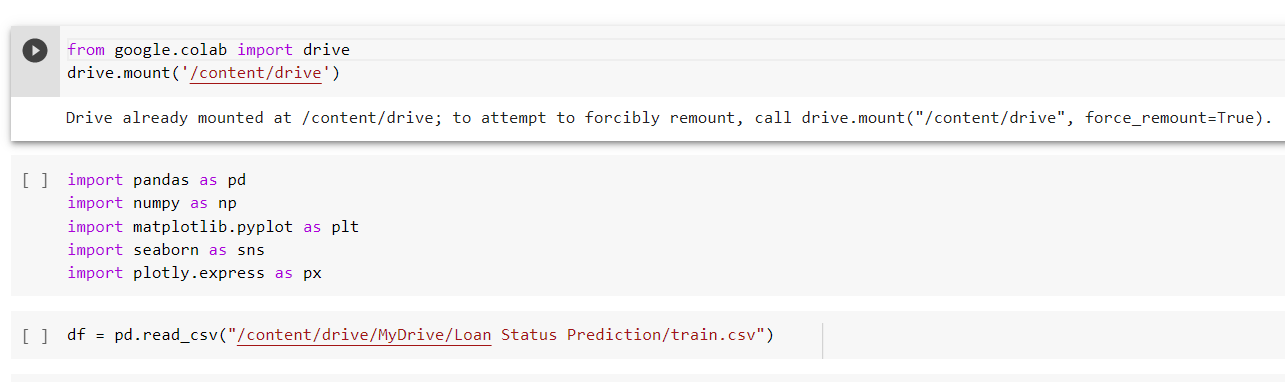
Aim:

* Loans are the core business of banks. The main profit comes directly from the loan’s interest. The loan companies grant a loan after an intensive process of verification and validation. However, they still don’t have assurance if the applicant is able to repay the loan with no difficulties.
* we'll build a predictive model to help determine if an applicant will be able to repay the lending company or not. We will prepare the data using Python and use various models to predict the target variable.

Table of Contents:

* Loading the data
* Classifying the dataset
* Visualization of the data
* Data cleaning
* Logistic Regression
* Prediction of dataset
* Prediction Accuracy

**Loading the dataset:**



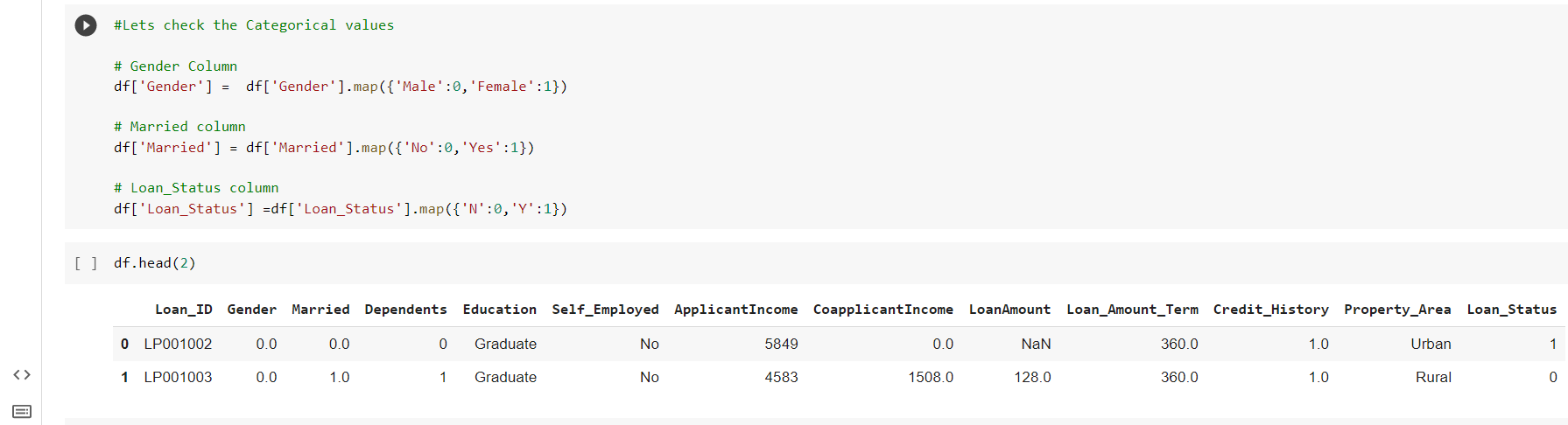
* Here,we mounted the data using Google colab ide and developed the data

.

