CREDIT EDA ASSIGNMENT

- By Manit Malhotra

Problem Statement

- Use data minimize the risk of losing money while lending to customers

- Application data contains information about the customers who have taken loans

- 1 indicates that they had difficulty paying back whereas 0 indicates that they paid all the dues on time

- Other data set contains details about previous loan applications, indicating whether the loan was approved, refused, rejected or unused

- Using these data sets we need to help make a smart decision that whether an application should be rejected or accepted

Assumptions

- 'XNA' in both the datasets means Not Available

- 'XAP' in both the datasets means Not Applicable

Overall Approach

- We import the data present in csv files into data frames

- We analyse data and impute all the blank values

- We clean all the incorrect data present in the files

- We checked for outliers and how to impute them

- Once the data is ready, we perform univariant and bivariant analysis on it

- Once we have performed analysis on both the data sets, we merge the data in a single data frame

- We perform analysis on the merged data to get more insights

Univariant Analysis

Number of Credits taken based on gender

From the graph we can observe-

▶ Females take more number of credits than males

▶ Females have more successful loan payments than males

▶ Females also made slightly more unsuccessful payments than males

Loans Based on GENDER
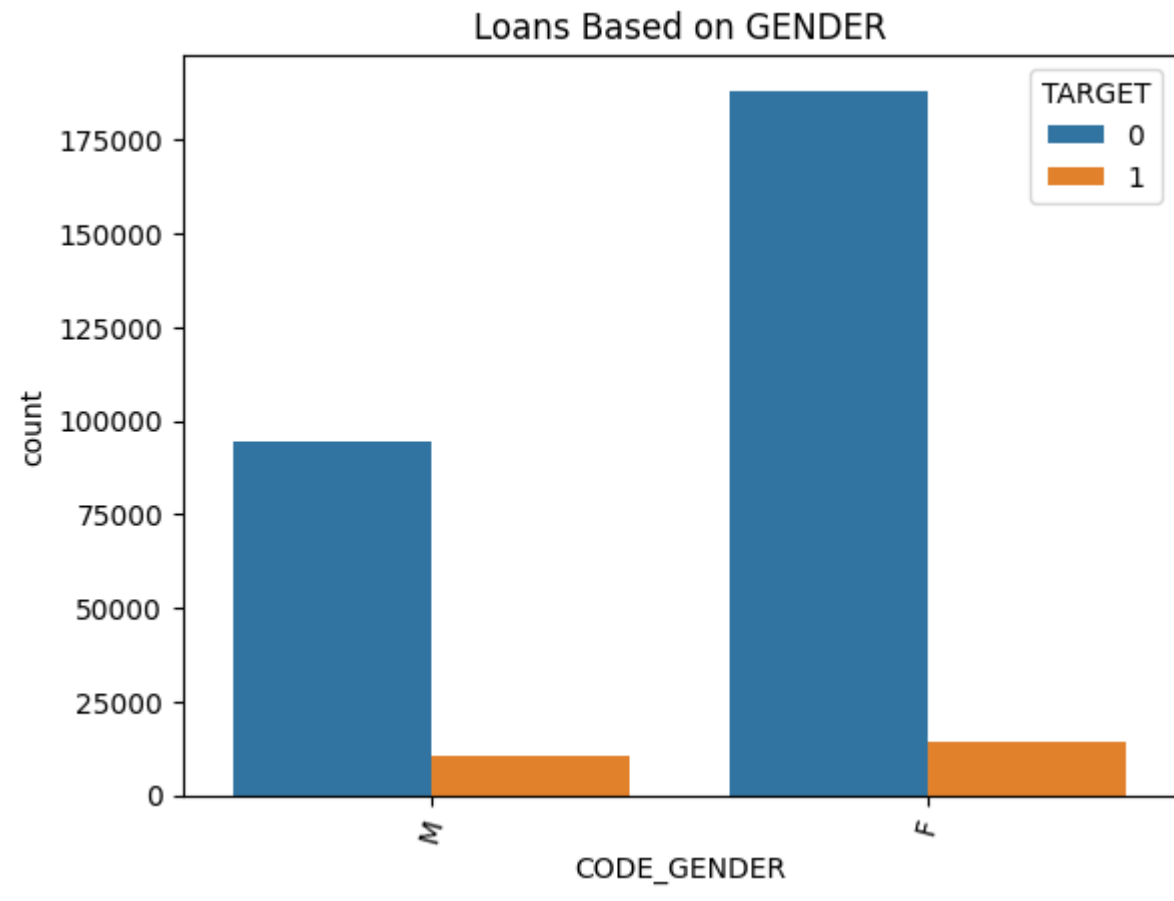
# Credits based on Income Type

From this we can observe the following-

▶ Customers whose income type is working take the most number of credits

▶ Customers with income type as working have the maximum number of unsuccessful payments

