**LOAN STATUS PREDICTION INSIGHTS**

**Feature Analysis:**

**LOAN STATUS**: As per the given data more than 70% of the people are approved for the loan.

**GENDER**: Approximately 80% of the applicants are men and only 20 % are female.

**MARITAL STATUS**: 65% of the applicants are not married and 35% were married.

**SELF EMPLOYED**: Only 13% were self-employed and rest of the 87% were not.

**CREDIT HISTORY**: 83% of the people had credit history i.e they previously availed loans and only 17% are applying for loan for the first time.

**DEPENDANTS**: 57% of them are not having any dependents. This might be because the spouse might be employed, and they might not have had children yet.

**EDUCATION**: 79% of them have Completed Graduation and only 21% are under graduates.

**PROPERTY AREA**: There are 3 types of areas: Urban, Semi-urban and Rural. Considering the cost and comfort levels and employment facilities most of the people chose Semi-urban than Urban and least preference was rural.

**LOAN AMOUNT**: The data is right skewed. i.e the density is more for some values and outliers were present which means in few cases the loan amount was very high.

**APPLICANT INCOME**: The applicant Income is also right skewed.

**Data Cleaning and Processing:** Data is cleaned by replacing the categorical null values with mode and numerical null values with median. Loan Amount and Applicant Income values are log transformed to reduce the skewness.

**Co-relation**: Income is directly proportional to Education. Property Area and Credit History are having high positive co-relation with Loan Status (target variable) comparing to others.

**Modeling:** Logistic Regression model is chosen to predict the Loan Status.

**Model Performance:**

**Accuracy** 0.7560975609756098

**Confusion Matrix** [[19 1]

[29 74]]

**Classification Report**

precision recall f1-score support

0 0.40 0.95 0.56 20

1 0.99 0.72 0.83 103

accuracy 0.76 123

**f1-score** : the harmonic mean between precision and recall, is between 0.8 and 0.9 (near to 1)so the model performance is good.