Credit EDA Assignment

By Mohammed Irshad

Problem Statement :

- The objectve of this analysis is to apply EDA techniques in a real business scenario, specifically in the context of risk analytics in banking and financial services.

1. To minimize the risk of financial losses while lending to customers.
2. the analysis seeks to identify patterns, trends, and potential risk
3. making decisions and reducing the likelihood of default or non-performing loans.

Introduction :

- Based on case study, there are three data sets, namely:
- • Application Data
- • Previous Application Data
- • Columns description Data Firstly,
- we have taken the ‘Application Data Set’ for our analysis. This contains all the information of the client at the time of application.

Data Cleaning :

* Data Cleaning is done for preparing data for analysis. The following steps are taken for Data cleaning:

1. Firstly, we have found out which all columns are necessary for our analysis.

2. Then, we have tried to find out the NULL values in each column.

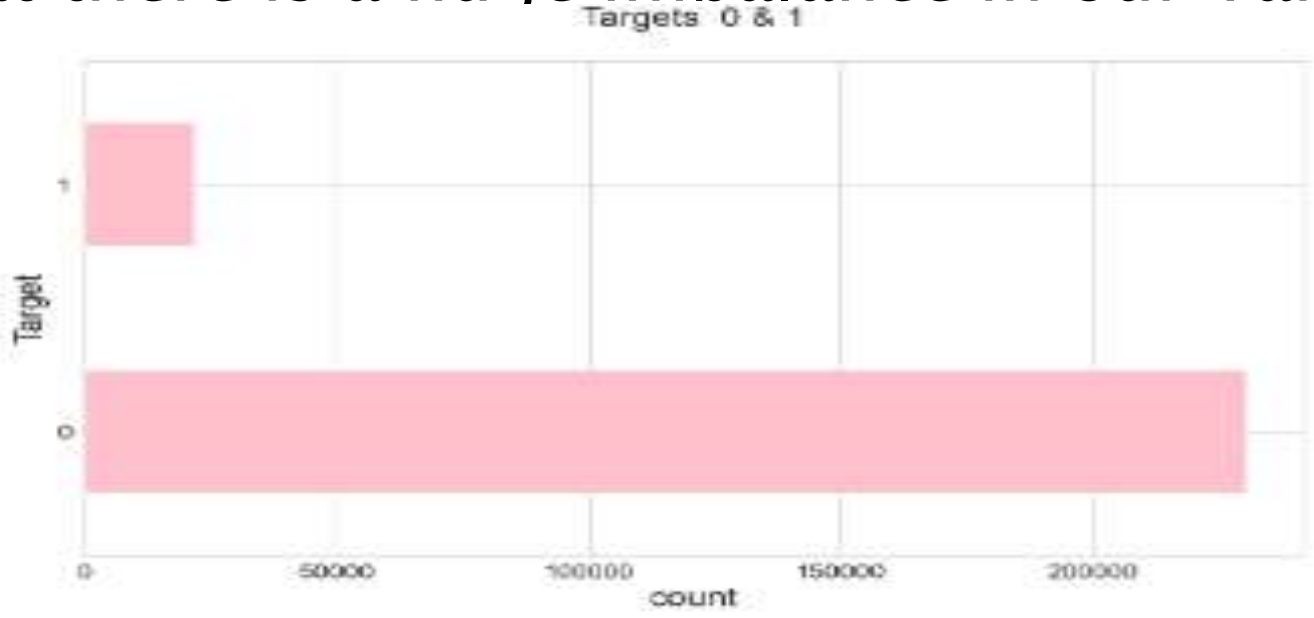
3. Dropping the columns where null values are more or equal to 40%