▶ Customers who are students or businessman have the least number of unsuccessful payments
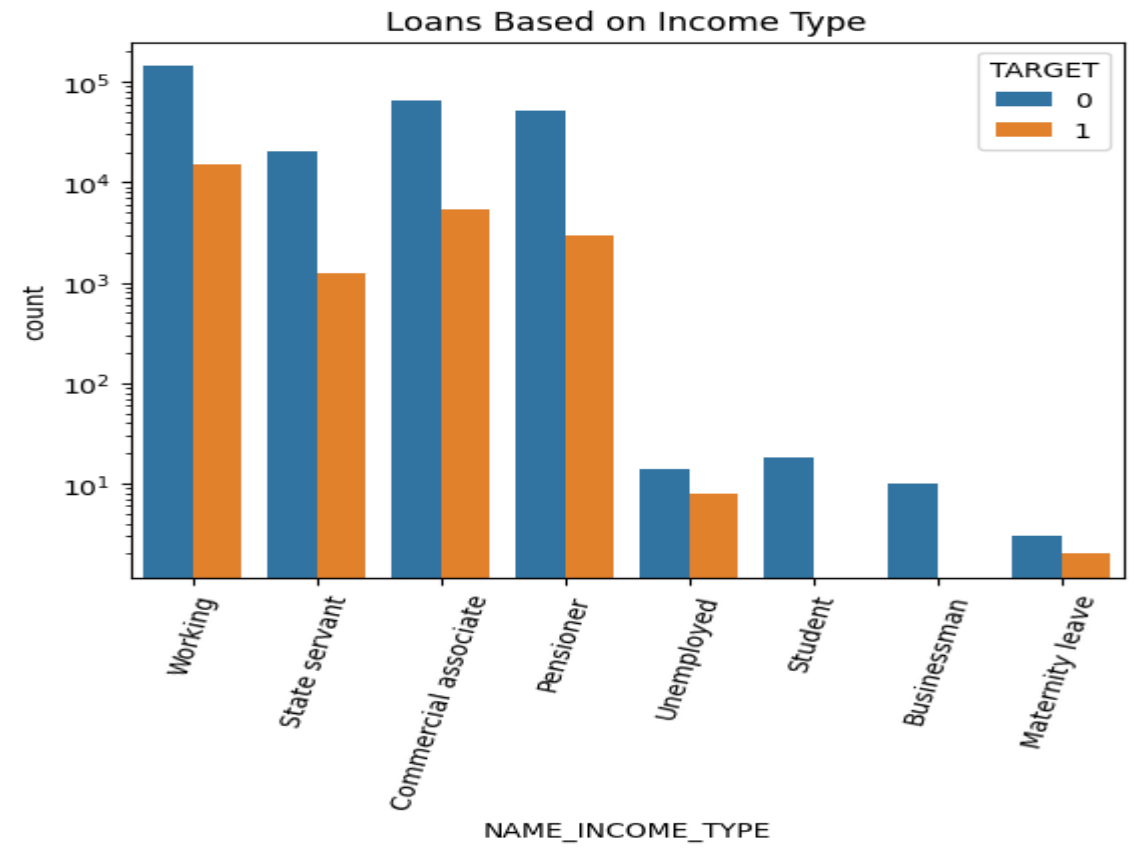
Credits based on ownership of real estate

From this we can observe the following-

▶ Customers who own a real estate take more number of loans

▶ Customers who own real estate have more successful payments
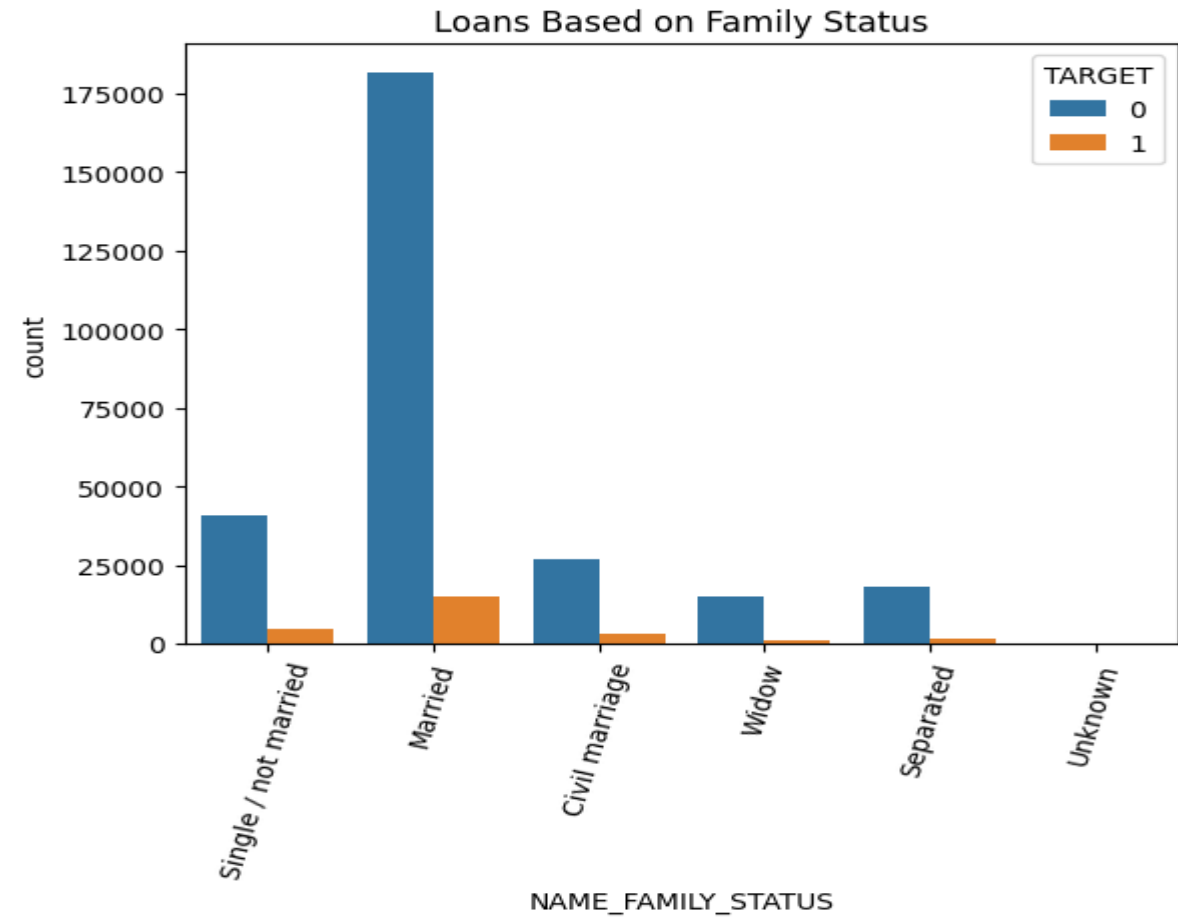

