META SCIFOR TECHNOLOGIES, BANGALORE.

AI INTERN

MINOR PROJECT-2

REPORT

(LOAN PREDICTION PROJECT)

By,

M.Rachel

CONTENT

1) PROBLEM STATEMENT

LOAN APPLICATION STATUS PREDICTION

This dataset includes details of applicants who have applied for loan. The dataset includes details like credit history, loan amount, their income, dependents etc.

Independent Variables:

- Loan_ID
- Gender
- Married
- Dependents
- Education
- Self_Employed
- ApplicantIncome
- CoapplicantIncome
- Loan_Amount
- Loan_Amount_Term
- Credit History
- Property_Area

Dependent Variable (Target Variable):

- Loan_Status

  To build a model that can predict whether the loan of the applicant will be approved or not on the basis of the details provided in the dataset.

2) DATA EXPLORATION

- Importing the Necessary Libraries such as Numpy, Pandas, Data Visualization libraries such as Matplotlib, Seaborn and Machine Learning Libraries such as Scikit-learn, Logistic Regression for Classification task, train_test_split for splitting the dataset as training set and testing set for Model Training, and Classification metrics such as accuracy score, Confusion metrics, ROC AUC curve

- Load the dataset using pandas, explore the data such as load first 5 rows of the data, checking the information of the data which shows the datatypes

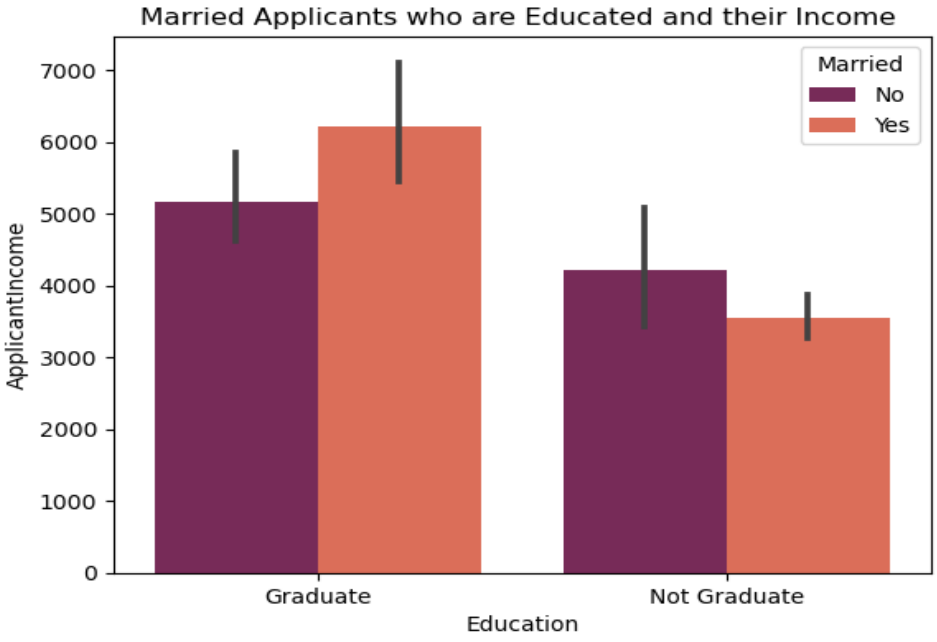
- Treating the null values by imputing with the mean value of the respective features .

- After all null values are treated in the data, lets check and treat the 0 values and treat the outliers.

