**Loan Application Status Prediction**

I am going to write about a complete end-to-end project for Loan Application Status Prediction

which should serve as a guiding path for many Data Science aspirants.

I have written down all the techniques in the form of sub-topics that I will be explaining one by one. And those sub-topics are as follows:  
  
1.      Problem Definition.  
2.      Data Analysis.  
3.      EDA Concluding Remark.  
4.      Pre-Processing Pipeline.  
5.      Building Machine Learning Models.  
6.      Concluding Remarks.

Let’s start with the problem definition or a short introduction on Loan Application Status Prediction the that I have chosen to elaborate.

**Introduction:**

* Load Application Status Prediction is a task that can be done based on historical information of the customer and bank. By checking the dataset already existed regarding the status of the Load Application and creating a model will help us to Predict the further Loan Application Status.
* Dataset includes details of applicants who have applied for loan. The dataset includes details like credit history, loan amount, their income, dependents etc.
* Will 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.

**Hardware & Software Requirements & Tools Used:**

### Hardware used:

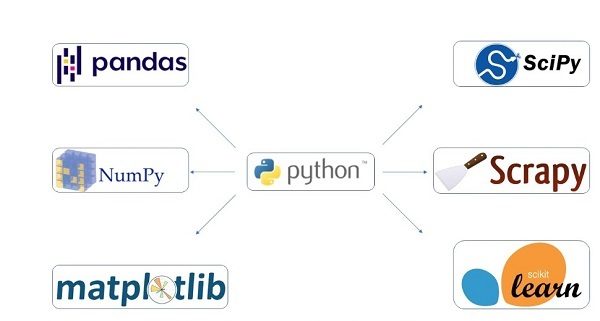
* Processor: Core i5 -10300H CPU @ 2.50GHz
* RAM: 8 GB
* Operating System: 64-bit
* ROM/SSD: 1 TB SSD
* Graphics: NVIDIA GeForce GTX 1650 Ti

**Software requirement**:

* Anaconda Navigator - Jupyter Notebook

**Libraries Used**:

* Numpy
* Pandas
* Matplotlib
* Seaborn
* Scipy
* Date Time
* Scikit Learn



Heading forward we will try to understand the problem statement and the dataset.

**1.Problem Definition**

# Problem Statement:

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

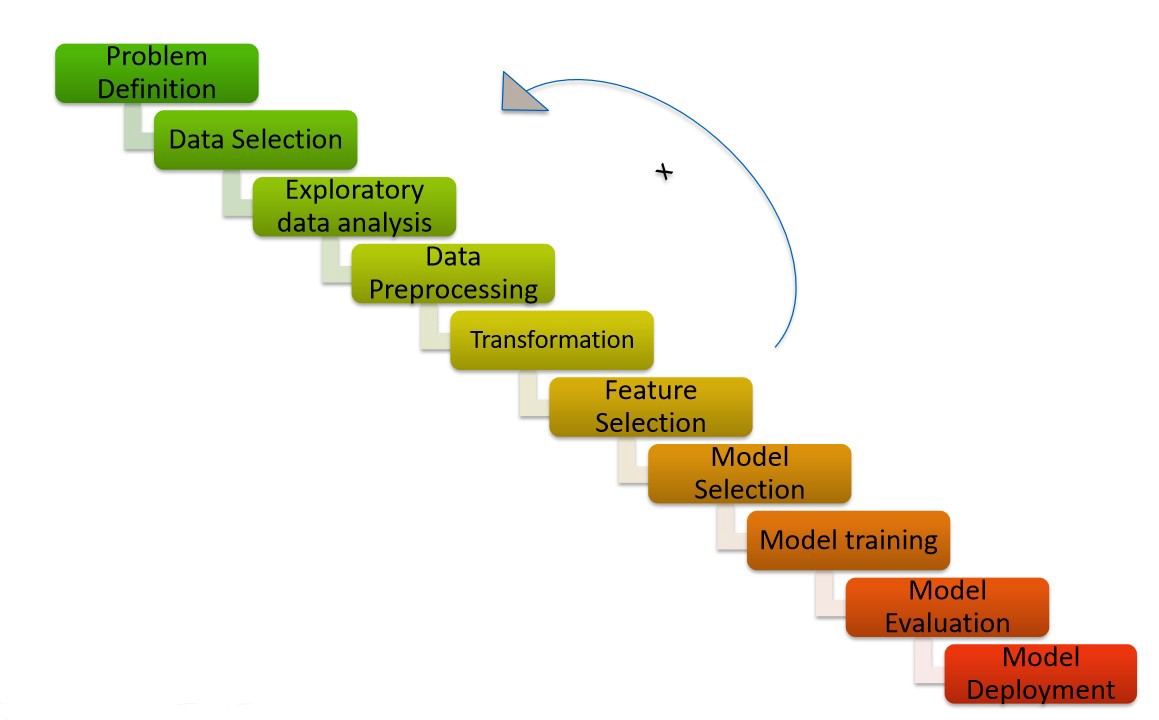
Independent Variables:

* 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

Let’s proceed step by step in the data analysis process.

**2.Data Analysis**

In order to build a Machine Learning Model, we have a Machine Learning Life Cycle that every Machine Learning Project has to touch upon in the life of the model. Let’s take a look at the model life cycle and then we will look into the actual machine learning model and understand it better along with the lifecycle as we move forward.



Now that we understand the lifecycle of a Machine Learning Model, lets import the necessary libraries and proceed further.

**Importing the necessary Libraries:**

To analyze the dataset or even to import the dataset, we have imported all the necessary libraries as shows below.

Pandas has been used to import the dataset and also in creating data frames.

Numpy has been used for numerical tasks.

Seaborn and Matplotlib have been used for Data Visualization.

Date Time has been used to extract day/month/date separately.

Scipy has been used in the Zscore method for removing outliers.

Sklearn has been used in the model building.

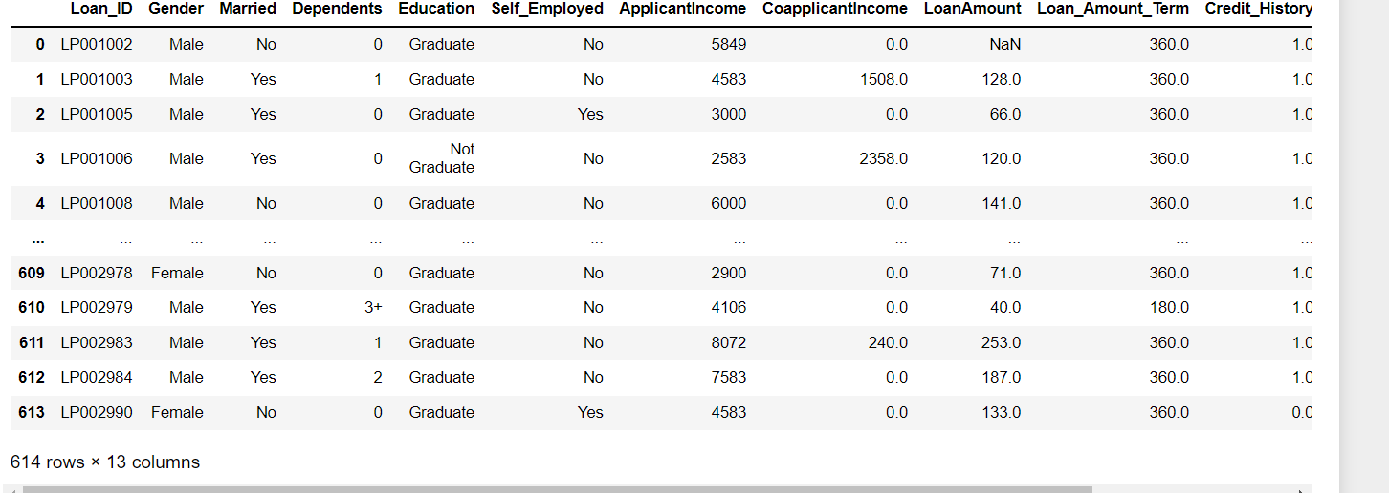
**Importing the Dataset**

Let’s import the dataset first.



Copied the raw data and saved it as a csv file on my local computer after which I imported the entire dataset on this Jupyter Notebook with the help of pandas.

I have imported the dataset which was in “csv” format as “df”. Below is how the dataset looks.



**3.Exploratory Data Analysis (EDA)**

Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods. It helps determine how best to manipulate data sources to get the answers you need, making it easier for data scientists to discover patterns, spot anomalies, test a hypothesis, or check assumptions.

EDA is primarily used to see what data can reveal beyond the formal modeling or hypothesis testing task and provides a provides a better understanding of data set variables and the relationships between them. It can also help determine if the statistical techniques you are considering for data analysis are appropriate. Originally developed by American mathematician John Tukey in the 1970s, EDA techniques continue to be a widely used method in the data discovery process today.

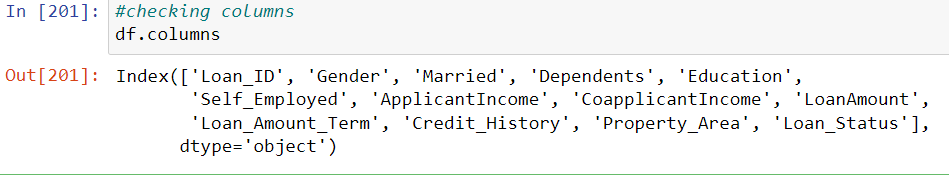
As per the lifecycle of the machine learning model we have already completed points 1 and 2. Now let’s move on to the point 3, 4, 5 and 6 which is the most crucial part of any machine learning model. If we prepare the dataset, analyze it and clean it in the best way possible the better model accuracy we will get, or the model can get over fitted or under fitted. We will discuss further all the steps that are used.

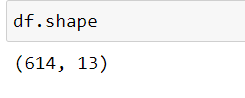
**Data Preparation:**

In this part we will firstly be exploring the data with some basic steps and then further proceed with some crucial analysis, like feature extraction, imputing and encoding.

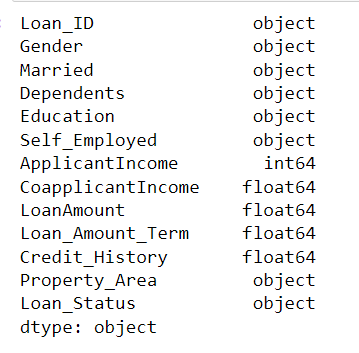
Let’s start with checking shape, unique values, value counts, info etc.

After doing the analysis if we find any unnecessary columns in the dataset, we can drop those column.

