INSIGHTS ON LOAN DEFAULT PREDICTIONS PROJECT

BY NAHID BANU K

Step By Step Process

Highlight two or more cells, right-click then choose "Merge Cells" to organize your table according to your needs!

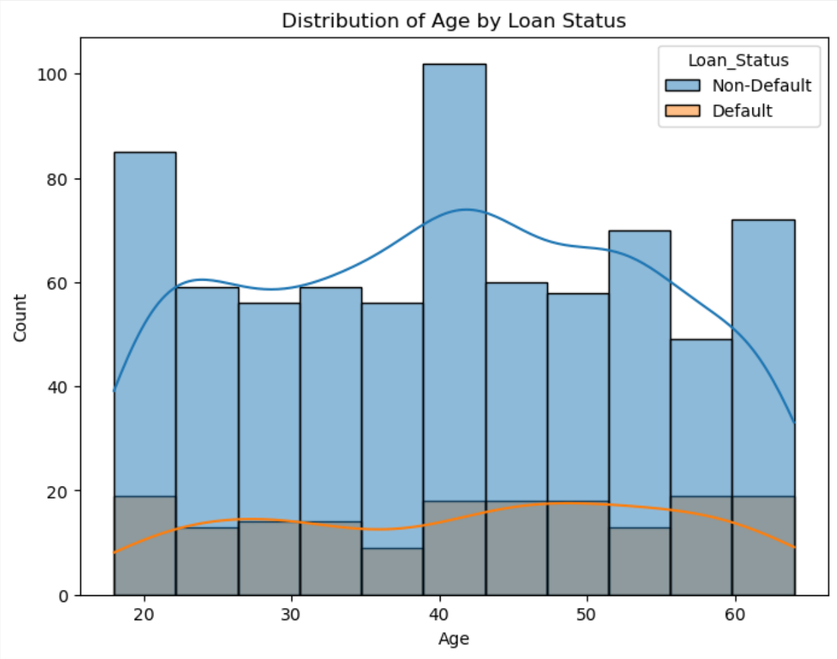
✓	Step 1	**Understand the data, Problem Statement given**
		Import the data and convert the datatypes to accessible format
✓	Step 2	**Check for any Null values, Outliers and Skewness if any**
✓		The Null values are as follows Gender 208 Employment_Status 94
✓	Step 3	Build a Machine Learning model to predict Gender and fill the null values with predicted values to the dataset and drop the null values of Employment Status using df.dropna() method
✓		Process EDA and visualize the data and impact of each column on the Loan Default status prediction
✓	Step 4	Build a MAchine Learning Model to Predict Loan Default Status
		As Loan Default Status is a class variable go with Classification Model
	Step 5	**Deploy the Model and build a streamlit application**
		pickle the best fit model and make prediction using unseen data/ test data

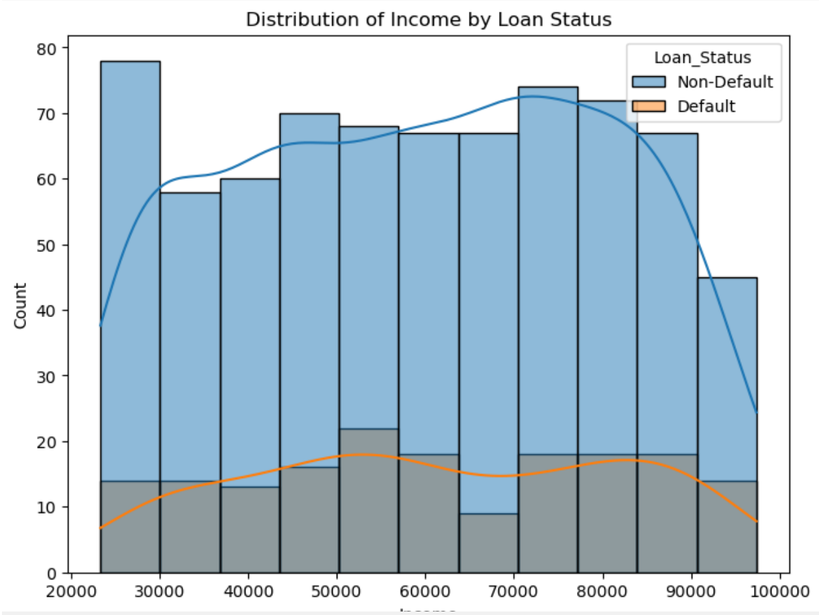
Build a Machine Learning model to fill Nan values of Gender