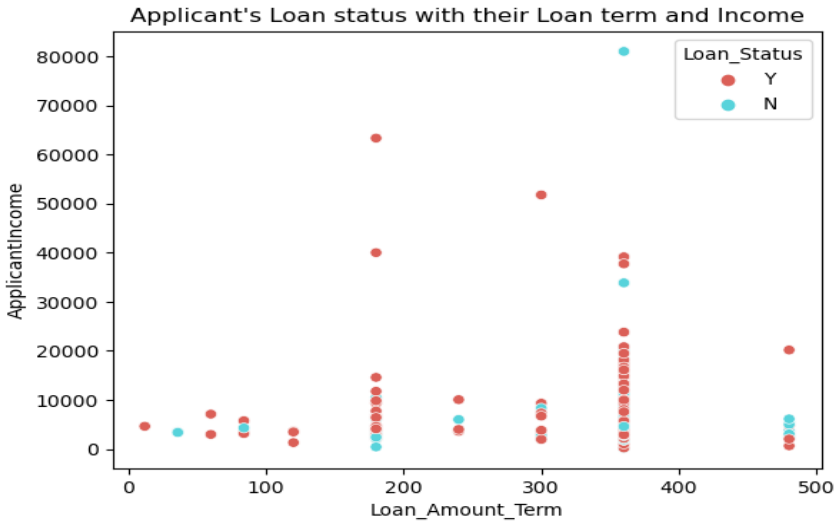
- Before Preprocessing let's visualize the data with features using different plots for better understanding of the data.

3) DATA VISUALIZATION

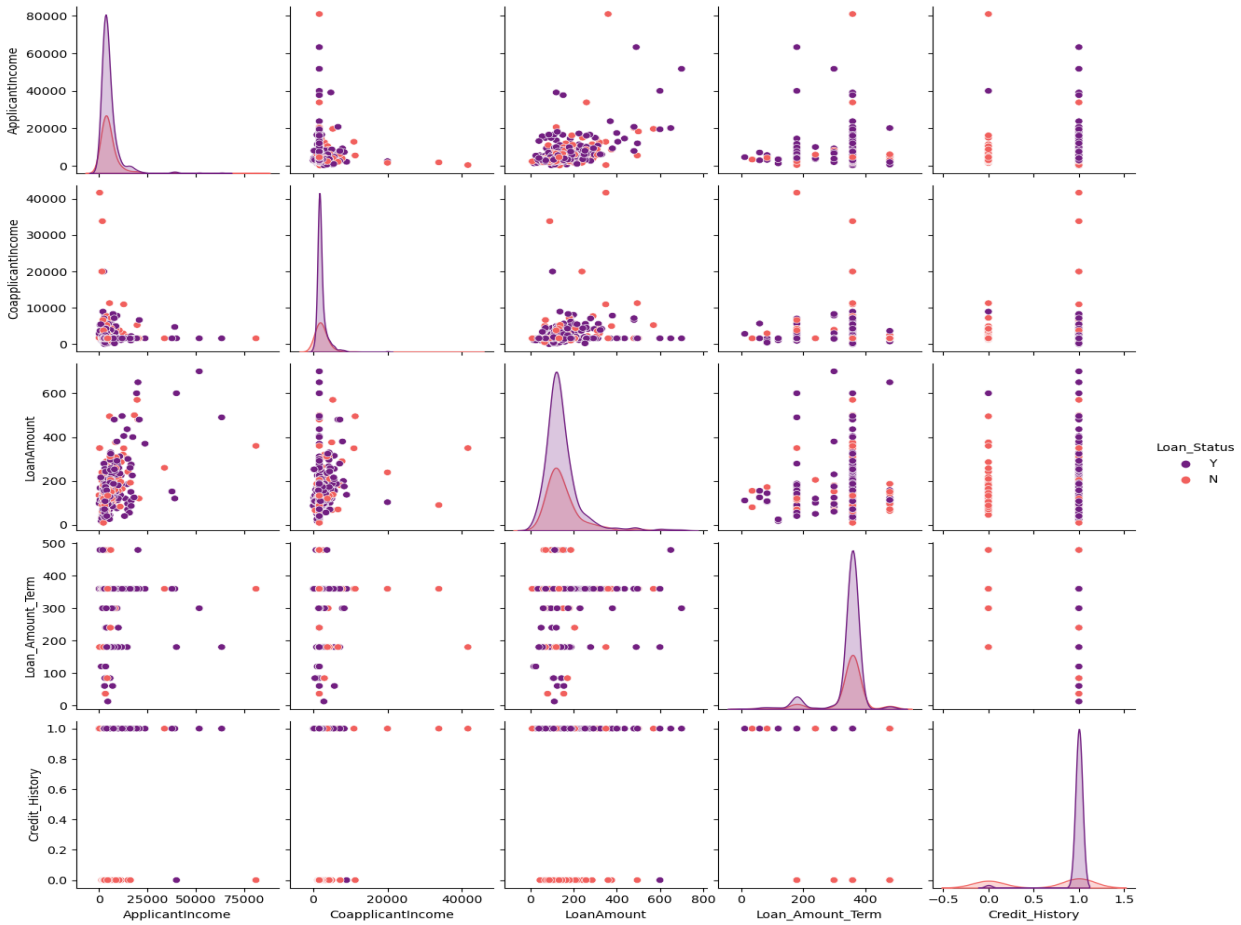
- Visualize with Education and Applicant Income with the Married Individuals.