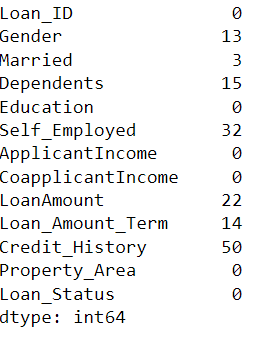




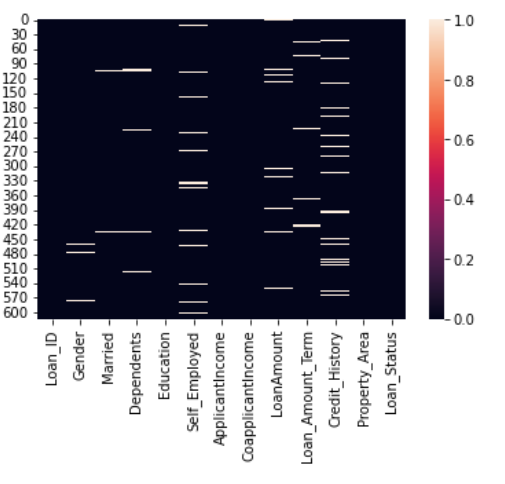




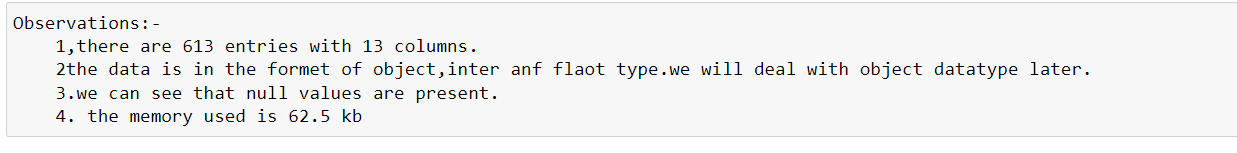


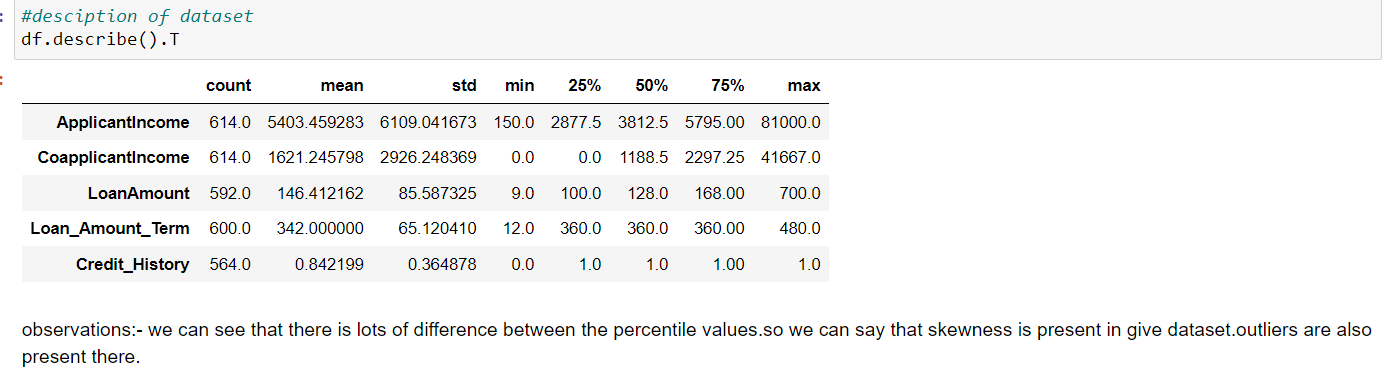


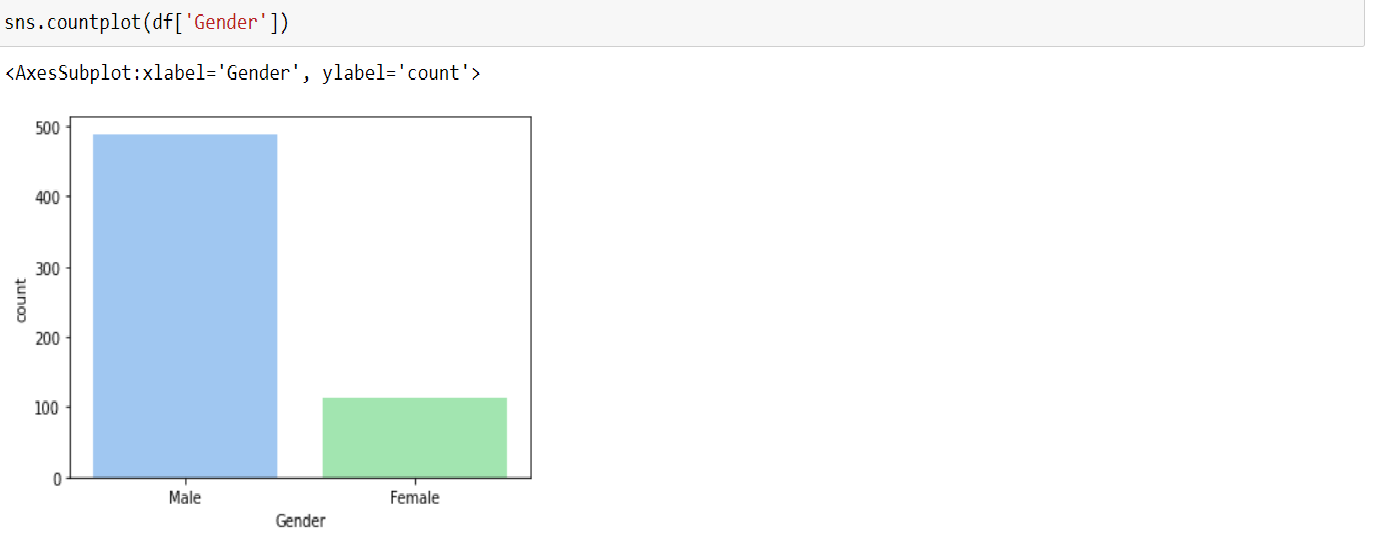


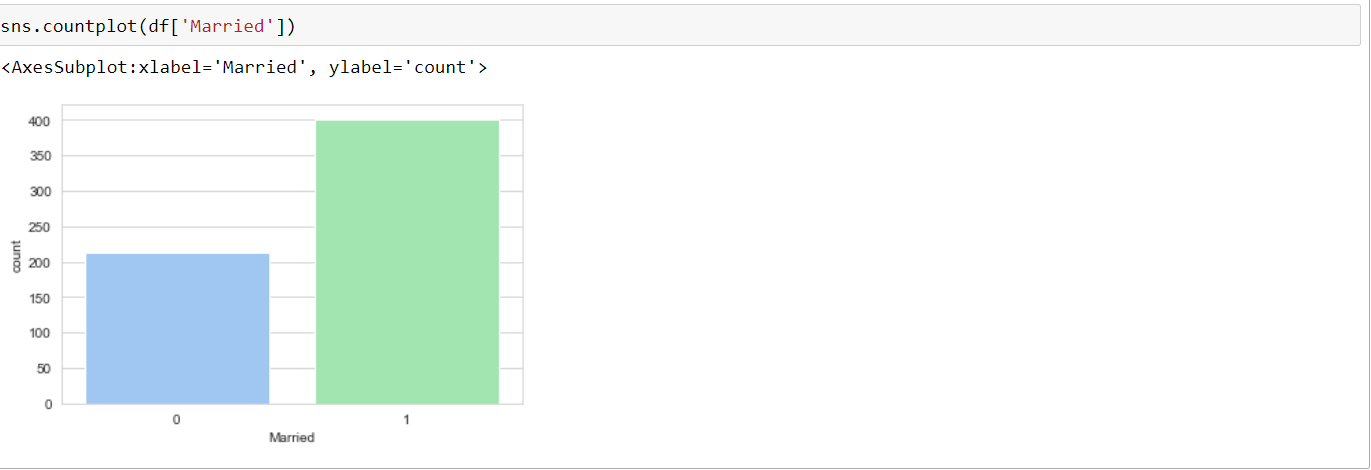


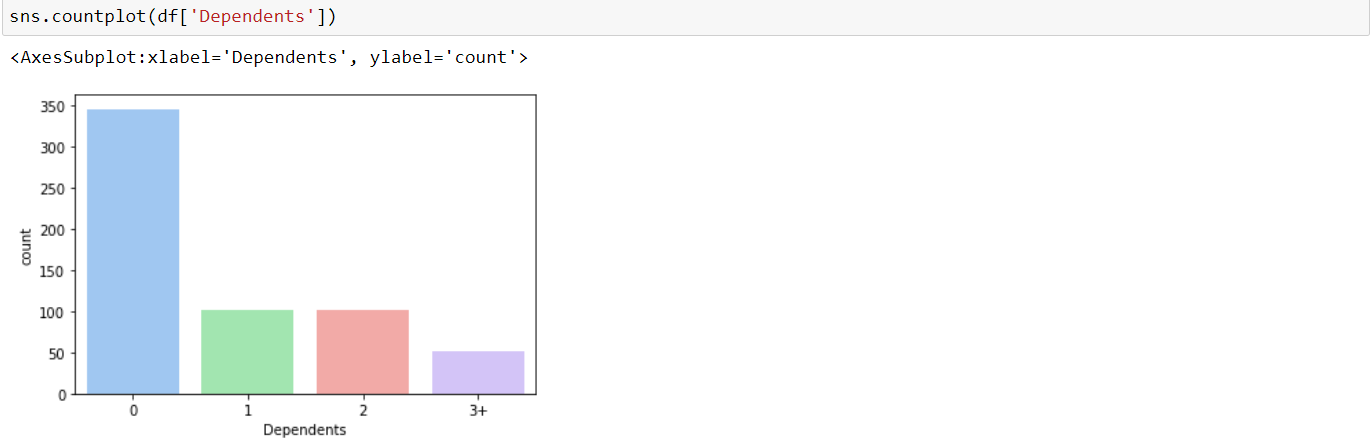


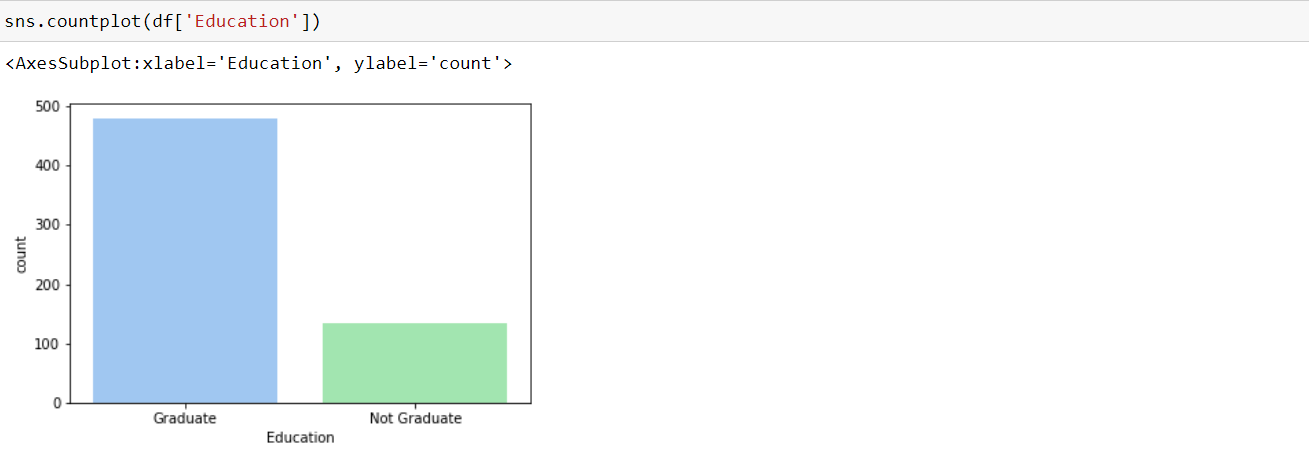


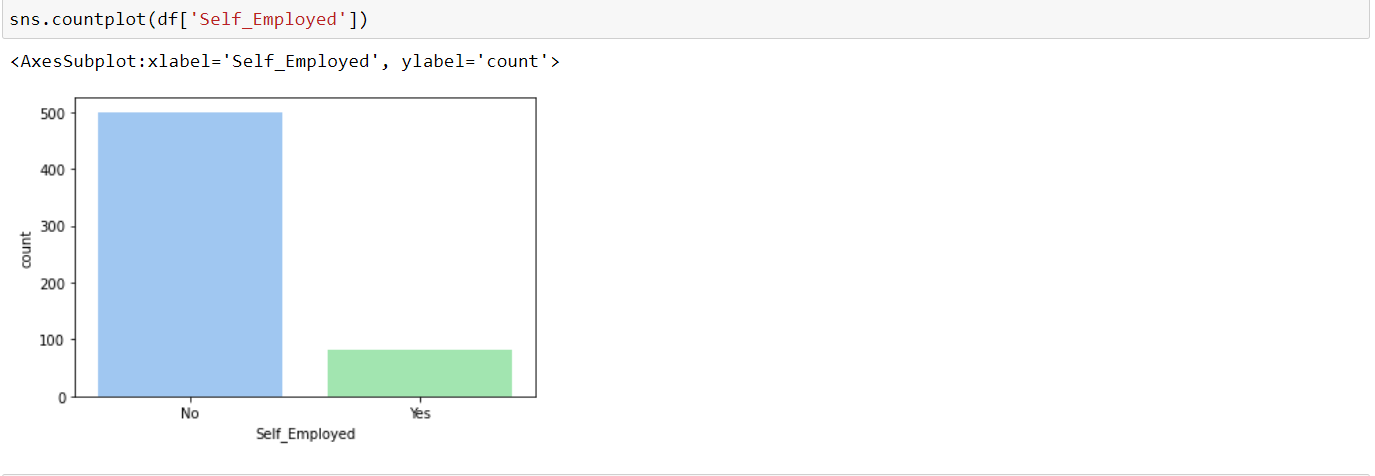


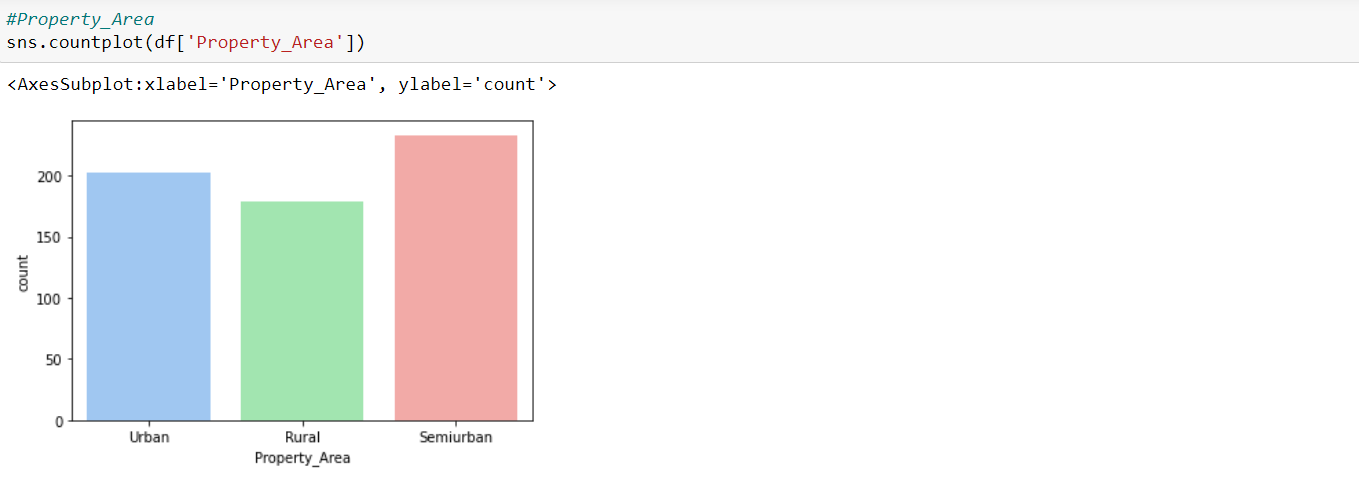


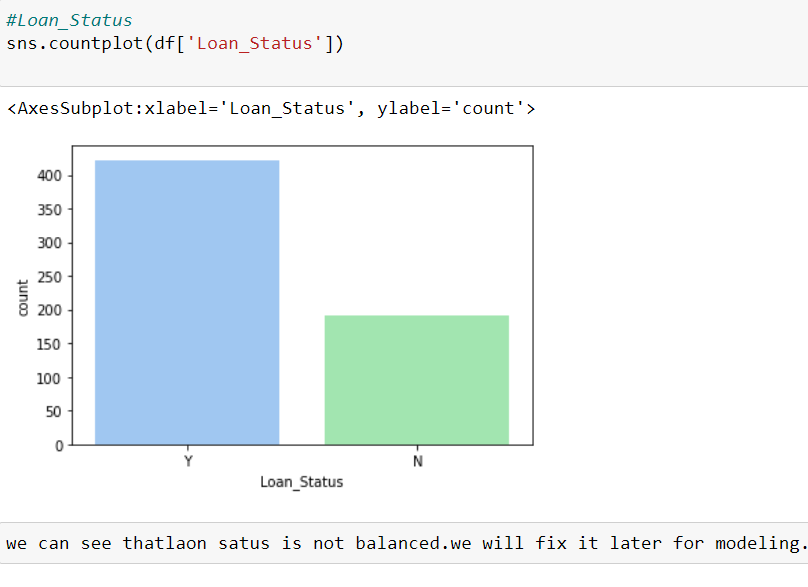




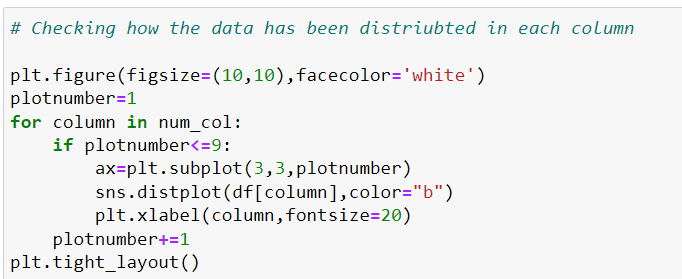


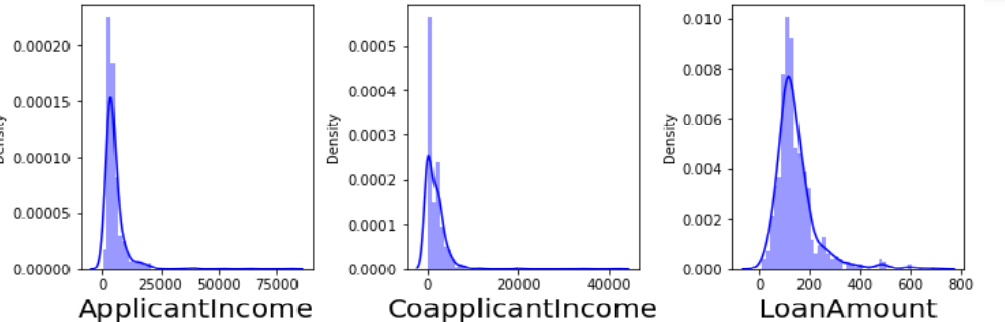


