LOAN DEFAULTER ANALYSIS

Pranav Belgaonkar

Project Description

Defaulter Characteristics Indentification

Consumers with low or no credit histories can turn into defaulters and lending companies can face huge losses if they do not manage their risks properly. Therefore Identification of applicants who are not capable of repaying loans is important so that they might be rejected. This project aims to find such applicants using exploratory data analysis.

Approach

01

Metadata Analysis

Understanding the characteristics of the data that has been provided

02

Handling Null values

The data is bound to contain a lot of empty fields that cannot be used and those filers have to be modified appropriately.

03

Feature Engineering and outlier detection

The data has to be modified to suit a certain use case by merging two columns or binning data for easier undersanding and so on.

04

Univariate analysis

Its major purpose is to describe; It takes data, summarizes that data and finds patterns in the data.

05

Bivariate Analysis

It involves the analysis of two variables (often denoted as X, Y), for the purpose of determining the empirical relationship between them.

Tech Stack

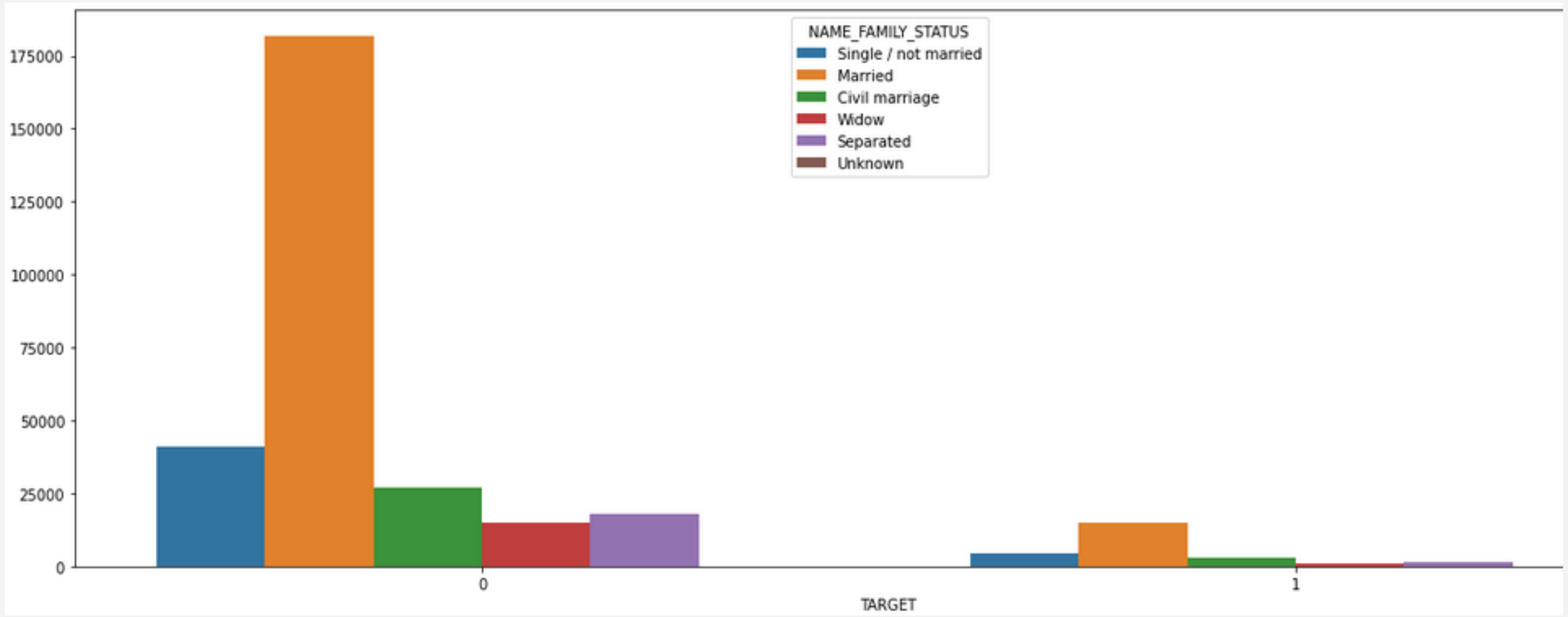
Numpy

Pandas

Seaborn

Google Colab

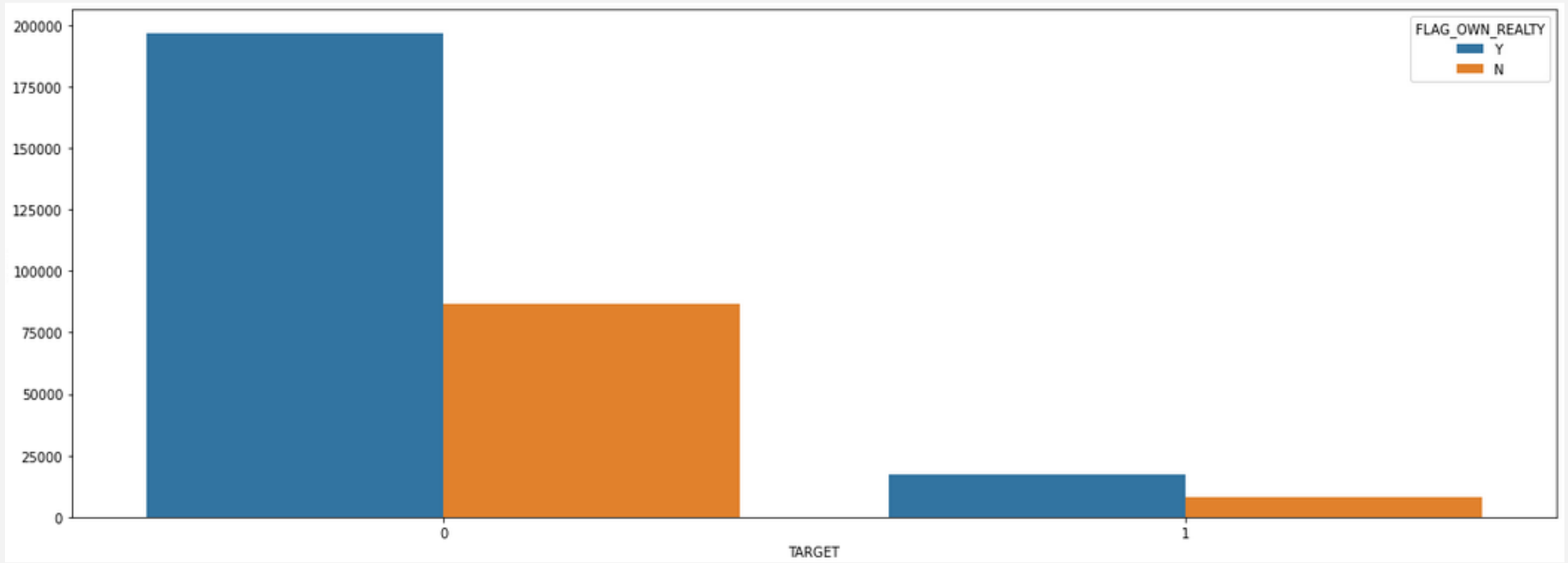
Important insights



1

Married people are less likely to default whereas single people are more likely to default. More indepth look into correlation between single people, their incomes and their default rate.