Loans Based on House Ownership
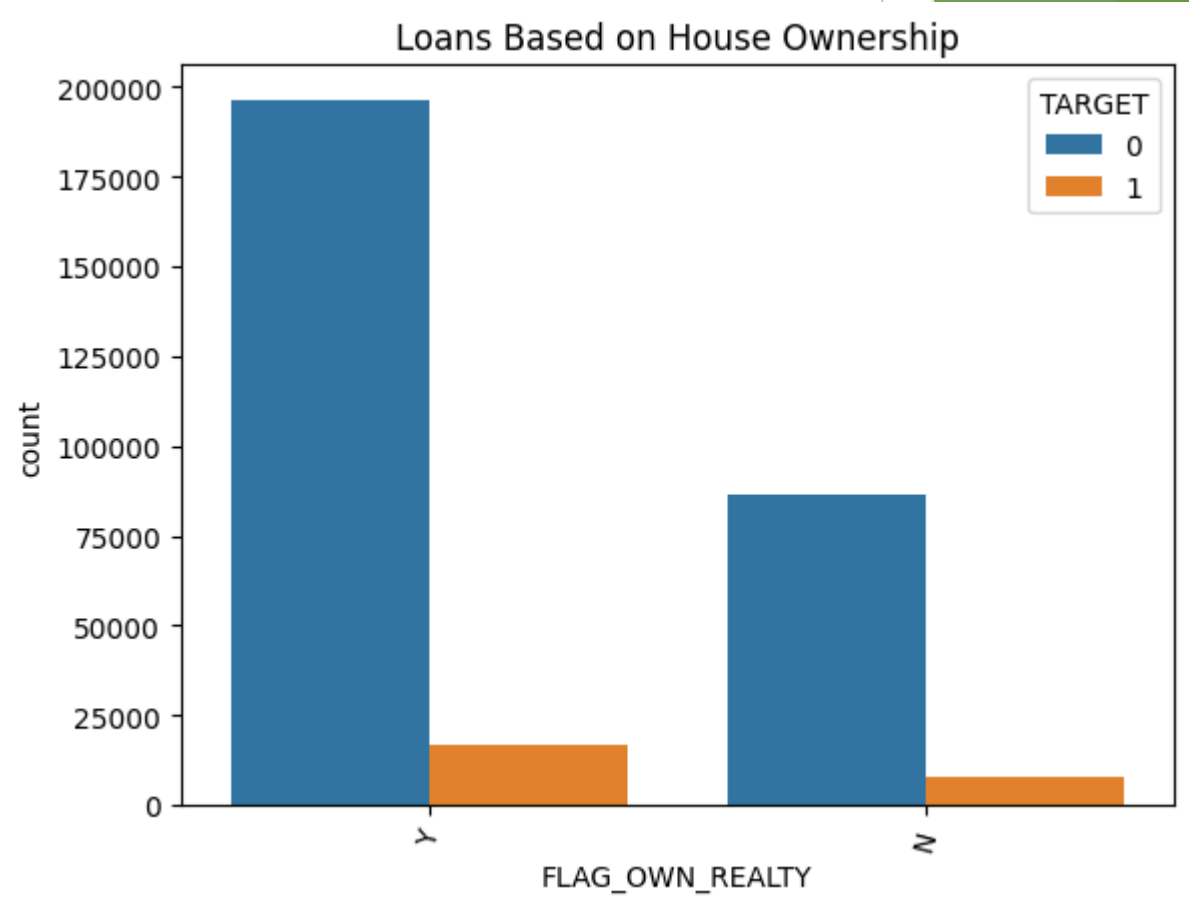
# Credits based on family status