Analysis and Data cleaning

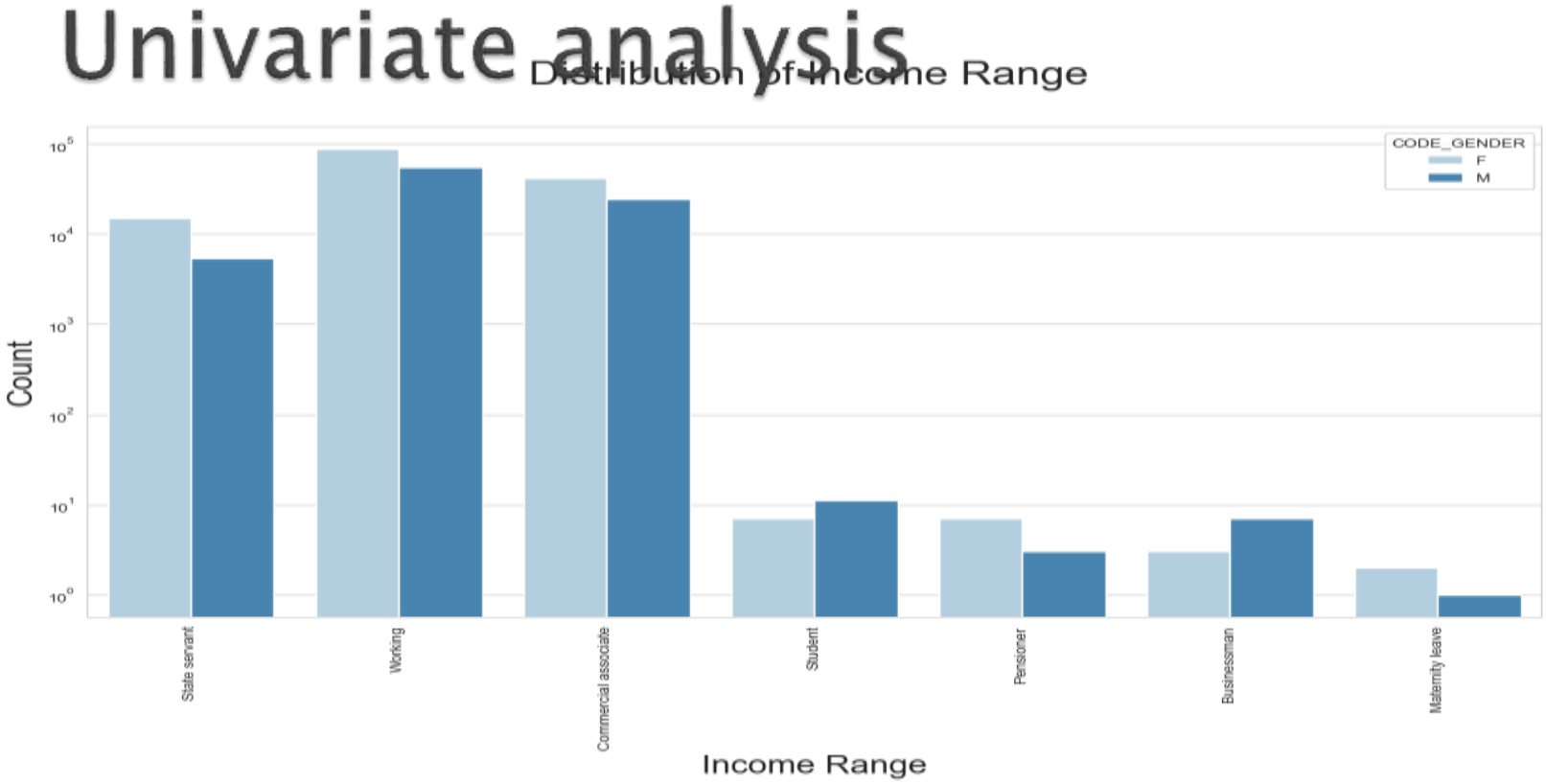
- Removing all columns with a null value
- Removing column with maximum no of missing value
- Replacing XAP/XNA with Na
- Finally, we are done with the identification of the outliers. We have removed them and plotted them again to observer the difference.

ANALYSIS OF APPLICATION DATA

- we have proceeded with the analysis of data.
- Checking the TARGET Variable If we plot the count of the TARGET Variable, we observe that there is a huge imbalance in our Target v



