Credit EDA Case Study

upGrad & IIITB | Data Science Program - May 2023

Submitted By : Nitish Kumar

Introduction

Welcome to the presentation on Exploratory Data Analysis (EDA) for risk analytics in the banking and financial services sector.

Scenario

- We are addressing the needs of a consumer finance company that specializes in providing loans to urban customers.
- The company faces two critical risks: approving loans for clients likely to default and rejecting loans for trustworthy clients.
- Our objective is to perform EDA to identify patterns that indicate potential loan payment difficulties among clients.
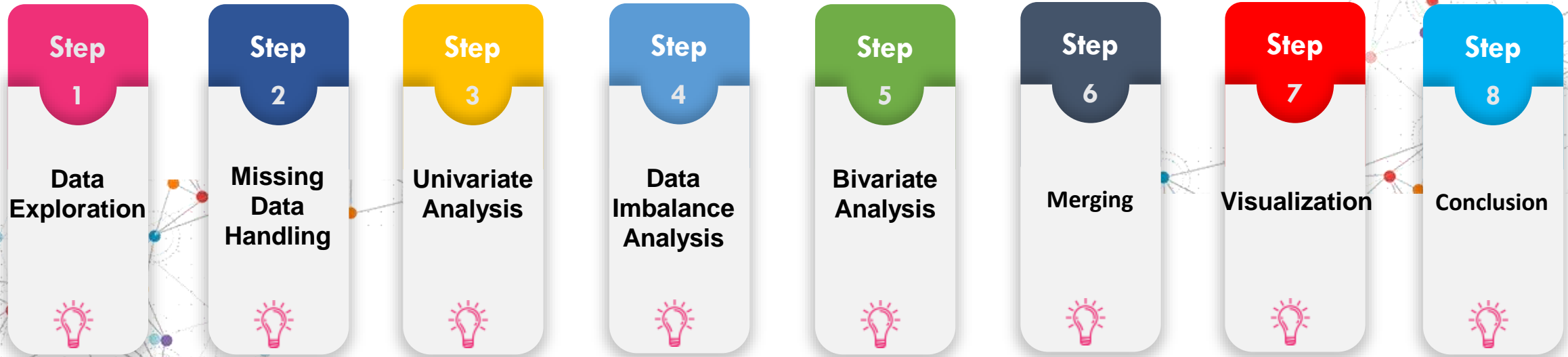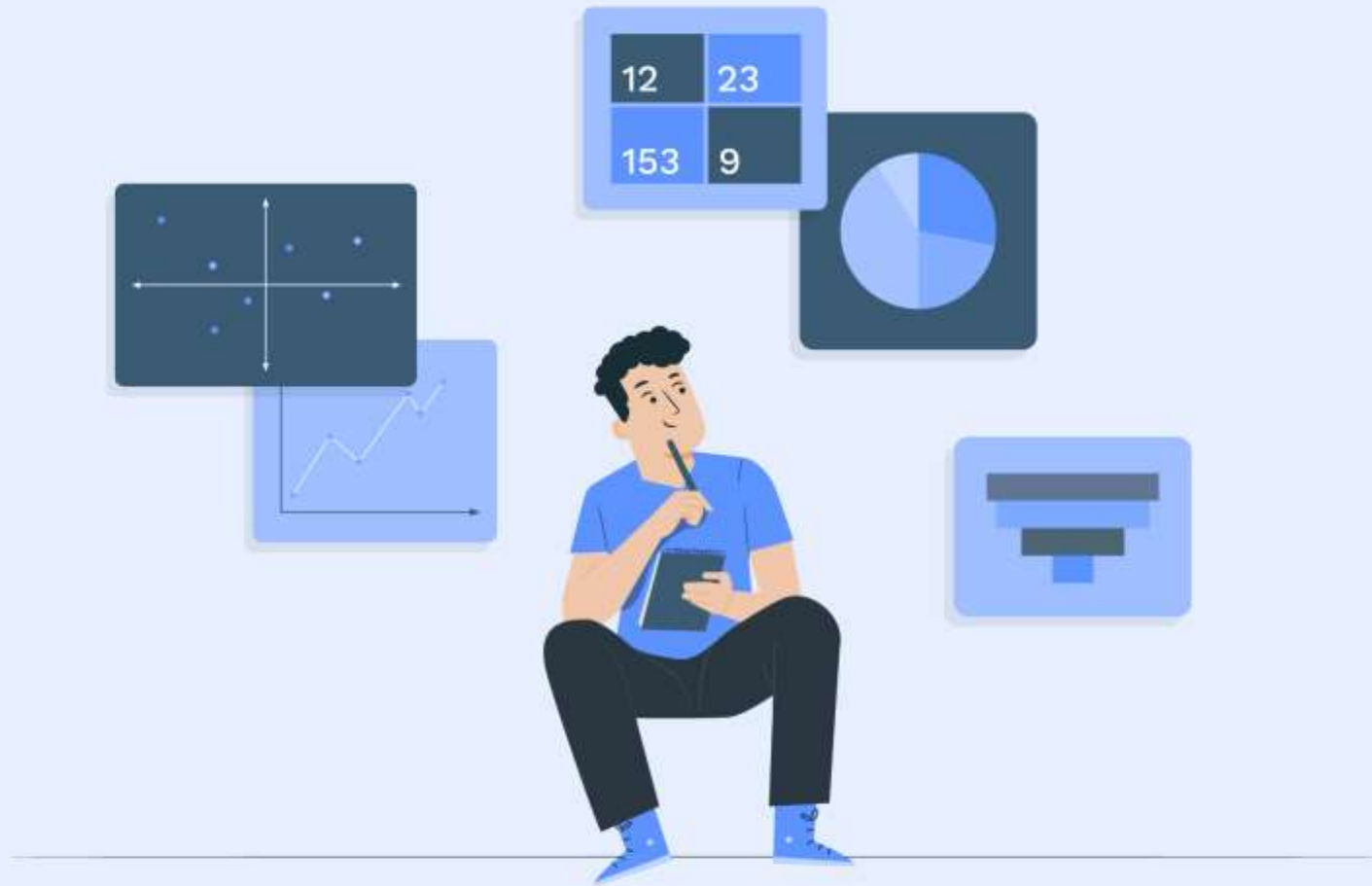
Problem Statement