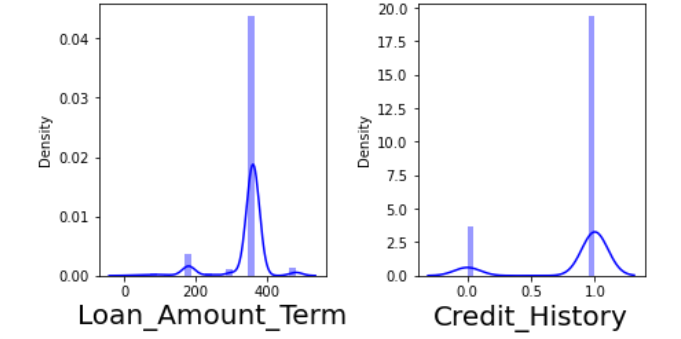


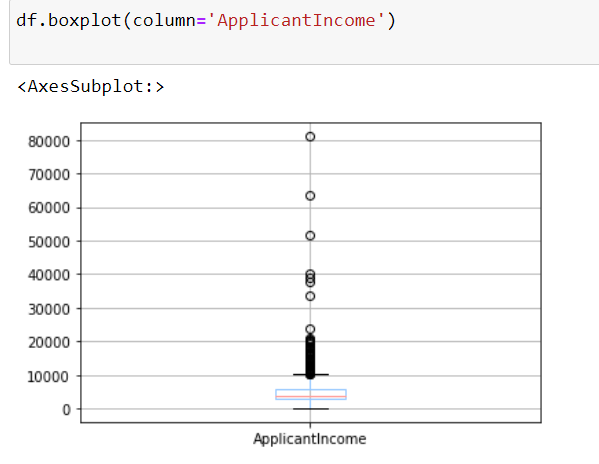


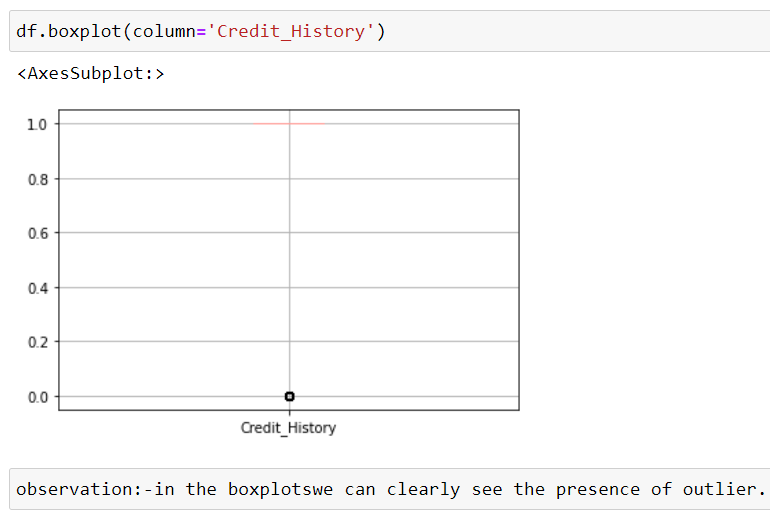


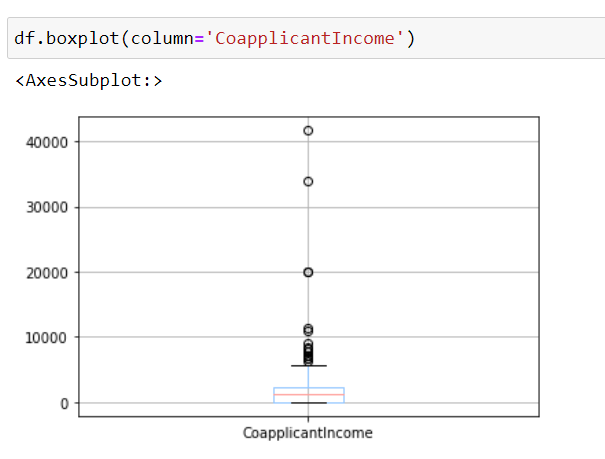




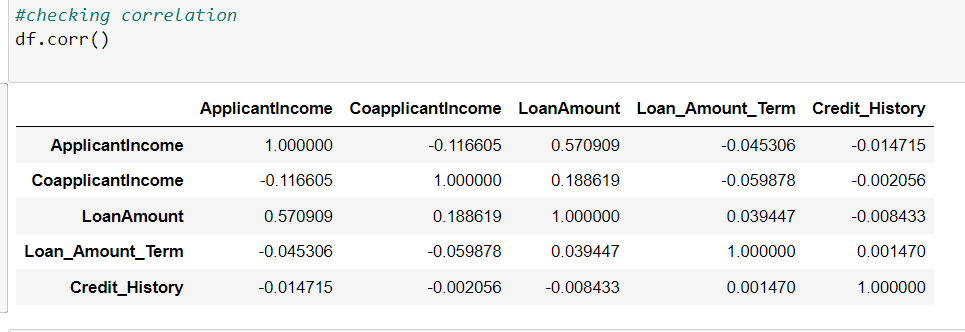


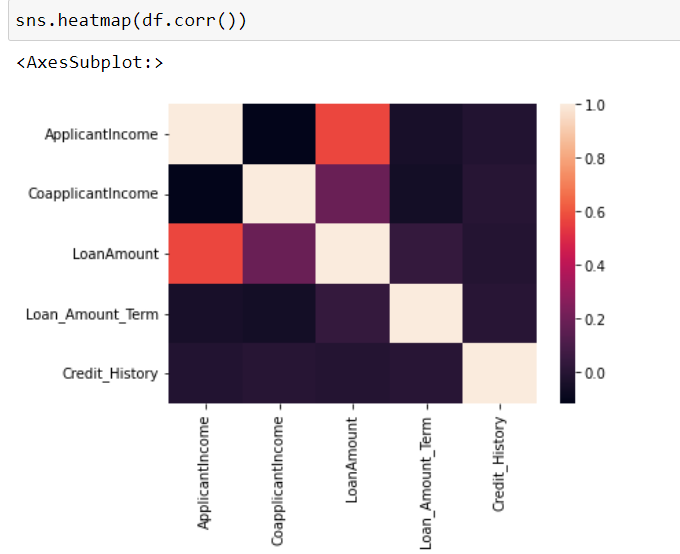






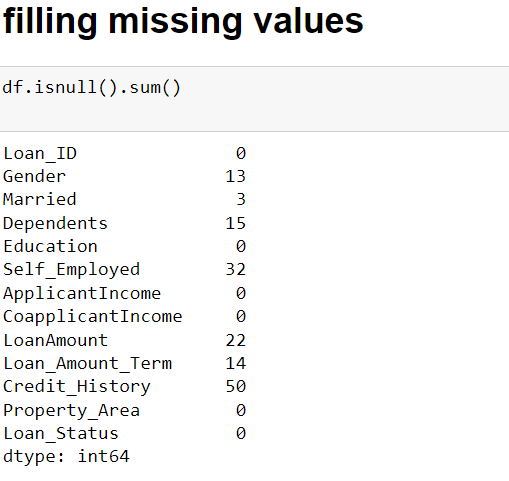


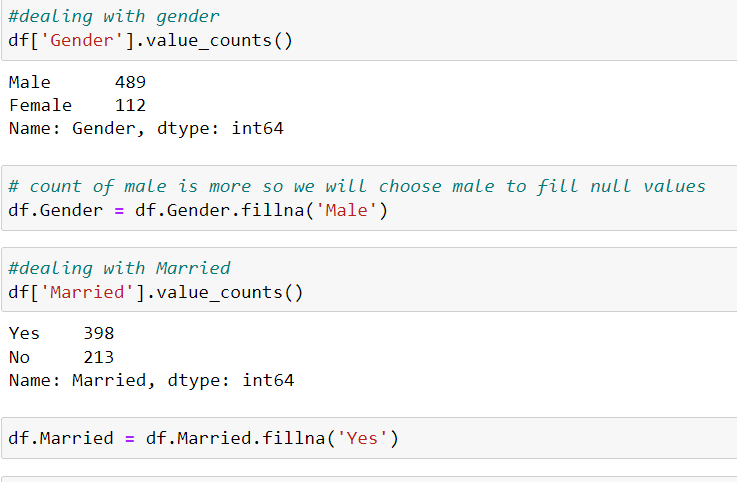


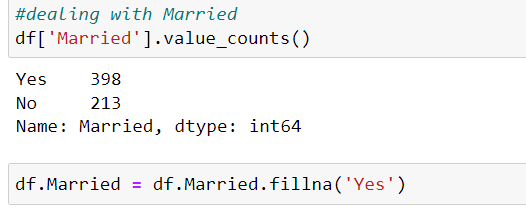


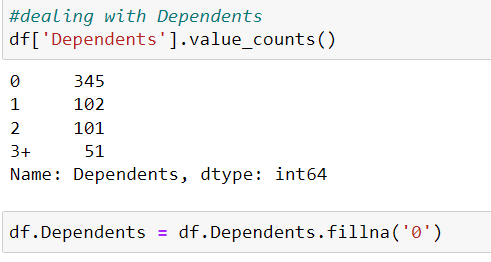
**4.Pre-Processing Pipeline**

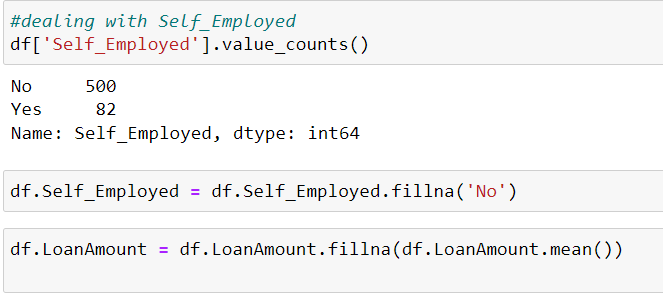
Data preprocessing in Machine Learning is a crucial step that helps enhance the quality of data to promote the extraction of meaningful insights from the data. Data preprocessing in Machine Learning refers to the technique of preparing (cleaning and organizing) the raw data to make it suitable for a building and training Machine Learning models. In simple words, data preprocessing in Machine Learning is a [data mining technique](https://www.upgrad.com/blog/most-common-examples-of-data-mining/) that transforms raw data into an understandable and readable format.

