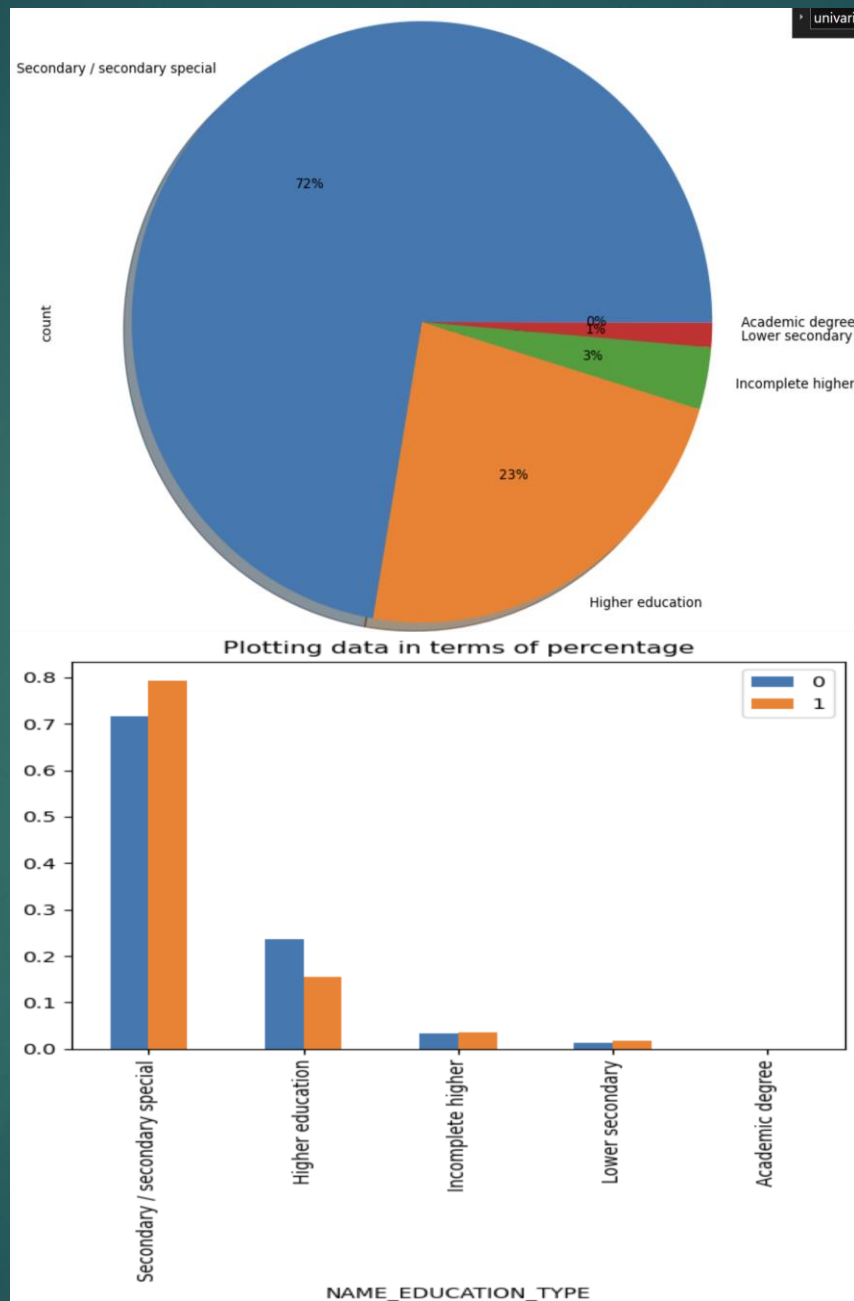
Loan providers should pitch or market the revolving loans more.



