**Answer Sheet :**

**Machine Learning Assignment Submission - (Parikshit Prajapati)**

Name : Prajapati Parikshit Virabhai

Phone No. : 9998428706

Email : parikshitprajapti2001@gmail.com

Use the [Oil Spill Dataset](file:///D:\MACHINE%20LEARNING%20ASSIGNMENT\oil_spill.csv) and solve the following question by using the

dataset, to download the dataset click on the dataset name.

**About Dataset :**

The dataset was developed by starting with satellite images of the ocean,

someofwhichcontain an oil spill and some that do not.

Images were split into sections and processed using computer vision

algorithms to provide a vector of features to describe the contents of the

image section or patch.

The task is, given a vector that describes the contents of a patch of a satellite

image, then predicts whether the patch contains an oil spill or not, e.g. from

the illegal or accidental dumping of oil in the ocean.

There are two classes and the goal is to distinguish between spill and

non-spill using the features of a given ocean patch.

* **Non-Spill: negative case, or majority class.**
* **OilSpill: positive case, or minority class.**

There are a total of 50 Columns in the Dataset, the output column is named

as a target.

**QUESTIONS:**

**Q1) Download the Oil Spill Dataset and perform Data cleaning and Data**

**Pre-Processing if Necessary.**

**Q2) Use various methods such as Handling null values, One-Hot Encoding,**

**Imputation, and Scaling of Data Pre-Processing where necessary.**

**Q3)Derive some insights from the dataset.**

**Q4) Apply various Machine Learning techniques to predict the output in the**

**target column, make use of Bagging and Ensemble as required, and find**

**the best model by evaluating the model using Model evaluation techniques.**

**Q5) Save the best model andLoadthemodel**

**Q6) Take the original data set and make another dataset by randomly**

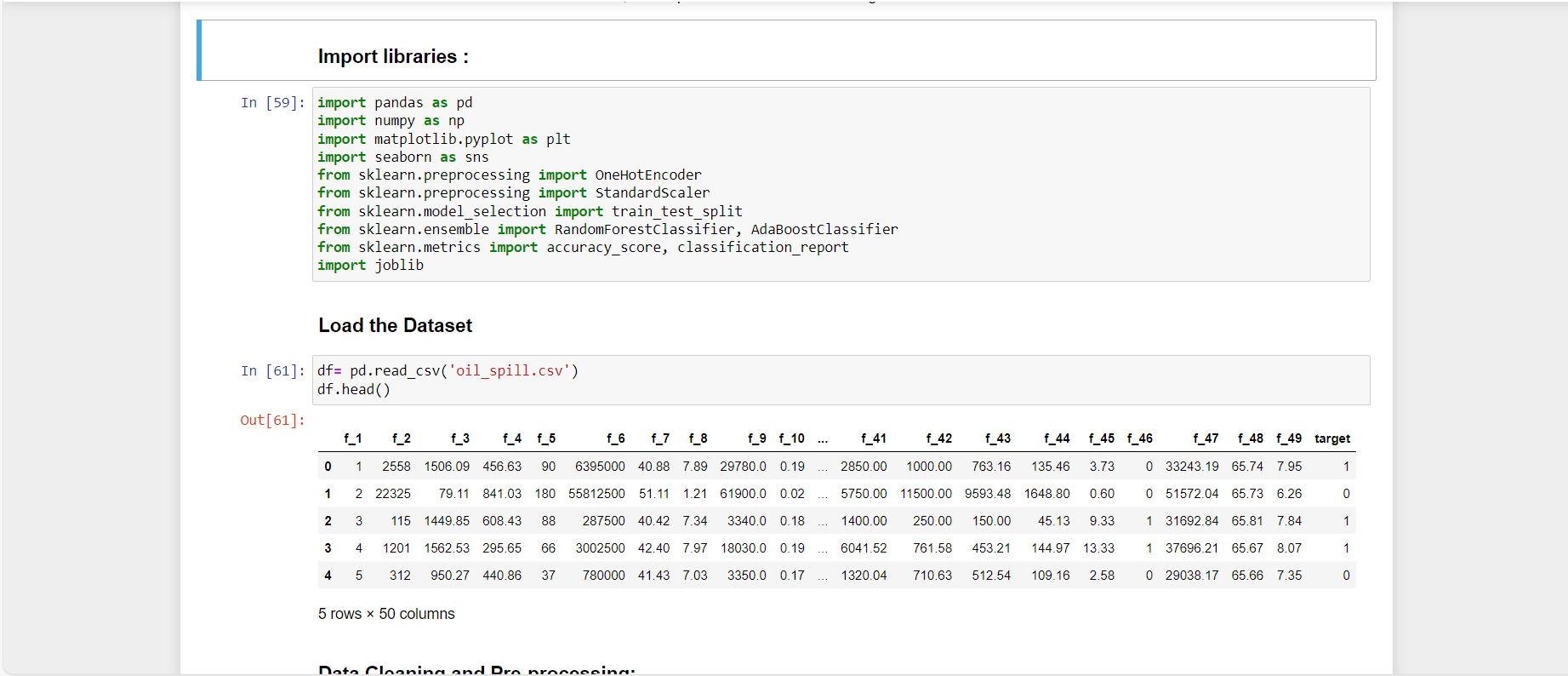
**picking 20 data points from the oil spill dataset and applying the saved**