From this we can observe the following-

▶ Married people tend to take more number of loans

▶ Widows and customers who have separated take the least number of loans

▶ Married people also have the highest number of unsuccessful payments

# Credits based on Occupation

From this we can observe the following-

▶ Laborers and sales staff take the most number of loans

▶ IT staff and HR staff take the least number of loans

▶ IT staff and HR staff customers have the least number of unsuccessful payments

▶ Laborers and sales staff have the most number of unsuccessful payments

Loans Based on Occupation Type
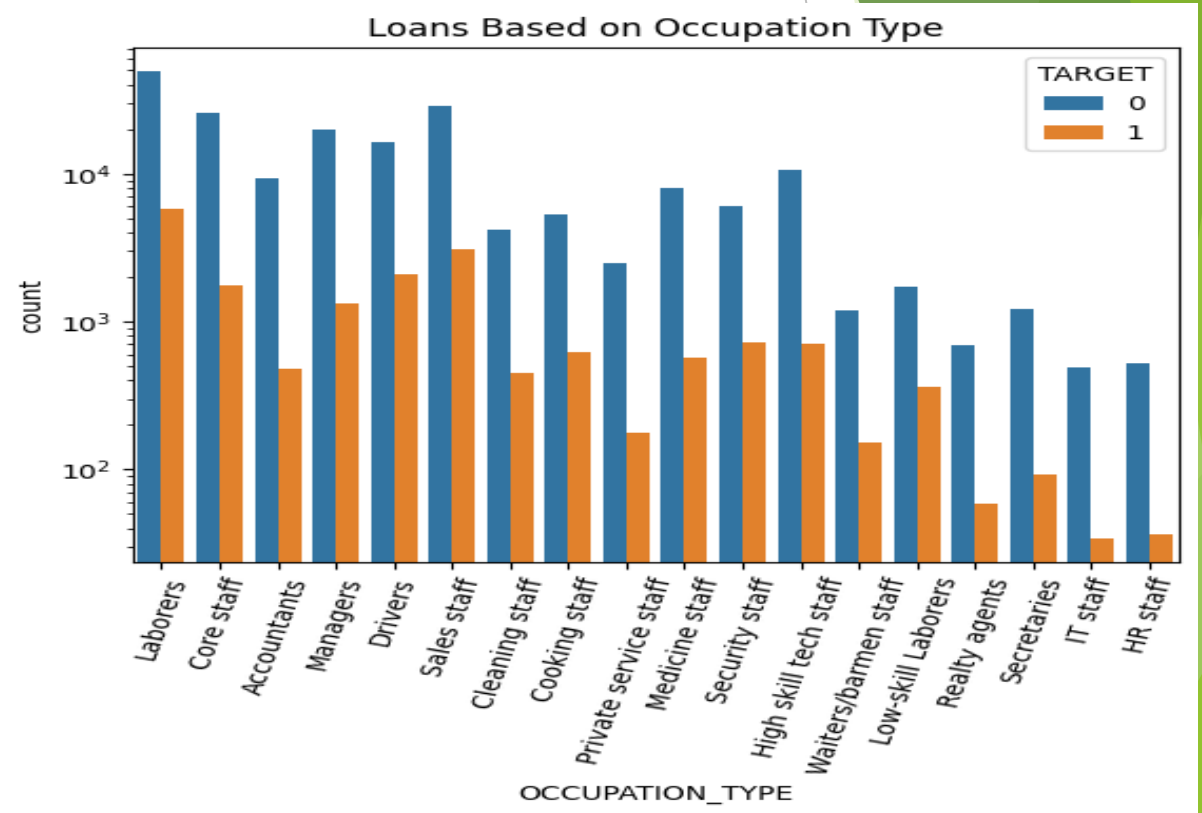
Credit based on Income Range
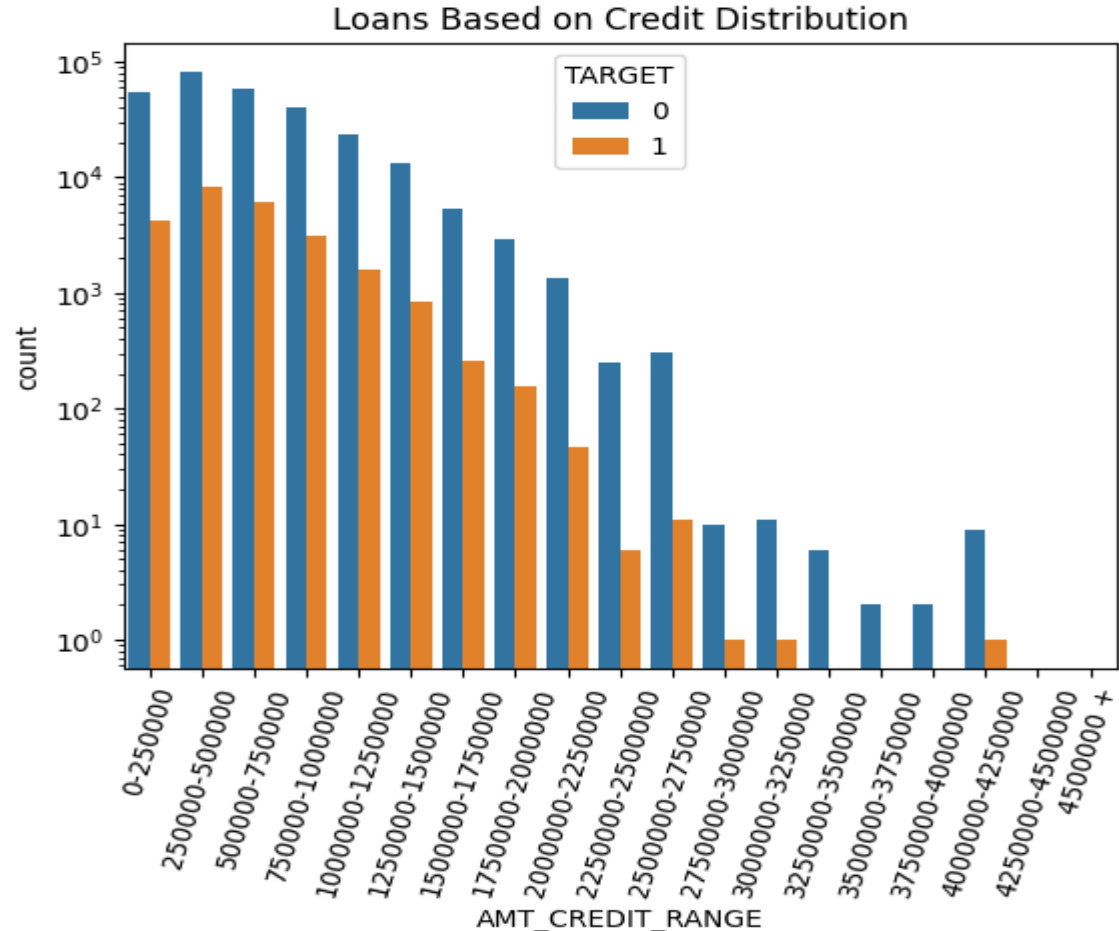
From this we can observe -