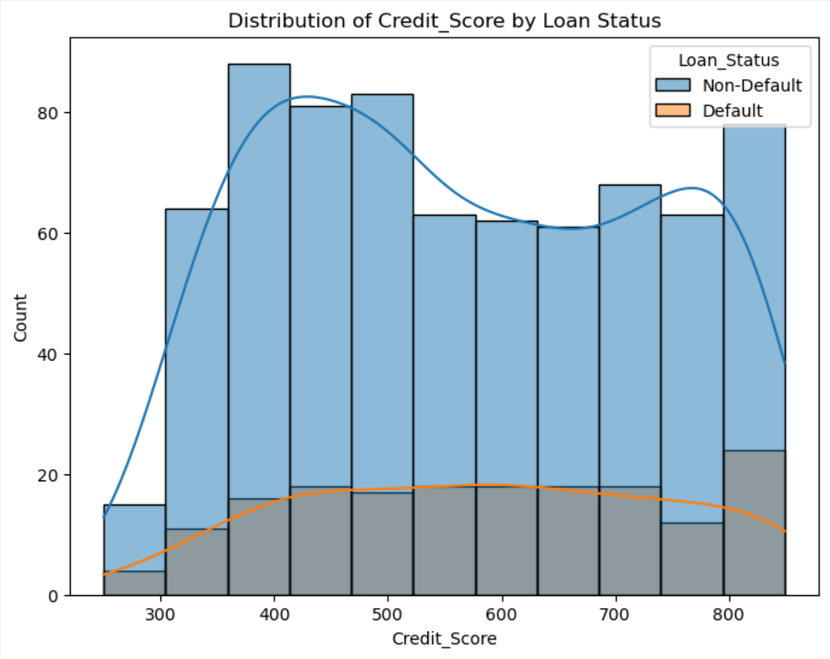
As Gender is a class prediction, we will go with Classification Model

Encode the **categorical data using One hot Encoding,** and fit the **Numeric Data using Standard Scaler** before feeding it into Machine Learning Model