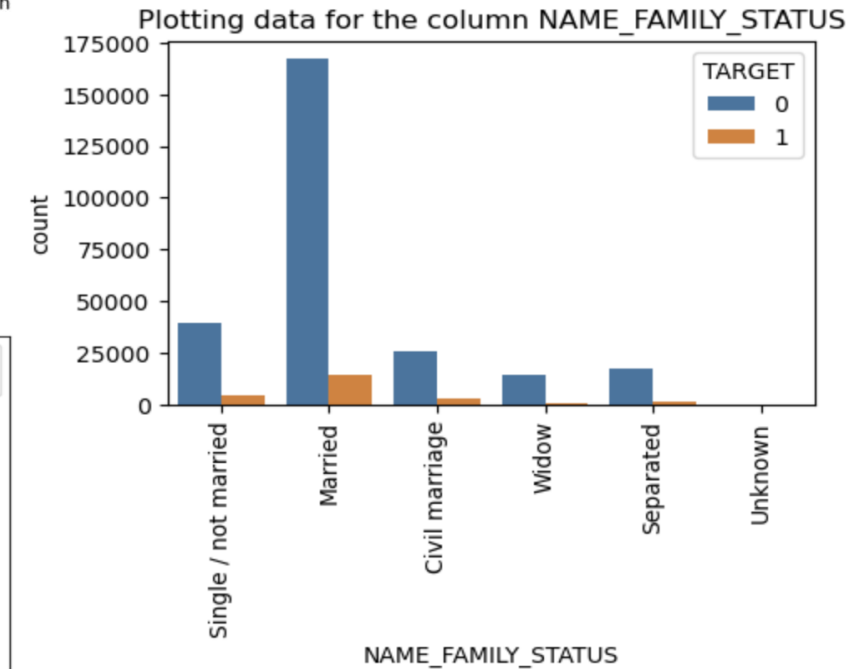
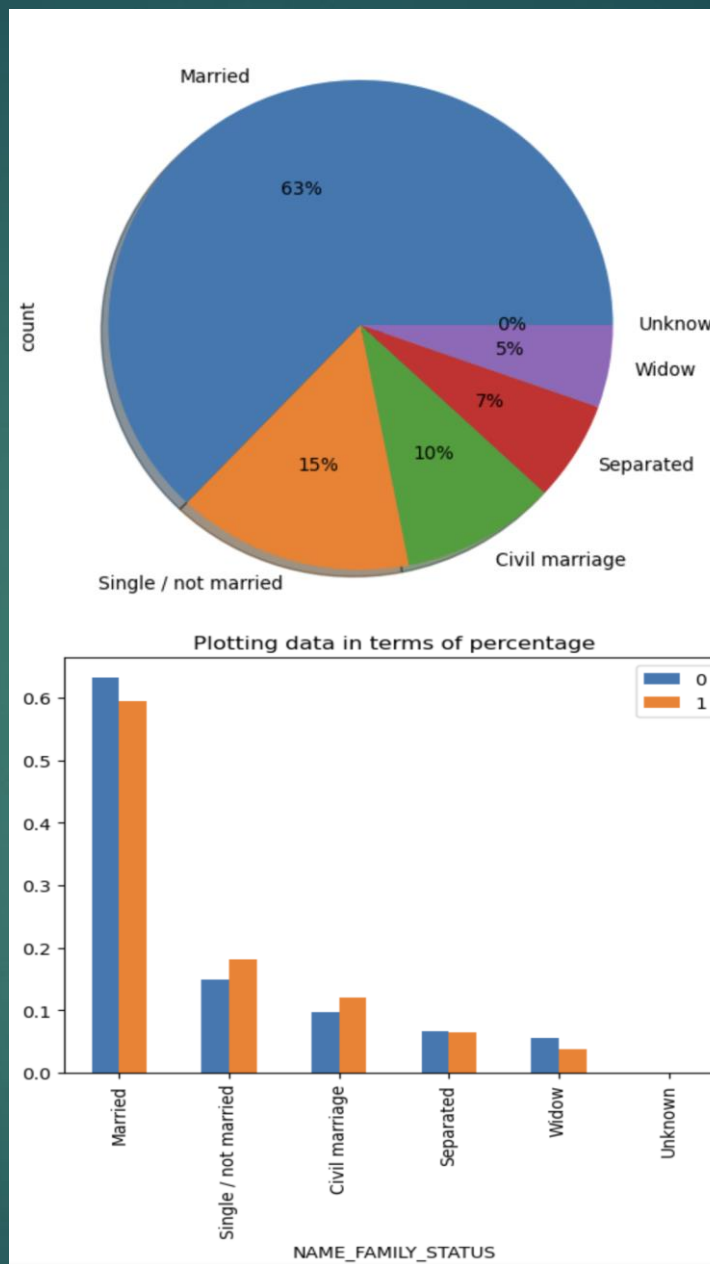
- ▶ Clearly Females are tends to apply loan more
- ▶ Even though the numbers of female applicants are much higher side Male applicants has the higher payment difficulty
- ▶ Loan providers should keep on focusing on female clients



- ▶ Maximum number of loans have been provided to the **Secondary/Secondary special** applicants.
- ▶ Analysis : which depicts **Secondary/Secondary special** class has the most requirements of money
- ▶ Higher education people have less difficulties in payment
- ▶ **Insight** : Better education provides better job which pays well