- The main challenge is to strike a balance between approving loans to minimize business loss and avoiding defaults to reduce financial losses.
- By understanding the driving factors behind loan defaults, the company can take appropriate actions, such as adjusting loan amounts, offering higher interest rates to risky applicants, or denying loans where necessary
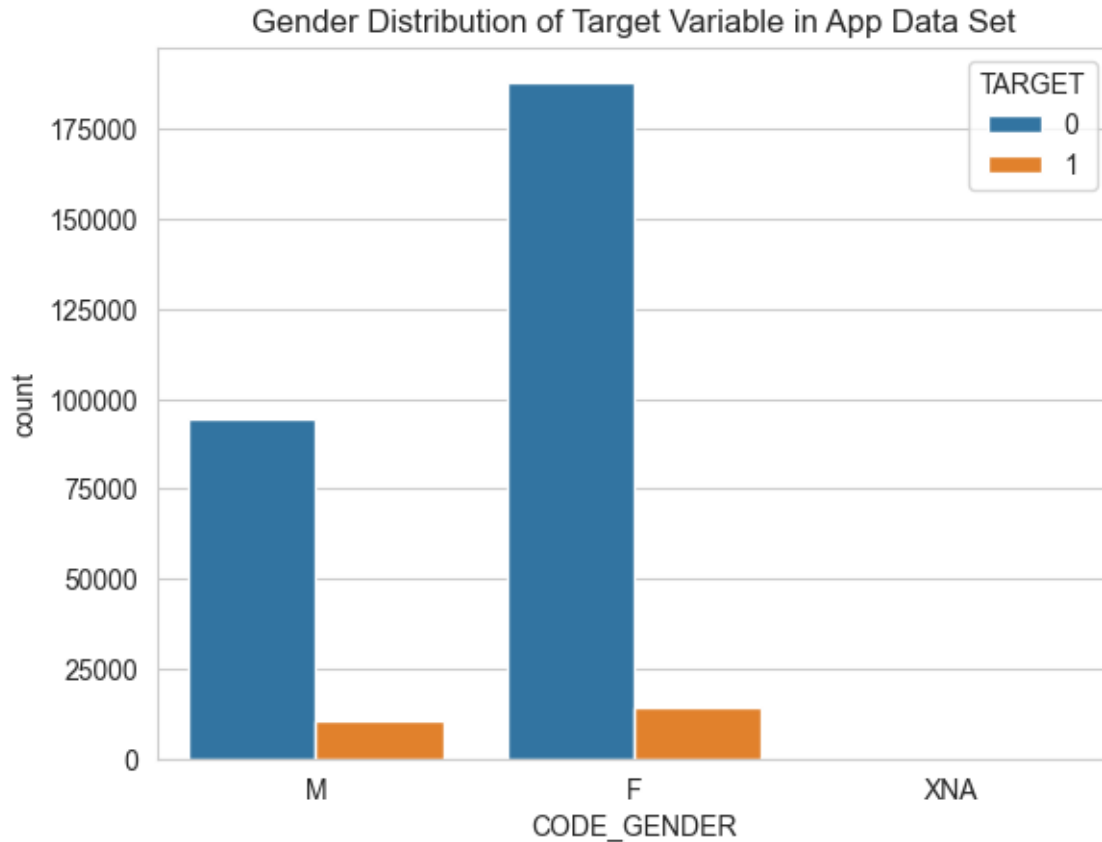
Approach

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8
Data Exploration	Missing Data Handling	Univariate Analysis	Data Imbalance Analysis	Bivariate Analysis	Merging	Visualization	Conclusion

Data Analysis | Visualizations

Gender wise Defaulters

Gender Distribution of Target Variable in App Data Set

It is evident from the graph above that men are more prone than women to default on their obligations. In other words, women would make better candidates for loan recipients

Men defaulted more because the bank had more female loan applicants than male applicants. Although there are fewer male applications, the number of defaulters is high or almost equivalent to that of female applicants; therefore, we may say that males default more frequently.
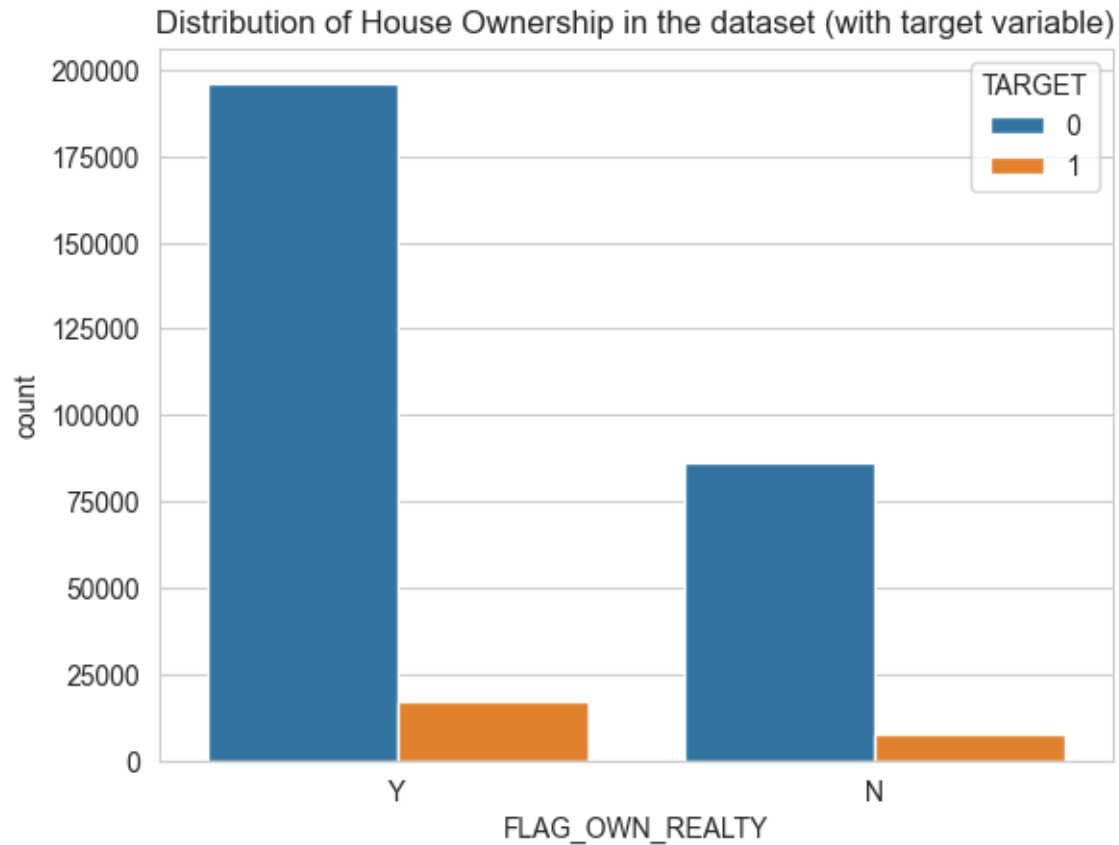
Defaulter against Car

People who possess a car have a higher percentage of defaulting on their loans. May be because of maintains .