* Python is the programing language along we used the libraries such as Pandas,Numpy,Matplotlib,Seaborn.

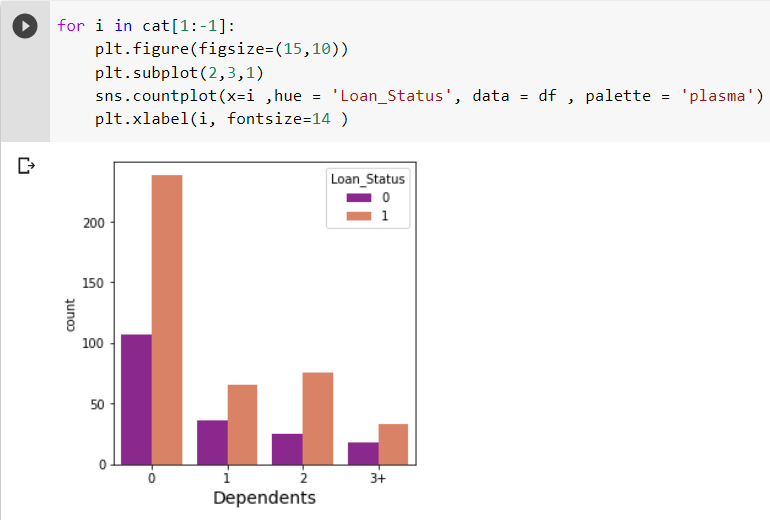
**Checking the categorical values:**

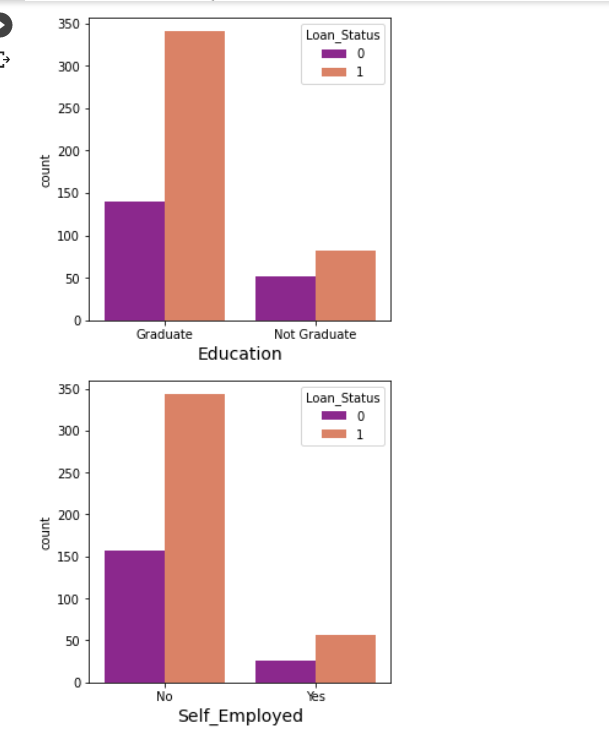
We need to check the categorical values as per the columns provided in the dataset. It helps to overcome the value error while we classifying the final set of values .Here, \_replace\_ can be used instead of ‘map’ as per the convenience if there is an error. The below figure shows the categorical values .