▶ Customers in income range of 0-500000 take the most number of loans

▶ Customers with income range of 2000000 - 2750000 and 3000000 - 4000000 usually repay their loans successfully

▶ Customers with income range of income range of 0-250000 have the most unsuccessful loan payments

Loans Based on Income Distribution
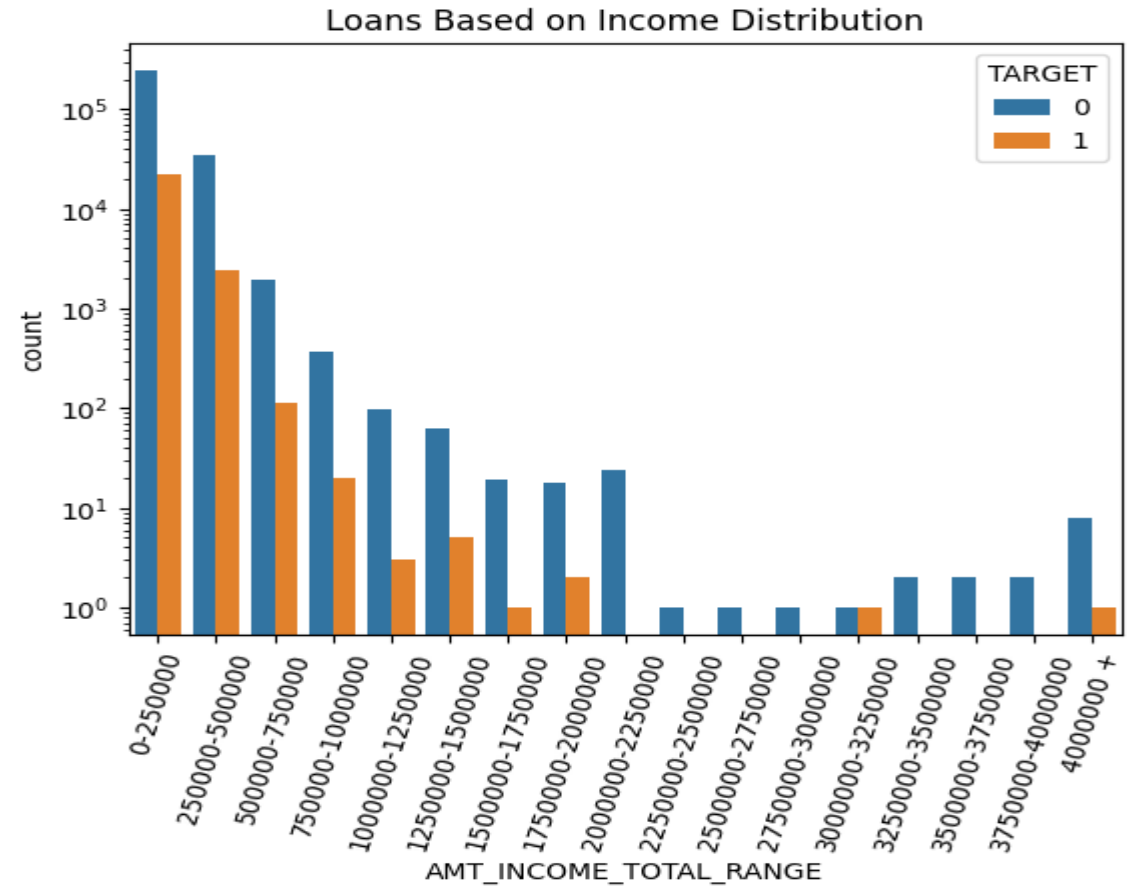
Credits based on the credit amount

- Customers whose credit range 3000000 - 3750000 usually don't have difficulties in paying back their loans

- Customers whose credit is in the range 250000 - 500000 take the maximum number of loans and have the highest unsuccessful loan payments

Loans Based on Credit Distribution

# Application Status based on Income Type

From this we can observe that -

▶ Working professionals have the highest number of approved loans followed by commercial associates and pensioners

