

CREDIT EDA ASSIGNMENT

By Gopal Yadav
DS C42(March Batch)

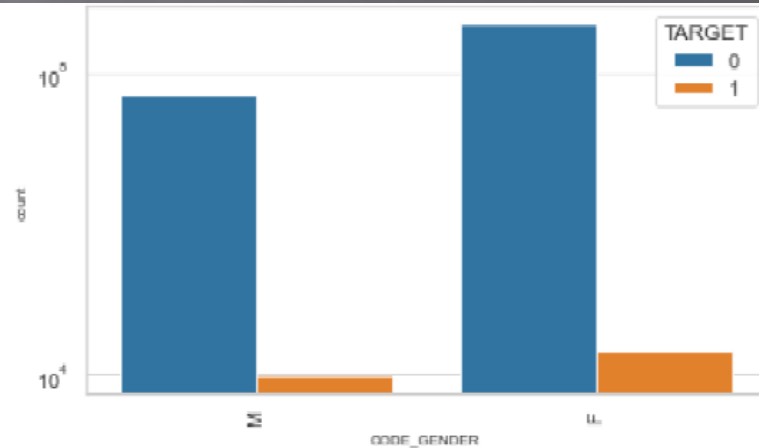
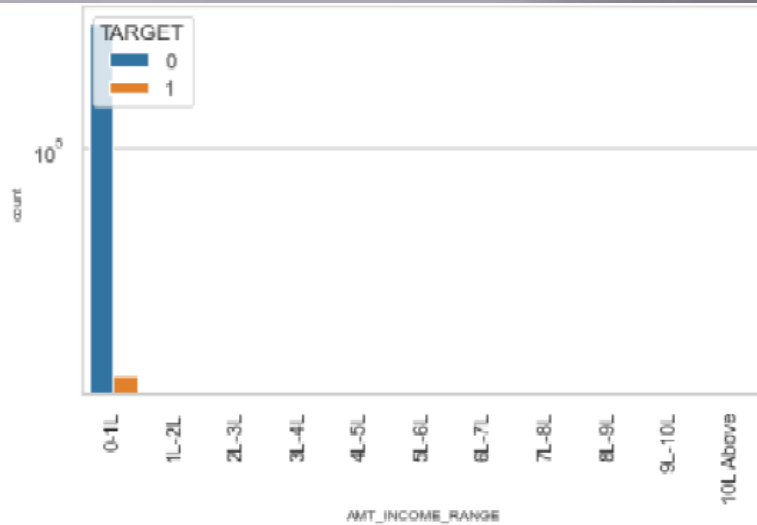
PURPOSE

- ▣ Credit EDA(Exploratory Data Analysis) helps us using risk analysis for different sectors like Banks or any Financial IT industry to take a decision for loan approval based on different applicant's.
- ▣ To minimize the risk of losing money while lending to customers.

STEPS

- ▣ Data understanding
- ▣ Checking data quality and data cleaning
- ▣ Check for imbalance data , univariate, bivariate analysis and correlation
- ▣ Merging of application data and previous data
- ▣ Observation