Univariate analysis

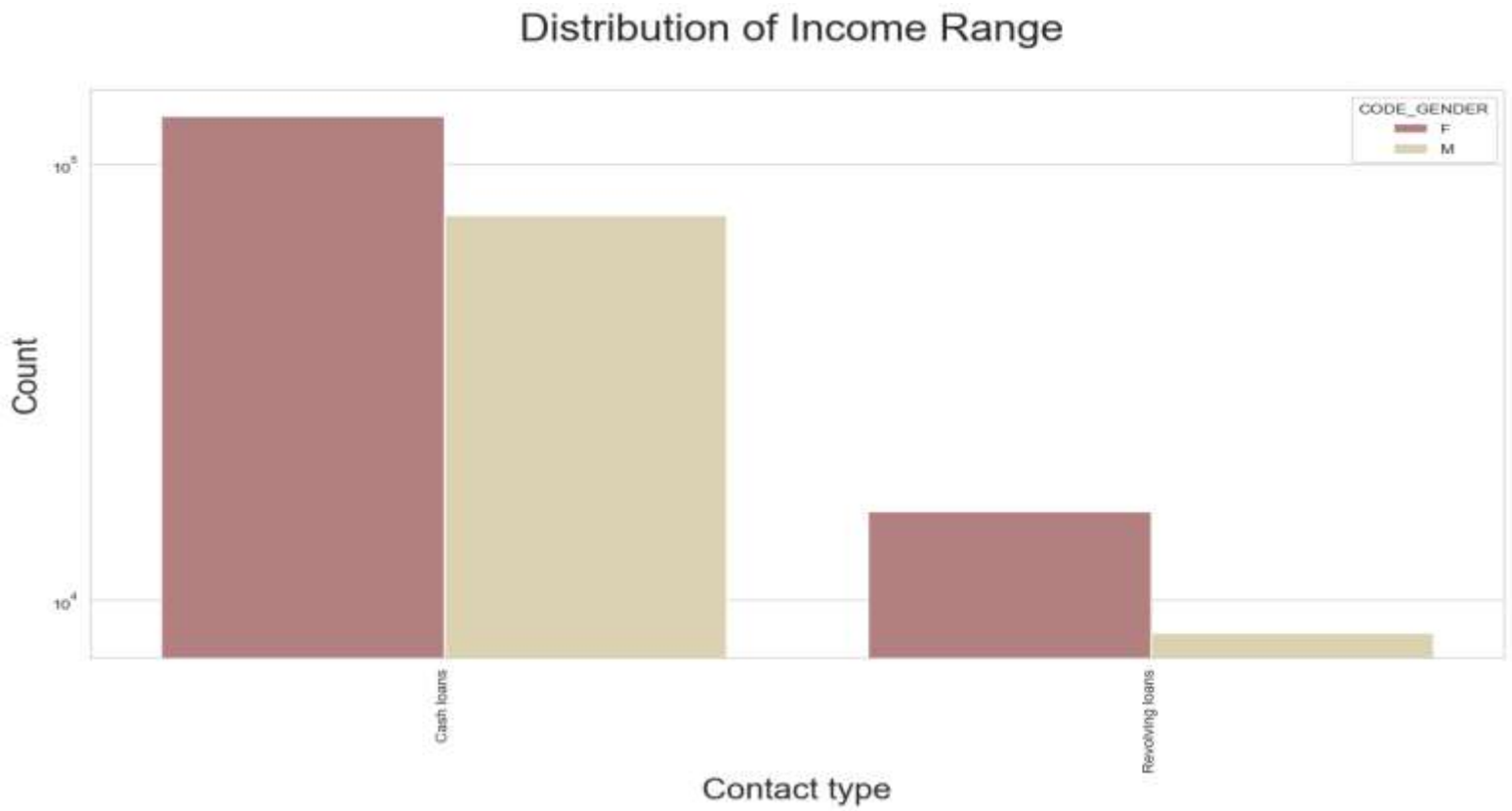


Conclusion from the graph:

1 It seems that working women have most credit than others

2. It seems that ‘State Servant’, 'Working’ and ‘Commercial Associate" have more credit counts Compare to others,

3. It seems that Women in Maternity leave’ has less credit in comparison to others.

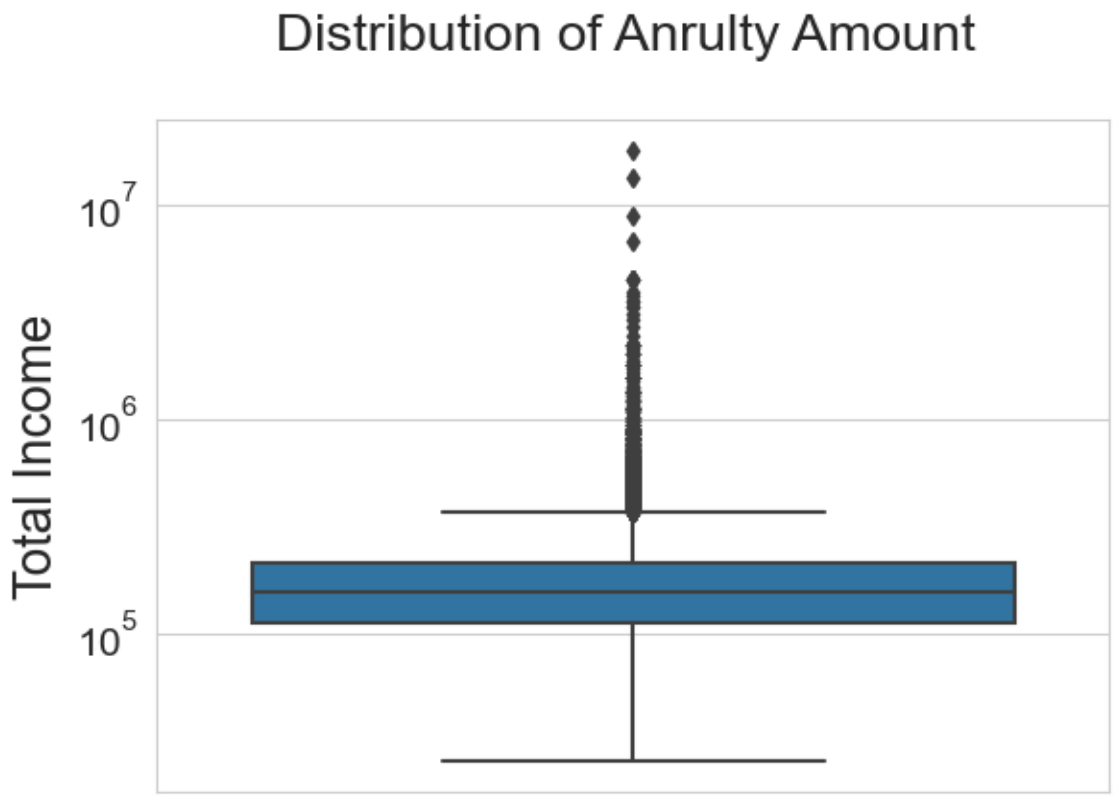


Conclusion from the graph:

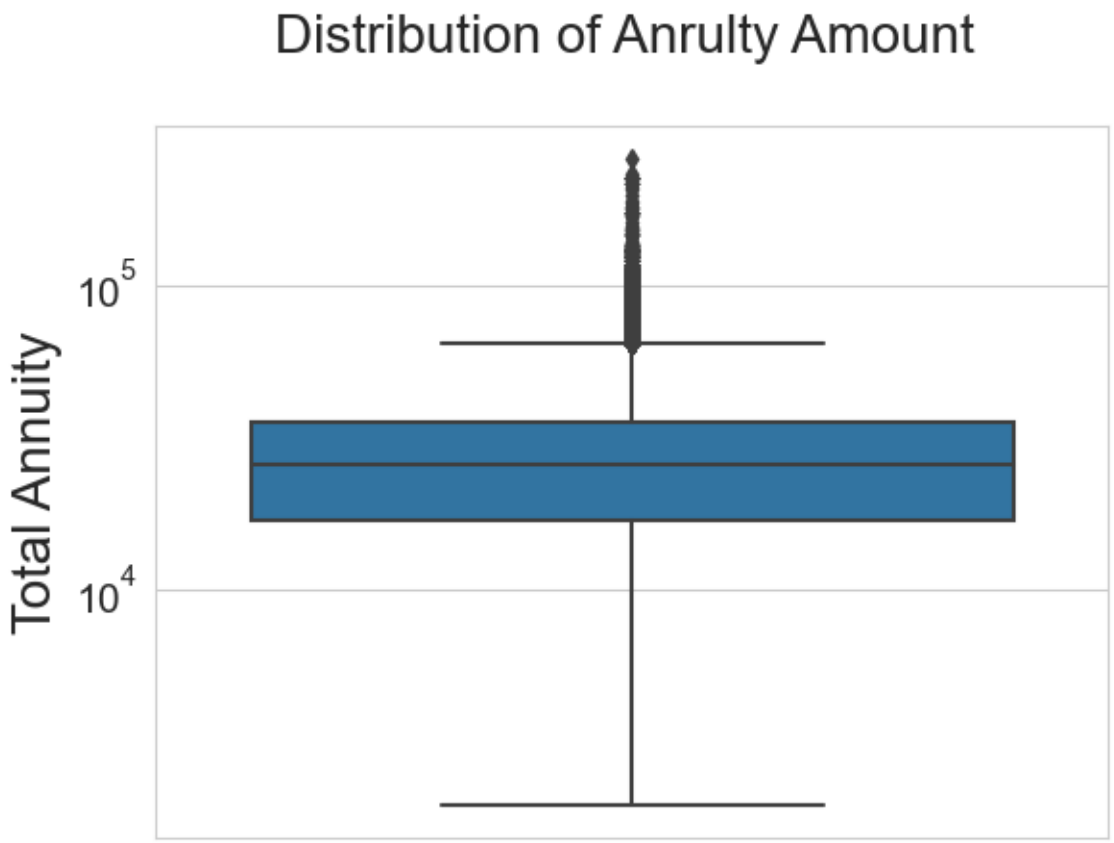
1. It seems that 'cash loans' is having higher number of credits than 'Revolving loans' contract type.
2. Also, female applies more for Credit.

