EDA ASSIGNMENT

Tushar Prajapati

LIST OF CONTENTS

INTRODUCTION

- This assignment gives a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.
- Nowdays, Fintech industry is on boom as loan takers are increasing and because of that defaulters are also increasing.
- The loan-providing companies find it hard to give loans to people due to their insufficient or non-existent credit history.
- In this assignment we have to analyse the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

PROBLEM STATEMENT

- Perform the EDA on Data.
- To understand and find the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. So that The company can utilise this knowledge for its portfolio and risk assessment.


NO
ASSUMPTIONS
WERE TAKEN.

APPORACH & METHODOLOGY

- Imported the required Libraries.
- Imported the data from given CSV Files.
- Checked the structure of Data to know about the data.
- Analyzed the Data
    - Checked for missing values in data.
    - Dropped the columns which had the more than 50% of it's value null because they could have lead to wrong analysis of data.
    - Dropped some unnecessary columns
    - Dropped some unnecesary values from some columns.

- Imputed the missing value for analysis by their median(for numerical columns)/mode (for categorical columns).
- Corrected the form of values in column.
- Checked the outliers in the data.
    - If there were outliers, I used flooring/capping method to deal with outliers.

•Did data imbalance check.

•Find the data imbalance ratio.

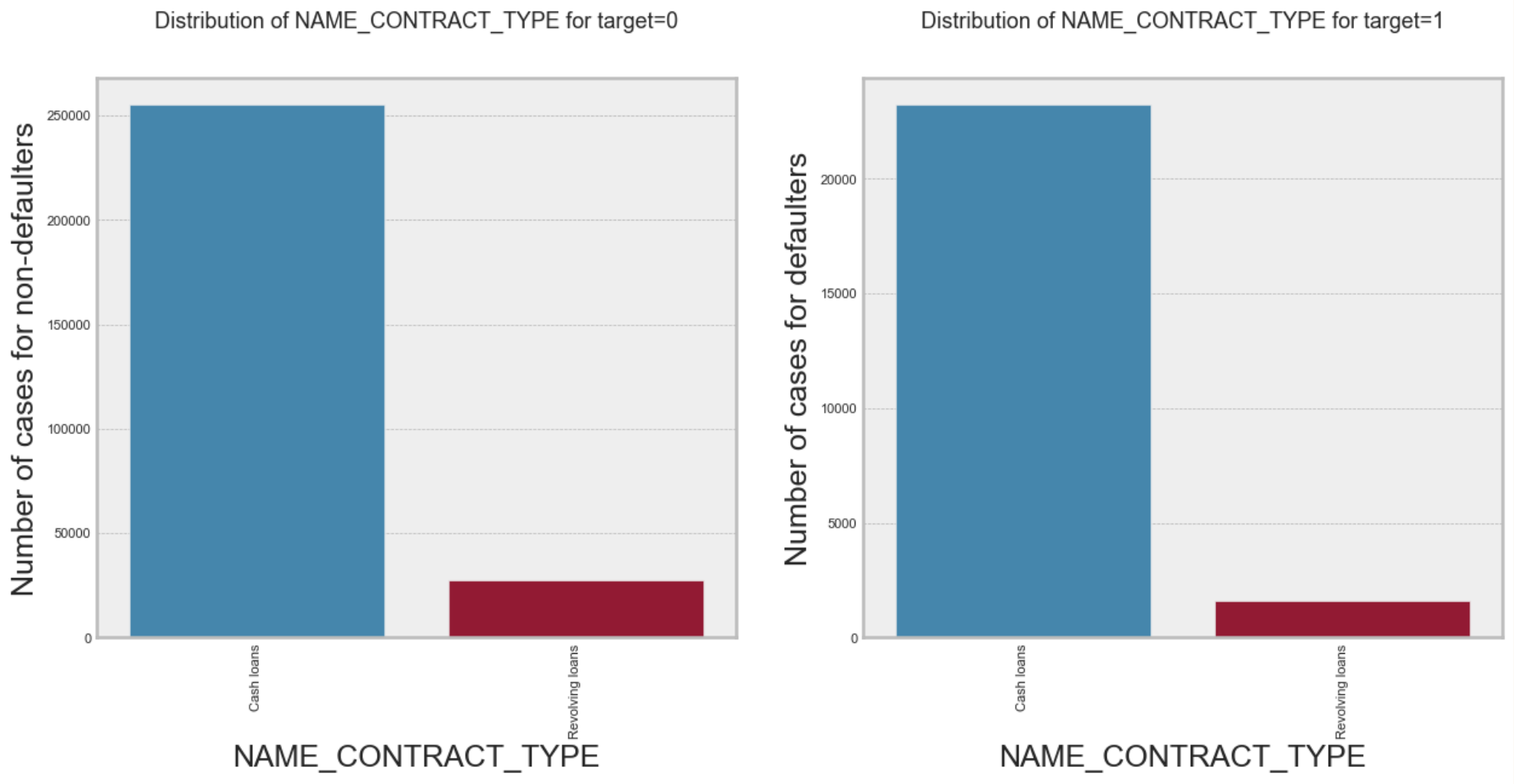
•This analysis is for the **'Target variable'** in the dataset (**clients with payment difficulties** and **all other cases**).