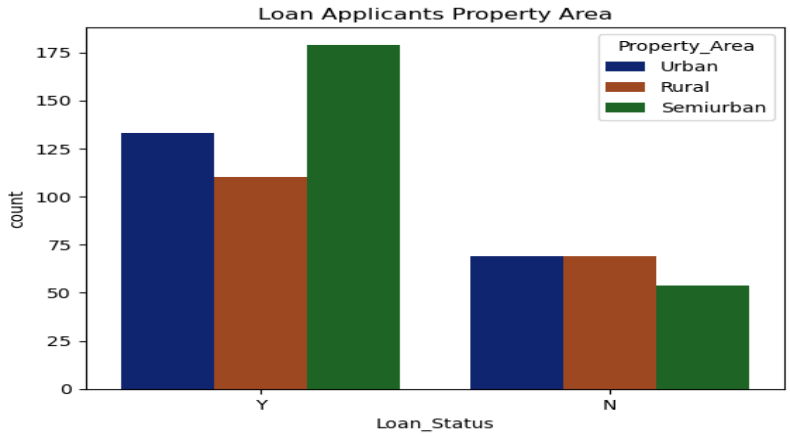
- Visualization on the Loan Amount Term and the Applicants Income with their Loan Status.



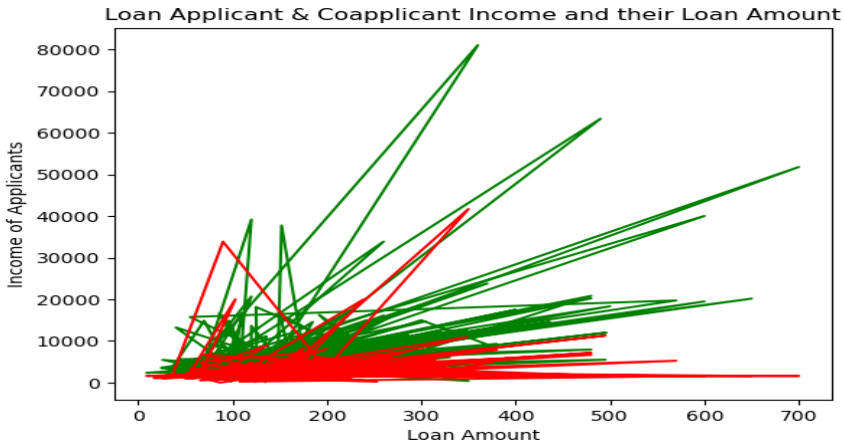
- Visualization on Pairplot of the dataset with their Loan_status



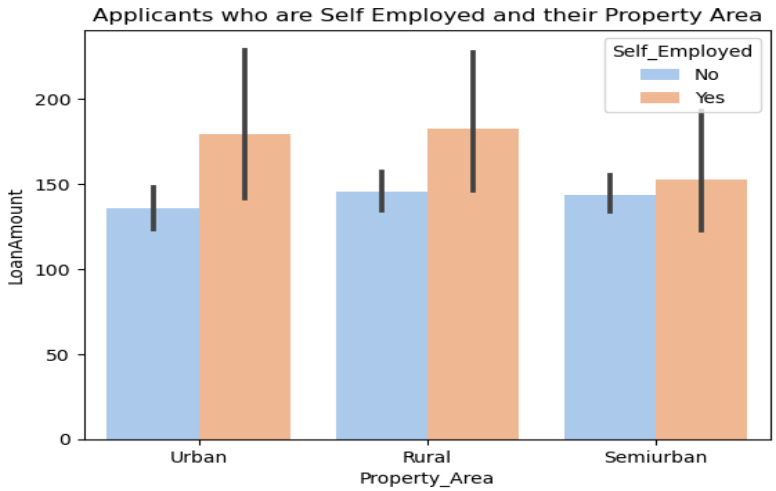
- Visualization of the Applicants with their Loan_status and property area