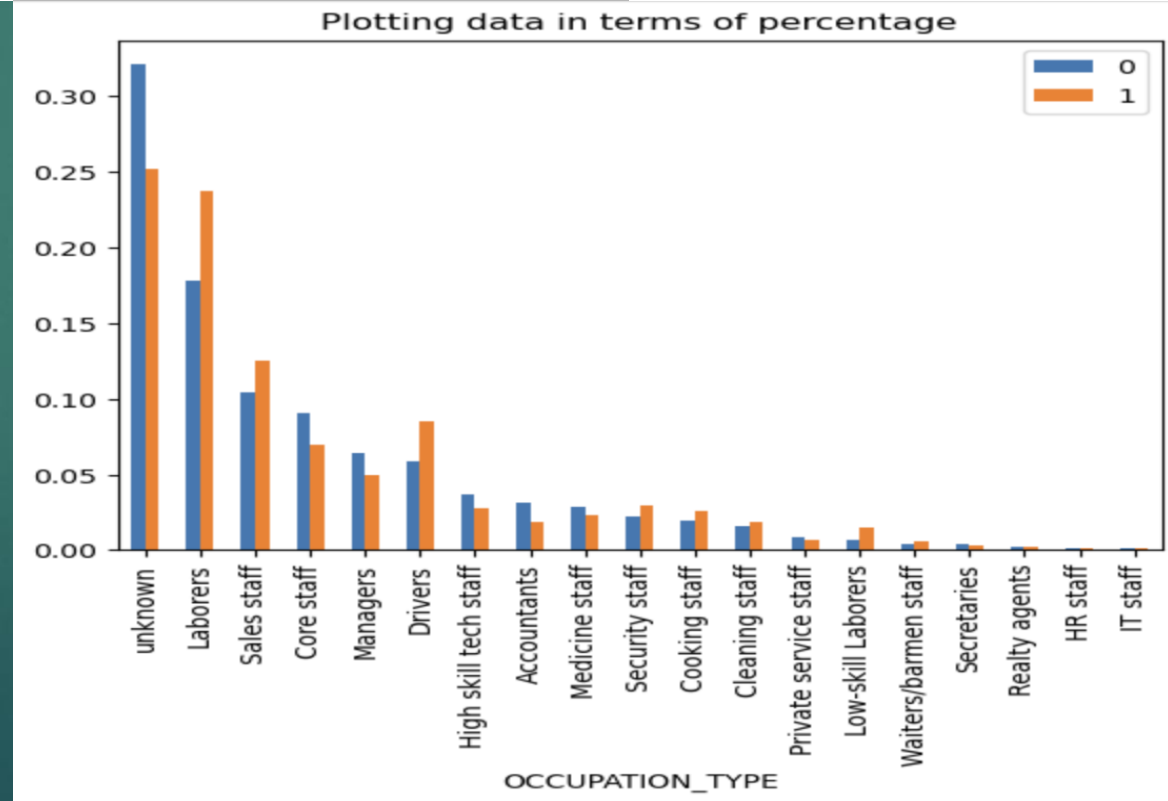


- ▶ Married people are in urgent needs of money hence they have taken more than 60% of loans
- ▶ Single/Not married and Civil marriage people surprisingly have more difficulties in paying off the loan
- ▶ Probably married people are frugal and tends to manage money wisely