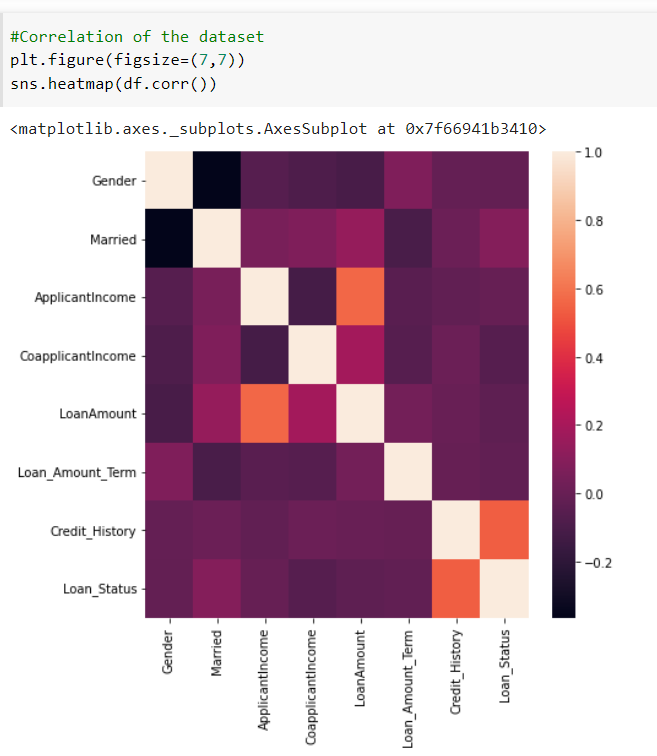
**Visualisation of Data:**

Here we used matplotlib and seaborn for the visualisation of dataset. The main feature of the datset is ‘Loan\_status’, based on that we plotted the figure comparing with other columns. The countplot shows the status of loan per Dependents, Self-Employed, Education.





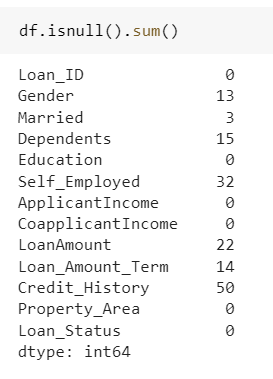
Then we used heatmaps for the correlation of the dataset.



**Data Cleaning :**

We must ensure that the values in the dataset has any empty spaces , hence it will show value error during the regression process. So, we need to clean the dataset with isnull() function to fill the empty spaces with zero.

The below figure shows the columns with empty spaces are filled by zero.



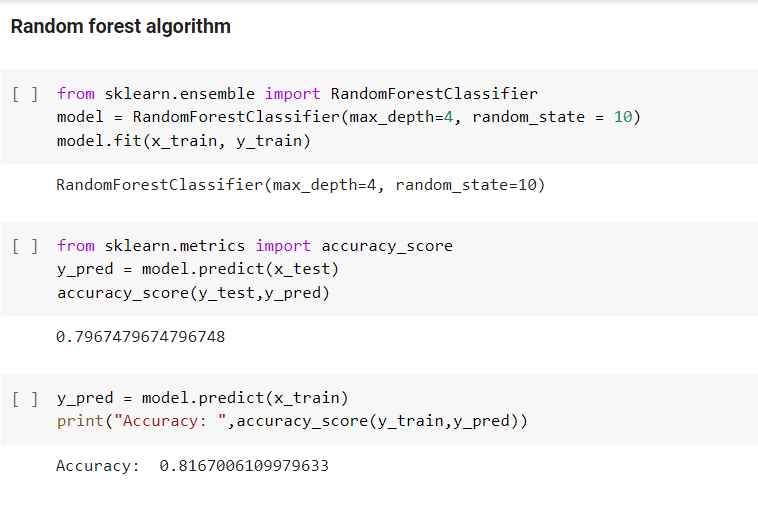
**Logistic Regression:**

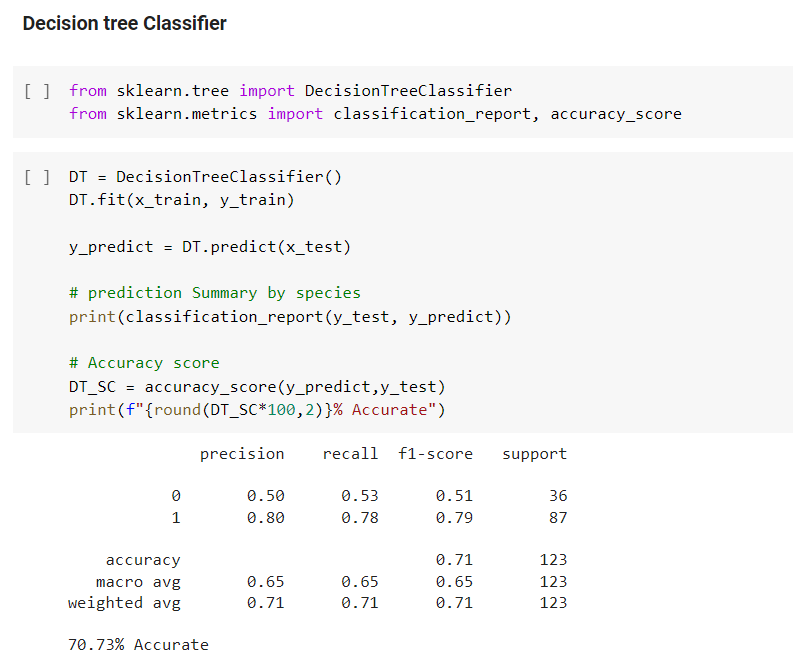
We need split the dataset to proceed the logistic regression . So,we used sklearn library for the Regression and accuracy of the model to be predicted.