Count of FLAG_OWN_CAR by Target Variable
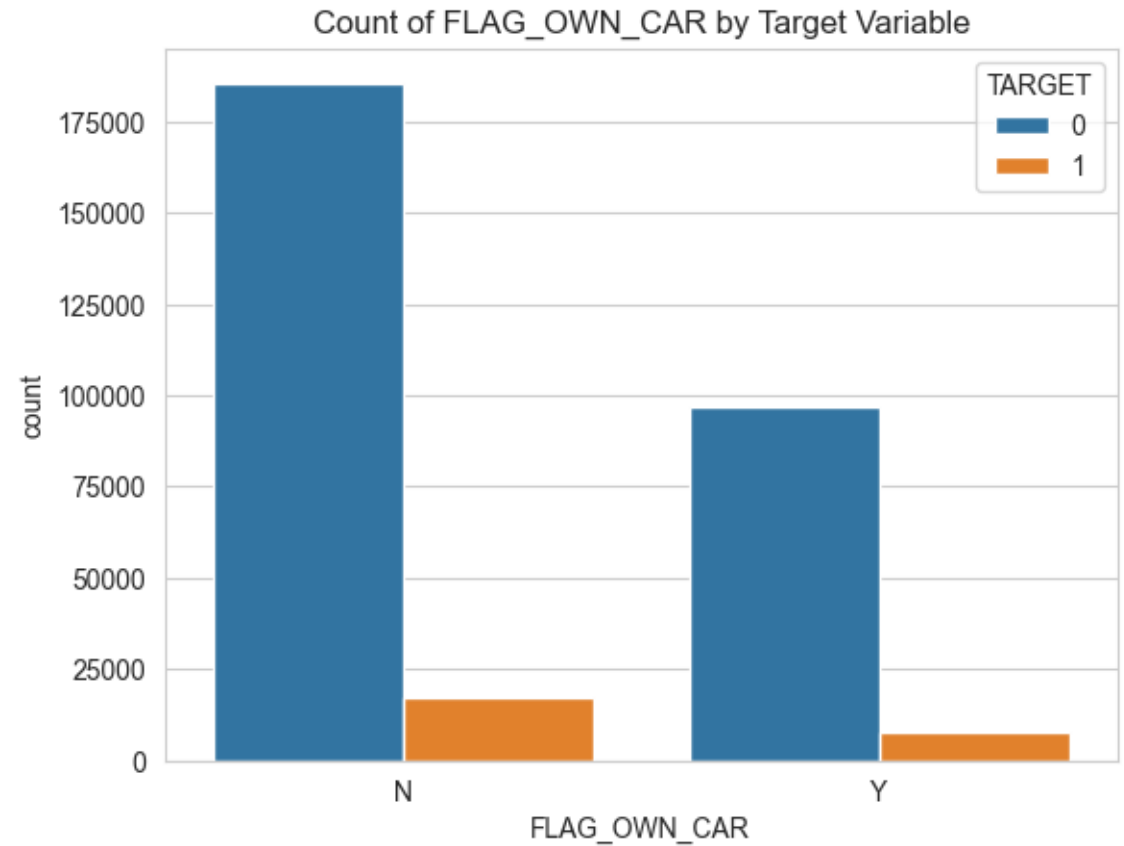
Defaulters statistics against Properties

Distribution of House Ownership in the dataset (with target variable)

Immovable property owners, such as people who own homes or apartments, are more likely to repay their debt than non-owners.

Those who don't own a house or an apartment could think that even if they don't pay their loan on time, the bank won't do them any harm, and they can easily leave whenever they want.

We could also assume that those who don't own homes must pay more rent and other costs.

Defaulter against Children count

Count of Children vs Defaulter Status in Application Data Set
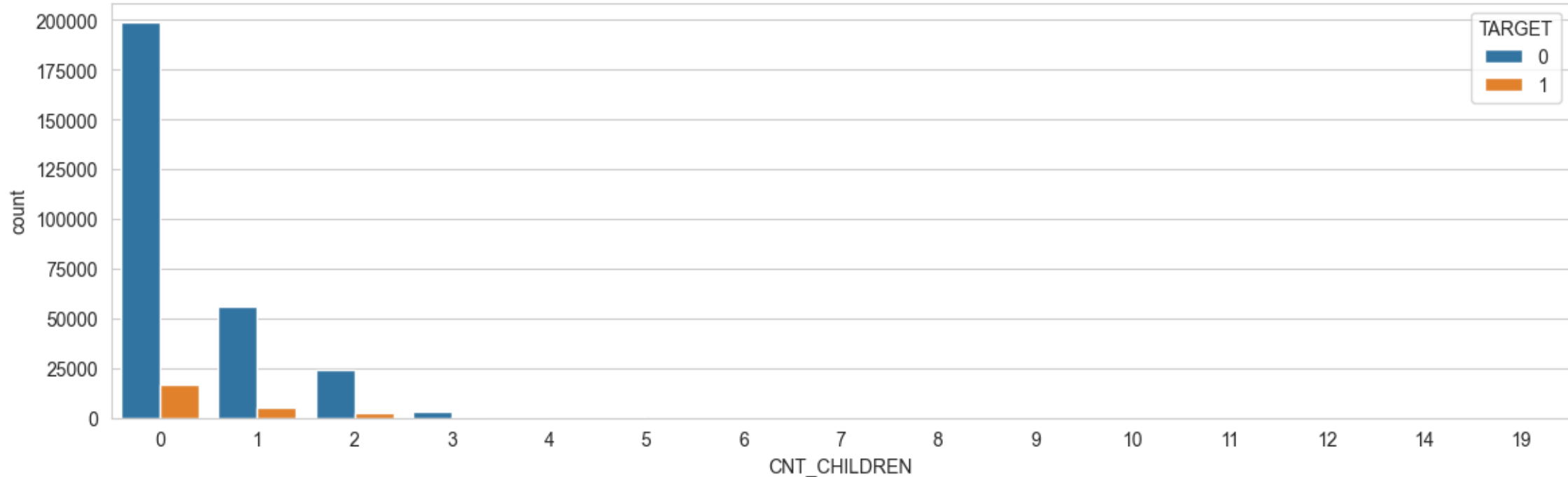
- When we did plotting with respect to application data, we came to know that the applicants with zero children were the highest among all (which means they applied more).