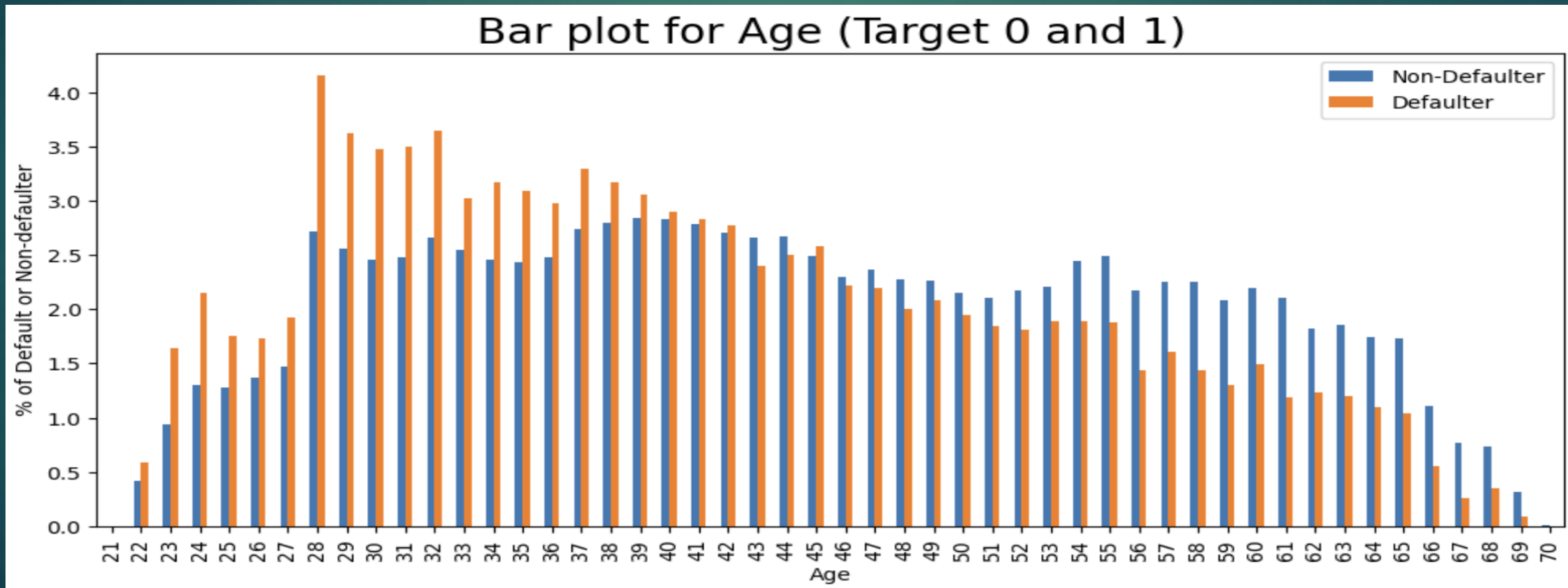


- 
- The left chart is a pie chart showing the distribution of OCCUPATION\_TYPE. The right chart is a bar chart showing the count of each occupation type, categorized by a target variable (0 or 1).
- Occupation Type Distribution (Pie Chart)**
- | OCCUPATION_TYPE          | Percentage |
|--------------------------|------------|
| unknown                  | 32%        |
| Laborers                 | 18%        |
| Sales staff              | 11%        |
| Core staff               | 9%         |
| Managers                 | 6%         |
| Drivers                  | 6%         |
| High skill tech staff    | 4%         |
| Accountants              | 3%         |
| Medicine staff           | 3%         |
| Security staff           | 2%         |
| Cooking staff            | 2%         |
| Cleaning staff           | 2%         |
| Private service staff    | 2%         |
| Low-skill Laborers       | 2%         |
| Waiters and barmen staff | 1%         |
| Realty agents            | 1%         |
| HR staff                 | 1%         |
| IT staff                 | 1%         |
- Occupation Type Count (Bar Chart)**
- | OCCUPATION_TYPE       | Count (Target 0) | Count (Target 1) |
|-----------------------|------------------|------------------|
| Laborers              | 48000            | 5000             |
| Core staff            | 25000            | 1000             |
| unknown               | 85000            | 5000             |
| Drivers               | 15000            | 2000             |
| Sales staff           | 28000            | 3000             |
| Cleaning staff        | 5000             | 1000             |
| Cooking staff         | 5000             | 1000             |
| Private service staff | 2000             | 1000             |
| Managers              | 18000            | 1000             |
| Medicine staff        | 8000             | 1000             |
| Security staff        | 6000             | 1000             |
| High skill tech staff | 10000            | 1000             |
| Waiters/barmen staff  | 1000             | 1000             |
| Low-skill Laborers    | 1000             | 1000             |
| Realty agents         | 1000             | 1000             |
| Secretaries           | 1000             | 1000             |
| Accountants           | 8000             | 1000             |
| IT staff              | 1000             | 1000             |
| HR staff              | 1000             | 1000             |