**Prediction of dataset:**

The model cannot be proceed through the single prediction .So,we used RT(Random forest Regressor) and DT (Decision Tree Classifier ) to predict the model with precise value .This is the best way to predict the individual is eligible for the loan approval or disqualified by the norms of the bank.





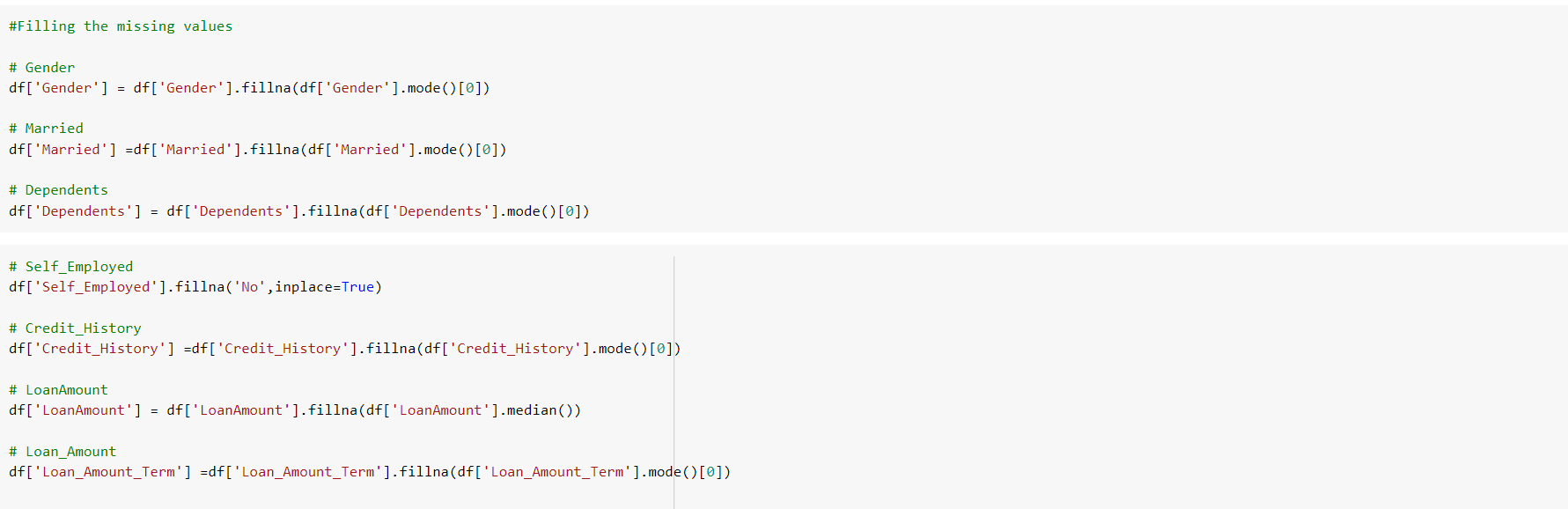
**Prediction report:**

* The predicted accuracy score of Logistic Regression = 81.26
* The predicted accuracy score of Random Forest = 81.67
* The predicted accuracy score od Decision Tree = 70.73

As per the report we conclude that Random Forest regressor predicted the data model precisely compared to others. It helps us predict the correct status for the loan approval.

**Key insight factor of the prediction:**

**This** factor is one which eliminates the common that everyone deals with during the prediction. Value error that string cannot be converted into float. Everyone must clean data to avoid these error at the final stage of the prediction. So I filled every columns with zero or null value while cleaning the dataset .It ensures the program to be run throughout the prediction.



THANK YOU