- Less child-parenting clients are more likely to repay their obligations. Lower the number of children, the greater the likelihood that the client will be the best choice.

Income distribution and with defaulter
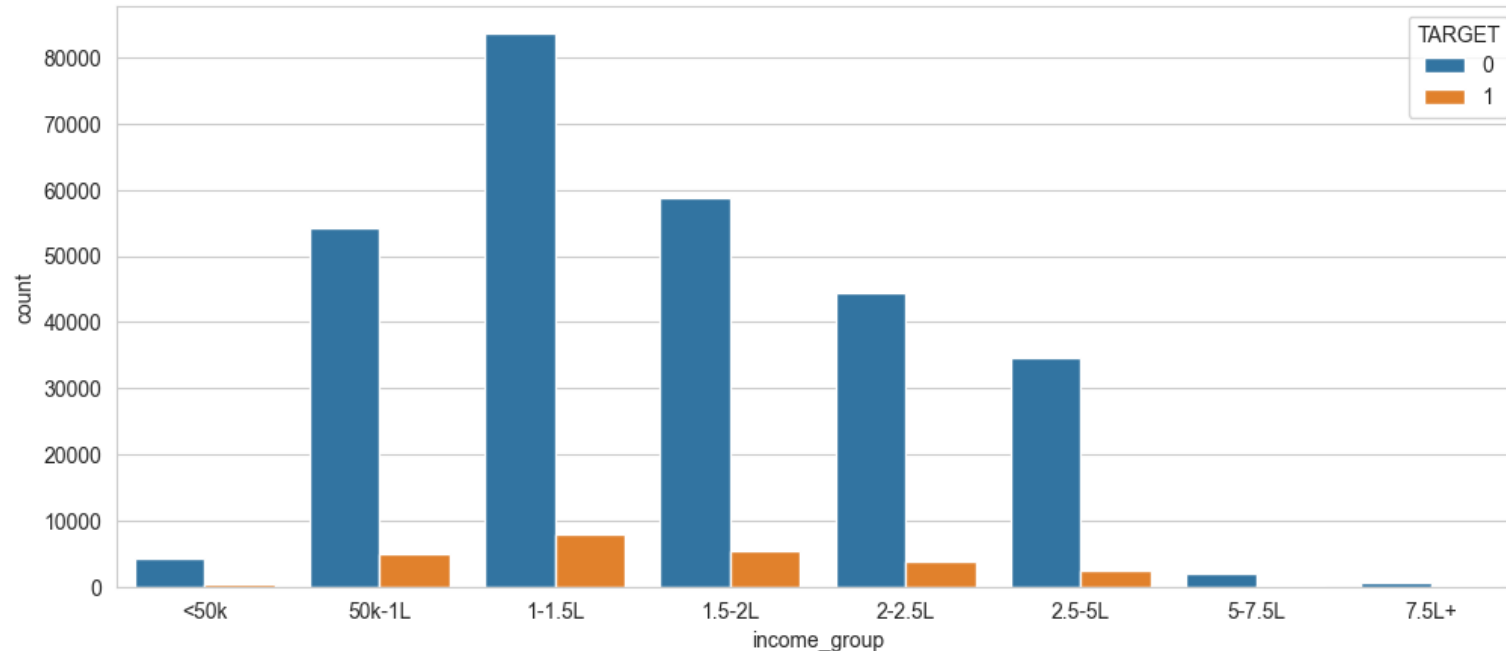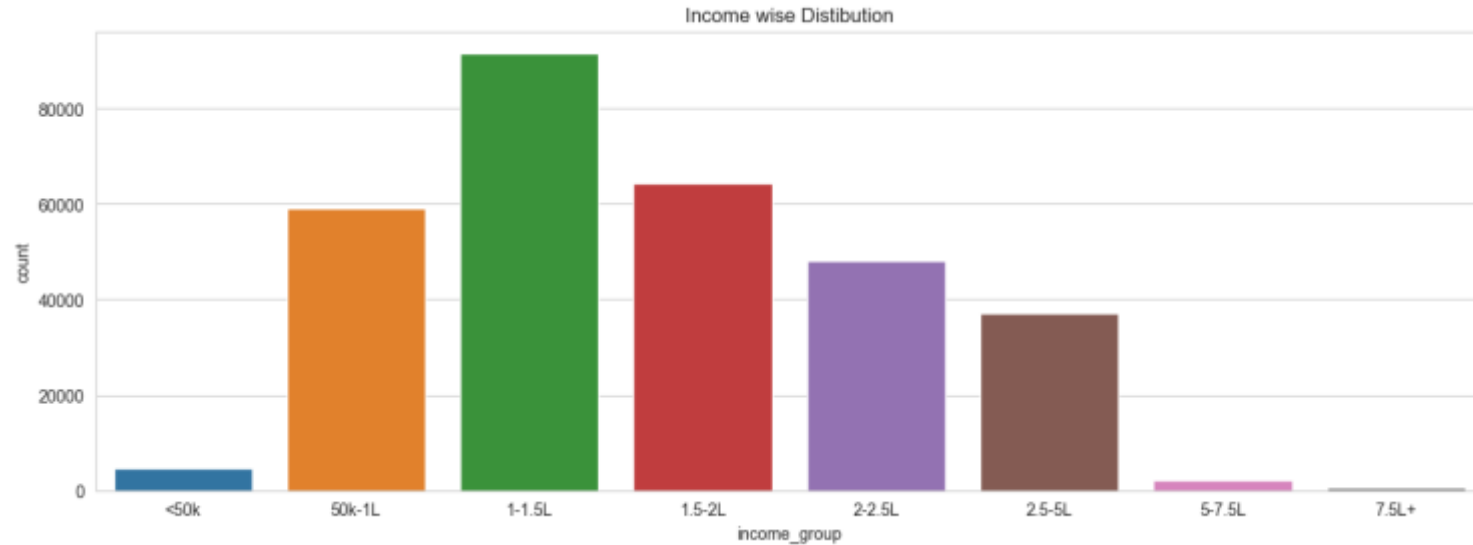
- The most number of people lies in the income range of 1-1.5L

- The income group that pays off their debts most effectively has an income between one and five lakh rupees.