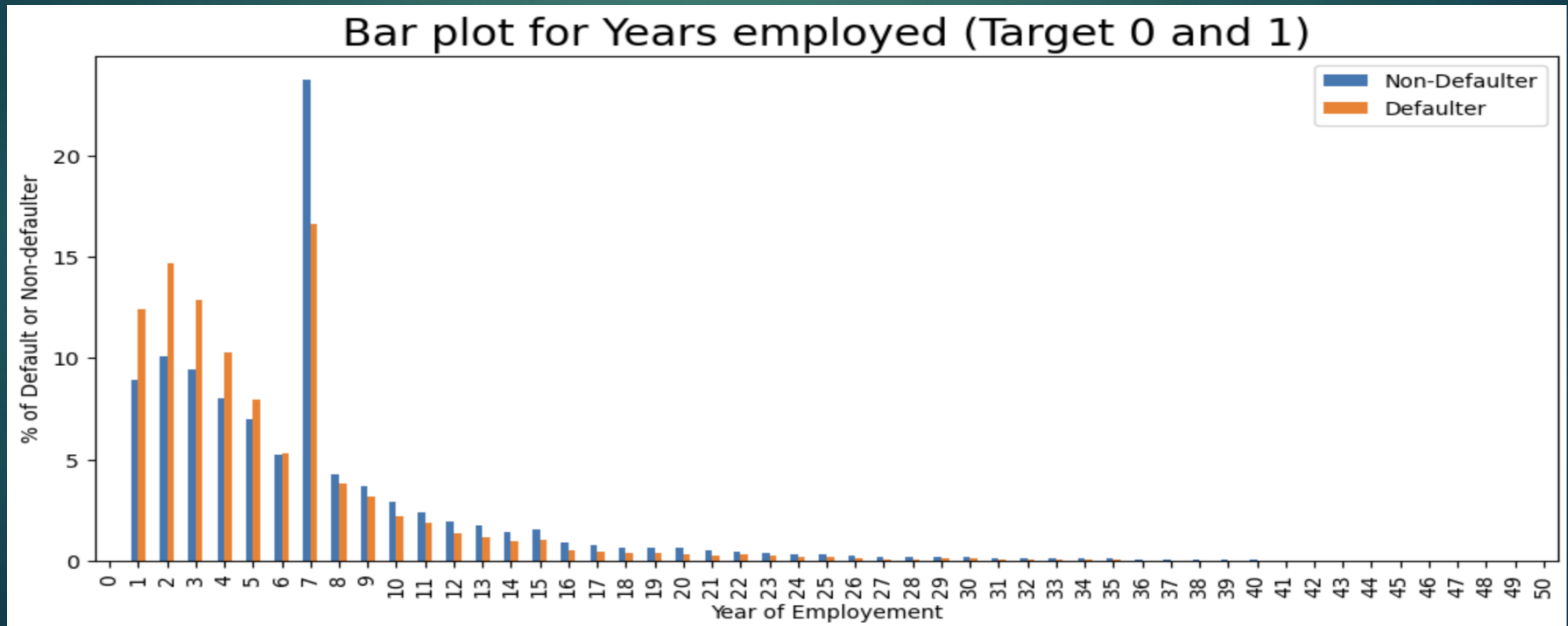




- ▶ 28-42 age group has the highest rate of defaulters due to less financial stability
- ▶ Old people has the better financial stability
- ▶ Loan providers should target age group more than 43

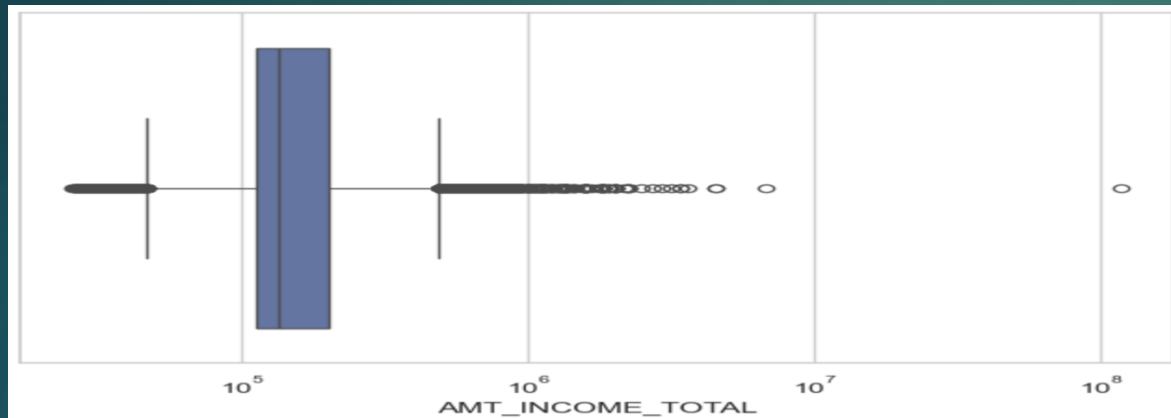


- ▶ At initial stage people tends to have more financial need hence is why the people work ex less than of 7 has higher tendency to default
- ▶ Loan providers should target people who has 7-10 years of work ex. Apparently people having 7 year of work ex are tend to have higher density in the dataset



# Univariate/Bivariate Analysis & Outliers for Numerical columns

- ▶ **AMT\_INCOME\_TOTAL** column had outliers and to tackle this I have created a separate column '**SALARY\_RANGE**' and limit the salary without losing any data



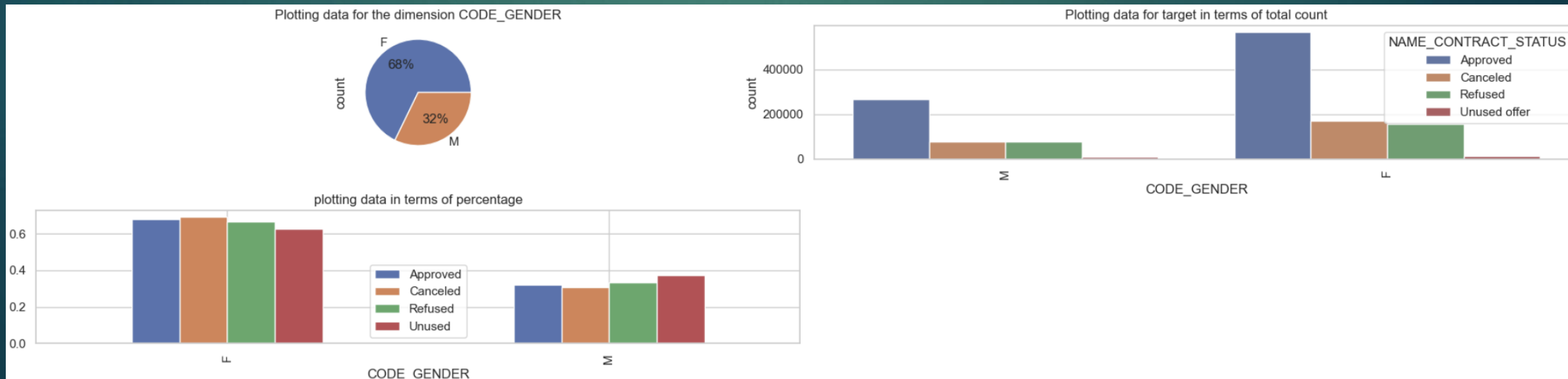
I would proceed with creating a categorical column by creating the bins