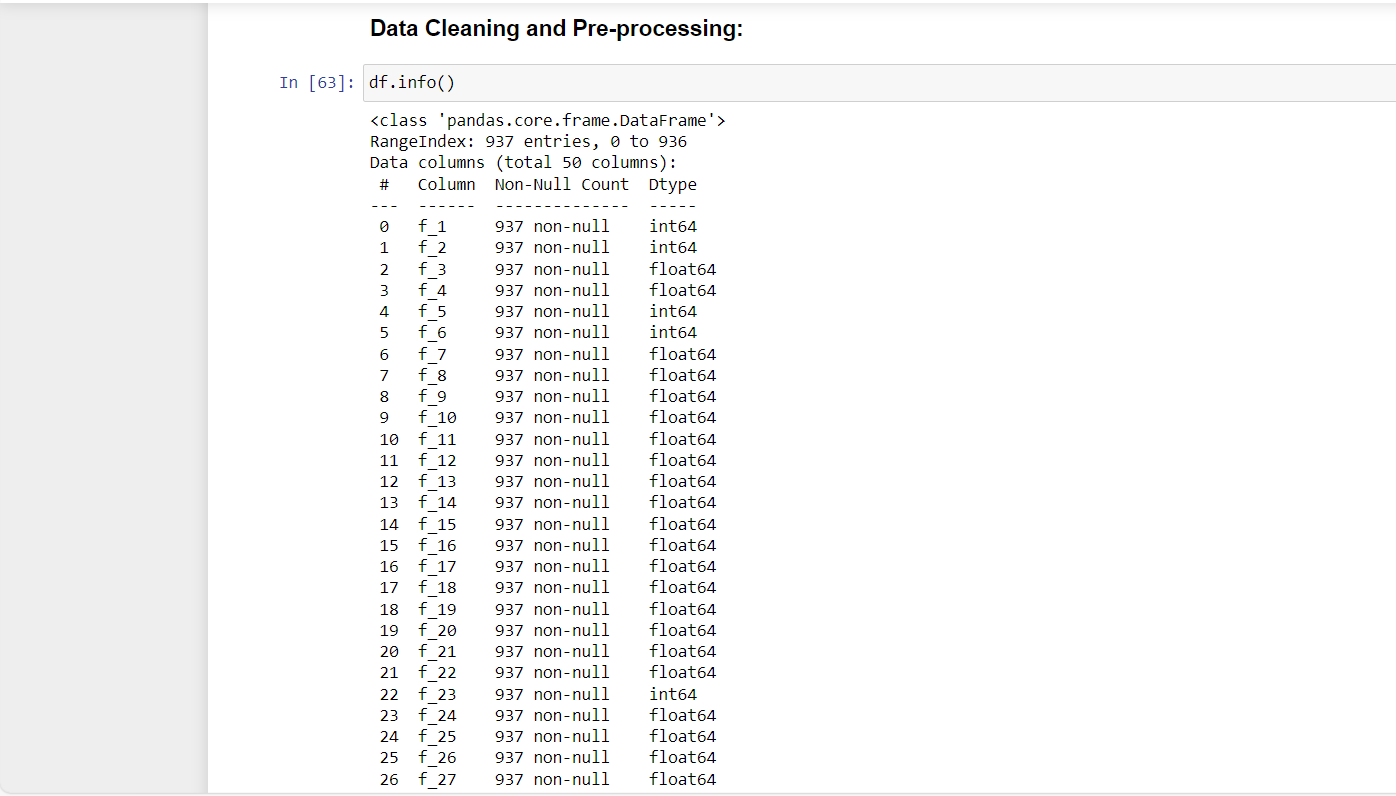
**model to the same.**

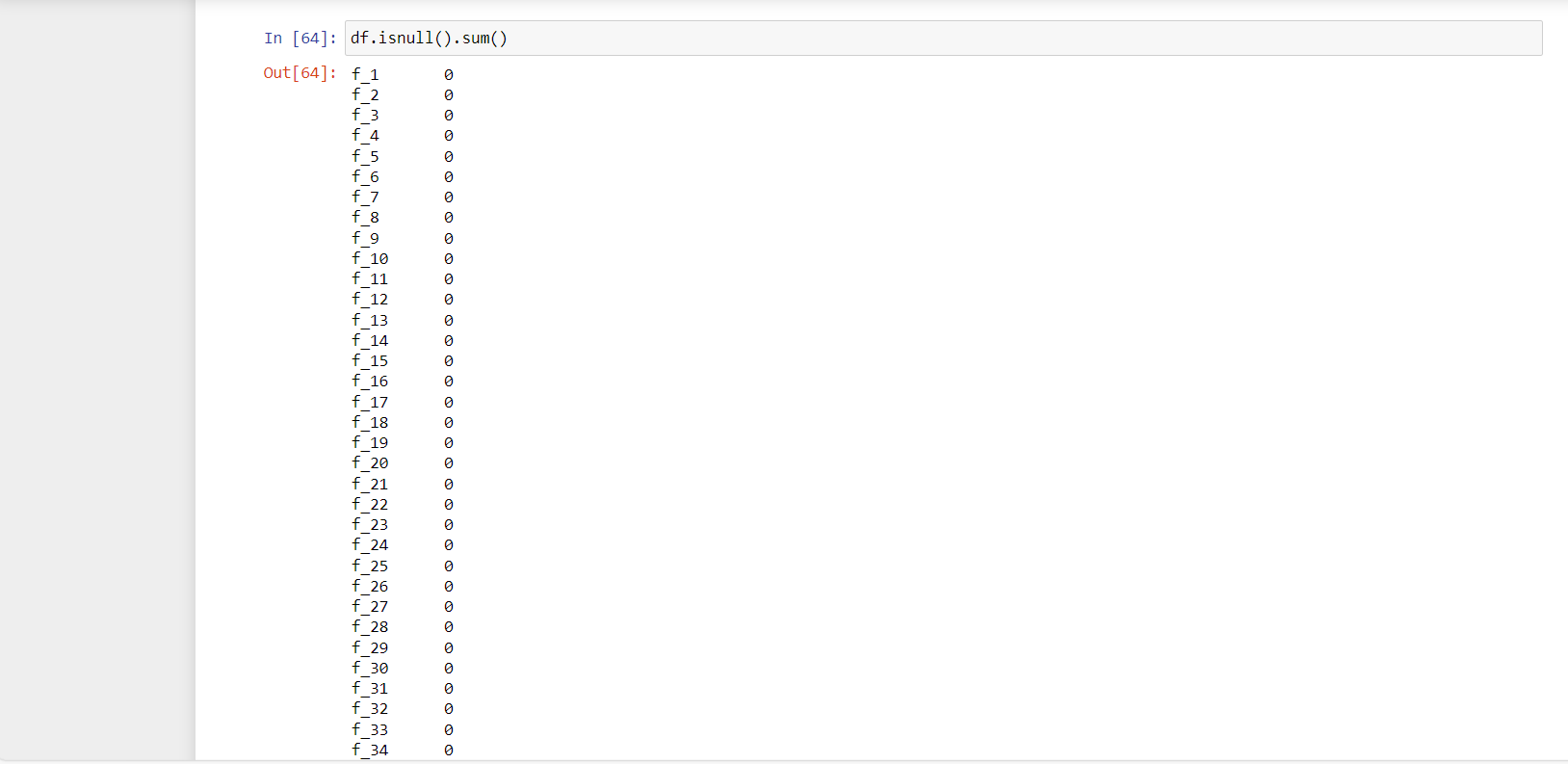
**NOTE : I HAVE COPYIED THE CODE FROM JUPYTER NOTEBOOK SO, THERE MIGHT BE SOME EXTRA LINES RIGHT IN SIDE , PLEASE IGNORE THEM**