Income type of defaulter

- Working individuals continue to be decent citizens by repaying their loans.

- But the defaults are also high in this category, which may be because the most number of applicants lies in this category and the change in defaults increases with the majority.
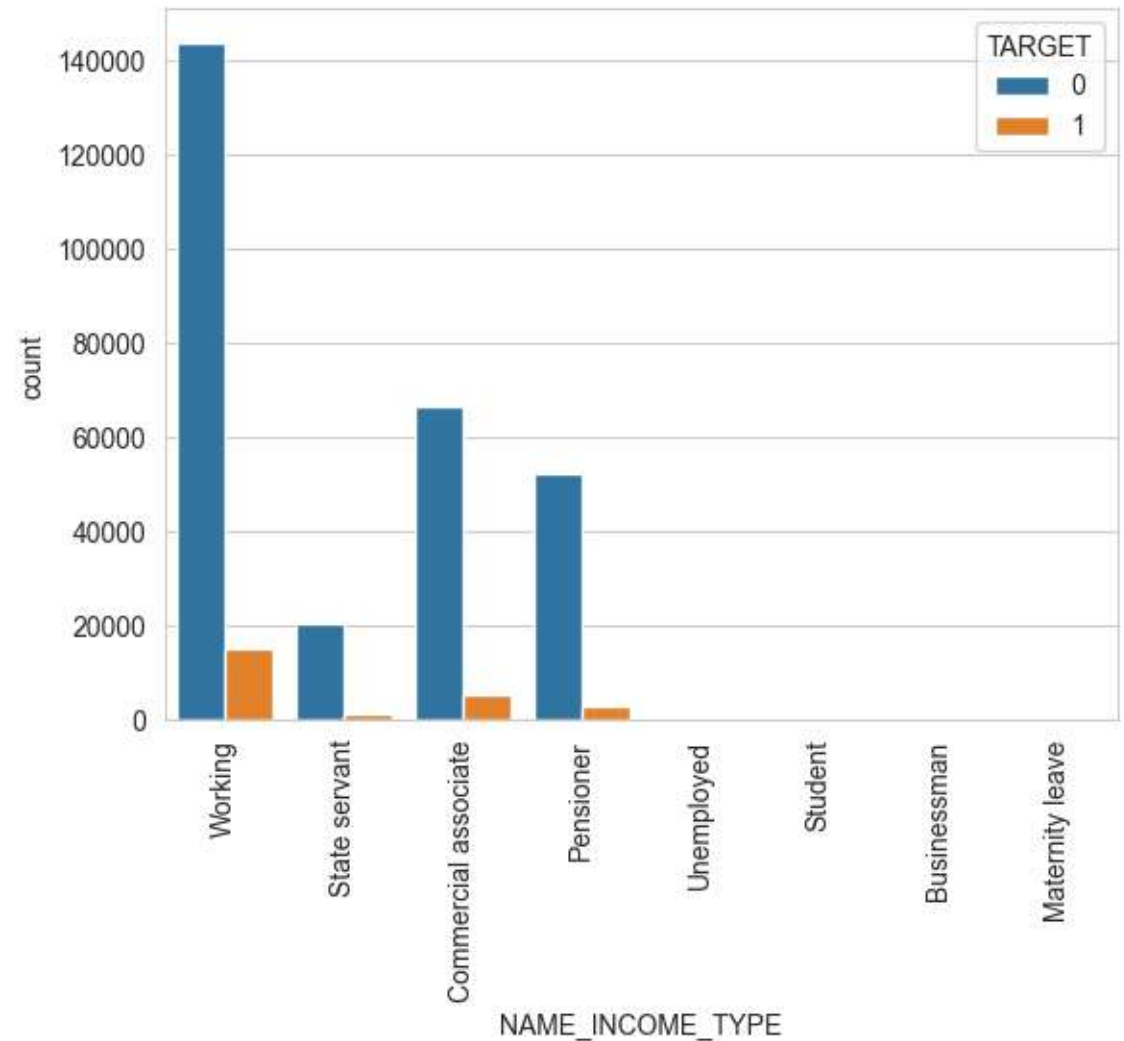
Accompany check with defaulters

NAME_TYPE_SUITE vs Defaulters Distribution in Application Data Set
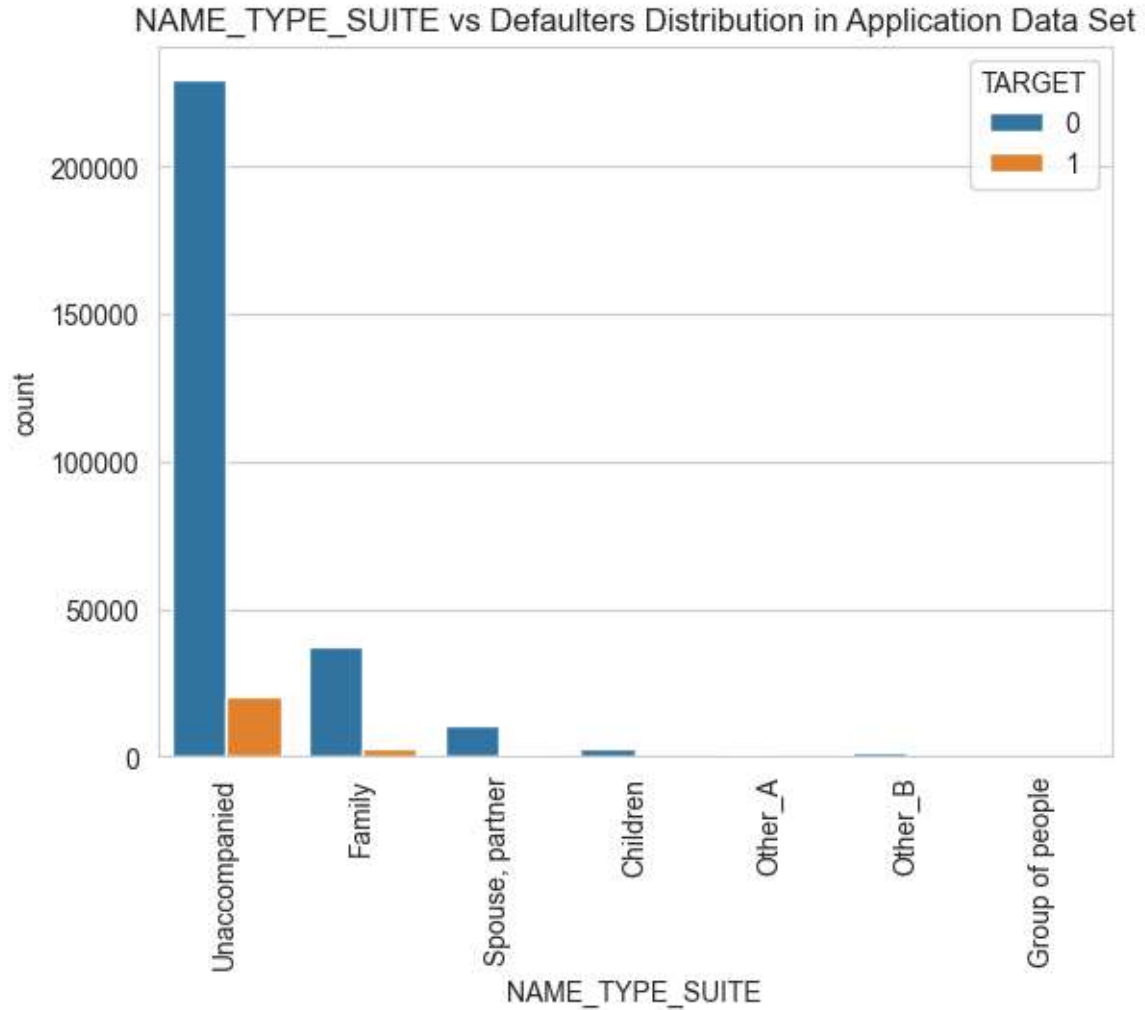
- The applicants who applied alone default more frequently, yet they are also consistent. This could be as a result of the highest number of candidates being turned down.