Exploratory Data Analysis





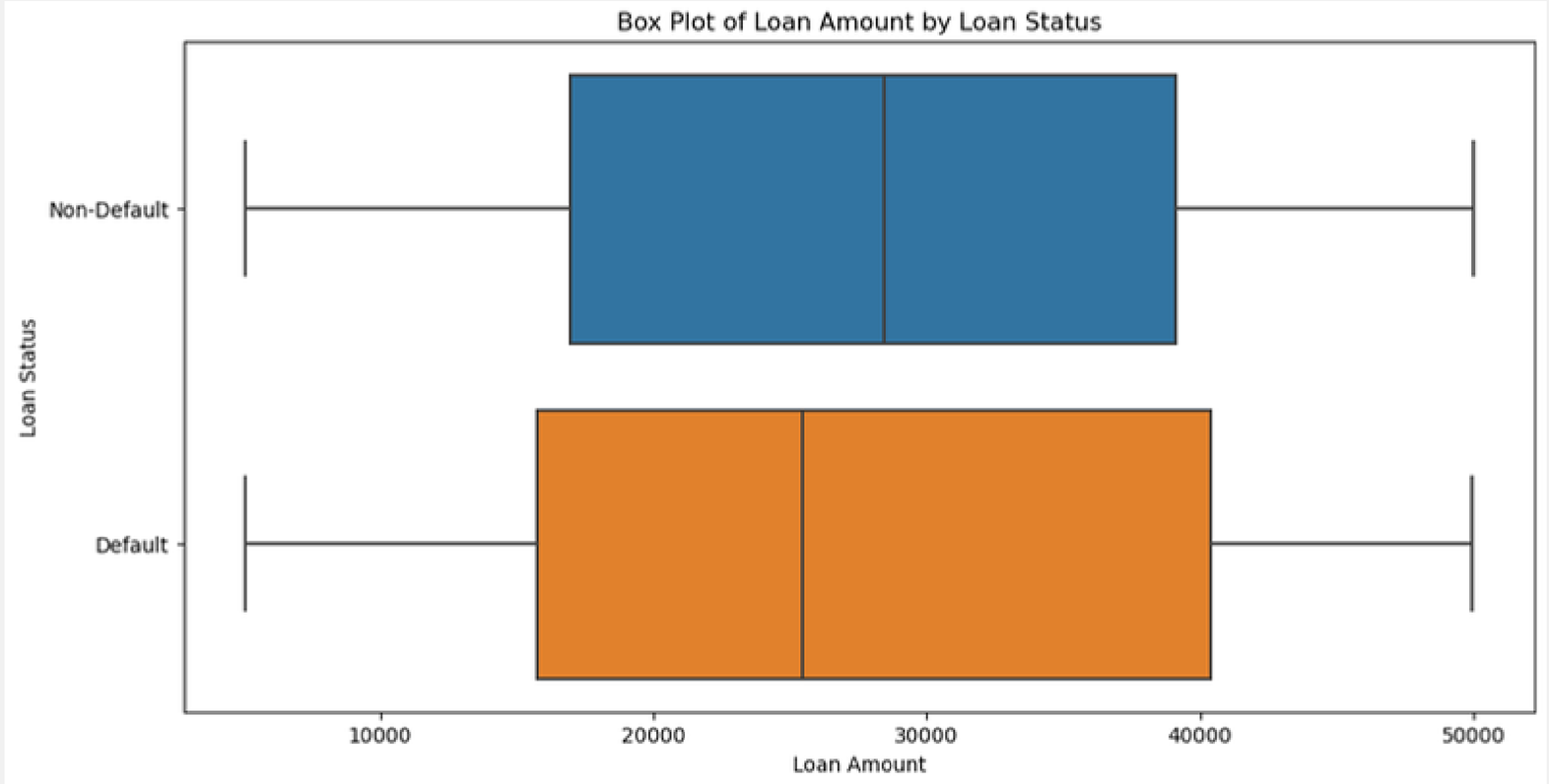


- Default Status is highest for age group of 45-50 and lowest at 20 s
- Default Status is highest for Income of 45000-55000 and lowest at 20000 &above 100000
- Default Status is highest for credit score of 400-680 and lowest for less than 300

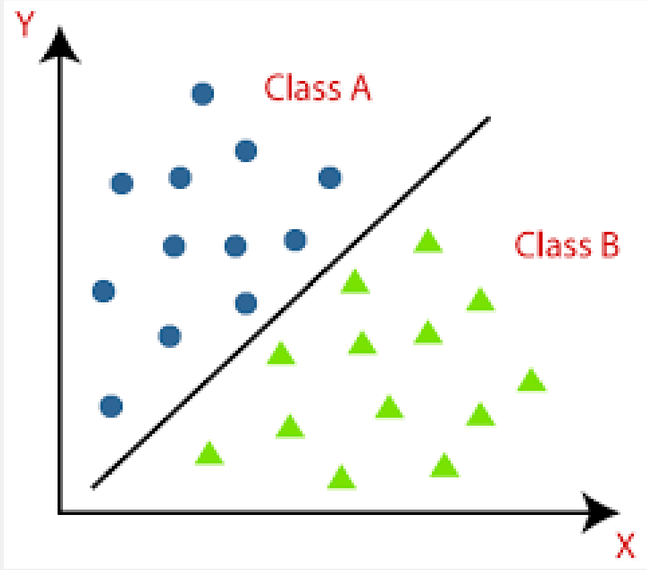
Total Loan_Amount and Loan_Status

Default: 27222.920278

Non-Default: 27883.202174



Status

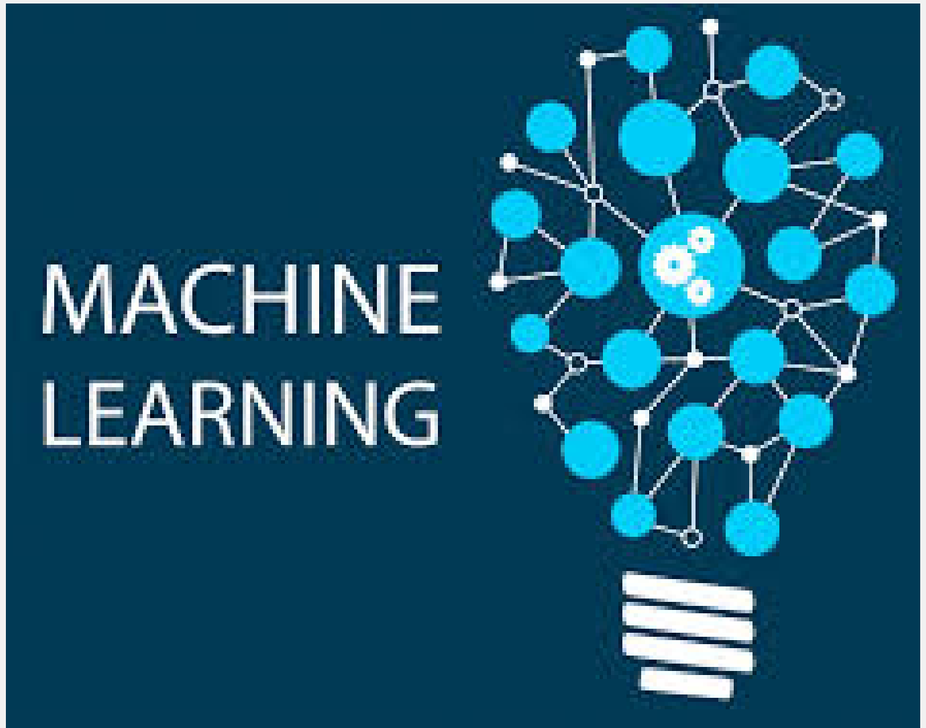


Encode the categorical Data : Gender and Employment Status using One Hot Encoding,
Location Data using Label Binarizer

Fit the Numerical Data using Standard Scaler.

