LOAN STATUS PREDICTION

USING CLASSIFICATION MODEL, LOGISTIC REGRESSION

PRIYADARSHINI G.
BADM, BATCH DM7

INTRODUCTION

In this analysis, the focus is on a dataset concerning Property Loan Status Prediction. The dataset includes information about applicants who have applied for loans based on various property and applicant-based metrics. The goal is to build a machine learning model that predicts whether a loan application should be approved or rejected. This is a binary classification problem, where the target variable is 'Loan_Status', indicating whether the loan was approved (Y) or not (N).

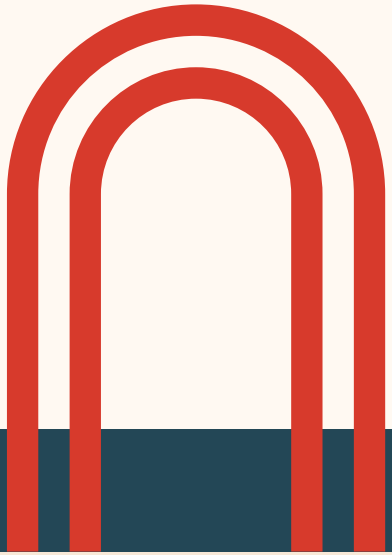
Problem Statement:

https://drive.google.com/drive/folders/1buvhyRpnXuaOz1eoyDWq5O9HpOqcAS09?usp=sharing

PERFORMANCE

A Logistic Regression model was utilized for loan status prediction. The dataset underwent preprocessing, encompassing the encoding of categorical variables and addressing missing data. Subsequently, the model was trained using the provided training data and applied to make predictions on the test data. The Logistic Regression model exhibited an accuracy score of approximately 76.92% on the training dataset.

This analysis of Property Loan Status Prediction, can be accessed on Kaggle via the following link: Kaggle Loan Status Prediction Dataset.