GRAPHS & INSIGHTS

An Example of Univariate Analysis according to Client with payment difficulties and all other cases (Target variable).

Insight

Revolving loans has less distribution in number of cases for defaulters compared to non-defaulters.

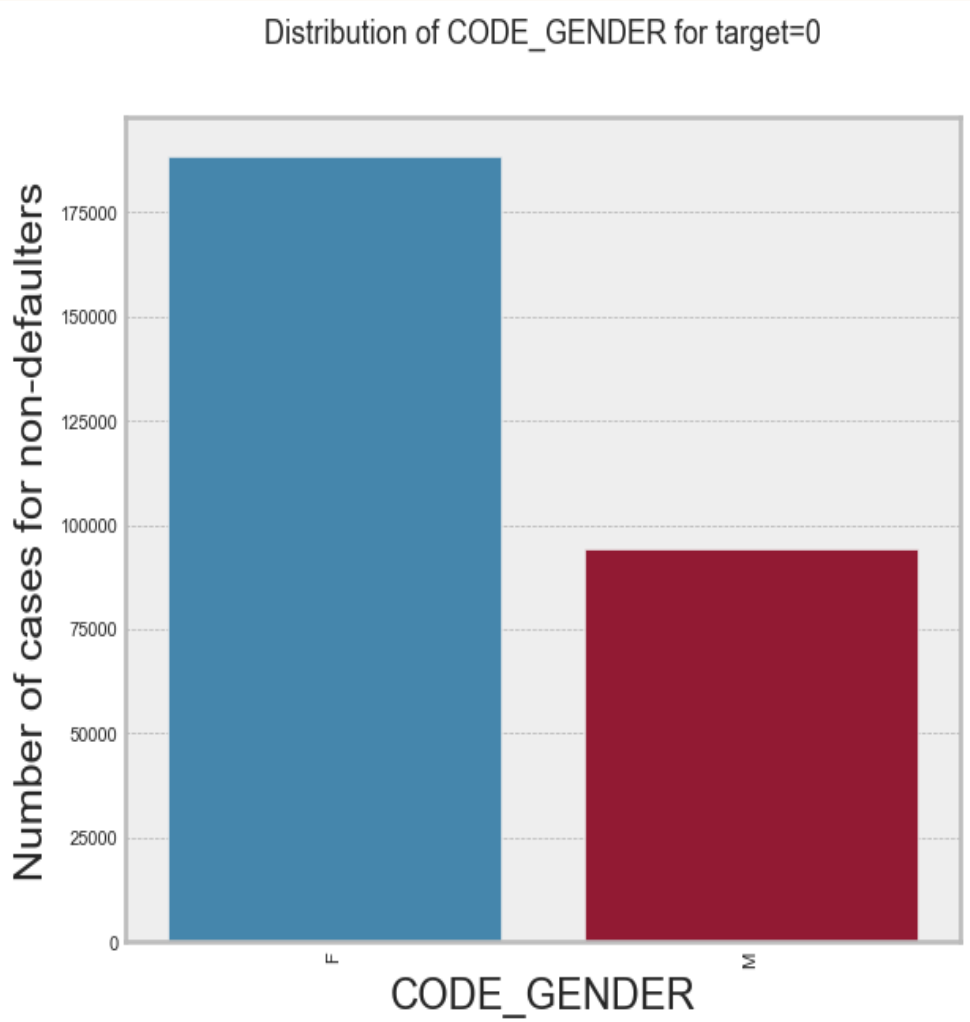


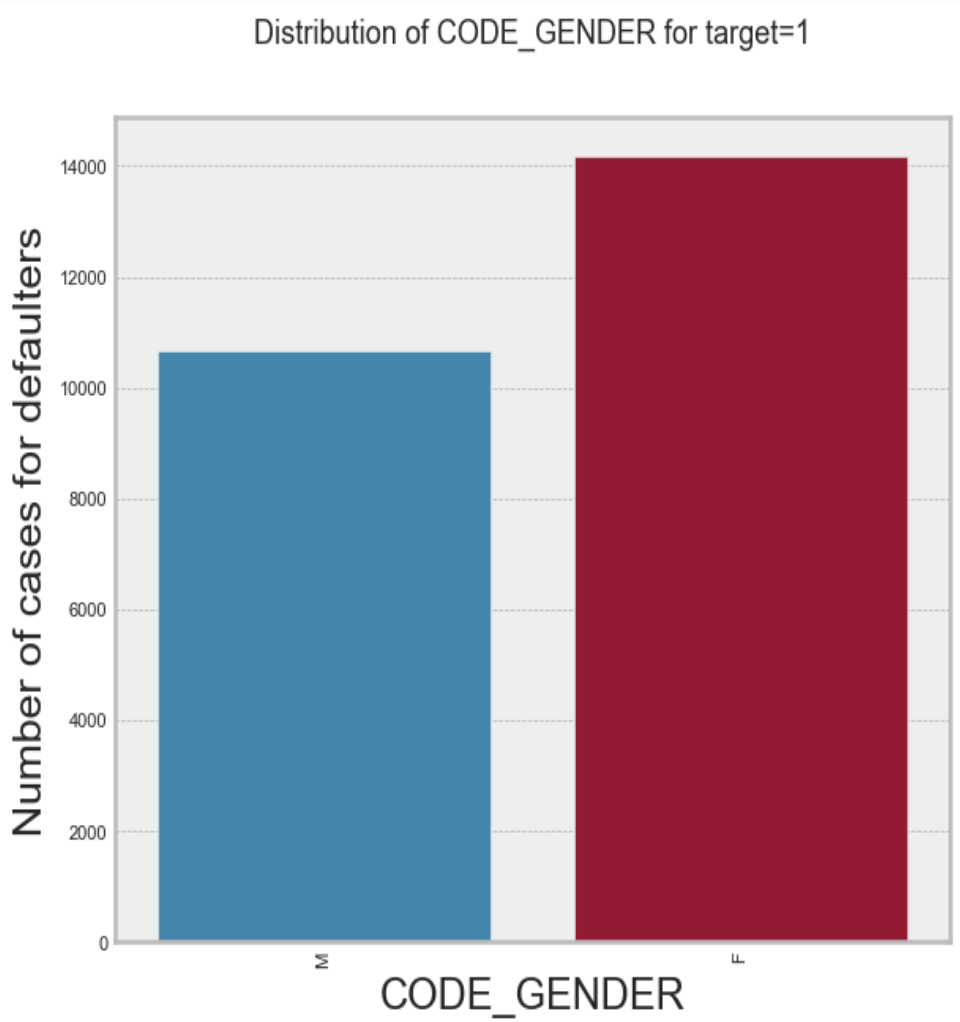
GRAPHS & INSIGHTS

An Example of Univariate Analysis according to Client with payment difficulties and all other cases (Target variable).