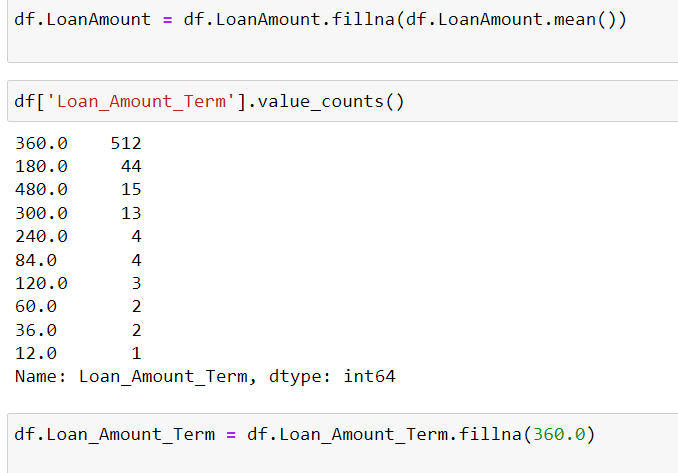


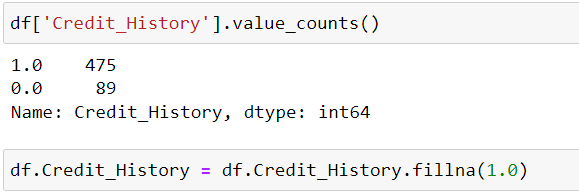


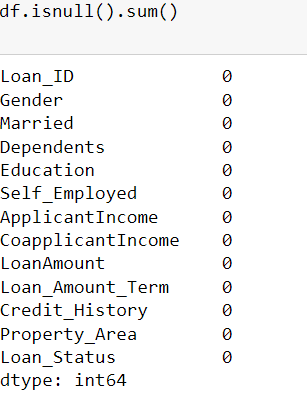


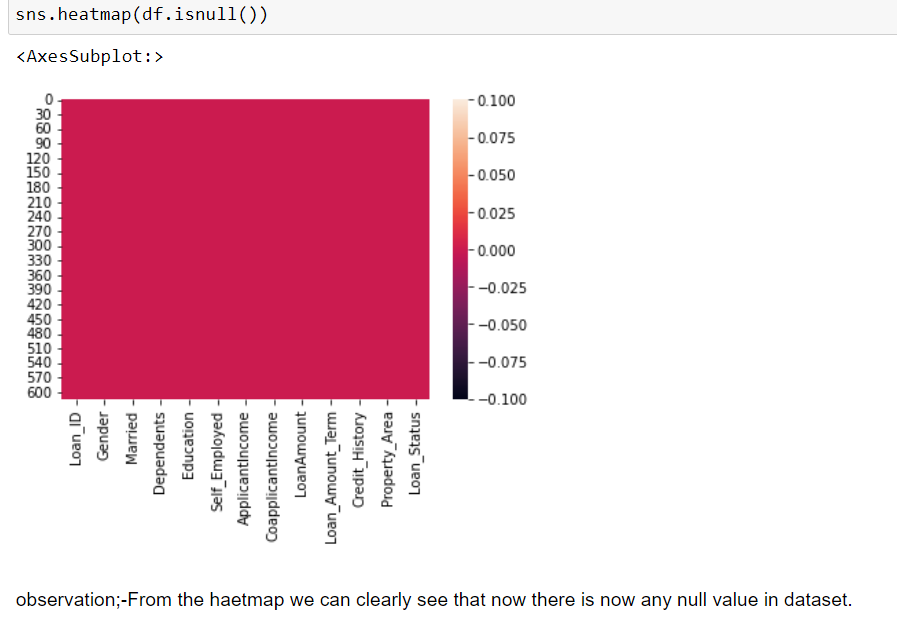


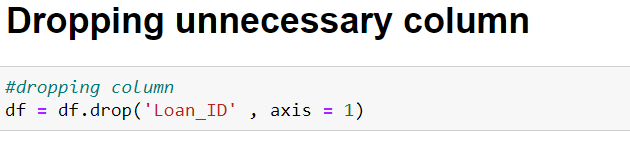






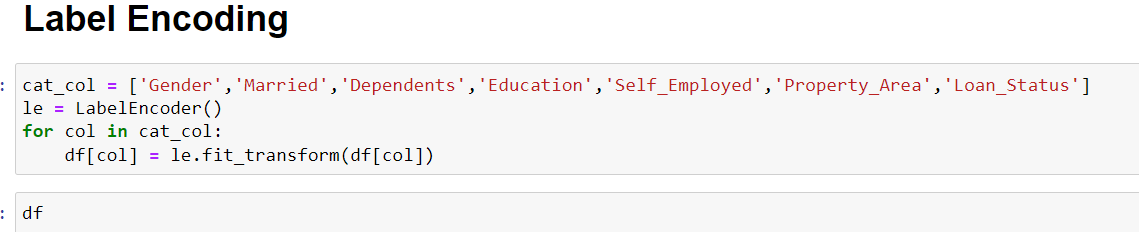


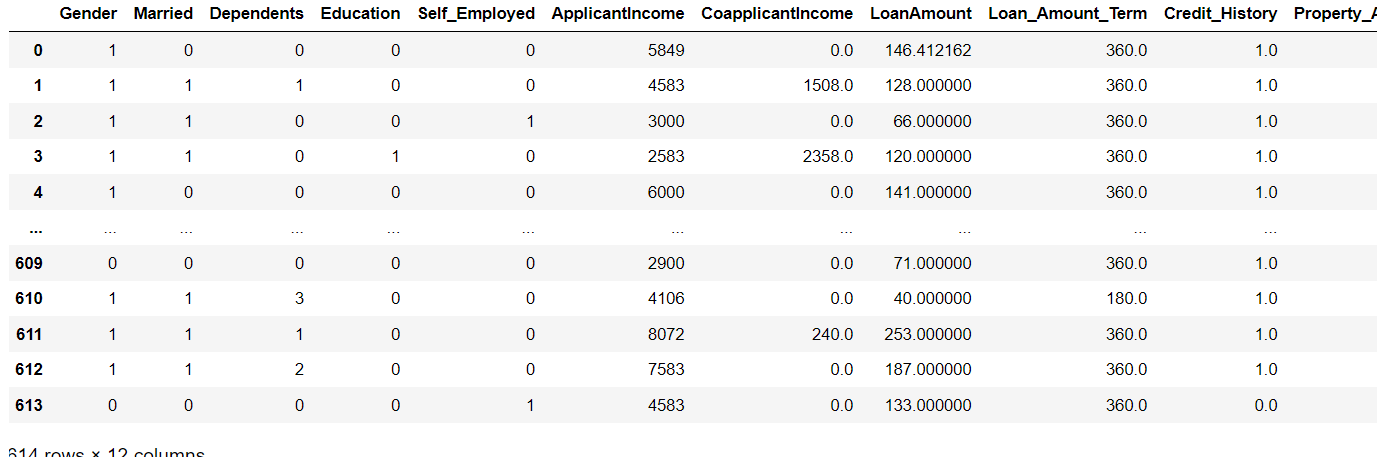




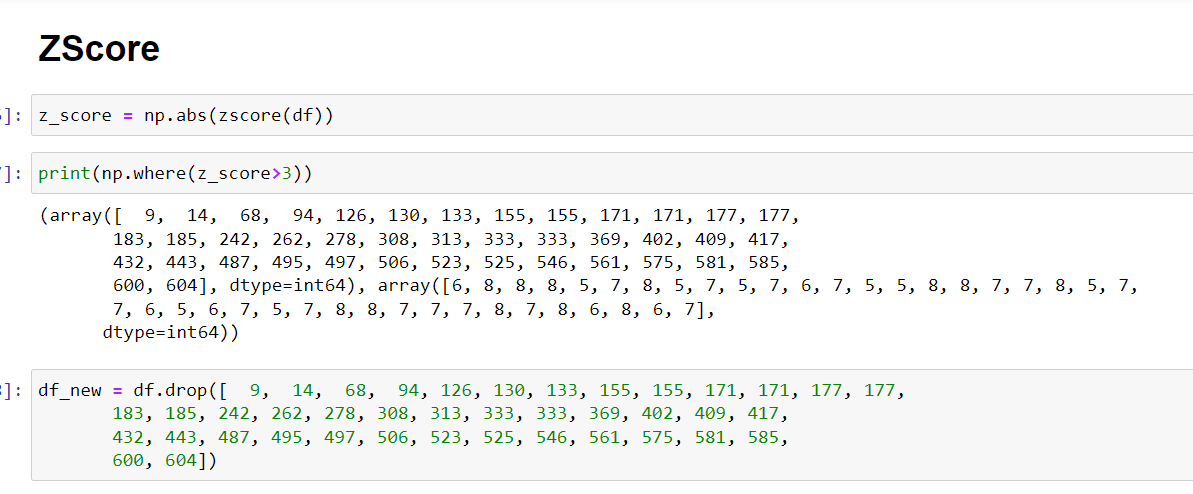
**Label Encoding**

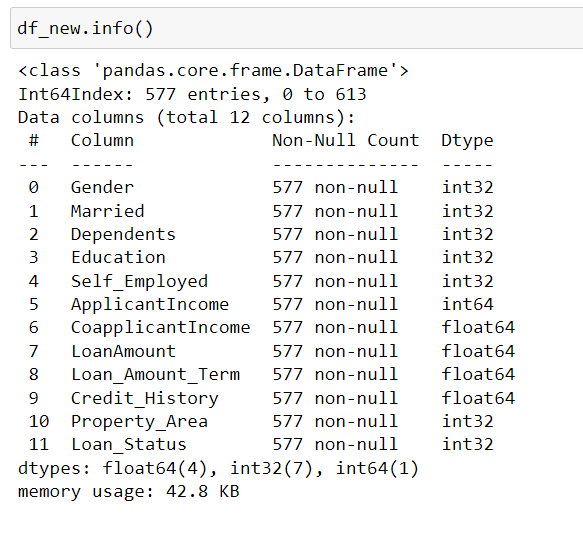
Label Encoding refers to converting the labels into a numeric form so as to convert them into the machine-readable form. Machine learning algorithms can then decide in a better way how those labels must be operated. It is an important pre-processing step for the structured dataset in supervised learning

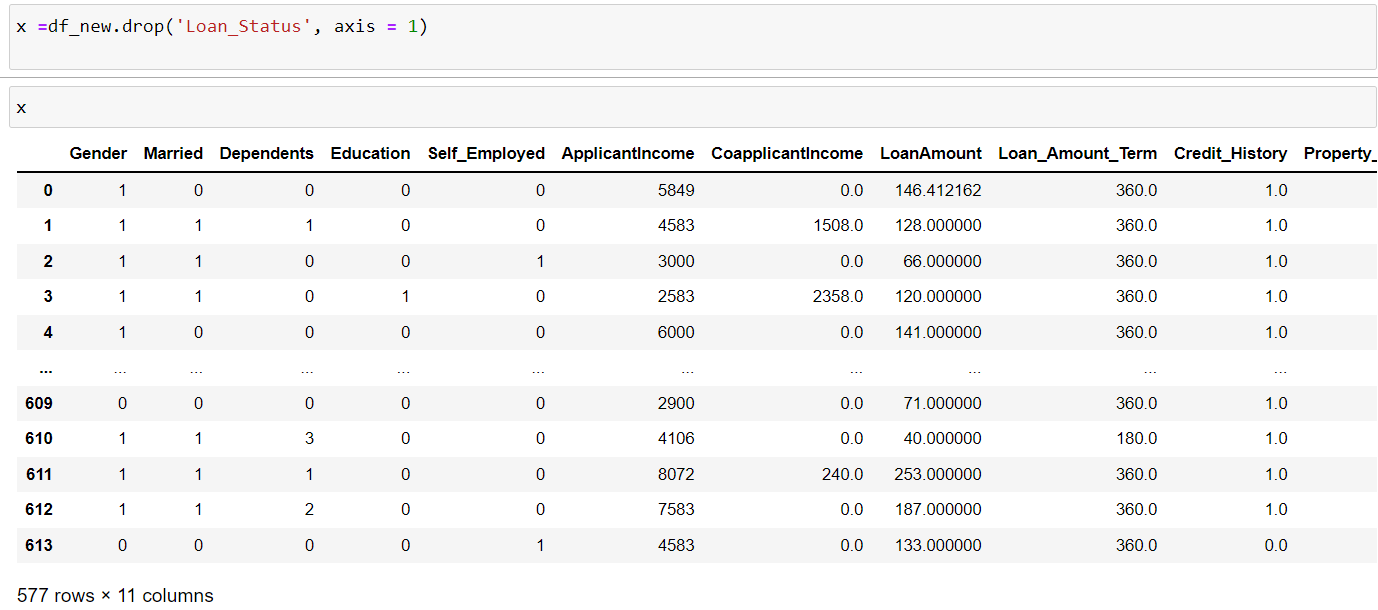


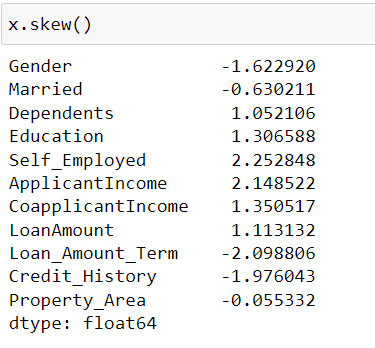


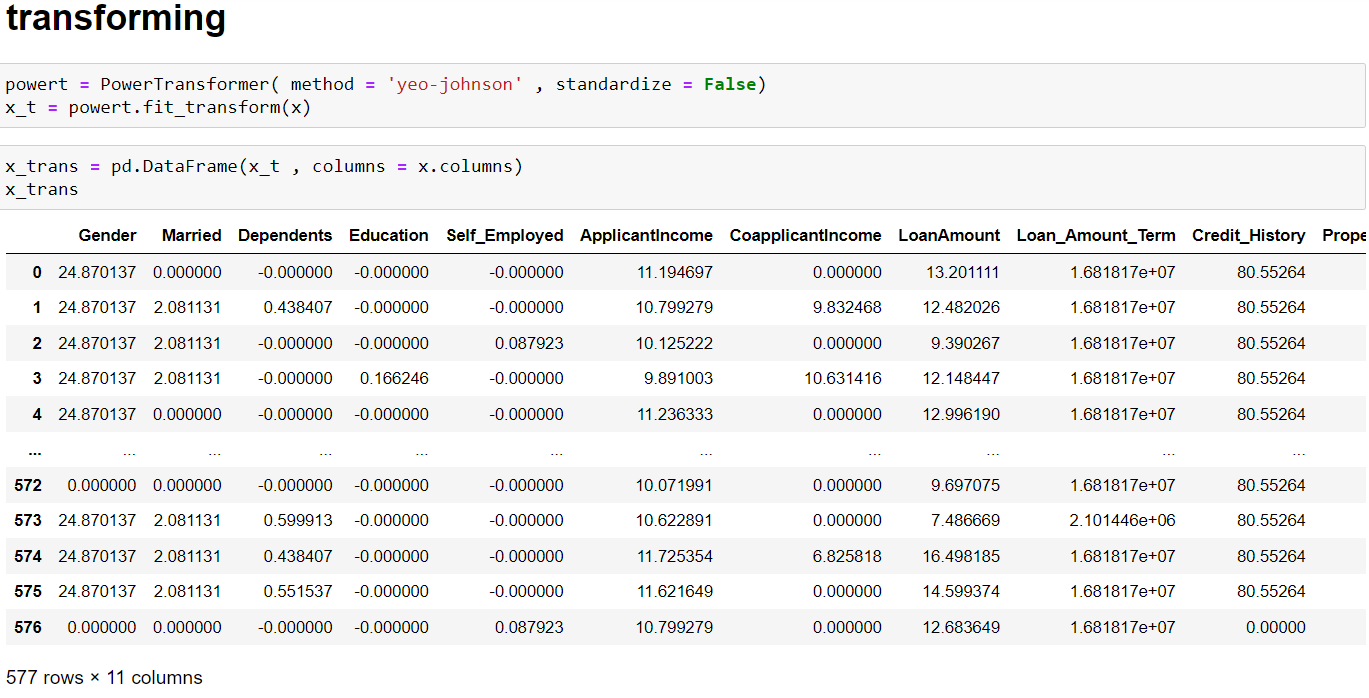
**Zscore:-** It is a strategy of normalizing data that avoids this outlier issue. In this technique, values are normalized based on the mean and standard deviation of the data.  The essence of this technique is the data transformation by the conversion of the values to a common scale where an average number/mean equals zero and a standard deviation is one. Technically, it measures the standard deviations below or above the mean.