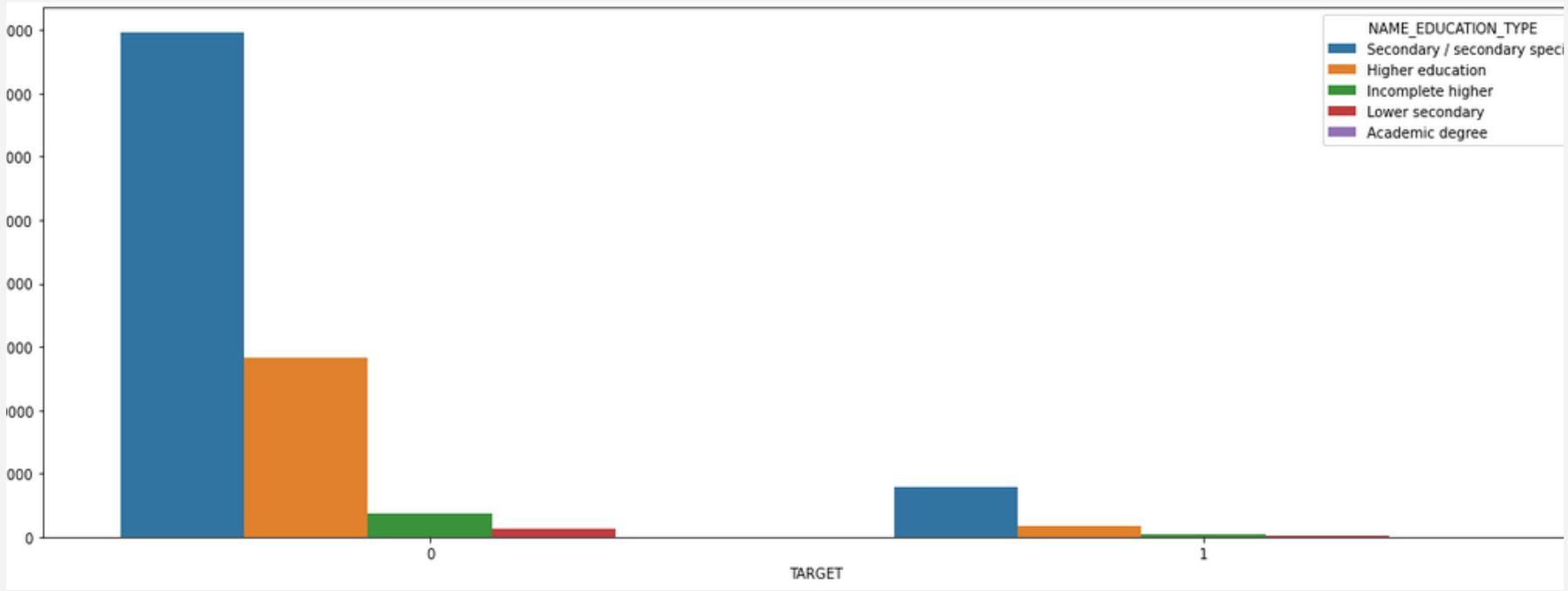
Important insights



2

Ownership of a house is not a good indicator whether a person will default or not.

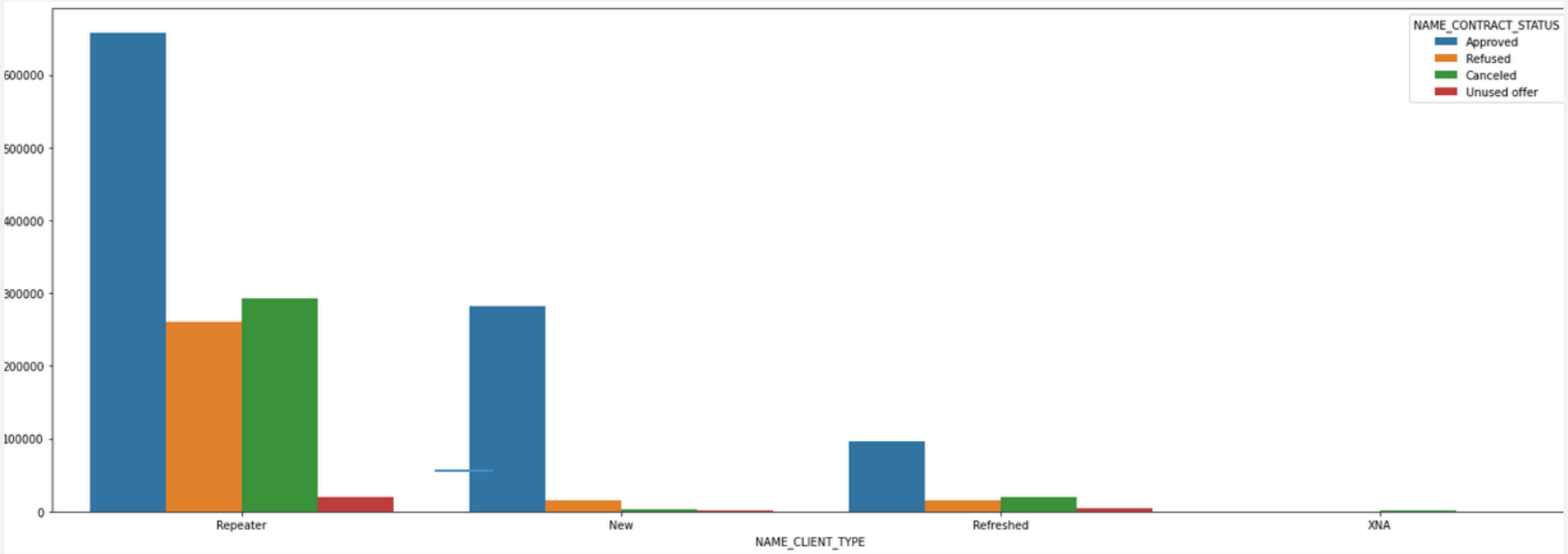
Important insights



3

Defaulters are more likely to have a secondary education and people who have higher education are less likely to default.

Important insights



4

New clients usually do not cancel their loans whereas repeaters and refreshed cancel their loans more than they are refused. Repeaters are also more likely to get refused. The reason can be because of issues during their previous loans.

Results

01 Analysis of a large real life dataset containing more than 16 lakh data points.

02 Realizing the obstacles that are encountered during EDA such as missing values and outliers.

03 creating visualizations that identify the behaviour of potential defaulters using univariate and bivariate analysis

All the insights can be found at the below link

https://colab.research.google.com/drive/1mCtJ-kakKS2ptCnrXZyUHi4fnKNzIp2s?usp=sharing