- 0-50000 - VL - very Low
- 50001-100000 - L - Low
- 100001-200000 - M - Medium/Moderate
- 200001-300000 - MH - Medium-High
- 300001-500000 - H - High
- 500001-1000000 - VH - Very High
- 1000001+ - UR - Ultra Rich

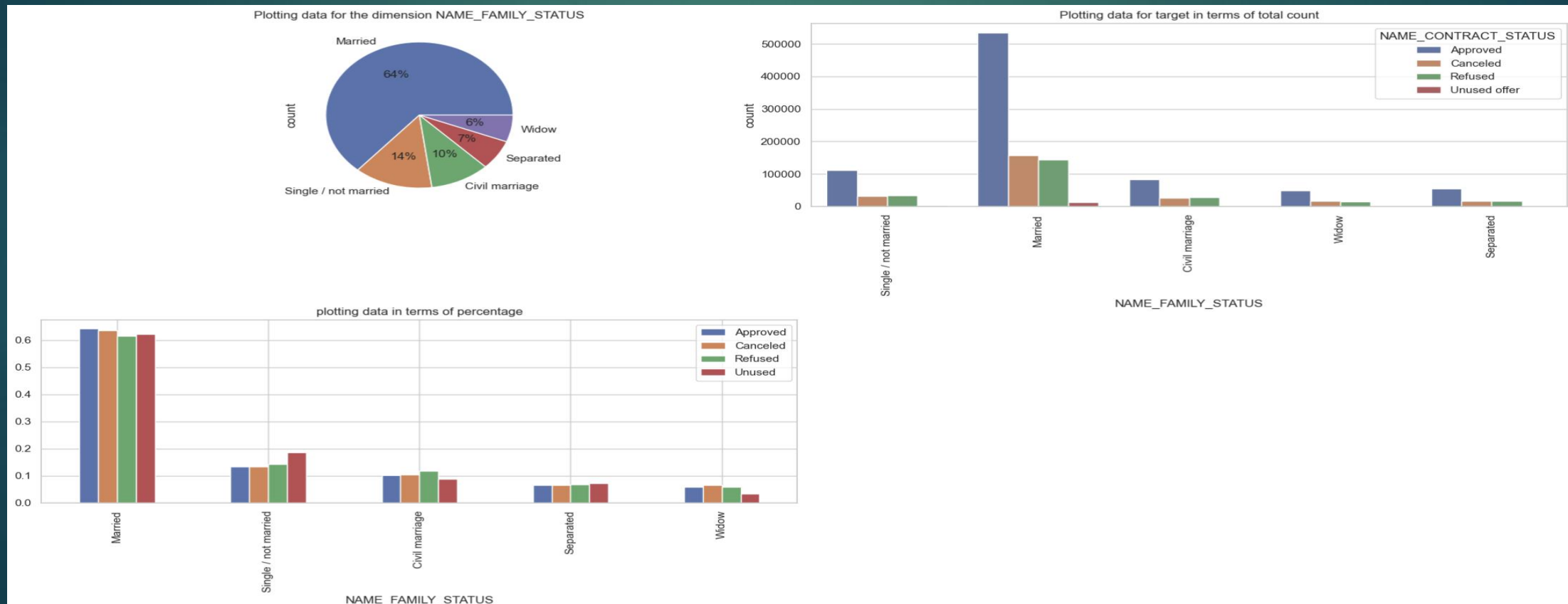


# Merged data

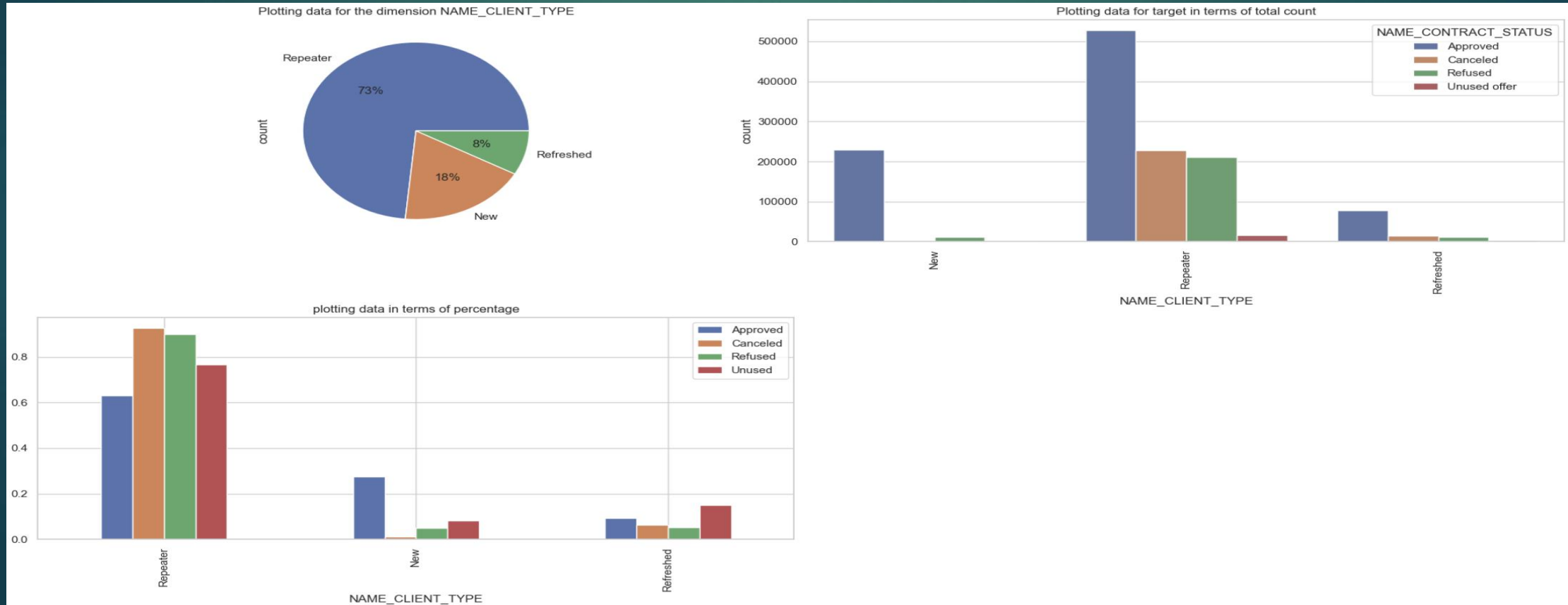
Sample has the most numbers of Female applicants hence in all cases such as Approved, Cancelled, Refused and Unused, Female has the higher rate compared to male applicants



1. Married peoples has the dominance in all categories
2. Single/ not married people has the highest % in unused category
3. Civil and single/not married people have the highest %rate of refused loan compared to others