▶ Customers who are either unemployed, students or on maternity leave have very less number of loans
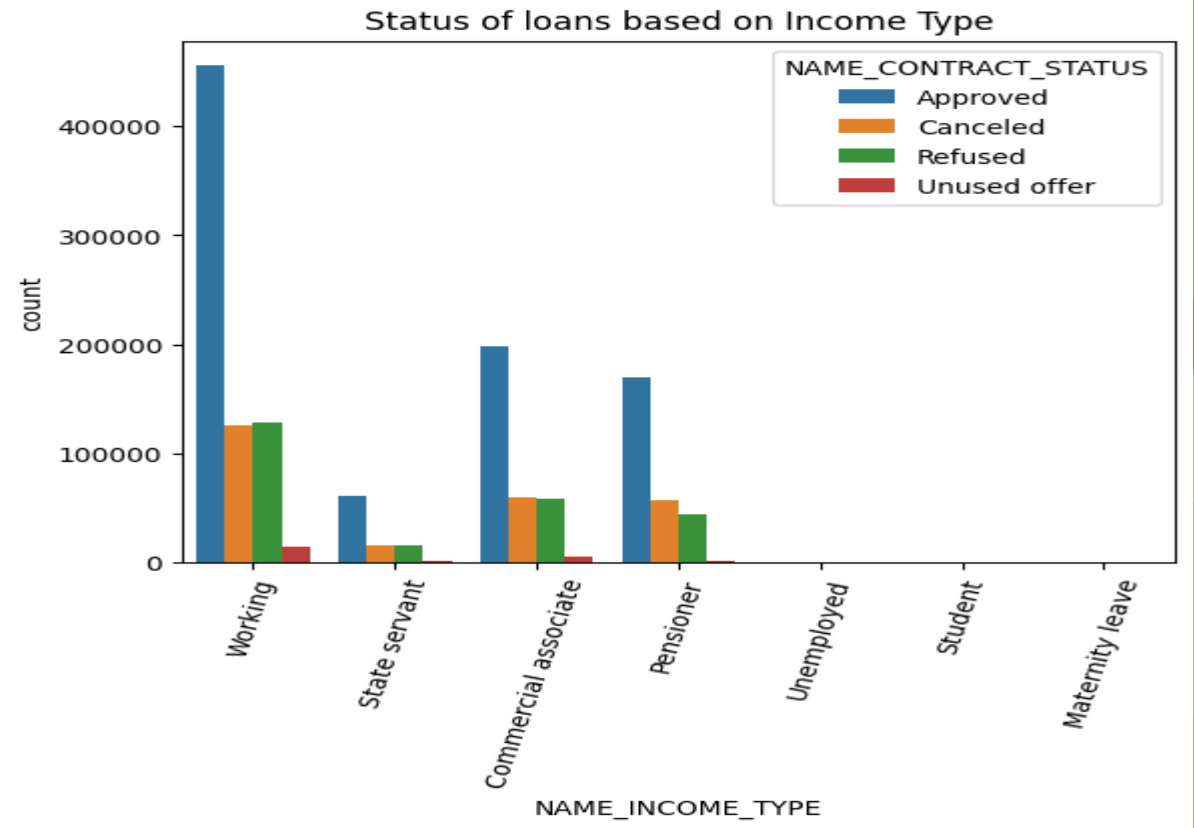
Credits for various purposes

From above we can observe that-

▶ Maximum number of loans are being taken for the purpose of repairs

▶ Repairs also have the highest number of defaulters as compared to others

▶ Customers who have taken a loan for the purpose of buying a garage usually don't face any difficulty in paying back the loan

Loan payments for various cash purposes
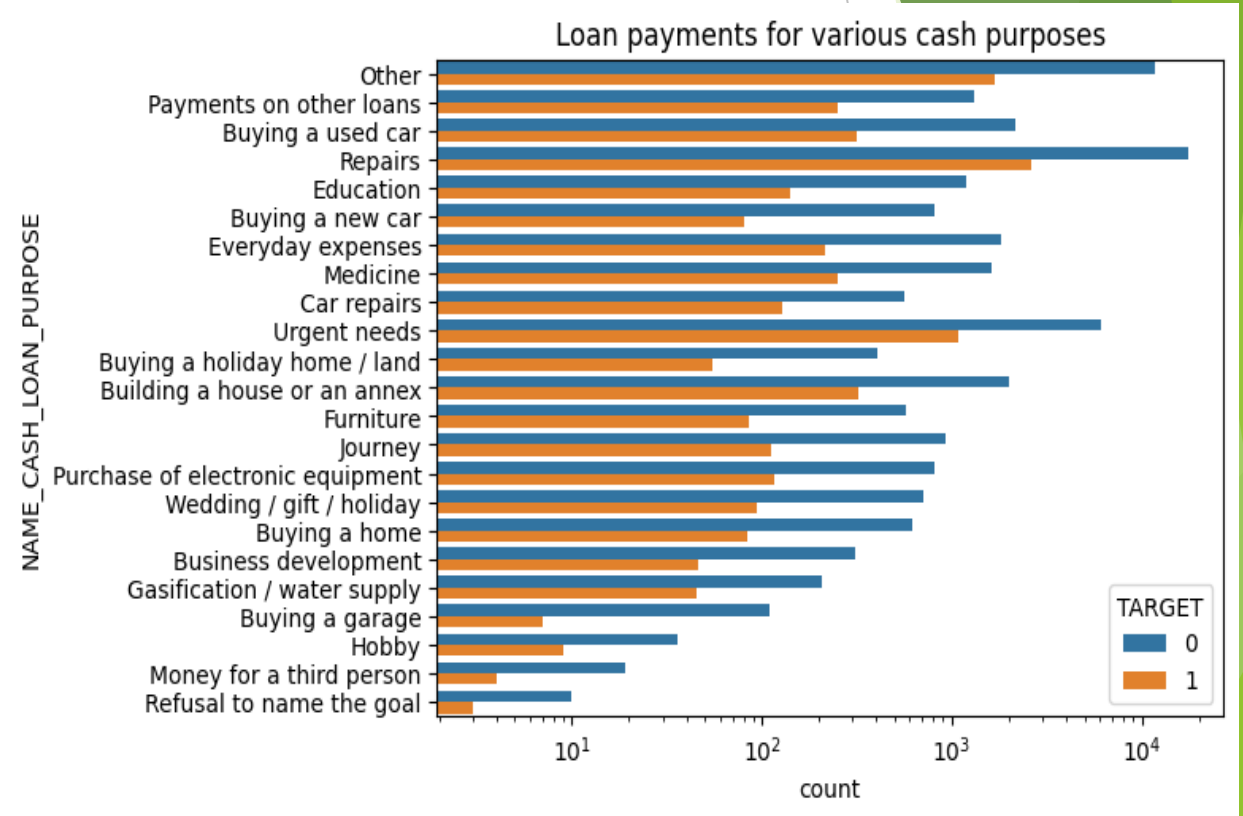
Credit for various cash purposes