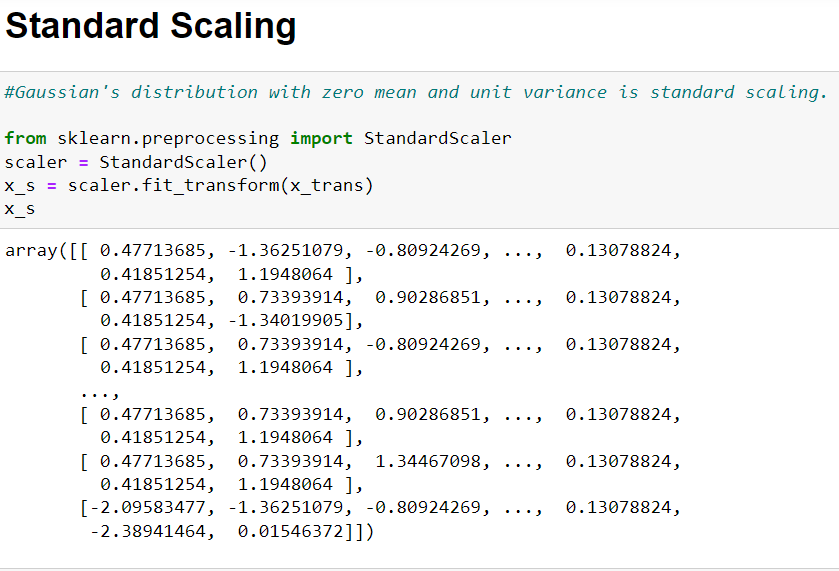


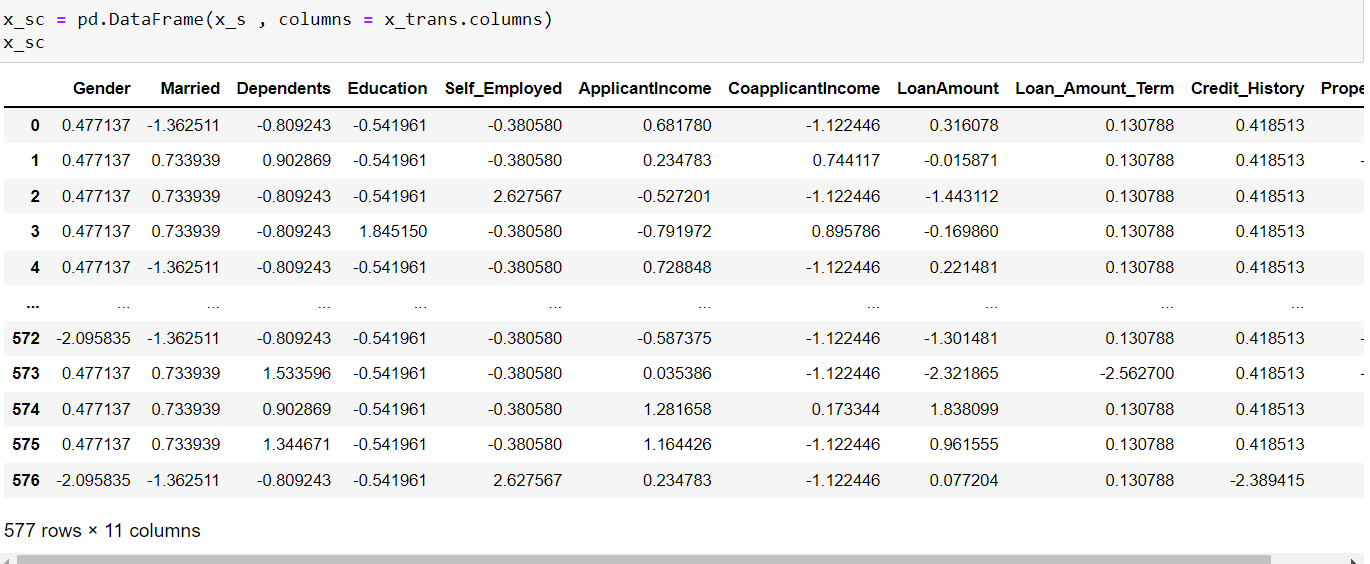


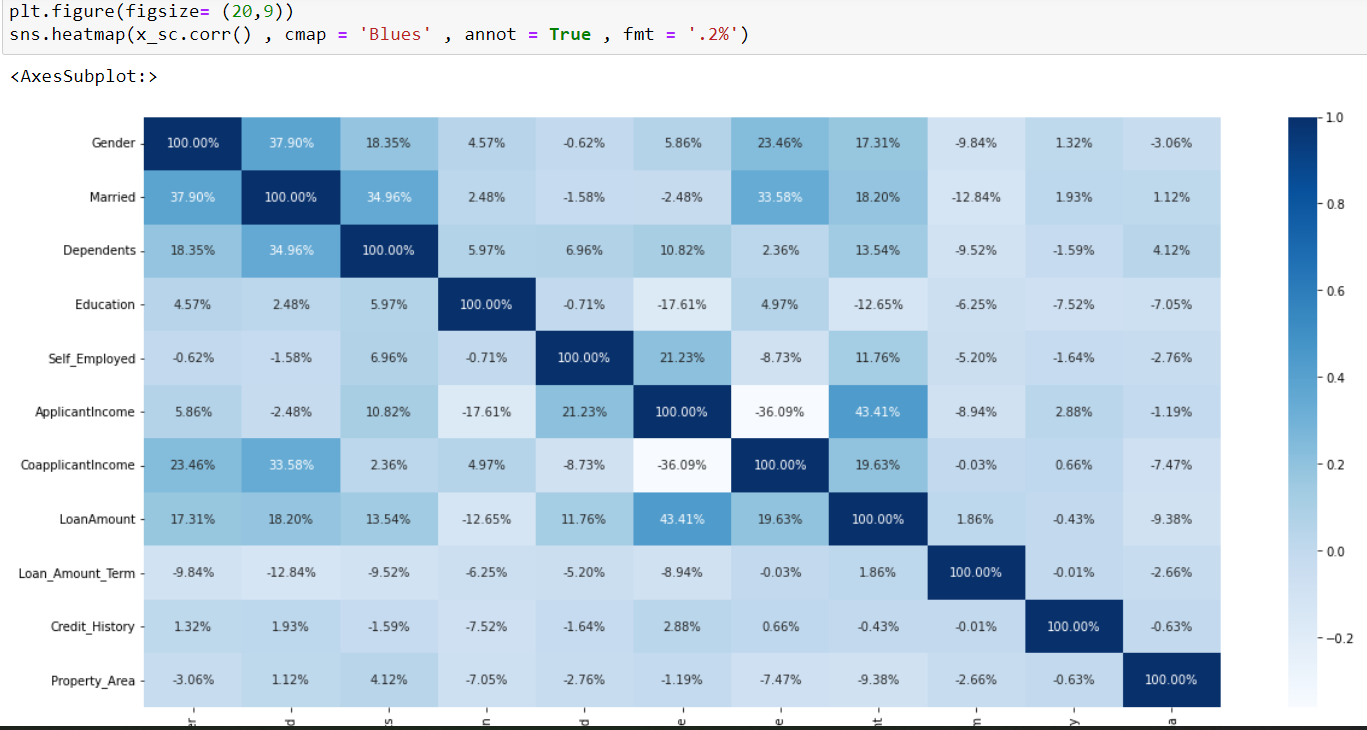


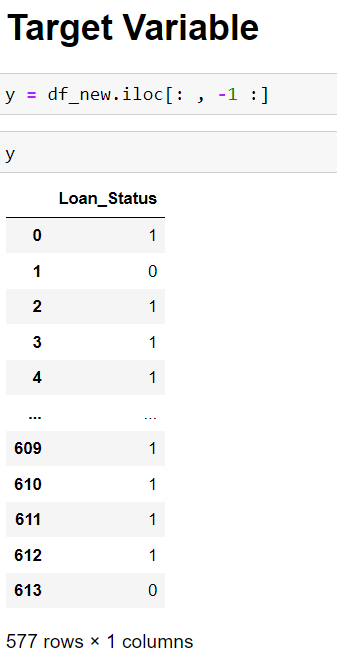


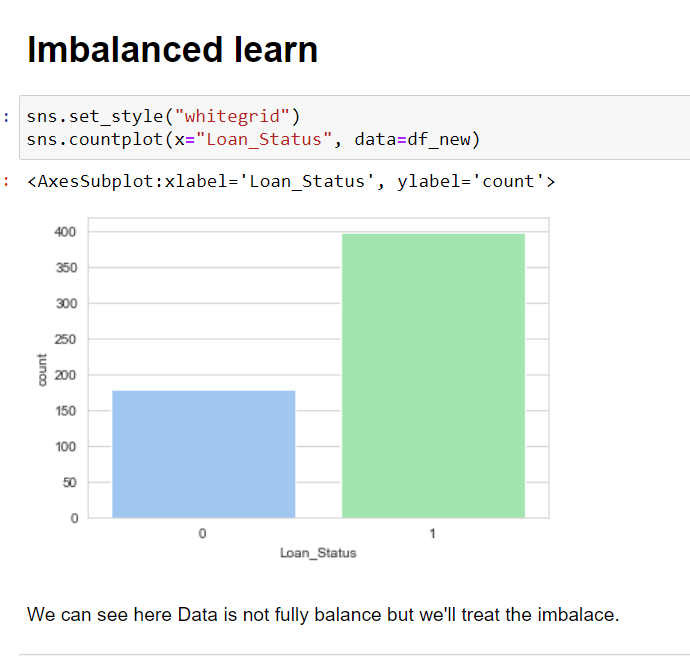


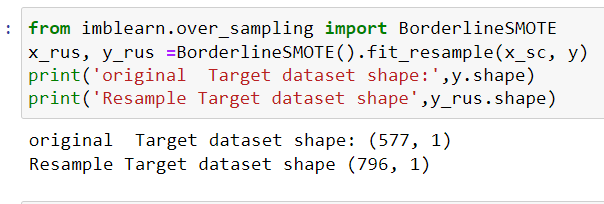






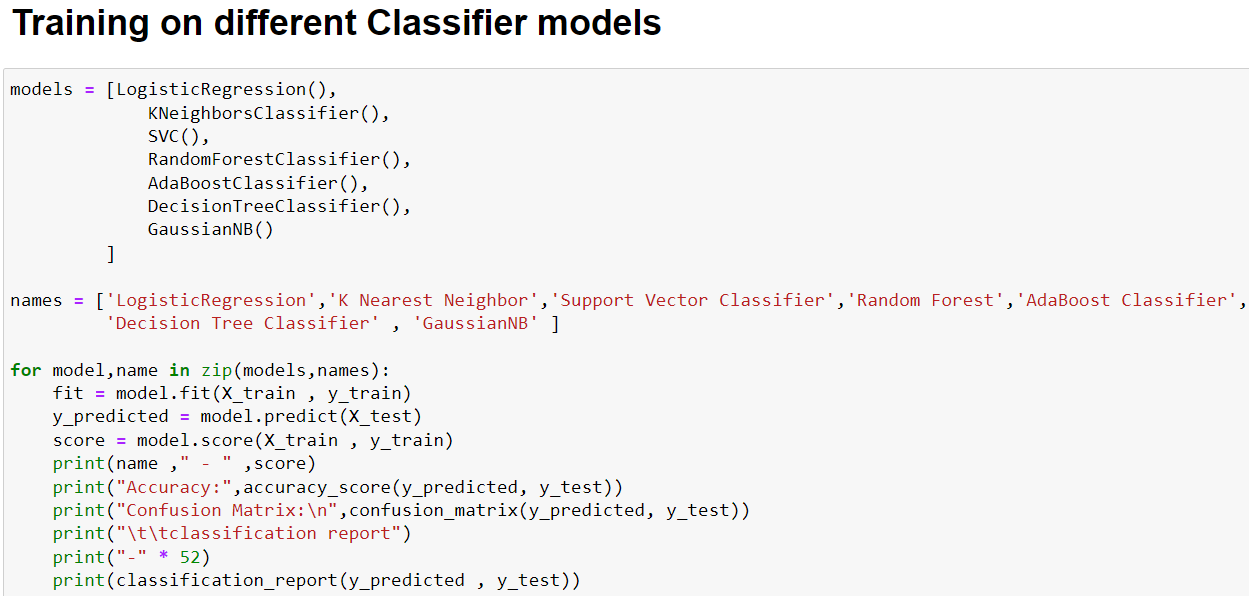


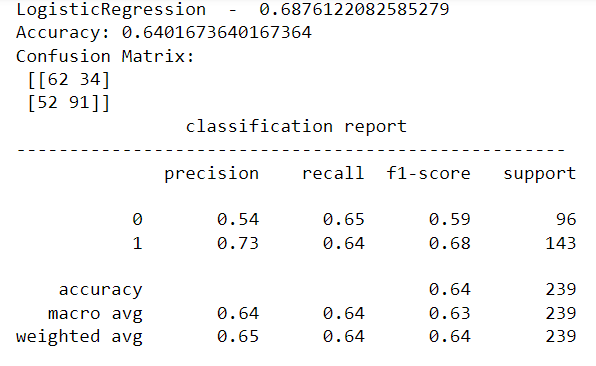


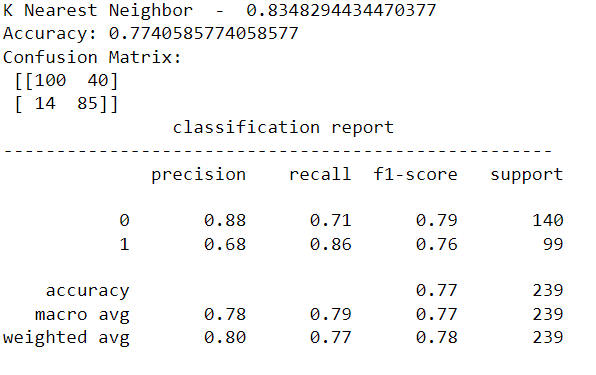


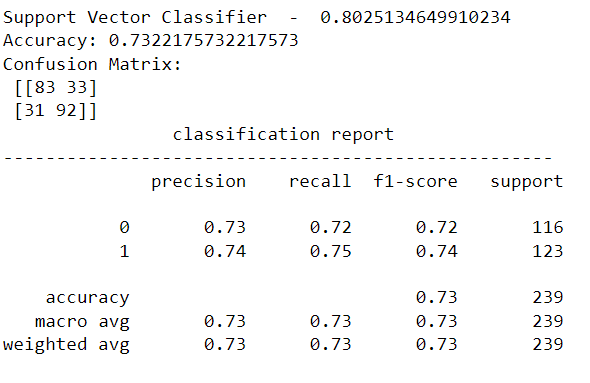
