Explanatory Data Analysis(EDA) Assignment

Problem Statement

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history

Observing the application:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

Contd..

Needs to understand the driving factors behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Assumptions

Assuming, if any column has more than 30% unique value is a numerical column else categorical column

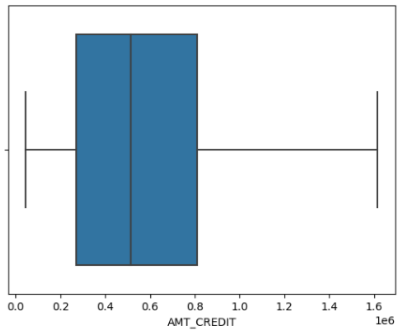
Approaches & Methodologies

For the Outlier, if any value is far from the IQR then its necessary to deal with it

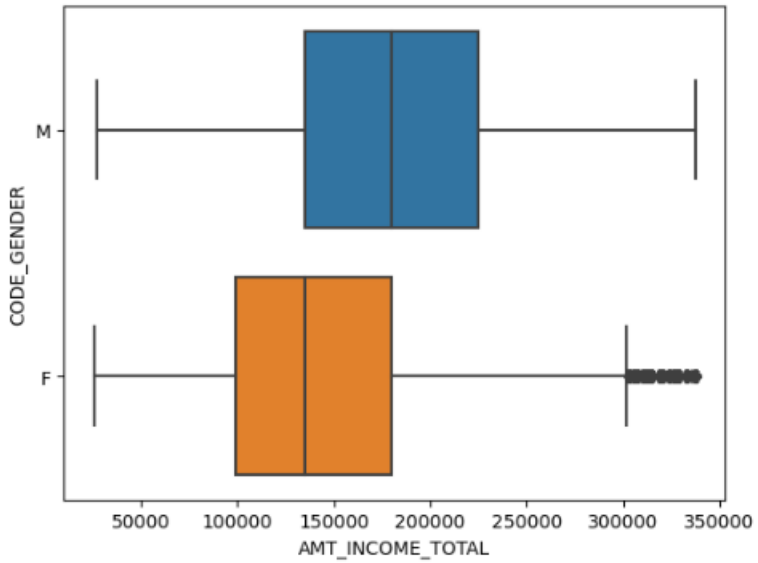
In order to deal with the outliers using Flooring and Capping method

Graph & Insights

After Dealing with ouliers(performed Flooring and Capping) the bar plot will be somewhat as shown :

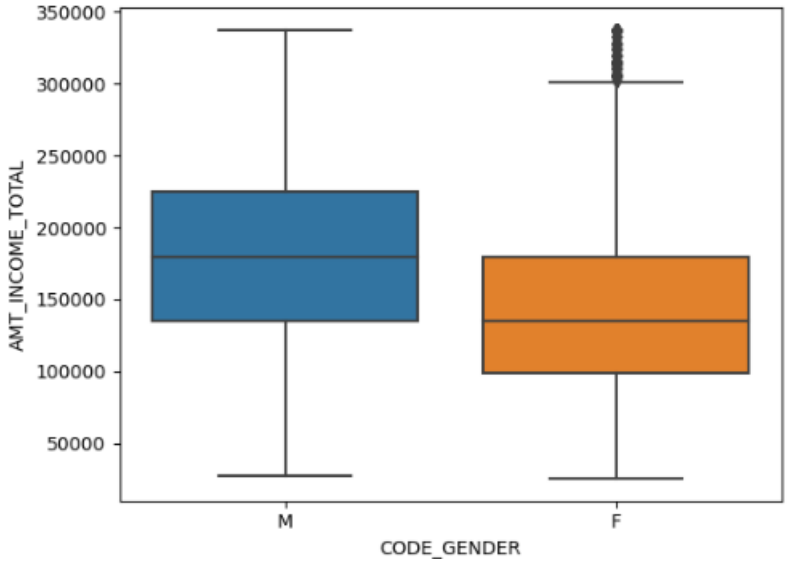


Income Amount on the basis of Gender

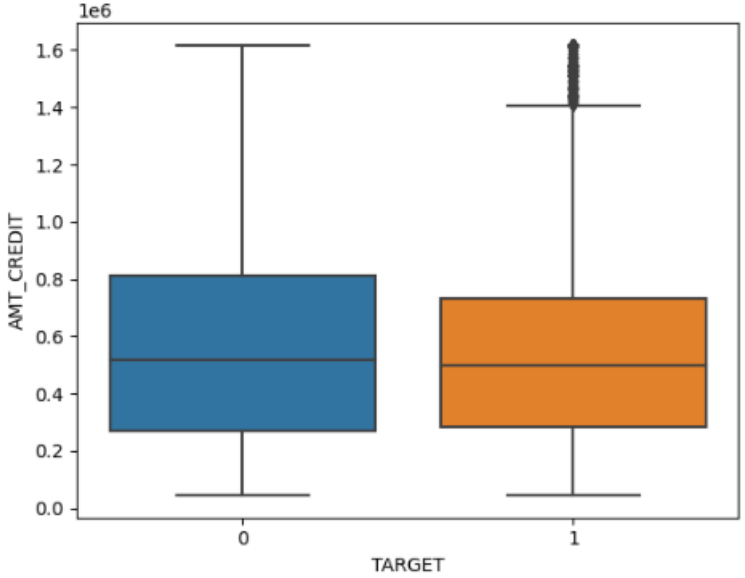


Graph

Income on the basis of gender

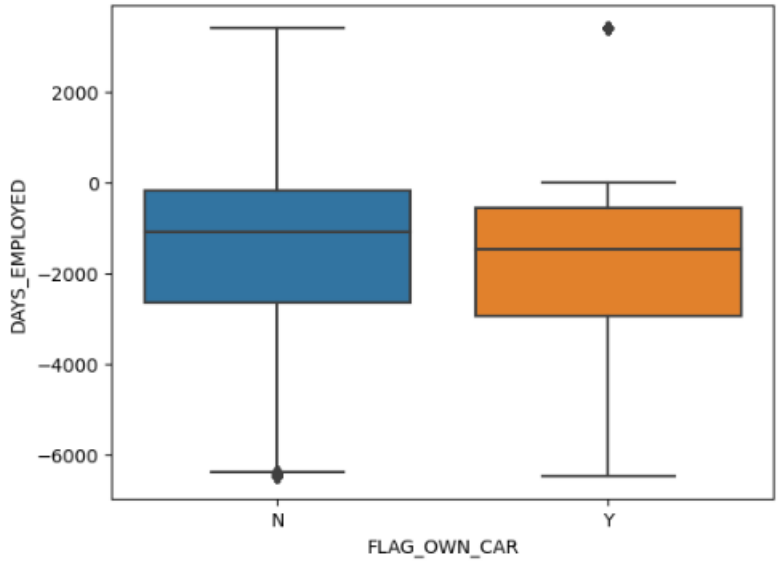


Amount credited vs Target



Graph

Days_Employeed vs Own a car



Conclusion or Recommendation

Data Imbalance is there in the data provided:

Approx -- 92% is non-defaulters

-- 8% is defaulters

It was expected to be data imbalance.