UNIVARIATE ANALYSIS OF APPLICATION DATA



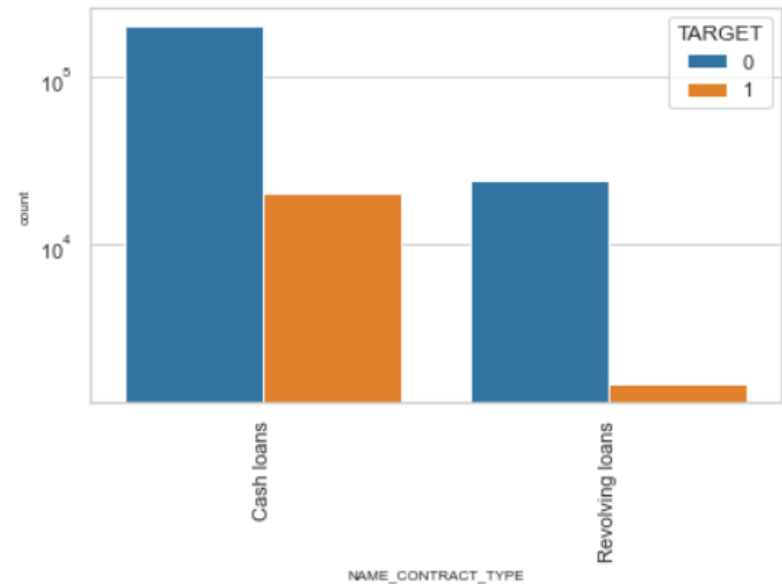
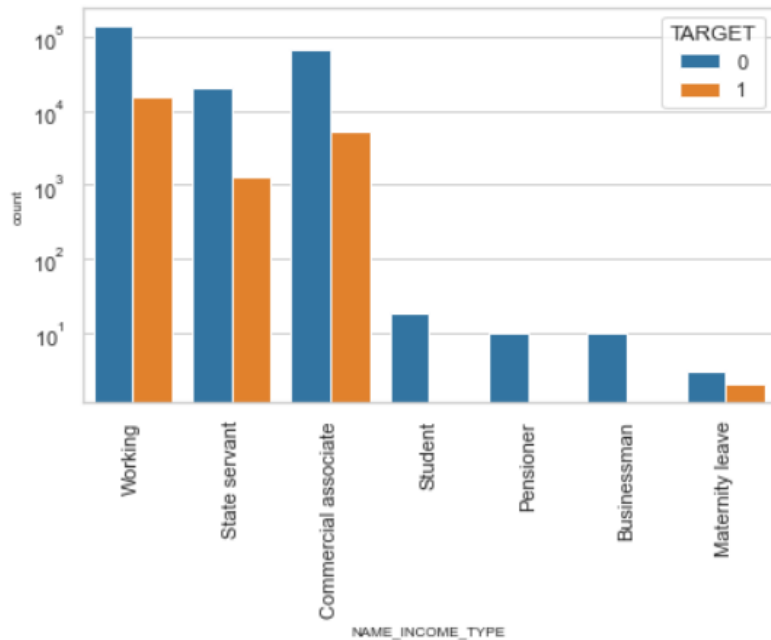
Observations:

AMT_INCOME_RANGE :

-> People in range 0-1L have high number of loan and also have high in defaulter. -> Income segment >5L has less defaulter.

CODE_GENDER:

-> The percentage of defaulters are more in Male as compared to Female.



NAME_INCOME_TYPE:

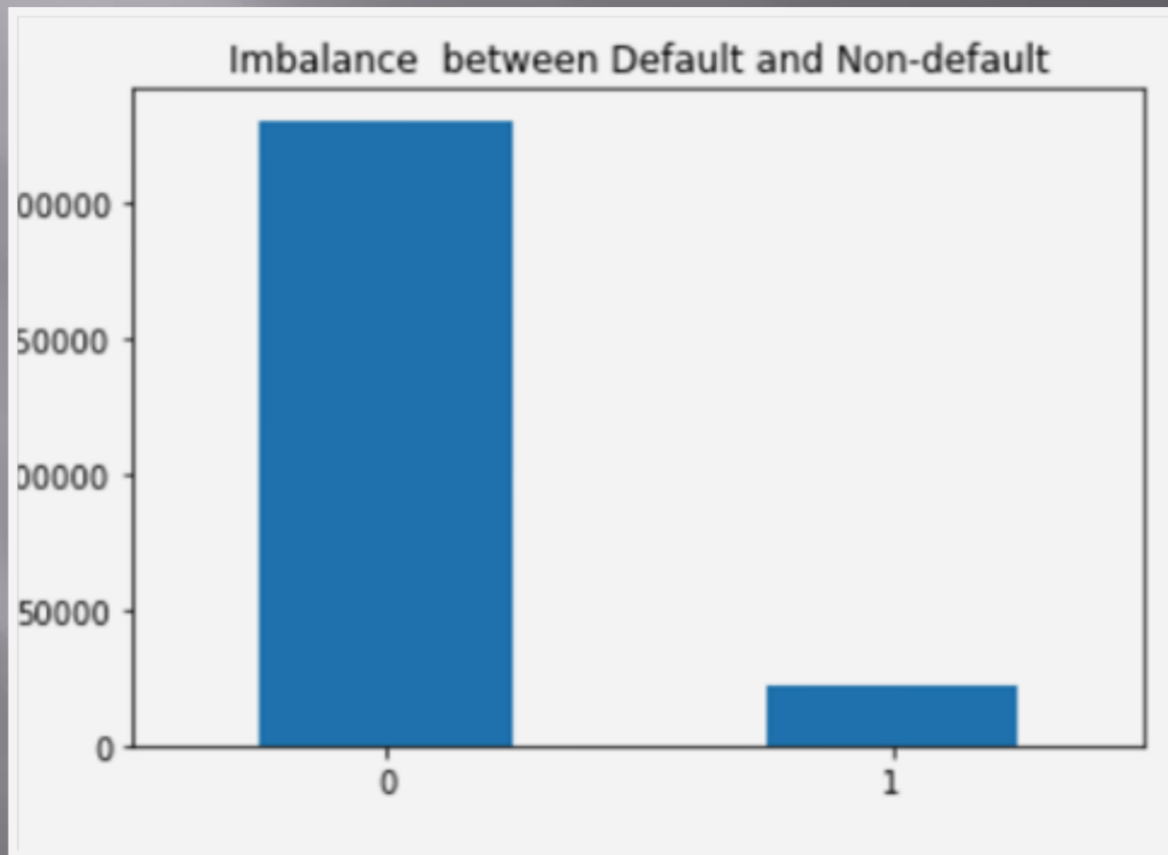
-> Student and business are little higher in percentage of loan repayment. -> Working, State servant and Commercial associates are higher in default percentage. -> Maternity category is significantly higher problem in repayment.

NAME_CONTRACT_TYPE

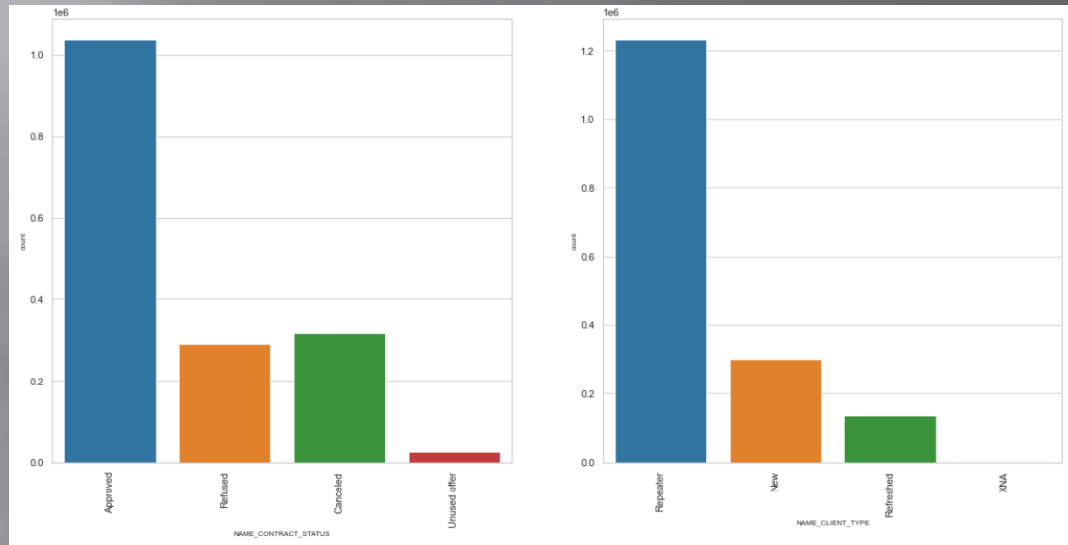
-> In contract type 'Cash loans' are high in number of credits than 'Revolving loans' contract type. -> IN above graph 'Revolving loans' is small amount as compared to 'Cash loans'