- Less defaults are made by candidates who were accompanied by a spouse or partner when they applied.

Occupation Type

Count plot of OCCUPATION_TYPE with appdata


Count plot of OCCUPATION_TYPE with Target variable
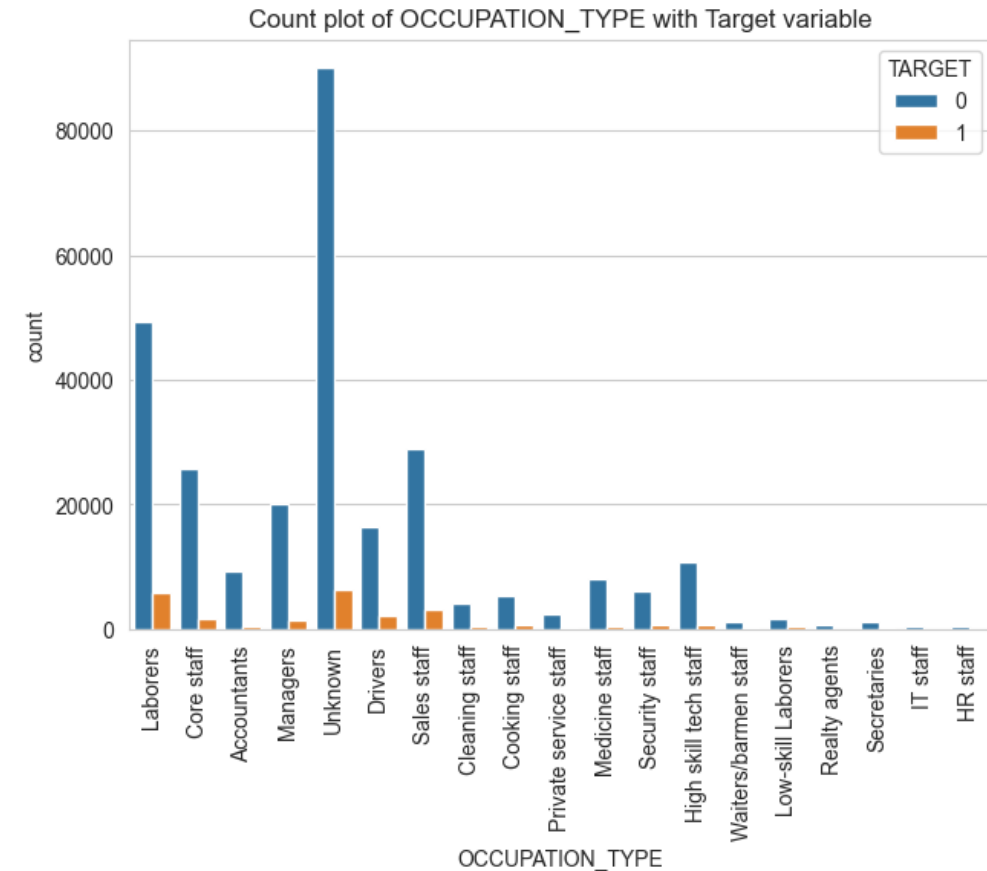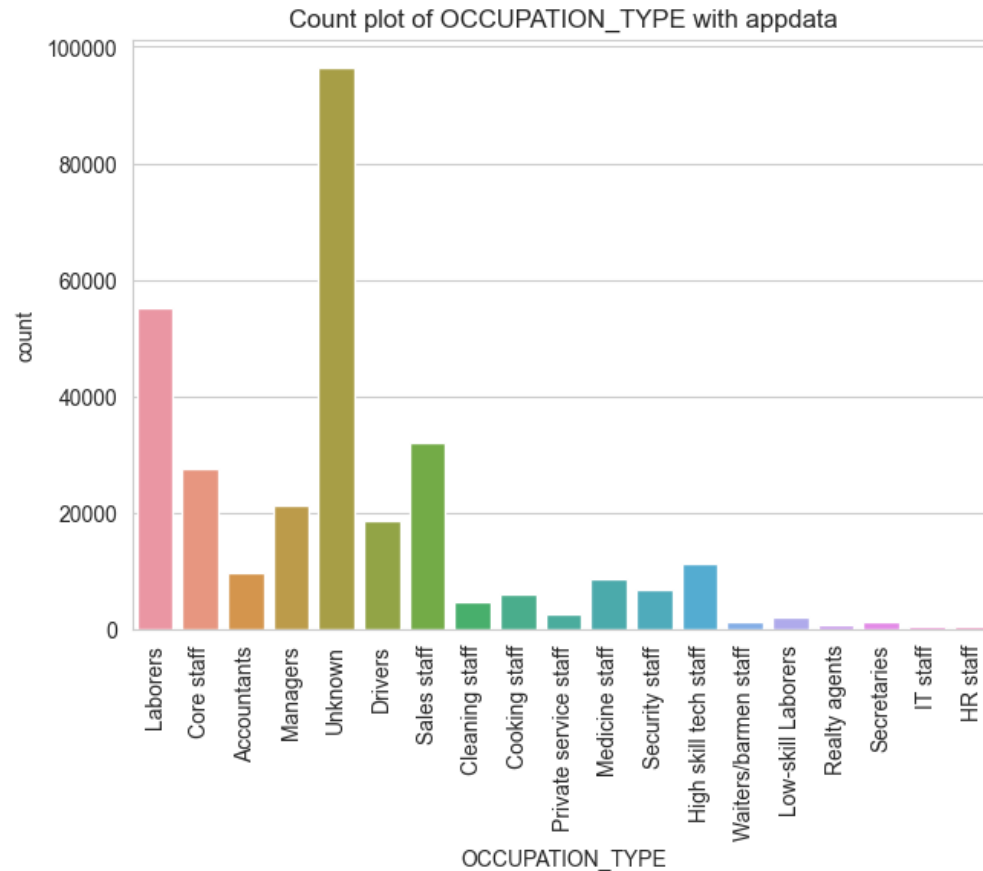
- This shows that the Business Entity Type 3 has applied most for the loan.