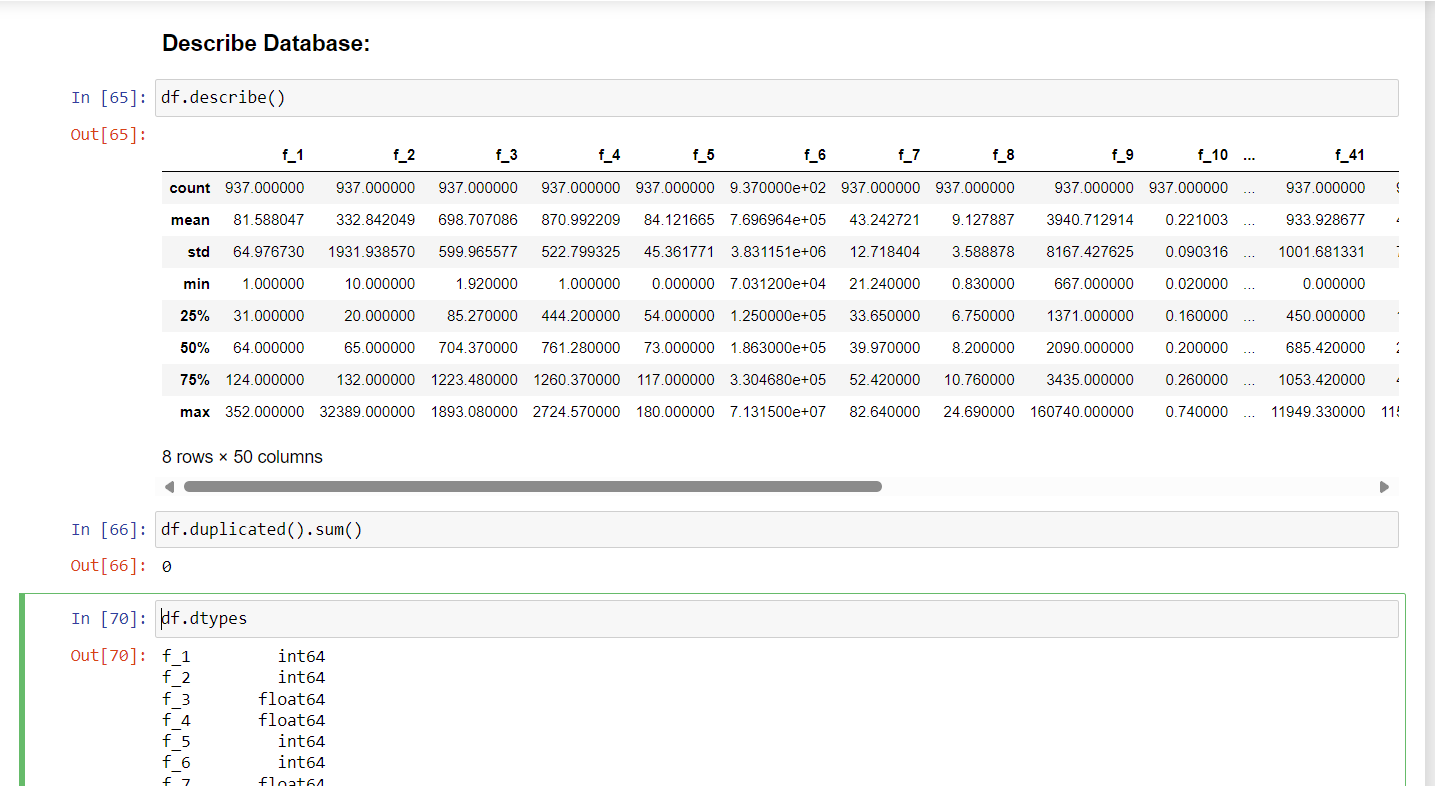
**Q1) Download the Oil Spill Dataset and perform Data cleaning and Data:**