From above we can observe that-

▶ Maximum number of loans are being taken for the purpose of repairs

▶ Repairs also have the highest number of defaulters as compared to others

▶ Customers who have taken a loan for the purpose of buying a garage usually don't face any difficulty in paying back the loan
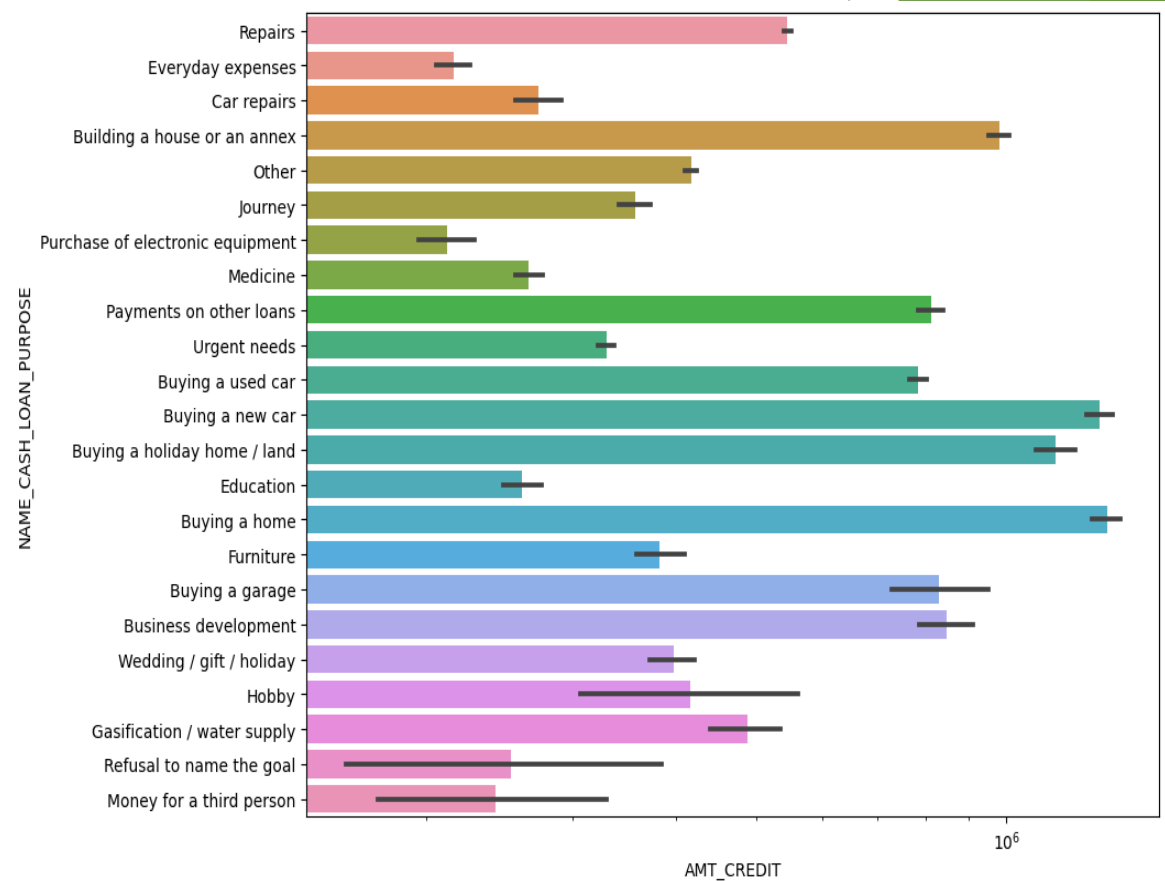

Loan payments for various cash purposes
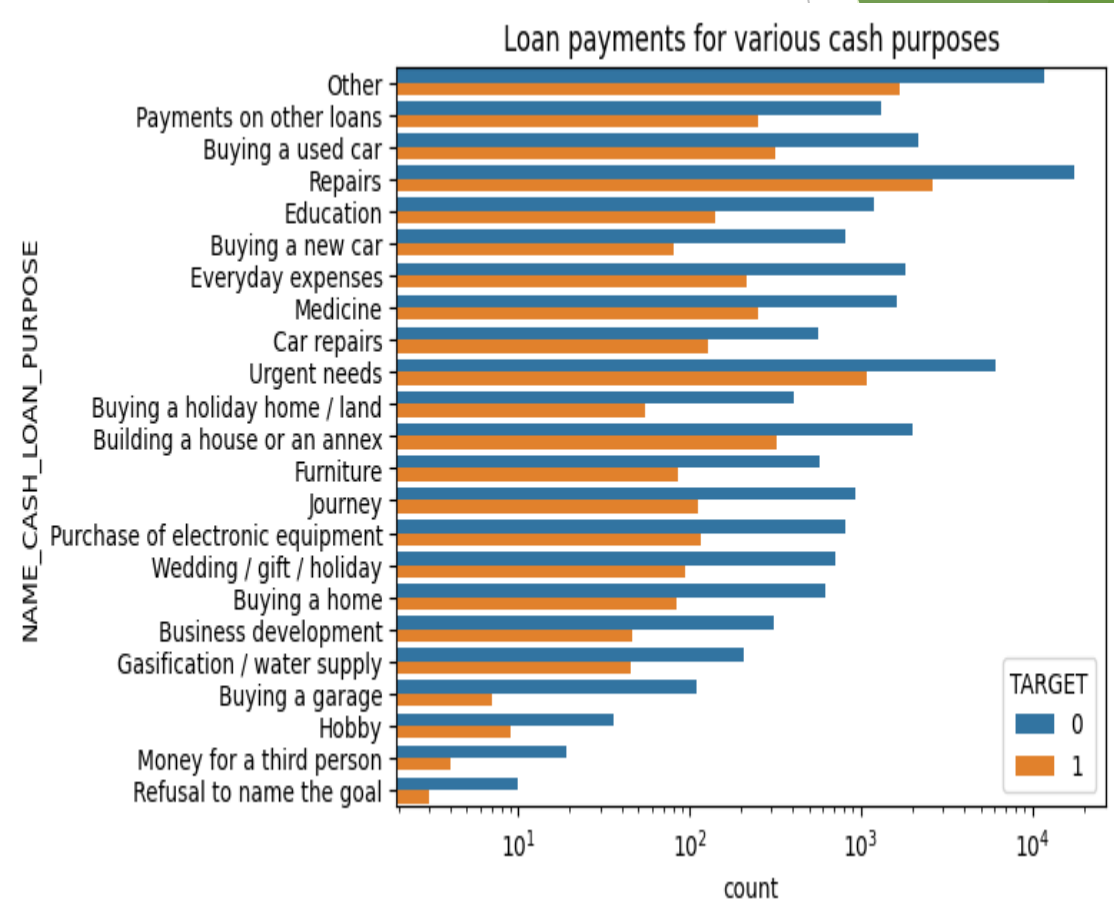
Bivariant Analysis

