EDA of Risk Analytics in Banking and Finance Services

v1.0

Created By:

Dipanjan Roy (Lead)

Ajeet Kumar Biswas

on

29/09/2021

Business Understanding:

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose we work for a consumer finance company which specializes in lending various types of loans to urban customers. We have to use EDA to analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

Business Objectives

This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (too risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants' using EDA is the aim of this case study.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e., the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

To develop your understanding of the domain, you are advised to independently research a little about risk analytics - understanding the types of variables and their significance should be enough).

Data Used for the Analysis:

Three datasets are used for this analysis. The details are as follows:

1. *application_data.csv contains* all the information of the client at the time of application. The data is about whether a **client has payment difficulties.**

2. *previous_application.csv* contains information about the client's previous loan data. It contains the data whether the previous application had been **Approved, Cancelled, Refused or Unused offer.**
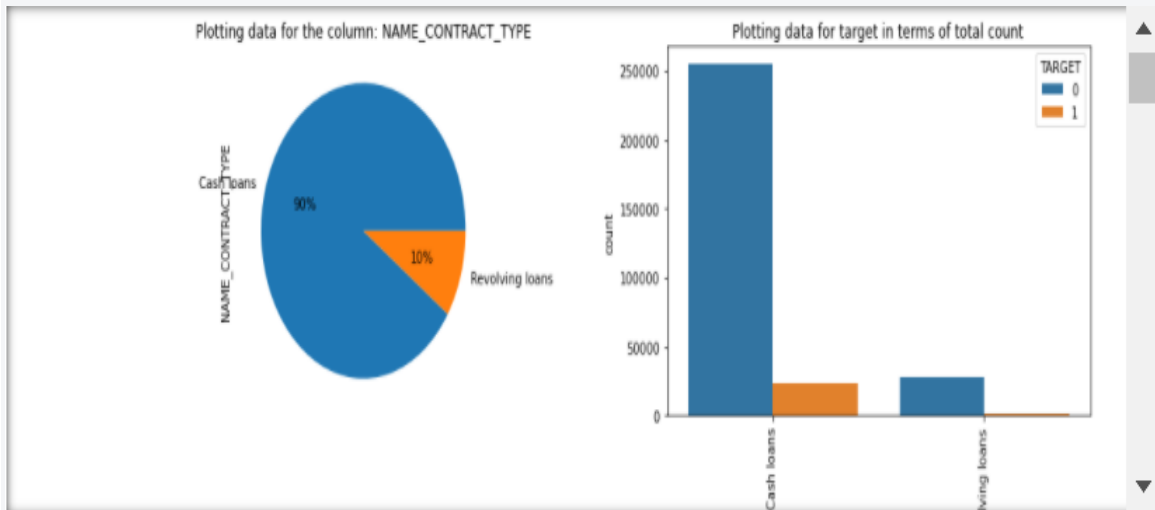
3. *columns_description.csv* is data dictionary which describes the meaning of the variables.
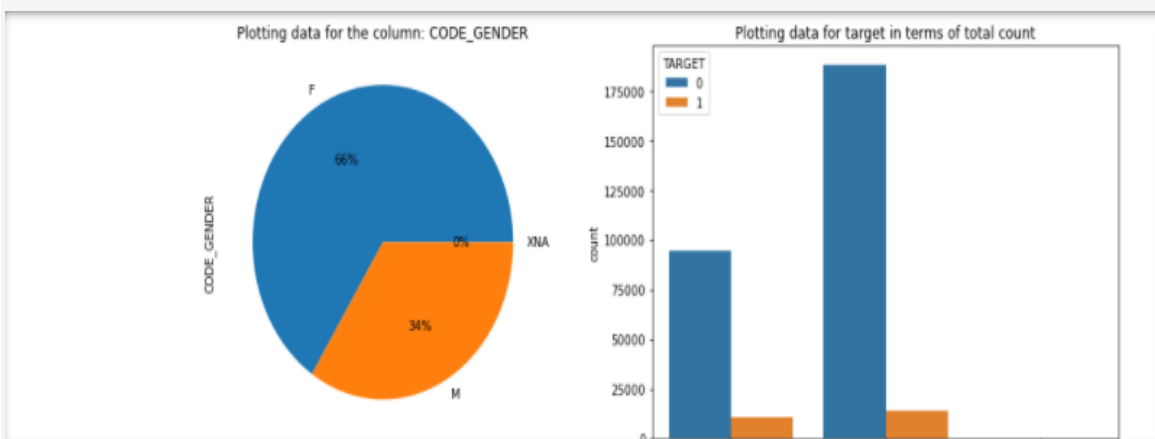
Insights:

1. NAME_CONTRACT_TYPE:

There are two types of loan cash loan and revolving loan. Number of cash loan is higher than revolving loan but defaulters are also higher in cash loan compare to the other.
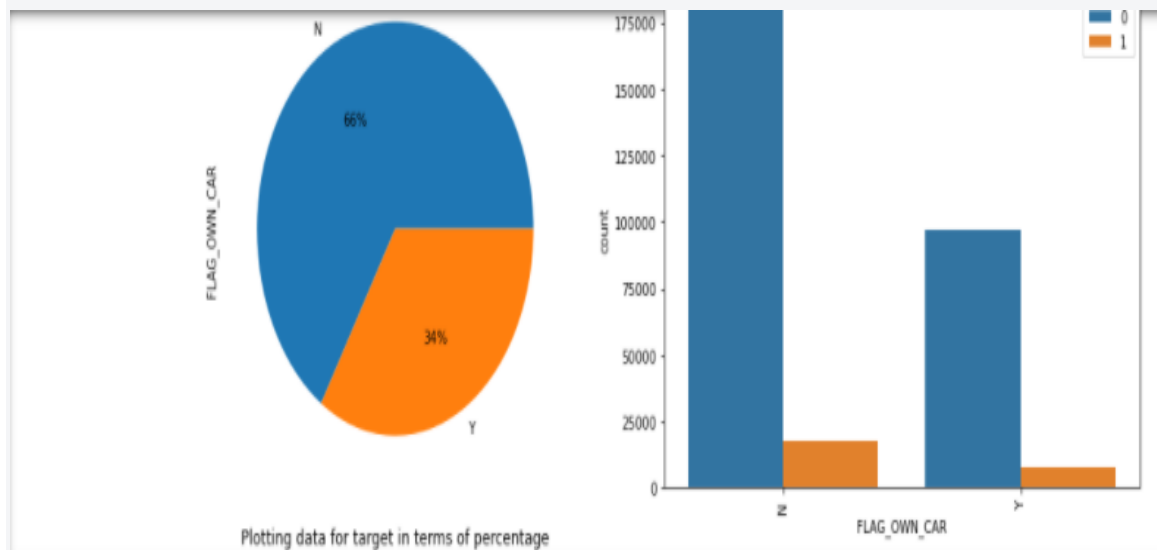

2. CODE_GENDER:

males take a smaller number of loans as compared to females, but the defaulter rate is higher in males compared to females.
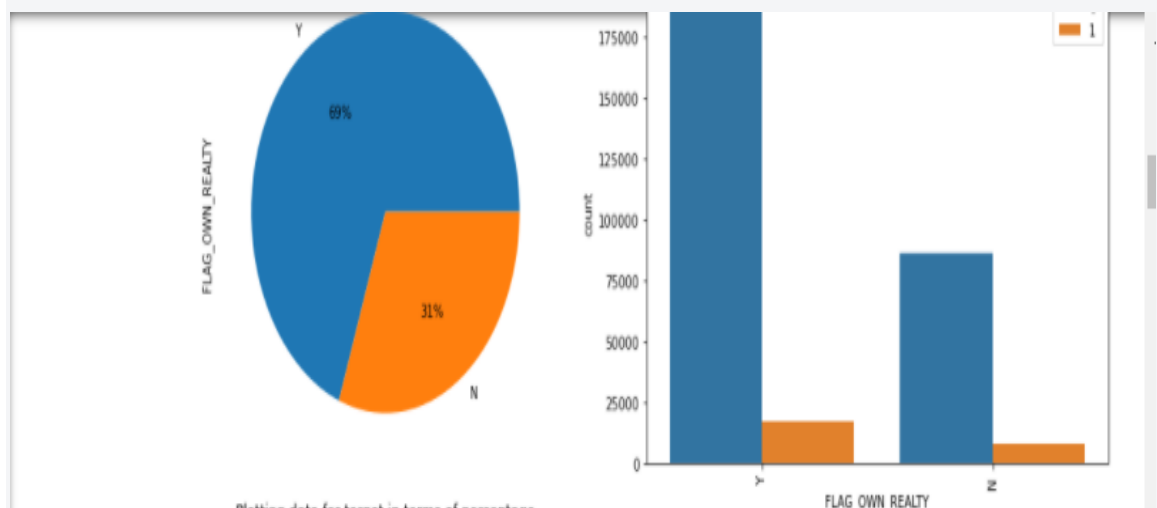
3. FLAG_OWN_CAR:

Client owning cars have more defaulting rate than the client not having cars.

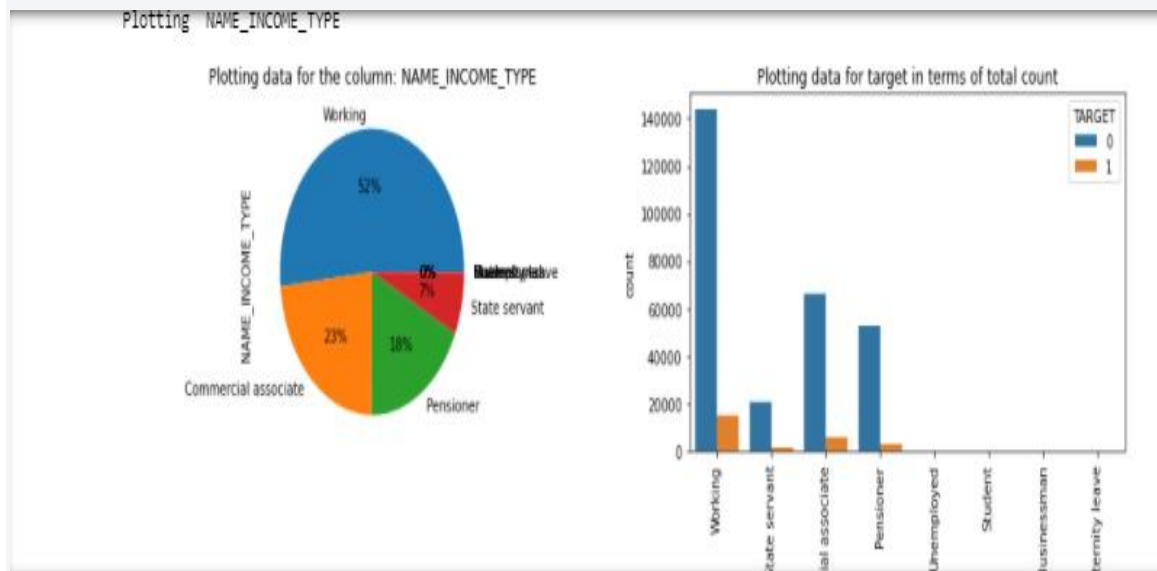Plotting data for target in terms of percentage