**Input :**

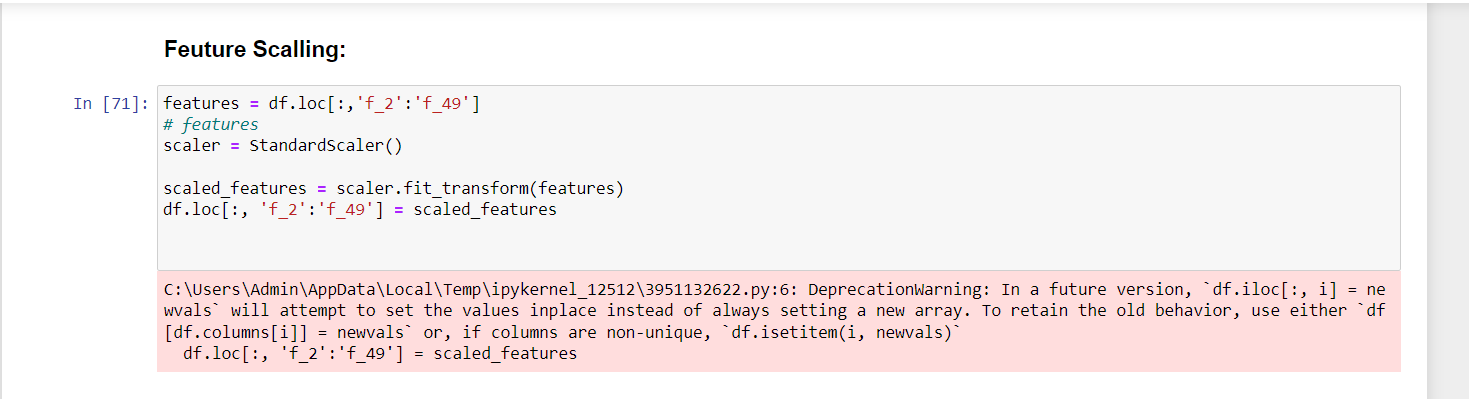
****

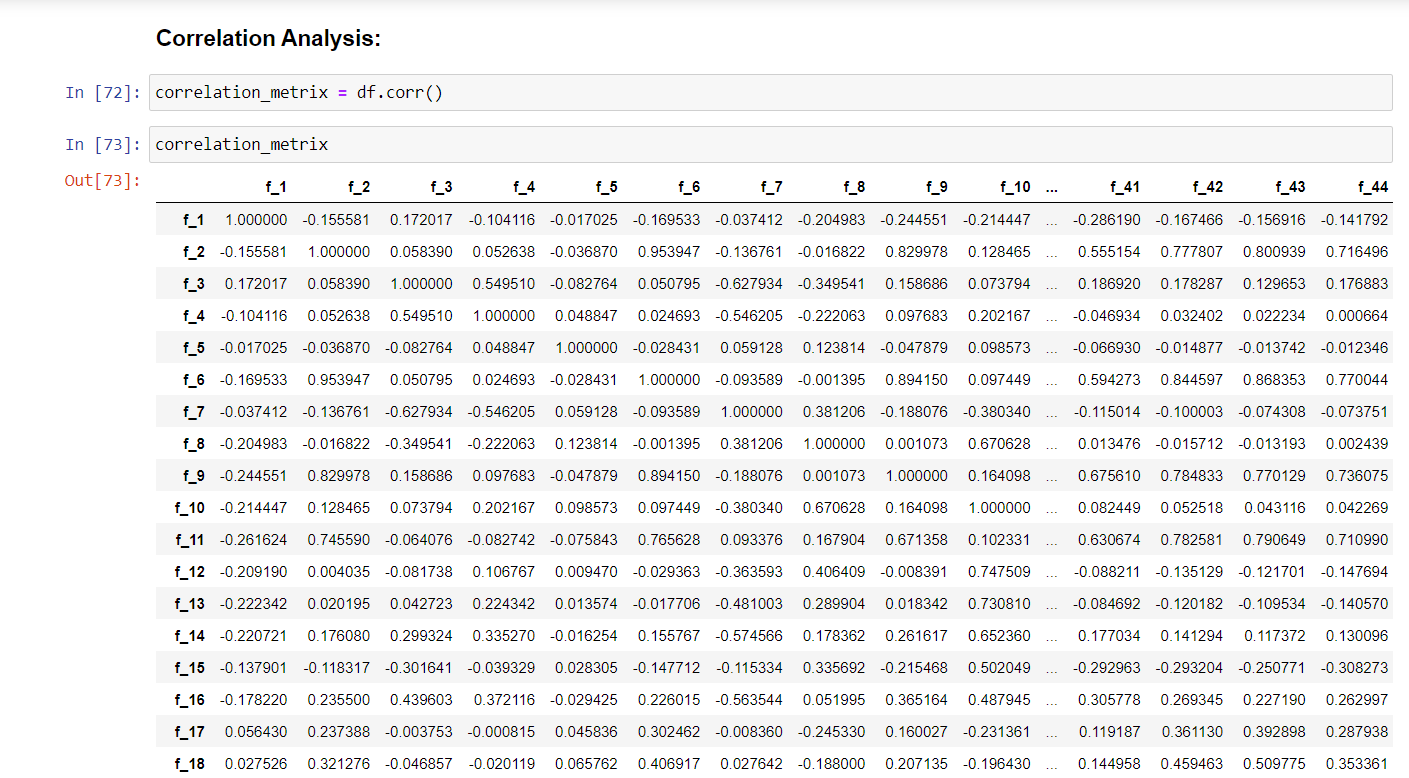
