

EDA Group Study

Business Objective:

In the pursuit to assess the probability of loan default with respect to a customer, and,

At the same time reducing the risk of financial loss to the business, in the event of rejecting a prospective non-defaulter's loan application, it is crucial to have detailed, and accurate information about the customer demographics and credit details of the loan applicant.

An effective statistical model is expected so that the current information and historical data, if available for the loan applicant, can be studied to classify the applicant as a potential defaulter or non-defaulter. The inferences drawn shall help the business to increase profitability, and reduce its non-paying liabilities.

Available Support:

Machine learning techniques to resolve the multiclass classification problem for predicting loan default propensity of a prospective customer who approaches the business client with a loan application.

Data availability-Business data used for the analysis comprises of two datasets that provide:

Customer's existing information collated at the time of his/her current loan application.

Customer demographics information collated at the time of his/her previous loan application.

EDA Group Study

Approach

The analysis is performed at two levels:

- First set of inferences drawn from data comprising details of customers' current loan applications.
- Second set of inferential study is done on merged dataset of the existing and previous loan application data of the customers. This validates the initial inferences drawn from the analysis of the first dataset.

Type of Analysis: Exploratory Data Analysis-

- The target data variable is bifurcated between clients with payment difficulties, and all other cases.
- The statistical assumptions shall provide top 10 correlations between the target outcome and the various reasons in the data that lead to either of the expected outcome.
- Data visualization supported by machine learning that provides better clarity to understand the interdependencies of these variables with each other.
- Extensive data driven study of the reasons of people taking the loans and ascertaining reasons for the default.
- Study and visualize the demographic nature of loan defaults.

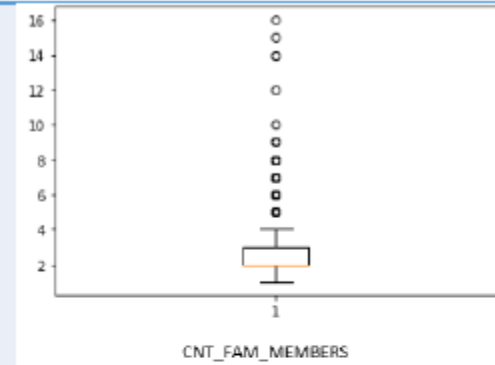
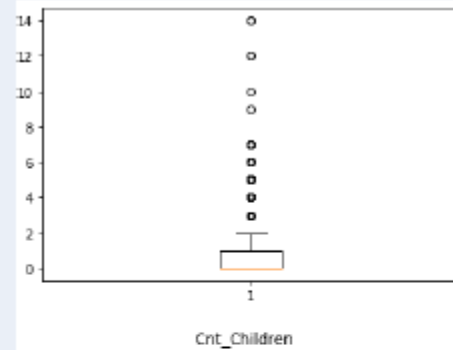
Challenges

Data redundancy in the form of-

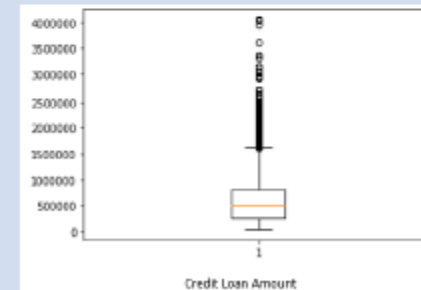
- Missing data
- Incorrect data formats used
- Non-standardized variable names
- Inadequate data dictionary that restricted the ease of understanding the data
- Non-scalability of data-No units of measurement for variables provided either in the datasets or in the data dictionary. Variables like 'Days_Birth', require further calculations to be better analyzed.

EDA Group Study

CNT_CHILDREN 7 CNT_FAM_MEMBERS- The family household of loan applicants varies between loan applicants who have no kids to those with 14 children.
The range fluctuates between that of 1 to 16 family members.

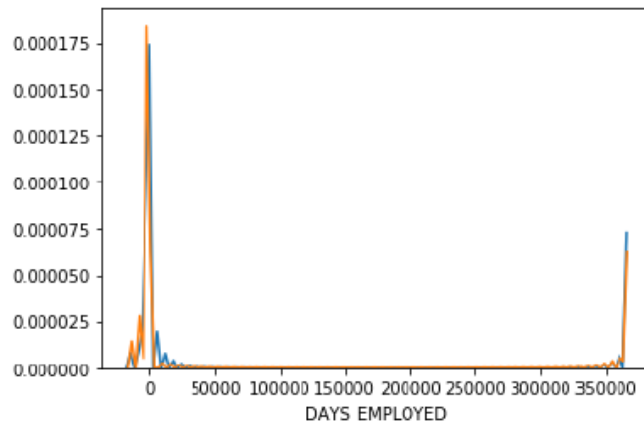
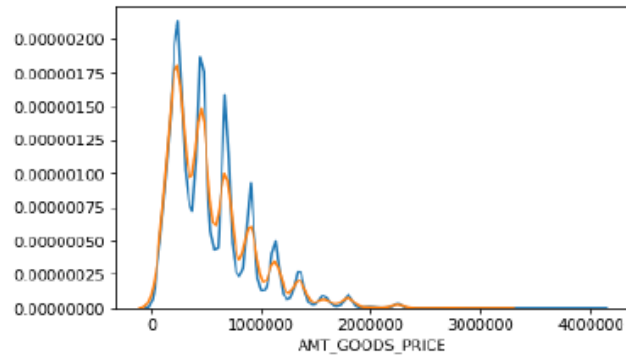