Finding Outliers

Univariate Analysis For Target=0



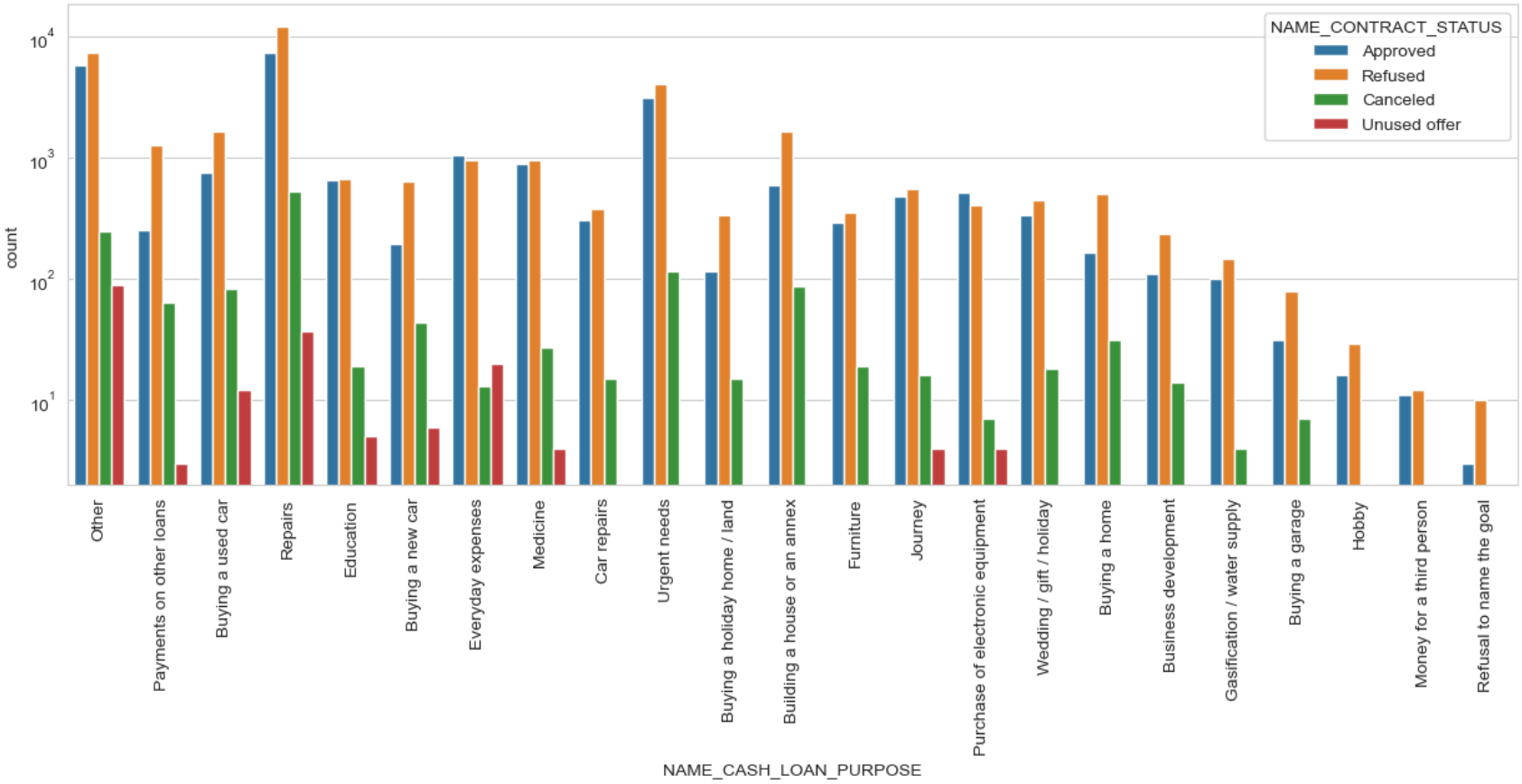
Conclusions from the graph:
1. There seems to be an equal distribution of the income amount of the client
2 also some of the outlayers present in the data set.



