Analytics Using Python

Team Name – Non - Coders

Team Members –

1. Shivang Jitendra Khare (PG – 2)
2. Ishojit Roy Choudhury (PG – 2)
3. Sahil Samir (PG – 2)
4. Garima Rathi (PGDM – Business Design)
5. Ladi Amarnath Patro (PG – 1)

Problem Statement

The Problem Statement opted by our team is **Analysis of Customer Eligibility for loans.** Customer Eligibility for loan is a major confusion for a consumer seeking loan. The loan limit as well as eligibility criteria has always been a major concern for an individual.

Analysis & Inferences

Through our analysis, our team has tried to find a relation between loan approval and multiple variables that can prove to be detrimental in getting a loan sanctioned. Our data set took us to quite an interesting insight which will be discussed further. Before jumping to the insights, let's see what variables we had. The major variables that were included in this analysis are Gender, Marital Status, Number of dependents, Education, Employment, Applicant Income, Co-applicant Income, Loan Amount, Credit History and Property Area. With these variables in hand, we went further to analyze as to what degree these factors impacted the Loan Status of a customer.

We first conducted a **Univariate Analysis.** The univariate analysis is done to analyze each variable independently. Through such an analysis, we inferred the following:

- **81.36%** of the applicants were **Male**
- **65.13%** of the applicants were **married**
- **14.03%** of the applicants are **Self-Employed**
- Most applicant did not have dependents
- **78.17%** of the applicants are Graduates.
- The Population of Semi-Urban was the highest.

On further analysis, the data set was found to be skewed rather than normally distributed. The major **outliers** were graduates with a high income.

We then went further with a **Bivariate Analysis.** For the bivariate analysis to succeed, our team made following Hypothesis:

- **Higher Income** leads to **Higher chances of loan approval.**
- **Urban residents** should have **higher** chances of **loan approval.**

- **Salaried employees** should have good chances of loan approval.
- Applicants with a good **Credit History** should have higher chances of **loan approval.**
- **Gender** plays a role in loan approval

Using this Hypothesis, our next step to analysis gave another interesting conclusion. The major inferences from the Bivariate Analysis are as follows:

- Proportion of male and female is almost same for approved and unapproved loans.
- People with **lower income** have less chances to get their **loan approved.**
- People with good **Credit History** have **higher** chances of getting their **loan approved**.
- **Semi-Urbans** have **better chances** of getting their **loan sanctioned.**
- Nothing significant can be said for **Self-Employed vs Loan Status.**

With the Bivariate analysis complete, we moved to find correlation between respective variables. To do so, our team decided to plot heat maps to show the level of correlation between different factors.

The heat maps further verified our previous conclusions. The information as given by heat maps are as follows:

- The strongest correlation was between **Credit History** and **Loan Status.**
- **Total income** had a good correlation with loan amount with **0.62** degree of correlation.
- **Applicant Income** also showed a fair degree of correlation with the loan amount.
- **Co-applicant** income had a very low correlation with the amount.

Solutions suggested

- Banks should create a software or a feature on their mobile app that can determine the chances of loan approval based on the analysis done using data available.
- Determine the loan amount & whether an applicant will be able to repay the loan that one has taken based on the various factors.