**5.Building Machine Learning Models.**

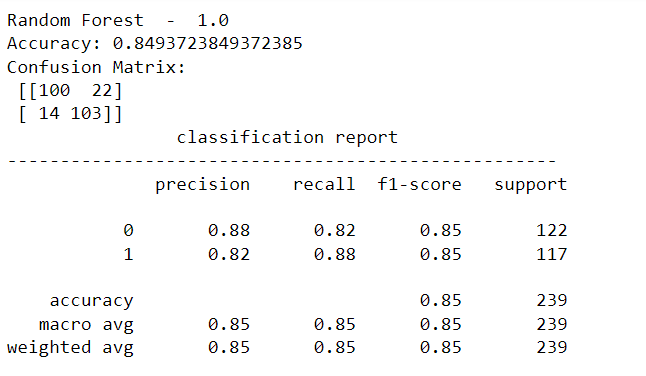


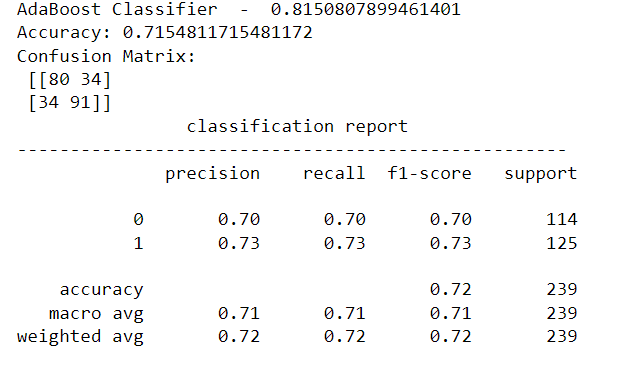


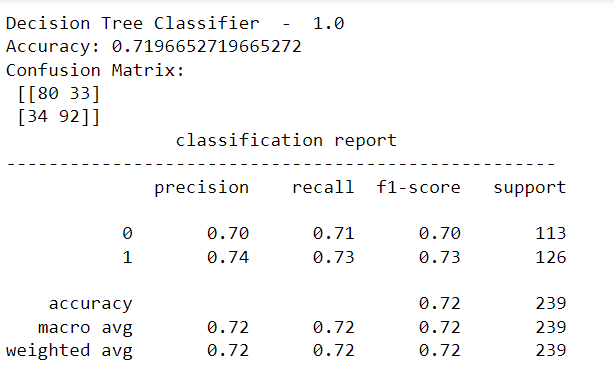


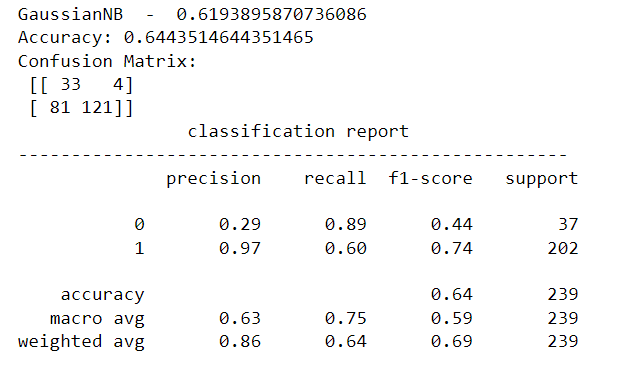






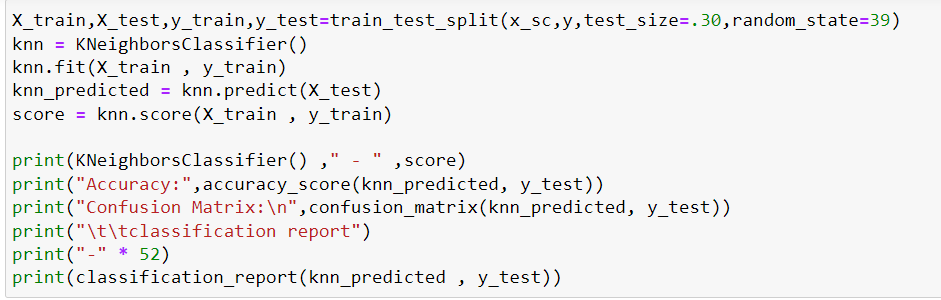


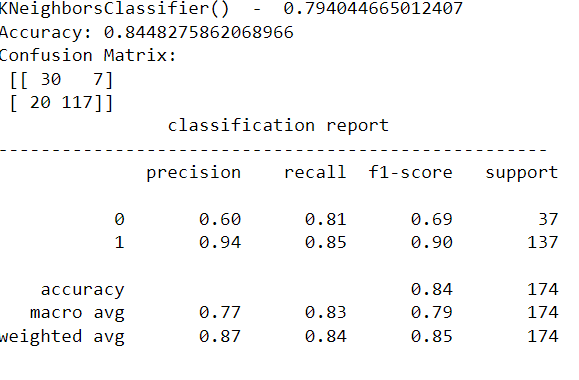


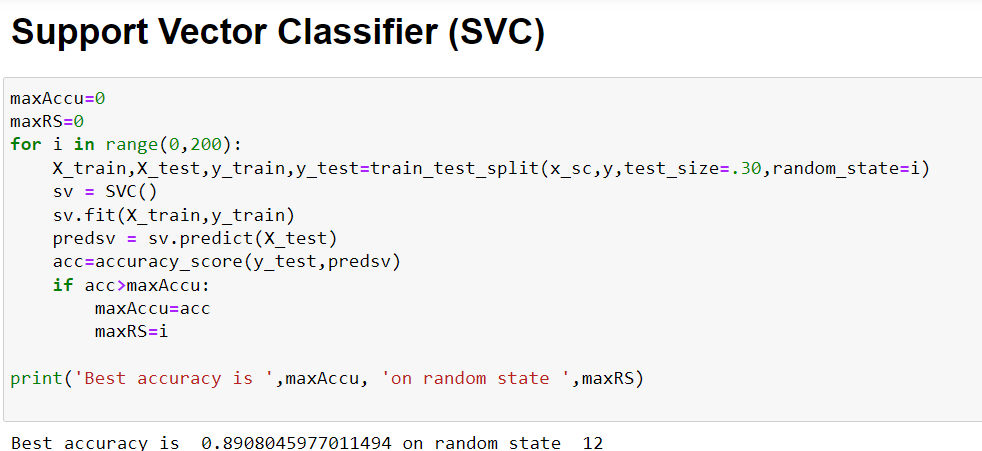


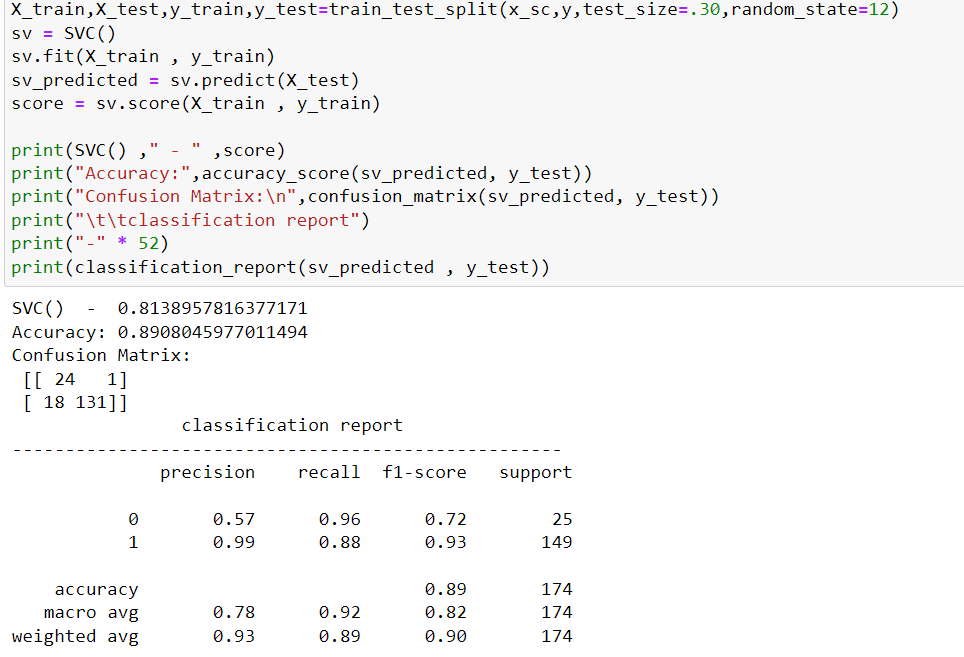
.