- Visualization of the Loan Applicant's and Coapplicant's Income



- Visualization of the applicants who are self employed their Property area and Loan Amount



4)DATA PREPROCESSING

- Label Encode all the Categorical Features.
- Plotting the distribution plot for checking all features to be Normal distribution.
- Plotting the boxplot for visualizing the outliers.
- Through quantile 25% (q1) and quantile 75%(q3) we calculate the IQR range and treat the outliers.
- Reindexing the data after treating the outliers of the respective features.
- Checking for Multicollinearity problem through correlation matrix, out of all features, since there is no feature that has more correlation so we further proceed with model building or training.

5)MODEL BUILDING OR TRAINING

- Splitted the dataset as X for Independent features and y for target variable as Loan_Status.
- Transforming the x features into scaled features through Standardization.
- Using train test split function Splitting the x scaled and y as training set and testing set with the parameters such as test size set to 25% .
- Now for Model Training we use Logistic Regression Machine Learning Algorithm, fitting the model with training sets such as x_train and y_train.
- Now predict the model with x_test and calculate the accuracy score with training as 82% and test score as 80%
- Through Classification metrics such as Confusion Matrix, precision, recall scores, ROC AUC score we can predict the Model Perfomance.

.6) CLASSIFICATION METRICS

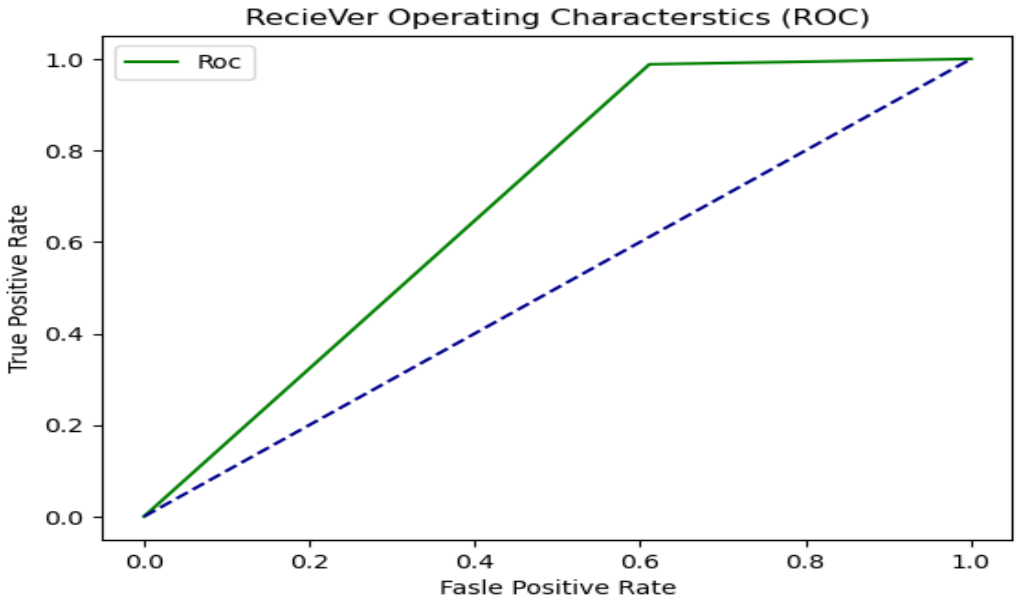
Accuracy score:

1. Training score - 82%
2. Test score - 80%

Metrics score:

1) Precision –> 0 – 0.93, 1 - 0.79
2) Recall –> 0 – 0.39, 1 - 0.99
3) F1 score - > 0 – 0.55, 1 – 0.88
4) Confusion Matrix: [[14, 22], [1, 84]]

5) ROC AUC Curve



Area covered: 68%