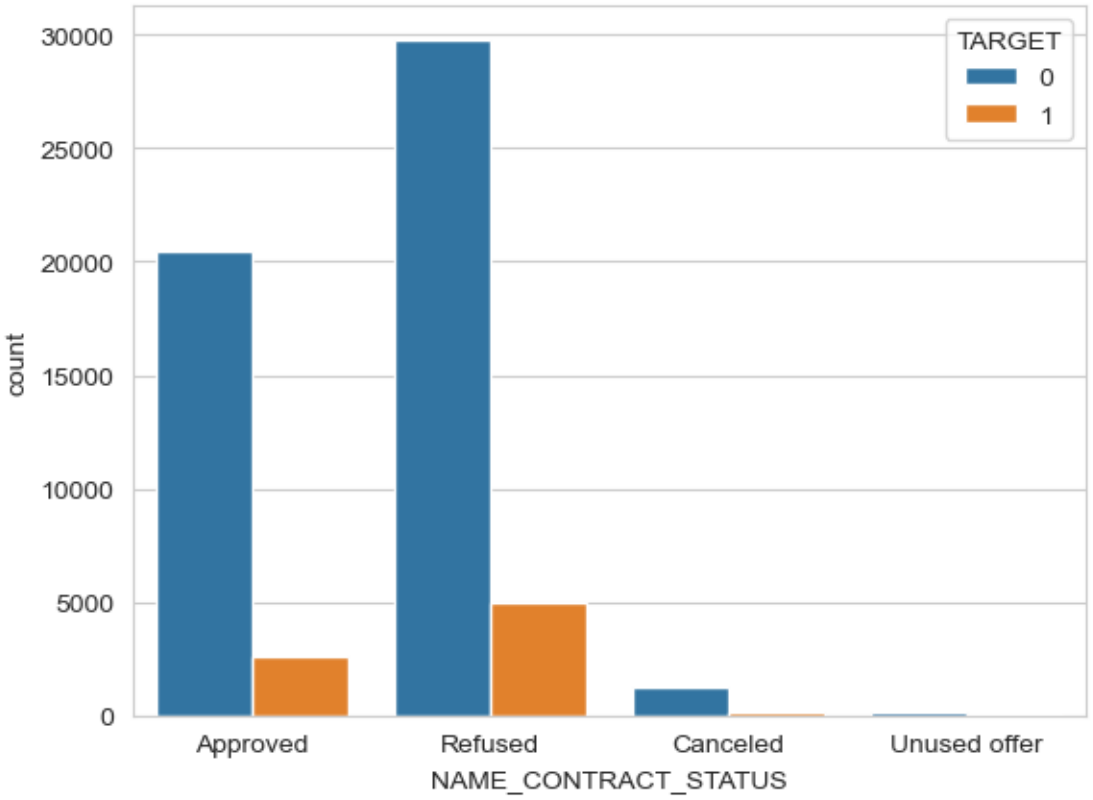
Conclusions from the graph:

1. The first quartile is bigger than the third quartile.
2. There seems some outers in the Annuity boxplot.

Merging the two datasets, i.e. application_dataset and previous_application



Conclusion from this graph:
For the repairing purpose customers had applied mostly prev. and the same puspose has most number of cancelation



Conclusion from graph:
most of the applications which were previously either canceled or refused 80-90% of them are repaire in the current data

Conclusion:

1.Cash loans are preferred by most customers and have a lower default rate.
2.Female borrowers show a lower default rate compared to males.
3.Working individuals, commercial associates, and pensioners have a better repayment track record.
4.The purpose of repairing has a high cancellation rate.
5.Previous unfavorable loan application outcomes do not necessarily indicate a 6.higher risk of repayment issues in the current dataset. These insights can inform risk assessment and lending strategies in the banking and financial services industry.

Thank You