Insight

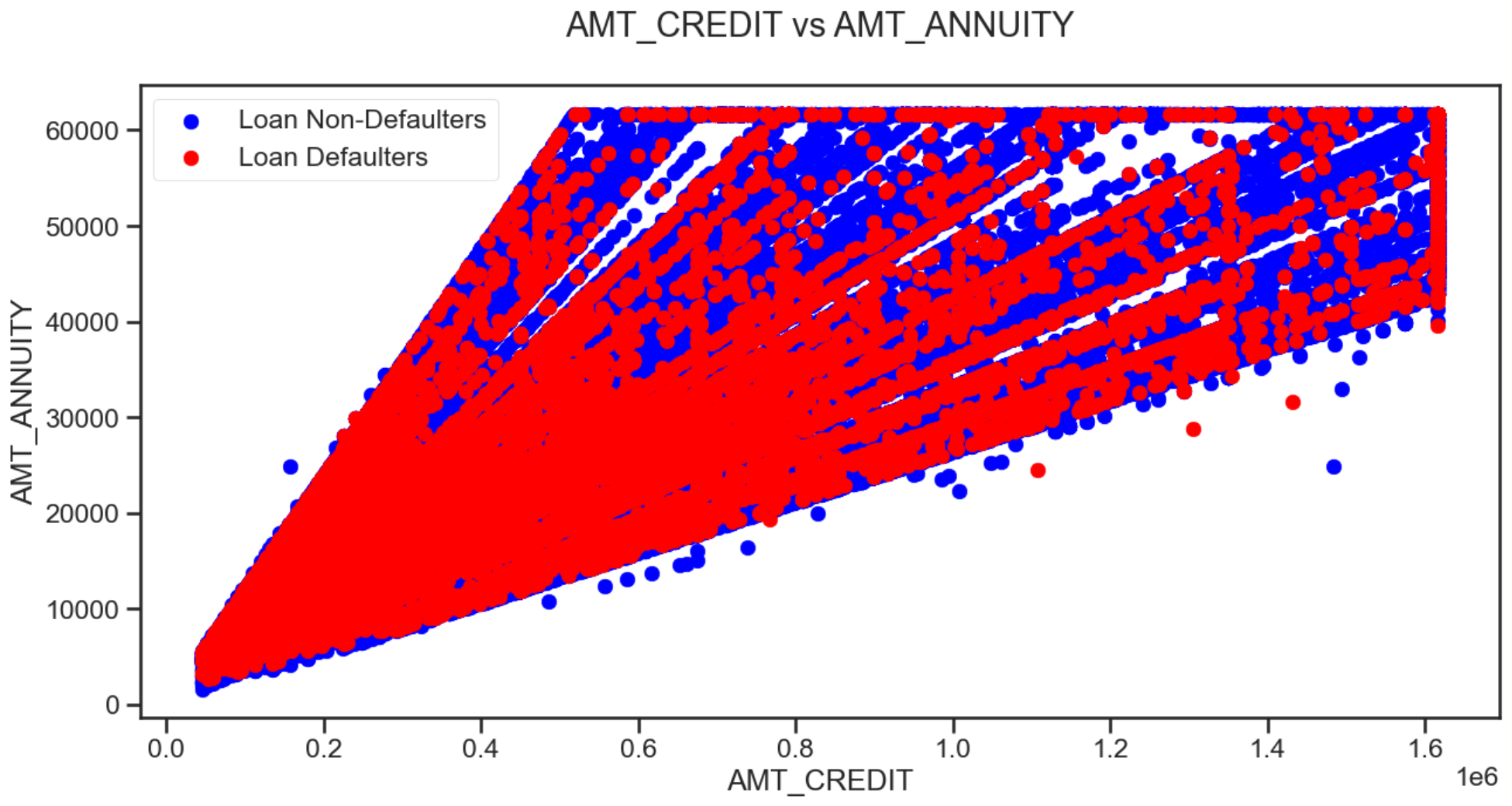
Male has more distribution in number of cases for defaulters compared to non-defaulters.





GRAPHS & INSIGHTS

An Example of Bivariate Analysis according to Client who are Loan Repayers and Loan



Insight

There is a decrease in defaulters when Amount Annuity increases. Most of the defaulters are having AMT_ANNUITY values less than 50000.

CONCLUSION

1. There are less defaulters when it comes to revolving loan.
2. Bank should target people who own cars.
3. Bank should target people who owns realty.
4. Bank should also consider people without any children.
5. Bank should target female client.
6. Bank should avoid Single/not married people.
7. Bank should avoid people living in apartments.
8. Bank should prioritize people with age between 40-50
9. Repairs loan purpose should be the lease priority.
10. People with Income Type 'Working' should be least priority.
11. Company should avoid people living in Co-op apartment.
12. Company should prioritize people living in Office apartment.