- Business Entity Type 3 defaults more, followed by self-employed compared to others.

Repeaters and status of loan
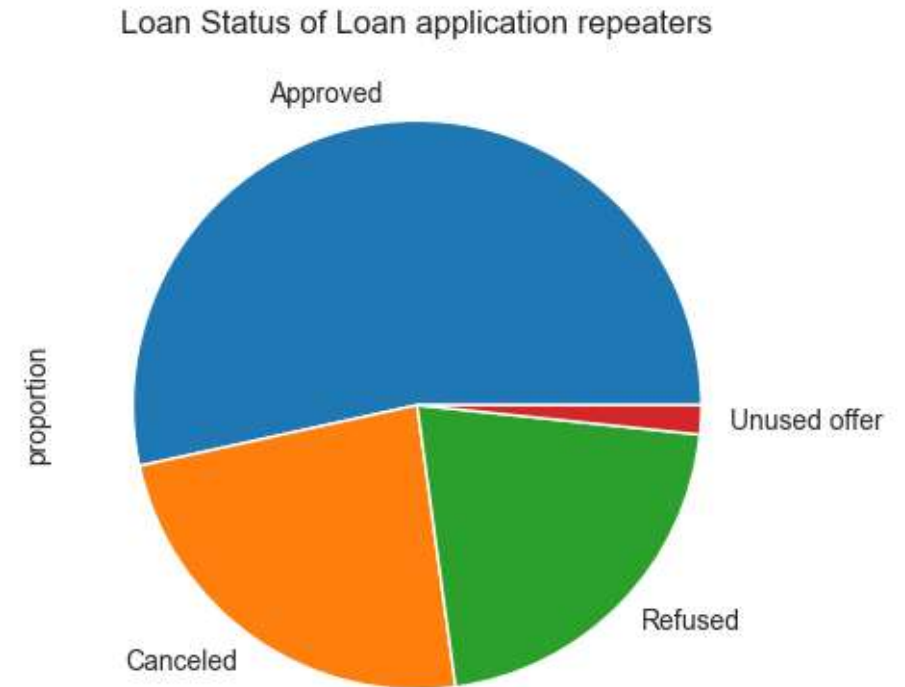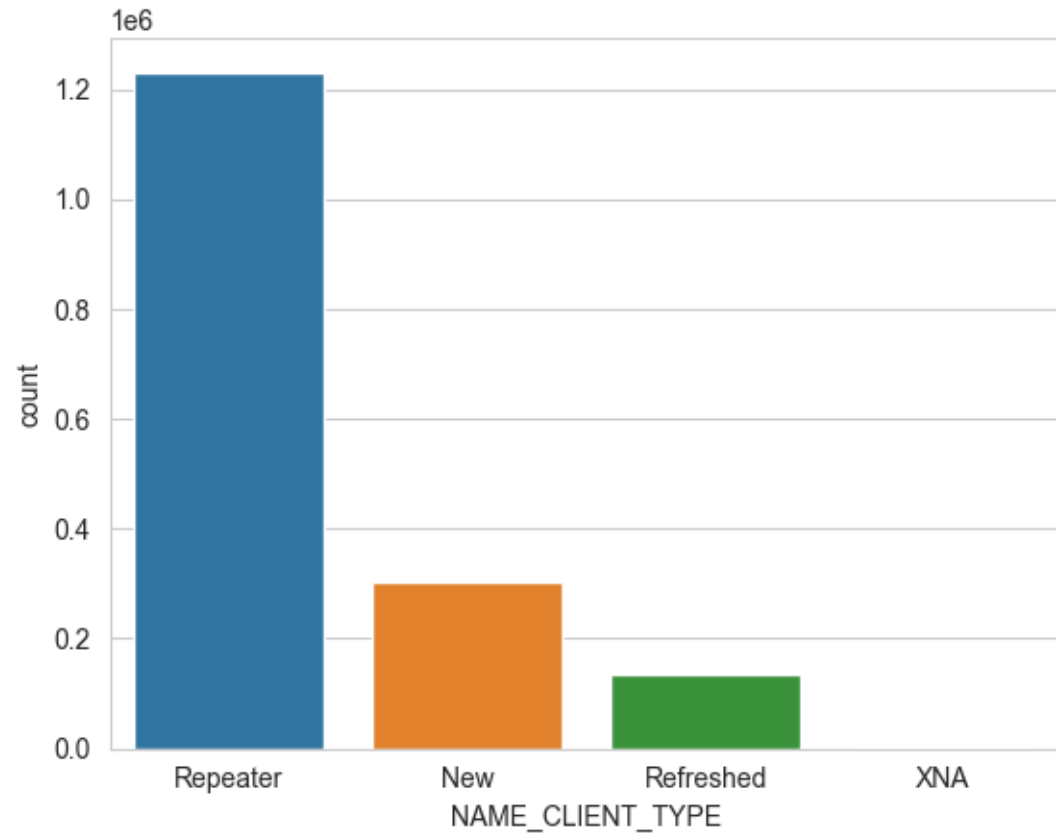
- People continue to seek for loans after receiving one. Repeat customers typically receive loan approval.

- People continue to seek for loans after receiving one. Repeat customers typically receive loan approval.