Imbalance in the Target variable column

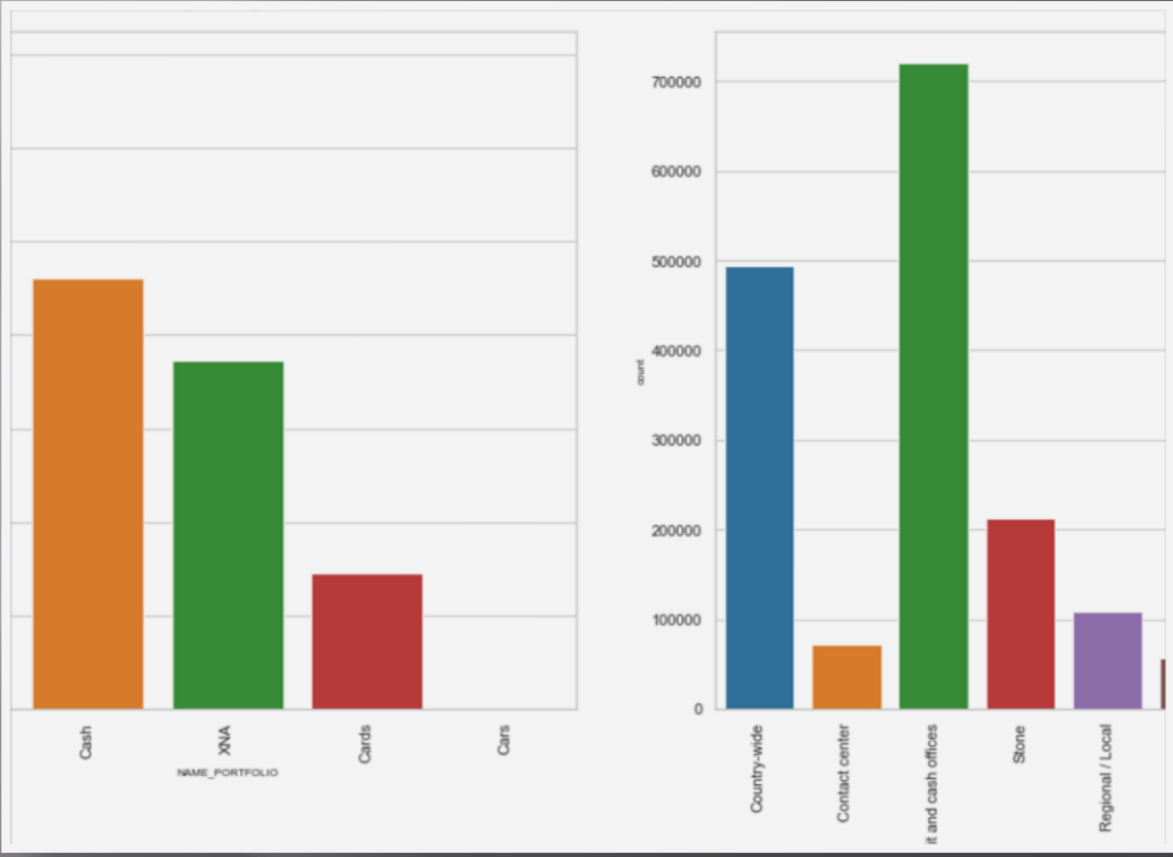
As we can see there is a large imbalance between the proportion of default (TARGET = 1) and non-default (TARGET = 0) in the dataset. The ratio of non-default to default is 11.4:1. Even if the default ratio is low, default of any application is costly and should be analyzed closely.