Credit Amount- The range of credit amount of the loan varies from 40k to 50lakhs



EDA Group Study

Relational Observations of other variables

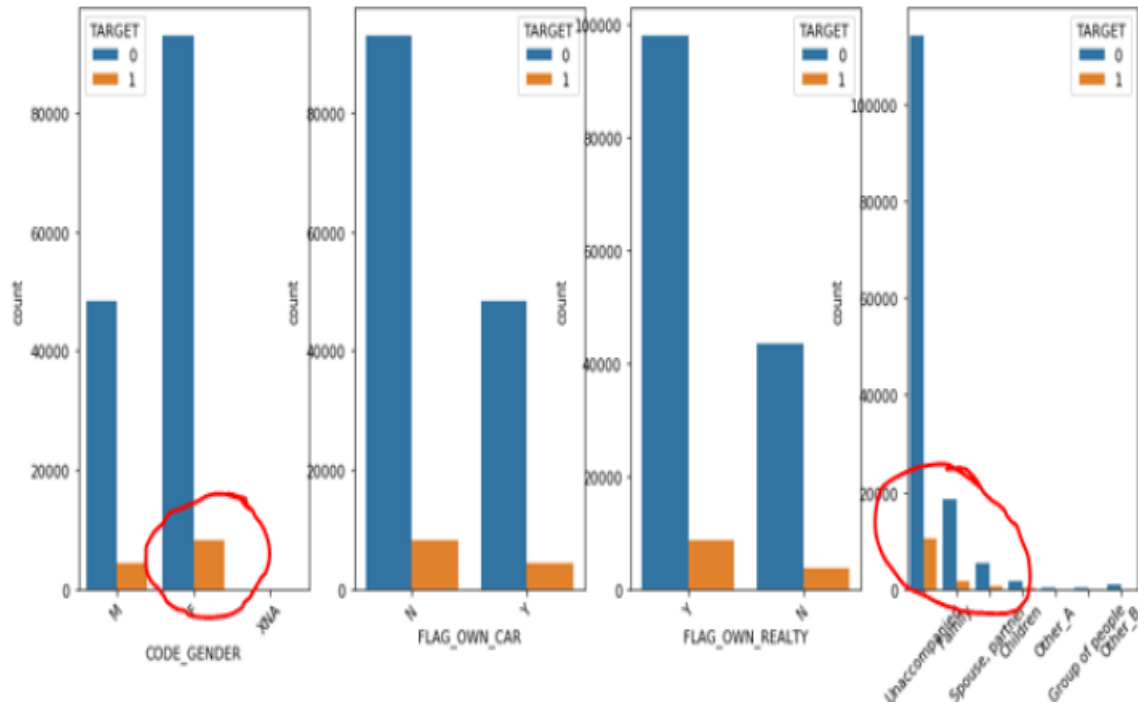


The inferences drawn by analyzing the continuous variables present in the data help in drawing following assumptions for customers who face difficulty in paying their loan installments state:

- Customers who face difficulty towards payments of loan installment (*indicated with the orange line*) have their consumer loan fall in the range where the price of the goods for which the loan is provided is upto 10lakhs.
- Study of customer's days of employment indicate that customers who are unemployed or have recently started in their jobs, and customers who have served long duration of more than 30,000 days as working professionals tend to find difficulty in paying their loan installments.
- The third visual denotes that customers who have no children

EDA Group Study

Relational Observations of other variables



The inferences drawn by analyzing the categorical variables present in the data state that customers who face difficulty in paying their loan installments state:

- More number of females than males default in being regular towards their loan payments as compared to males
- Customers who are not accompanied by their family members are also prone to become loan payment defaulters.

EDA Group Study

Inferences

- It is important that lending organizations collect educational information of the loan applicants. It can be inferred that customers with an intention to default, try to submitting educational credentials that can be used to track them in the event of loan amount collection.
- Customers with higher days of employment could be shifting towards retirement and hence might eventually start facing financial challenges.
- Small scale business owners who require low range consumer loan of upto 10 lacs can be those who might not have well laid businesses. Hence their probability of making loss in business is higher and hence they tend to default towards their loan payments.

Thank You