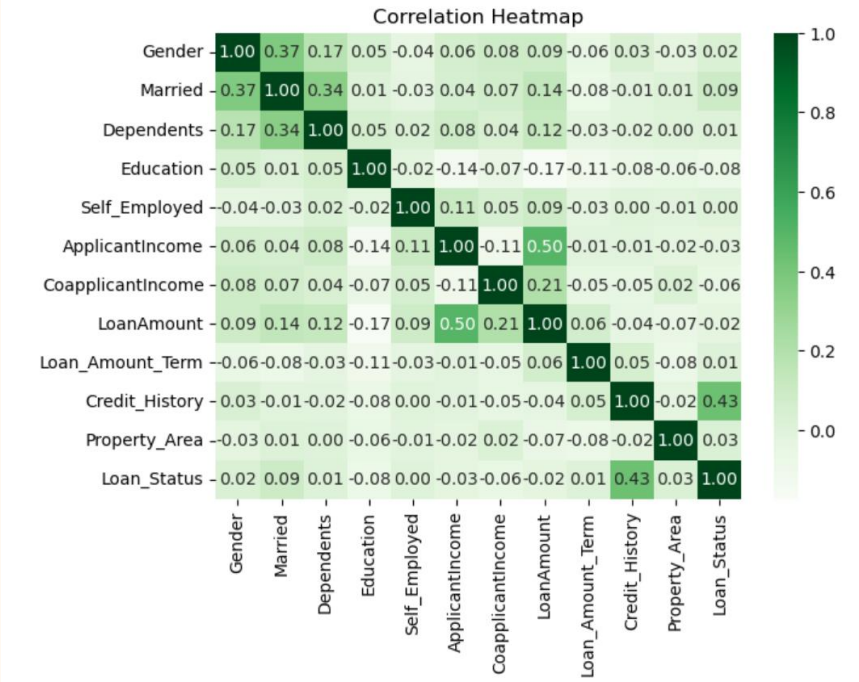


KEY INSIGHTS

1. CORRELATION ANALYSIS:

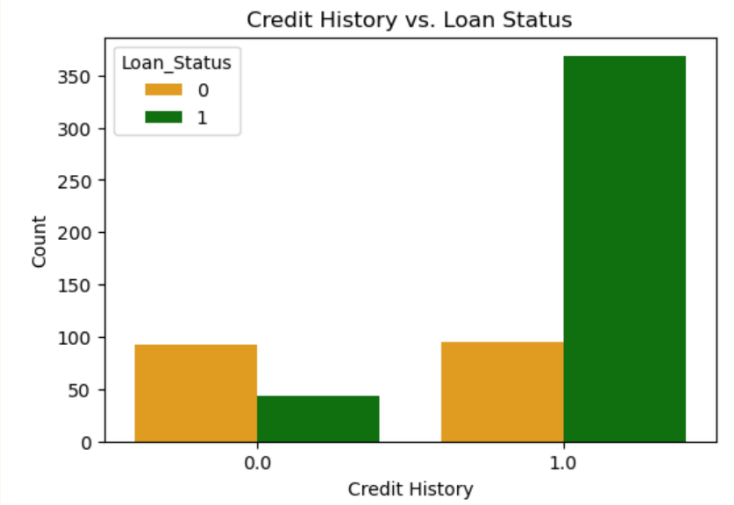
The correlation heatmap discloses that 'ApplicantIncome' displays a moderately significant positive correlation (0.50) with 'LoanAmount'. This indicates that higher applicant incomes are associated with larger loan amounts.

Furthermore, 'Credit_History' demonstrates a meaningful positive correlation (0.43) with 'Loan_Status'. A favorable credit history seems to positively influence loan approval.



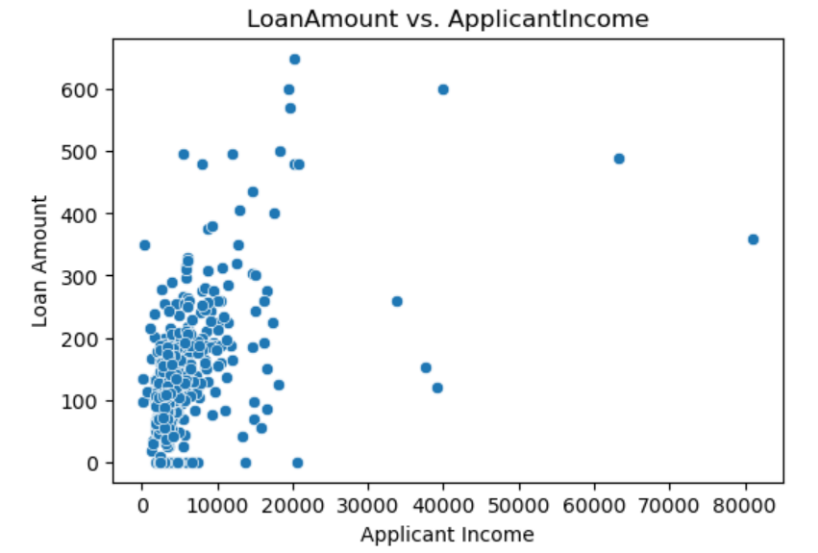
2. IMPACT OF CREDIT HISTORY:

The count plot of 'Credit_History' against 'Loan_Status' reveals that applicants with a positive credit history (1.0) are more likely to have their loan applications approved. This reaffirms the correlation highlighted in the heatmap.



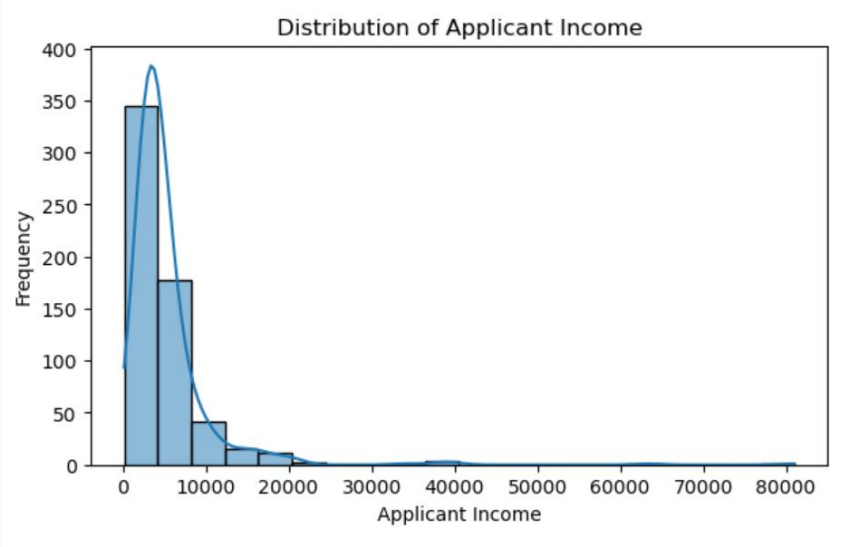
3. APPLICANT INCOME AND LOAN AMOUNT:

The scatter plot depicting 'ApplicantIncome' against 'LoanAmount' suggests a positive correlation between these two variables. As applicant income increases, loan amounts also tend to rise.



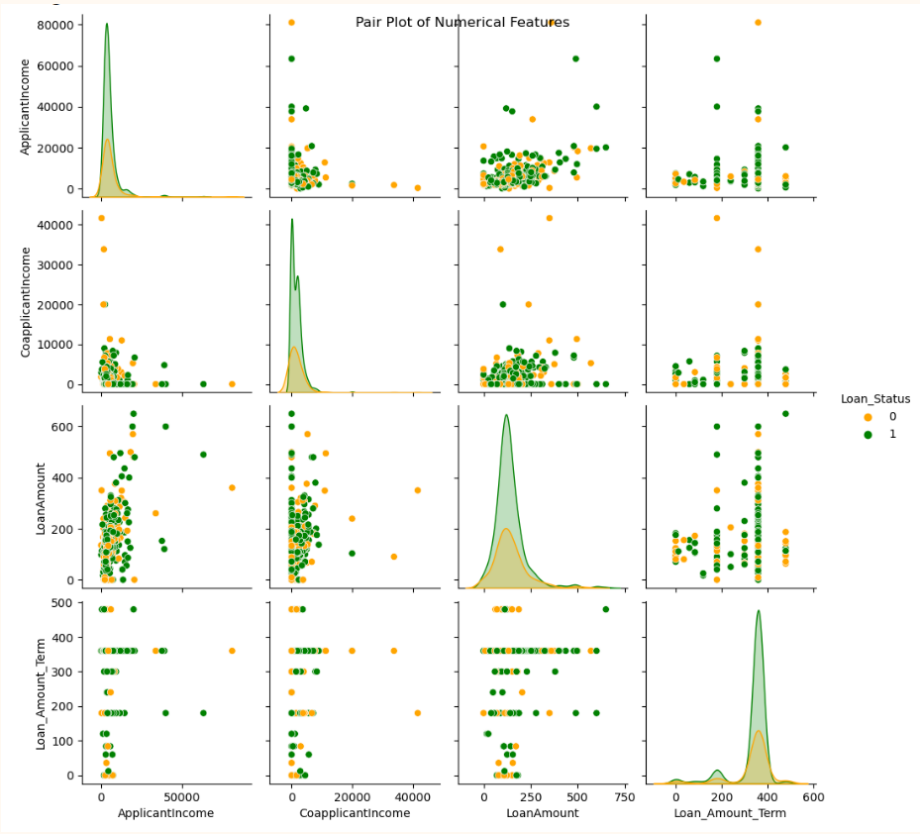
4. DISTRIBUTION OF APPLICANT INCOME:

The histogram illustrates the distribution of 'ApplicantIncome'. It showcases a right-skewed distribution, indicating that most applicants have lower income levels.



5. PAIR PLOT ANALYSIS:

The pair plot of numerical features such as 'ApplicantIncome', 'CoapplicantIncome', 'LoanAmount', and 'Loan_Amount_Term' offers a visual insight into relationships between pairs of features. The hue distinguishes loan statuses, facilitating the identification of patterns within the data.



CONCLUSION

This analysis focused on predicting loan status using a Logistic Regression model. The dataset was explored to uncover correlations between various factors and loan approval. The model achieved an accuracy of approximately 76.92% on the training data. Key insights include the positive correlation between applicant income and loan amount, as well as the influence of credit history on loan approval. The analysis provides valuable information for improving loan approval predictions and can be further explored using the provided Kaggle link: Kaggle Loan Status Prediction Dataset.