Use SMOTE to deal with imbalanced data.
Train Classification model and I found Gradient Boosting Classifiaction as best fit model and hence proceeded forward with it.

Pickle the model for further use and deploy it on streamlit as to make predictions based on user input



Loan Default Prediction

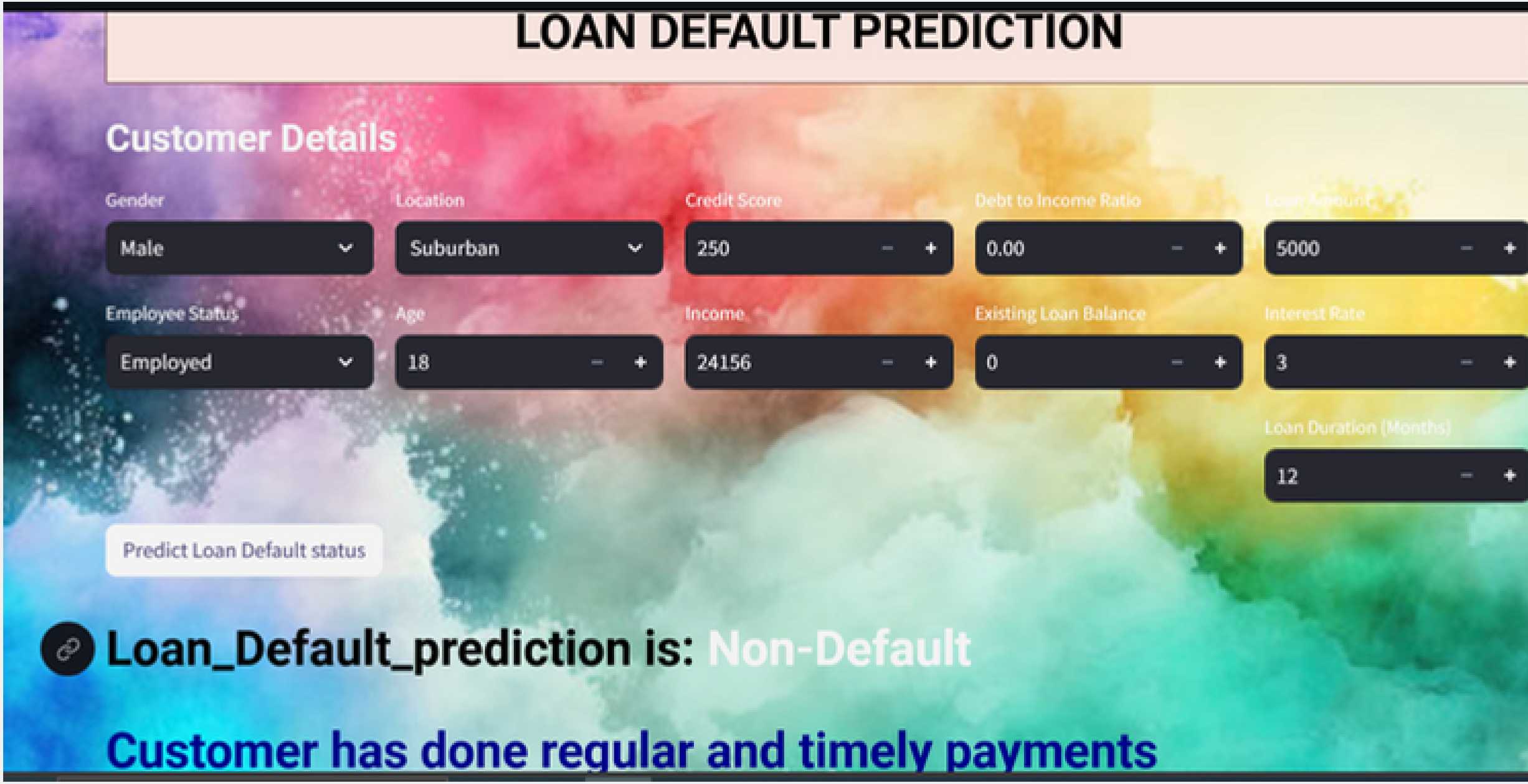
Home

Predictions

LOAN DEFAULT PREDICTION

Monitoring existing loans and identifying customers at risk of default helps banks manage and mitigate credit risk.

Age: Age can be a significant factor in determining financial behavior and risk assessment. Younger


LOAN DEFAULT PREDICTION
Customer Details
Gender
Male
Location
Suburban
Credit Score
250
Debt to Income Ratio
0.00
Loan Amount
5000
Employee Status
Employed
Age
18
Income
24156
Existing Loan Balance
0
Interest Rate
3
Loan Duration (Months)
12
Predict Loan Default status
Loan_Default_prediction is: Non-Default
Customer has done regular and timely payments

Thank You
For Your Time!