Loan Case Study

Madhur Ramachandruni

Project Description: This project's intention is to get familiarized with data analysis techniques and the libraries used. We will be doing a case study on Loans to get familiarised with the seaborn techniques.

Approach: I used an approach wherein I first tried all the functions I knew and applied them for simple applications later I used the relevant functions which were required for the problem.

Tech – Stack Used: I used Jupyter Notebook for this project.

Insights – After implementing data analysis techniques to the given data set, we can say that:

- We first take the application_data.csv and start our analysis.
- There are 122 columns having various data types like object, int, float and 305711 rows.
- There columns having negative, positive values which includes days. fixing is required
- There are columns with very high values, columns related to Amount (Price).
- We also find the correlation b/n Source & Target.
- We use itertools as it is an efficient method for looping plotting subplots.
- After removing unnecessary, irrelevant and missing columns. We are left with 46 columns.

- We do imputing for relevant missing columns wherever required.
- We plot a graph to understand the type of occupation. After plotting we know that Highest percentage of values belongs to Unknown group and Second belongs to Laborers.
- If days are in negative which isn't possible so we use absolute function to correct that.
- We convert days, months in terms of years for better understanding.
- Now we take the previous_application.csv and follow the same steps done earlier.
- There are 37 columns having various data types like object, int, float and 1670214 rows.
- There columns having negative, positive values which includes days, fixing is required.
- We plot a graph to understand distribution of AMT_ANNUITY. There is a single peak at the left side of the distribution and it indicates the presence of outliers and hence imputing with mean would not be the right approach and hence imputing with median.
- We perform imbalance plotting (Repayer vs Defaulter) to understand the data.
- We plot graphs to understand the gender of the loan applicant.
- The number of female clients is almost double the number of male clients.
- Based on the percentage of defaulted credits, males have a higher chance of not returning their loans about 10%, comparing with women about 7%

Link for the pdf file and the code file:

https://drive.google.com/drive/folders/1kR4U45lGXXUtvOLptMuzg0FLgMT75w2w?usp=sharing