Merged data with Gender and loan status
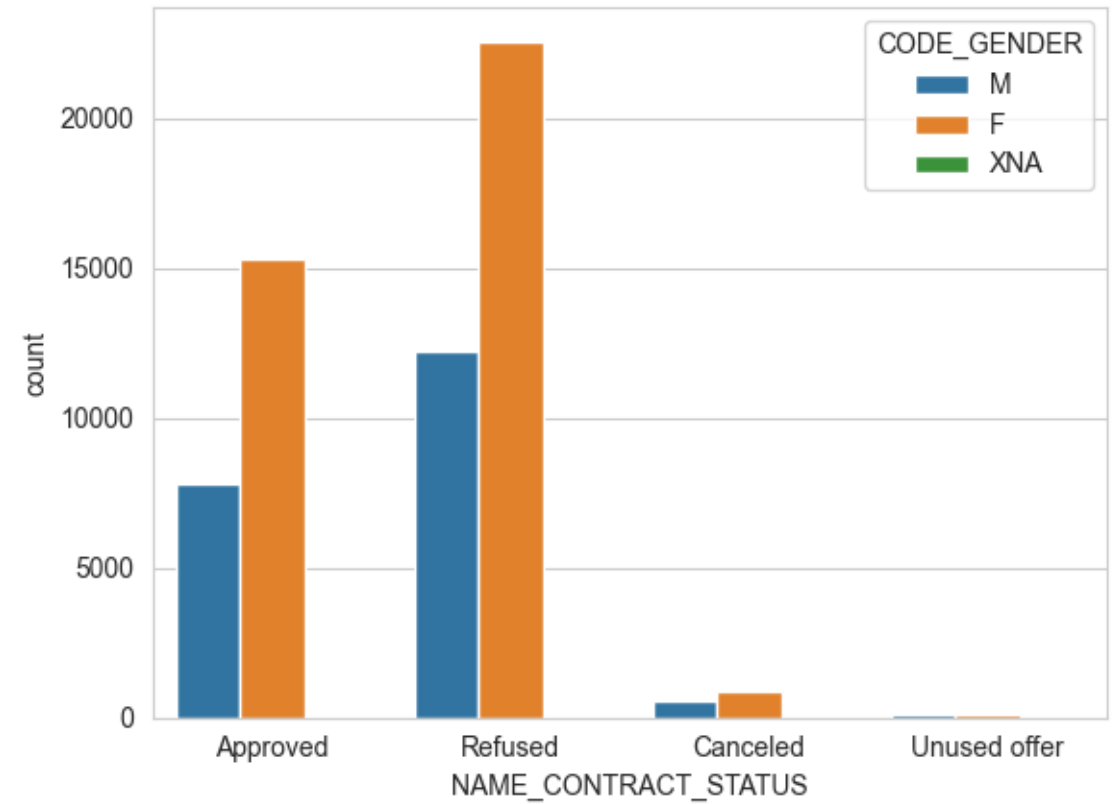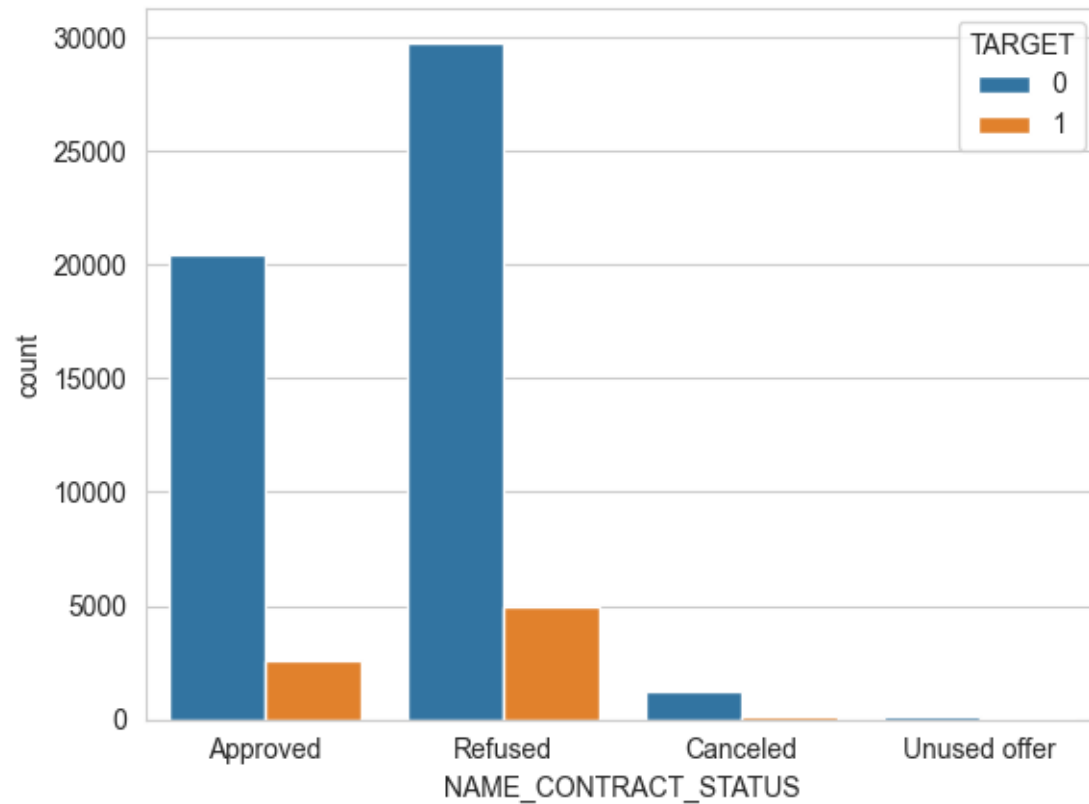
- The majority of defaulting clients are those whose prior applications were turned down. The best citizens, however, are also those whose prior applications were denied.
- Female application used to be refused more by the bank.

Conclusion

There may be many opportunities and new insights that have yet to be mentioned or discovered, but as of this writing, I have made a few suggestions and observations.

- Males are more prone than females to default on payments. In other words, women would make better loan recipients.
- Owners of cars are more likely than non-owners to default on their debts. May be because high maintenance.
- Owners of immovable property, such as houses or apartments, are more likely to pay off their debt than those who do not.
- Less child-parenting clients are more likely to pay off their obligations.
- The most effective income group is that of those who wipe off their debts and earn between 1-1.5 lakh.
- Working individuals continue to be decent citizens by repaying their loans.
- More defaulters are the people with lower income range.
- Repeat customers typically receive loan approval. Keeping customers is therefore crucial.
- The majority of defaulting consumers are those whose prior applications were denied.