# Purpose of loan VS credit amount

From above we can observe the following-

▶ Least credit was taken for everyday expenses and purchase of electronic equipment

▶ The highest amount of credit was taken for buying a new home and car

Loan purpose VS Goods Price

From this we can observe the following-

▶ Customers who took a loan for buying a new car had the most number of successful loan payments

▶ Customers who took loan for buying a holiday home/land had the most number of unsuccessful loan payments
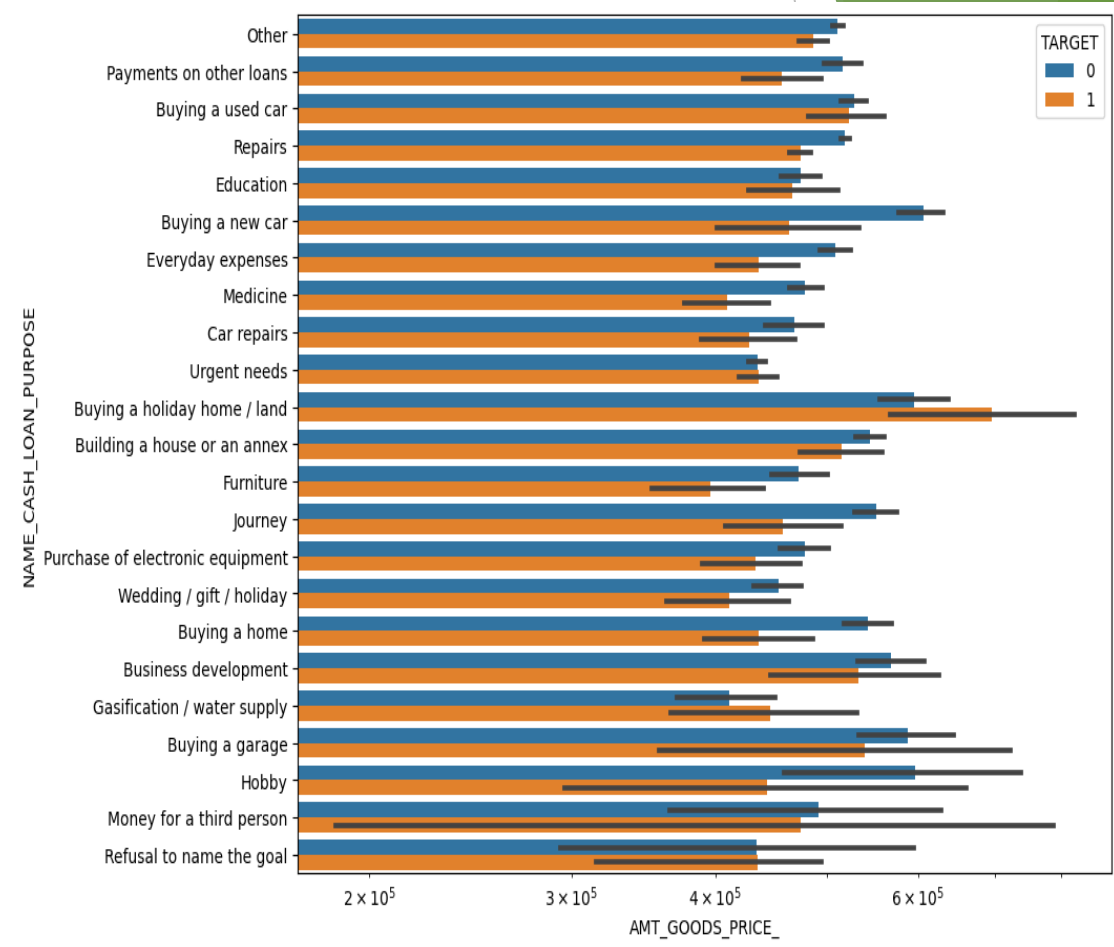
Conclusion-

- Banks should avoid giving loans for the purpose of buying a holiday home/land as they have high unsuccessful rate

- Banks should give more loans to students and businessman as they have the least unsuccessful payments

- Banks should give more loans for buying house and cars as the credit amount taken for them is high and also the successful payment percentage is decent