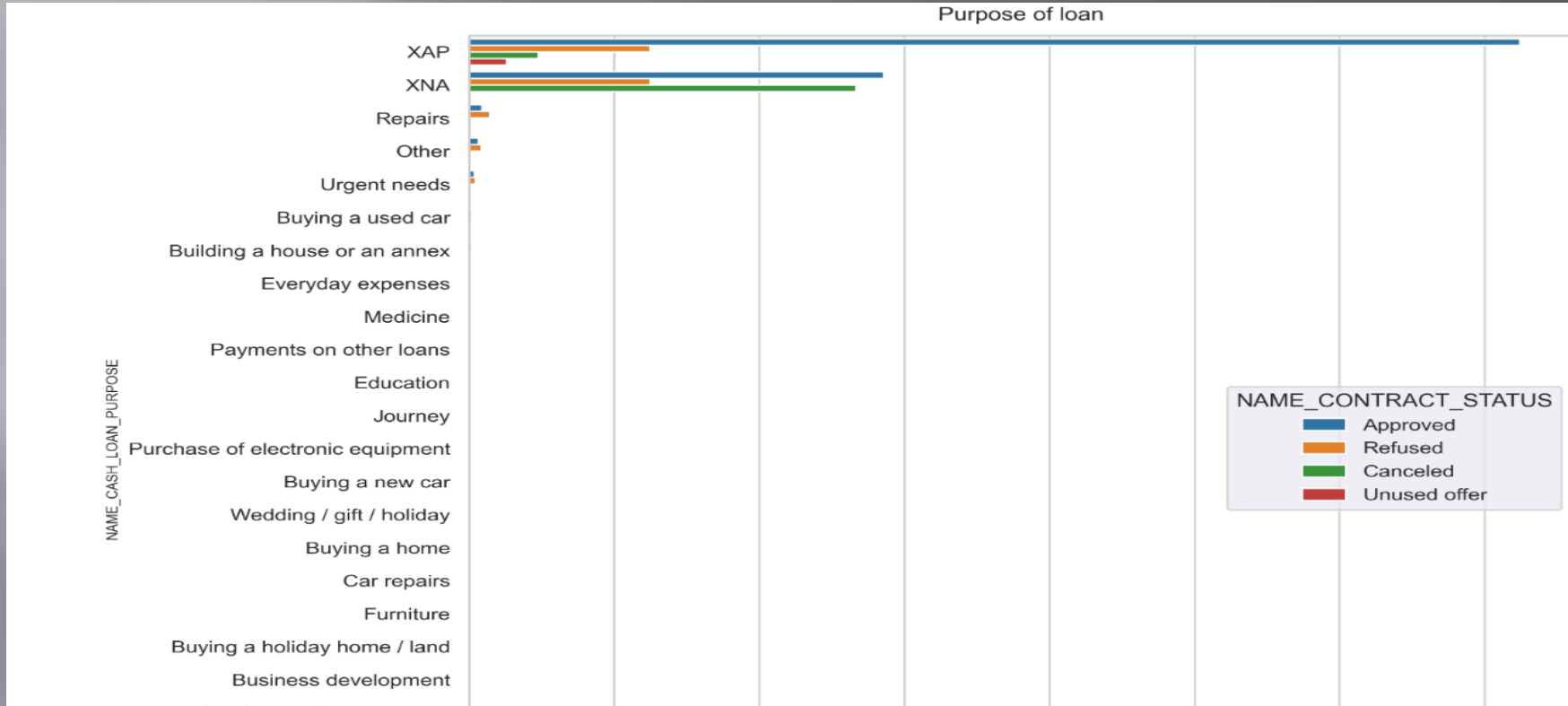
Univariate Analysis on categorical columns



Here we have performed univariate analysis on categorical columns .



Univariate analysis on Purpose of Loan



CONCLUSION

- Banks must target more on contract type 'Student', 'Pensioner' and 'Businessman' for profitable business.
- Banks must focus less on income type 'Working' as it is has most number of unsuccessful payments in order to get rid of financial loss for the organization