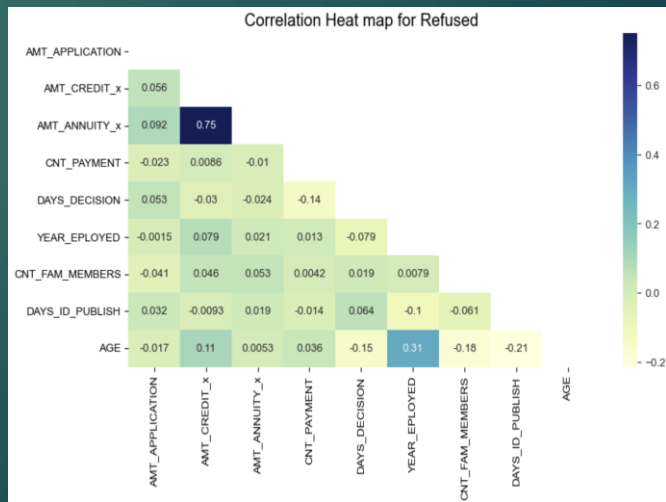
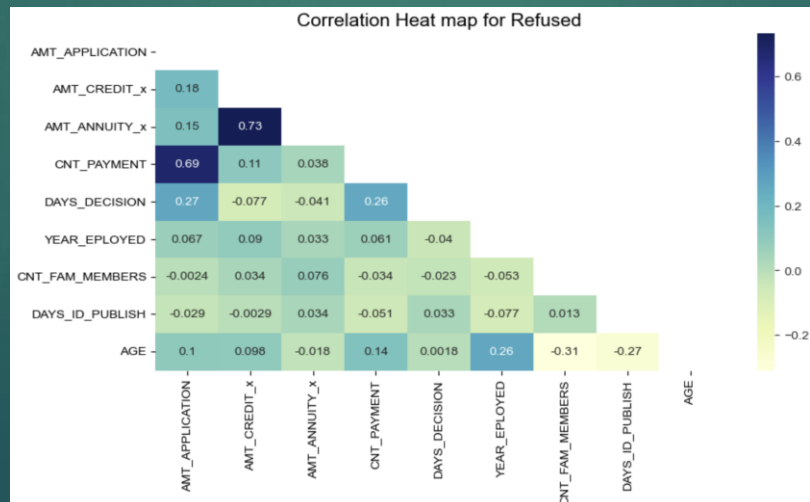
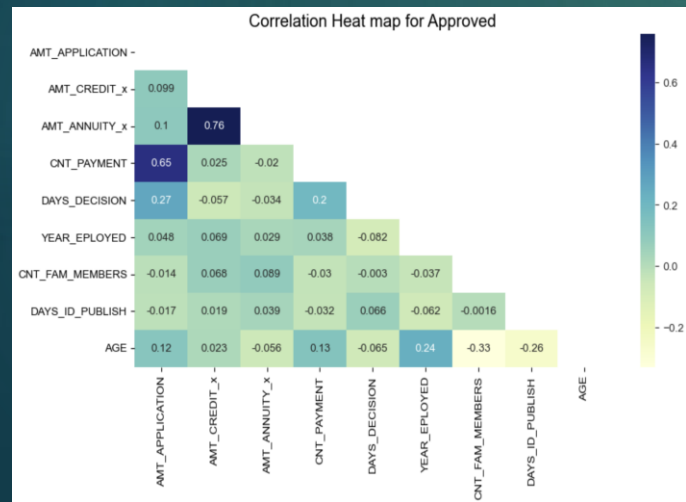
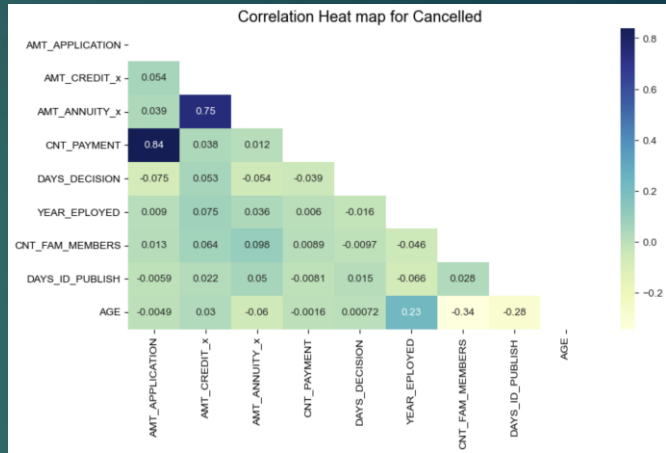
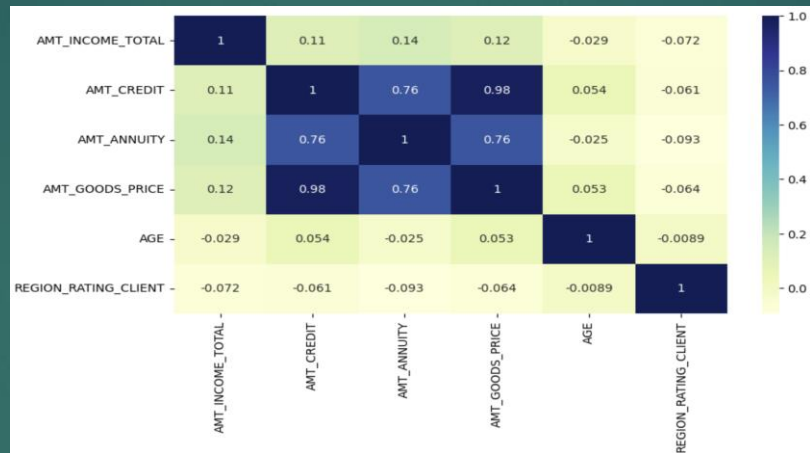
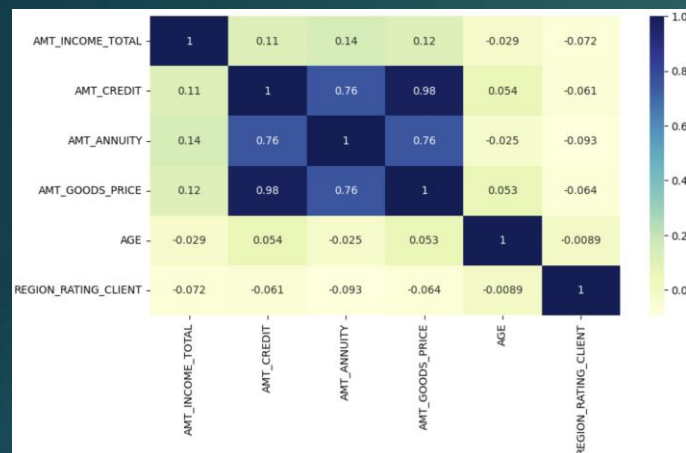


- ▶ Loan Repeater has the dominance in the loan application
- ▶ Repeater has the highest % of Refused and cancelled rate.
- ▶ Refreshed has the highest rate of unused if we consider the numbers of applications for each category
- ▶ New has the highest rate of approval if we consider the numbers of applications for each category





# Correlation



# Suggestions:

- ▶ Loan providers should target more the education type 'Higher education' and try to avoid Secondary/secondary special, incomplete higher or lower secondary as they face payment difficulties.
- ▶ Avoid 'Working' clients as they have high percentage of payment difficulties. Rather focus on Commercial associate, pensioner and State servant.
- ▶ Should focus on applicants from housing type 'House/apartment' Yes as they are having less payment difficulties.
- ▶ Loan Providers should keep on focusing on Female clients, as they have the low percentage of default rate compared to Male applicants.
- ▶ Loan Providers should track the occupation type more seriously as 'Unknown' type has the highest % of successful loan payment history. If we can reveal the OCCUPATION\_TYPE for those class Loan Providers can easily target the group.
- ▶ Loan Providers should focus on the client from age group of 43 to 70 as they will be financial stable and shows less paying difficulties.



# Thank You!

[linkedin.com](#) ----- > Souvik Sarkar