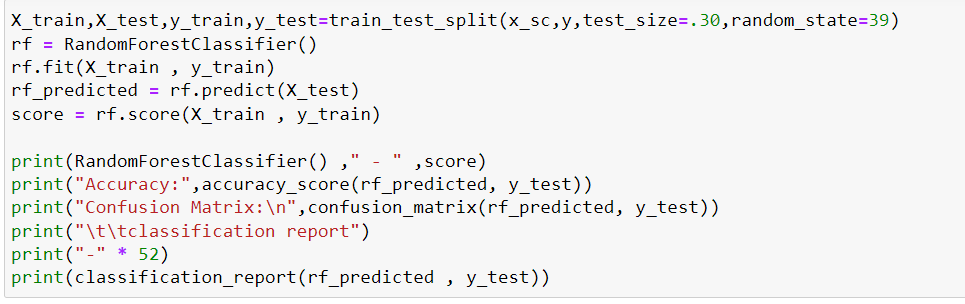


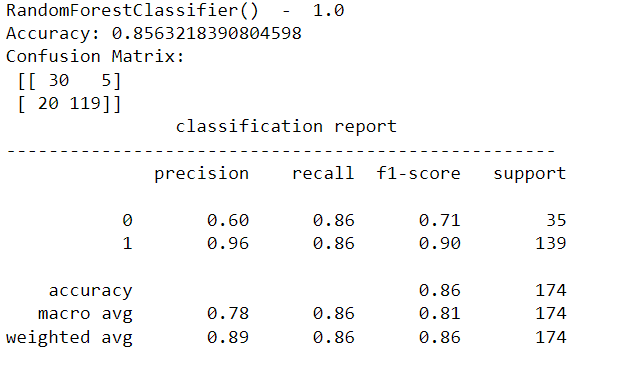




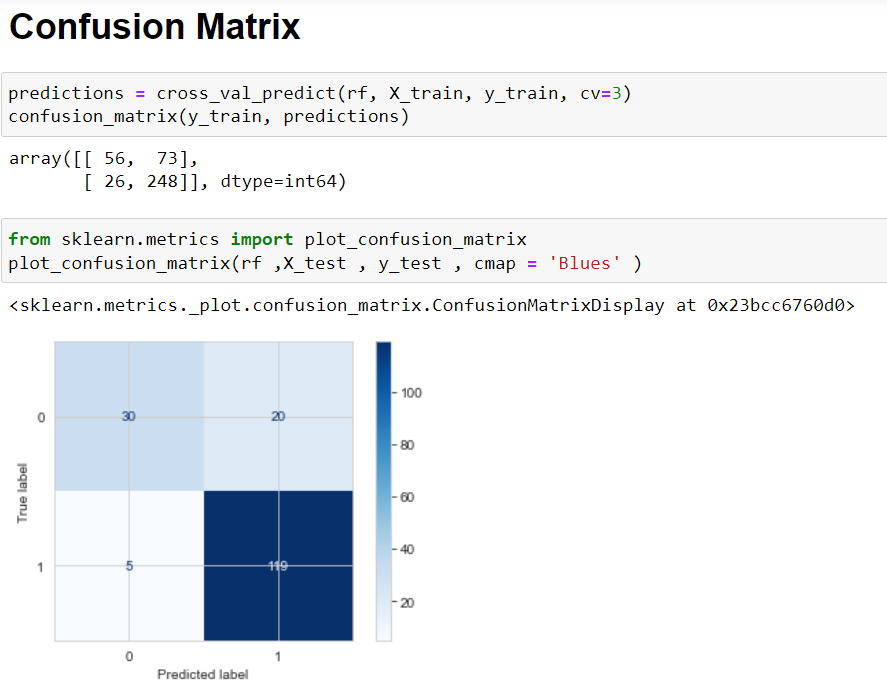


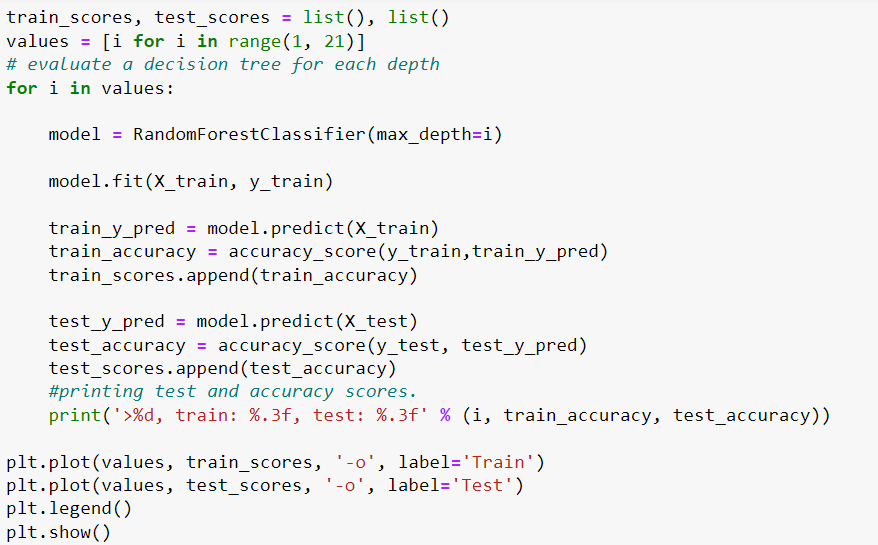


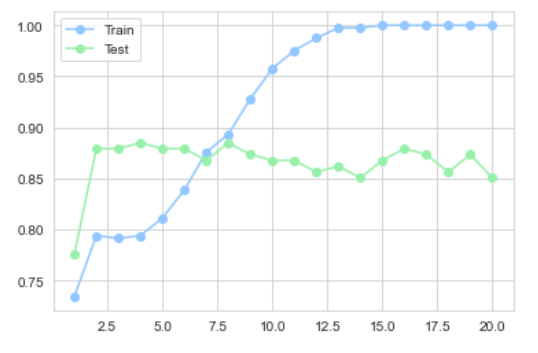


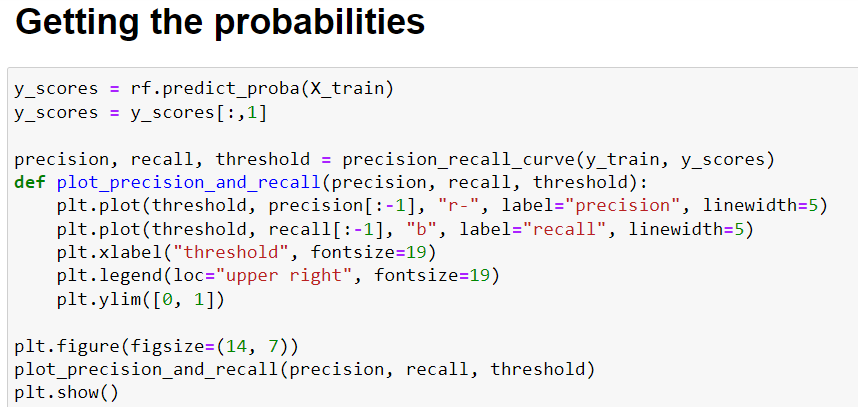


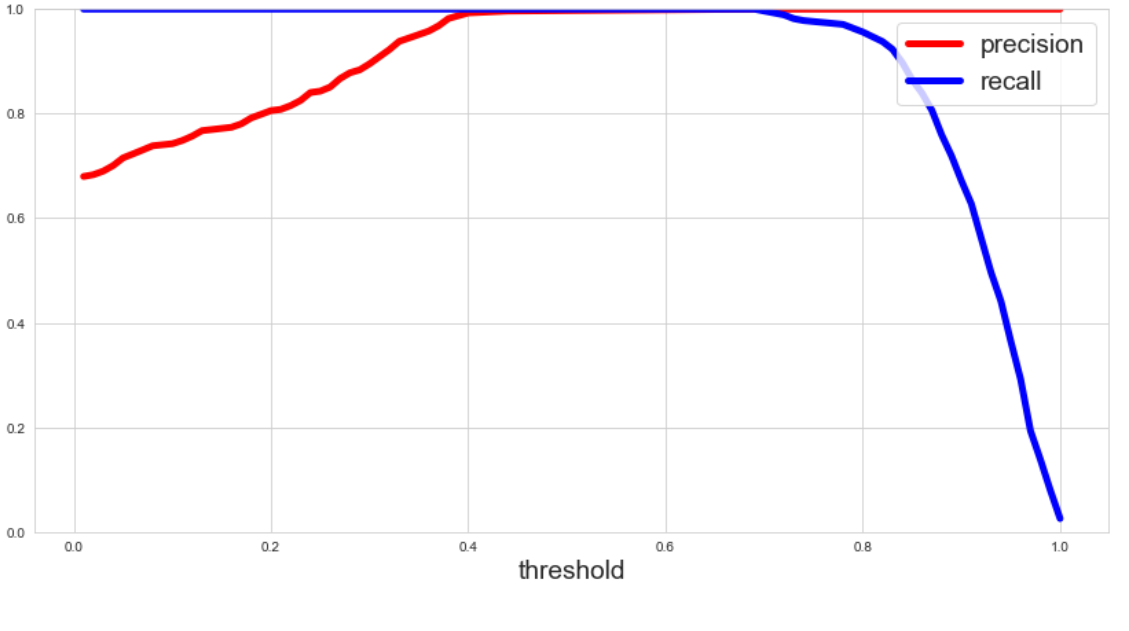
I am choosing random forest classifier for modelling a loan prediction model due to its higher score value.

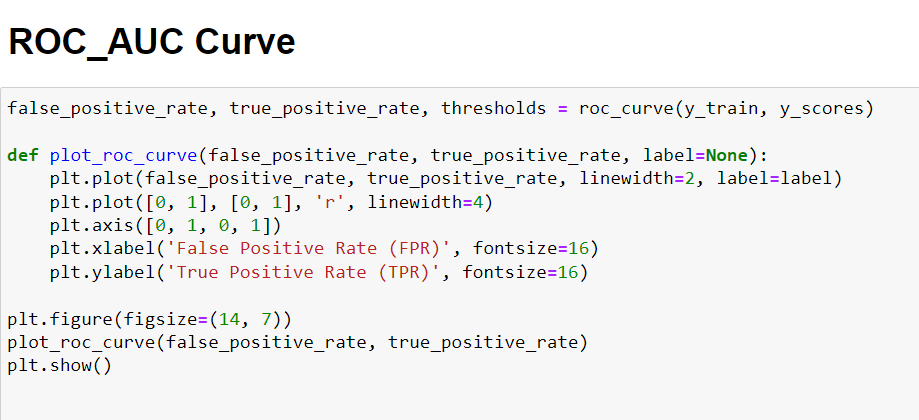


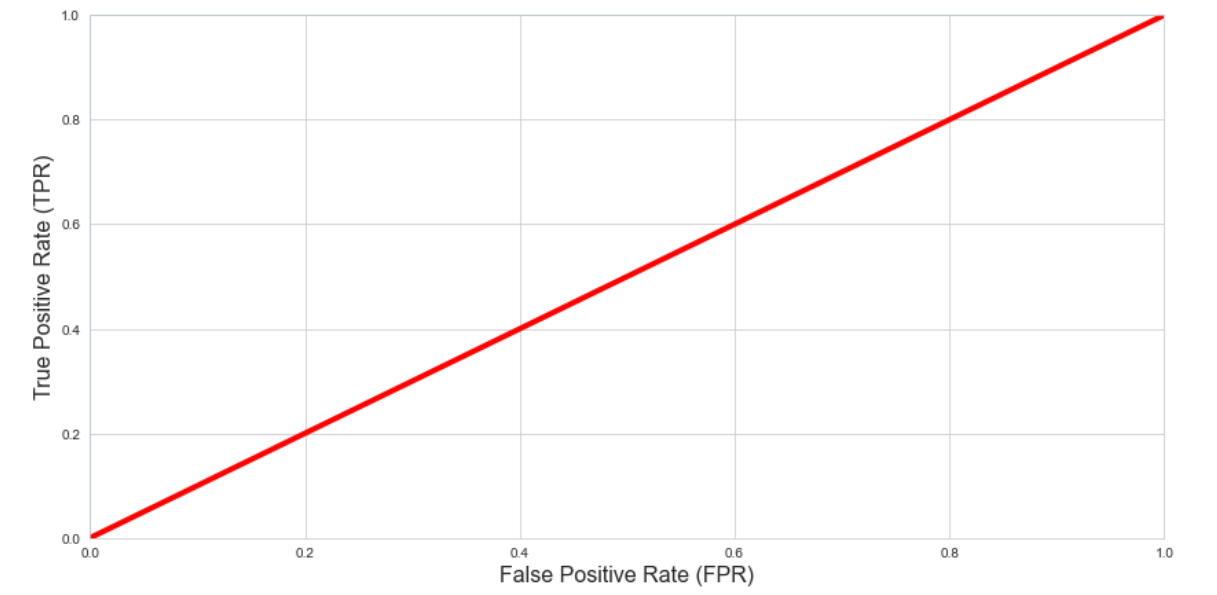














1. **Concluding Remarks**

In the beginning of the blog, we have discussed about the lifecycle of a Machine Learning Model, you can see how we have touched based on each point and finally reached up to the model building and made the model ready for deployment.

Let’s take a quick recap on all the steps that we went through starting from understanding the Problem Definition then going through the Data Analysis and EDA processes. We went through the necessary Pre-processing Data steps before the final Building Machine Learning Models step came into picture.

Iin every model building problem Data Analysis and Feature Engineering is the most crucial part.

You can see how we have handled numerical and categorical data and also how we build different machine learning models on the same dataset.

Using this machine Learning Model, we can predict whether the Loan is give or not .