4. FLAG_OWN_REALTY:

Clients owning realty property tend to default more than client not owning property.
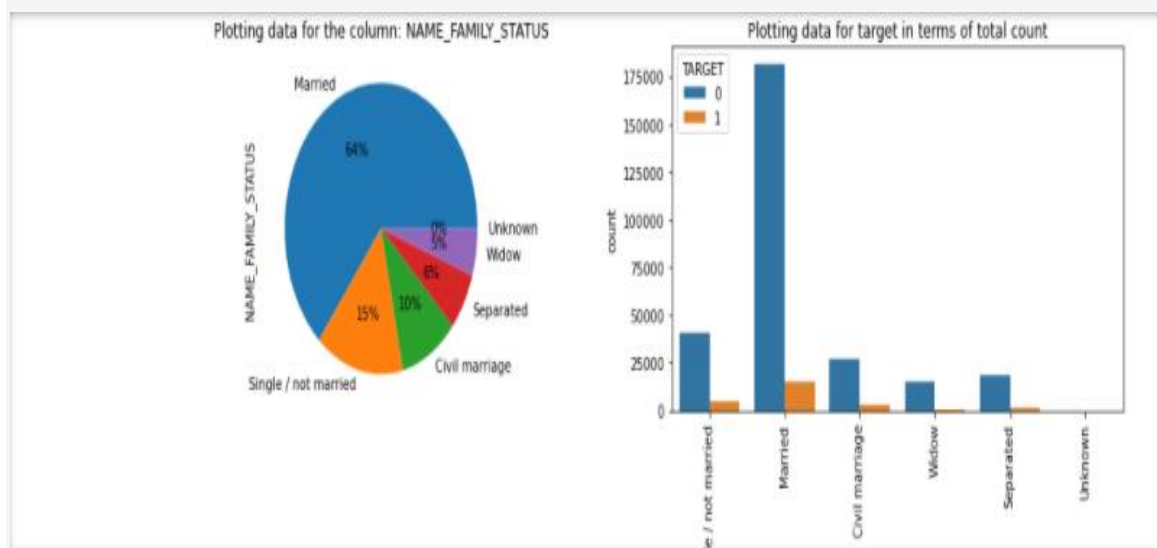
5. NAME_INCOME_TYPE:

Pensioner defaulter is lower than non-defaulter.
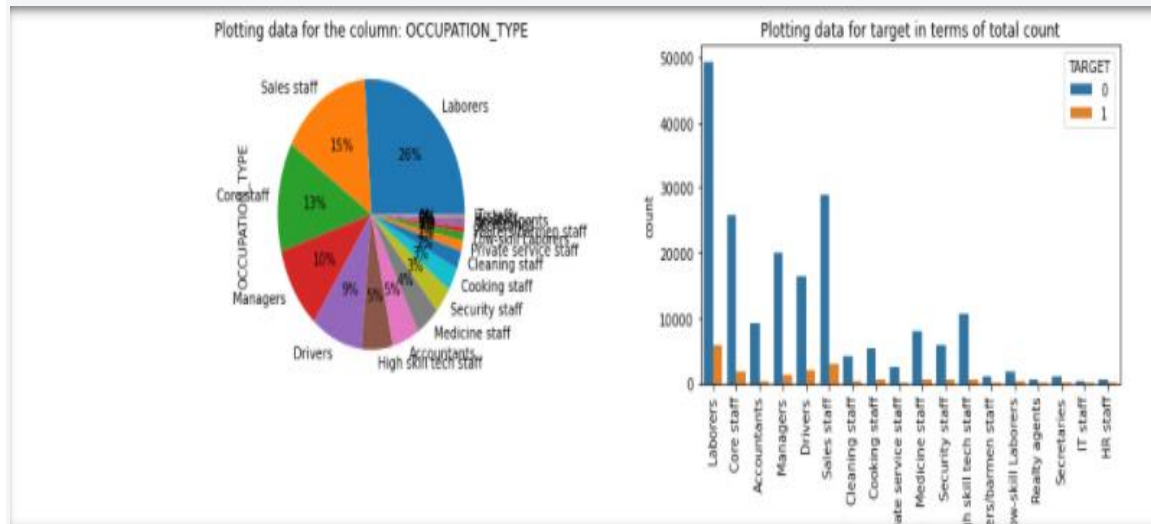

6. NAME_FAMILY_STATUS:

Most married people apply for loan, and mostly they are not defaulters.

Single and civil marriage turns out to be more defaulter.

7. OCCUPATION_TYPE:

Laborers and different categories of staffs mostly take the loan, but the managers and the high skilled tech staffs are most reliable.

8. NAME_EDUCATION_TYPE:

Most of the client take loan for secondary education followed by higher education.

But the default rate in secondary education is much higher than higher education.