****

****

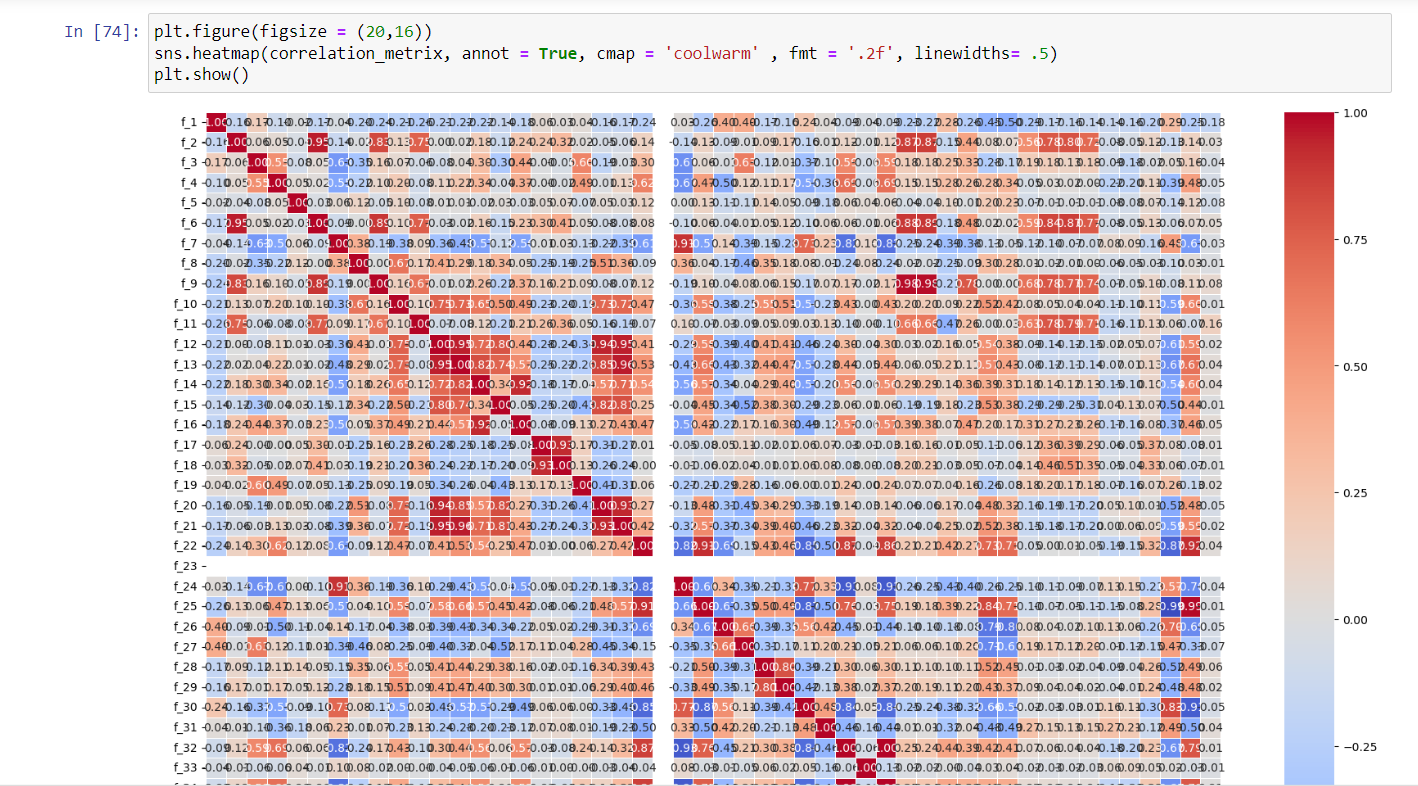
****

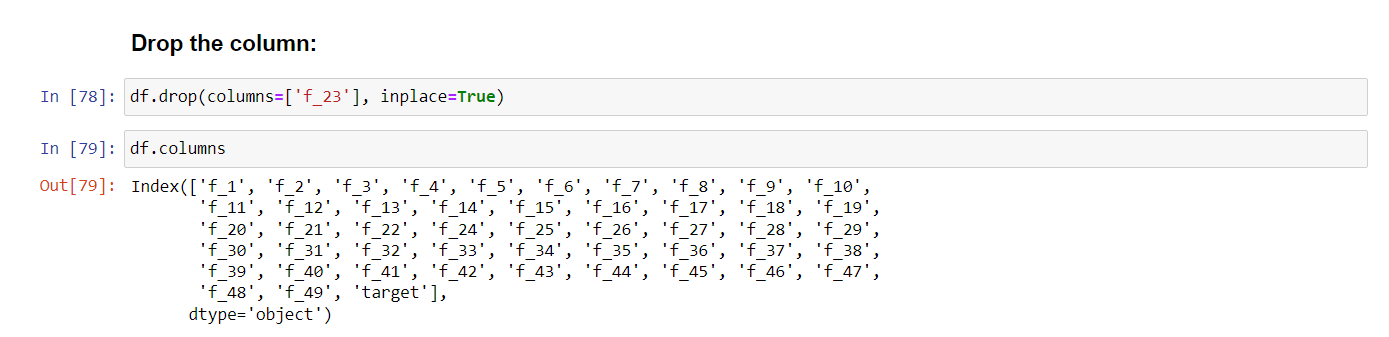
**Q2) Use various methods such as Handling null values, One-Hot Encoding, Imputation, and Scaling of Data Pre-Processing where necessary.**

****

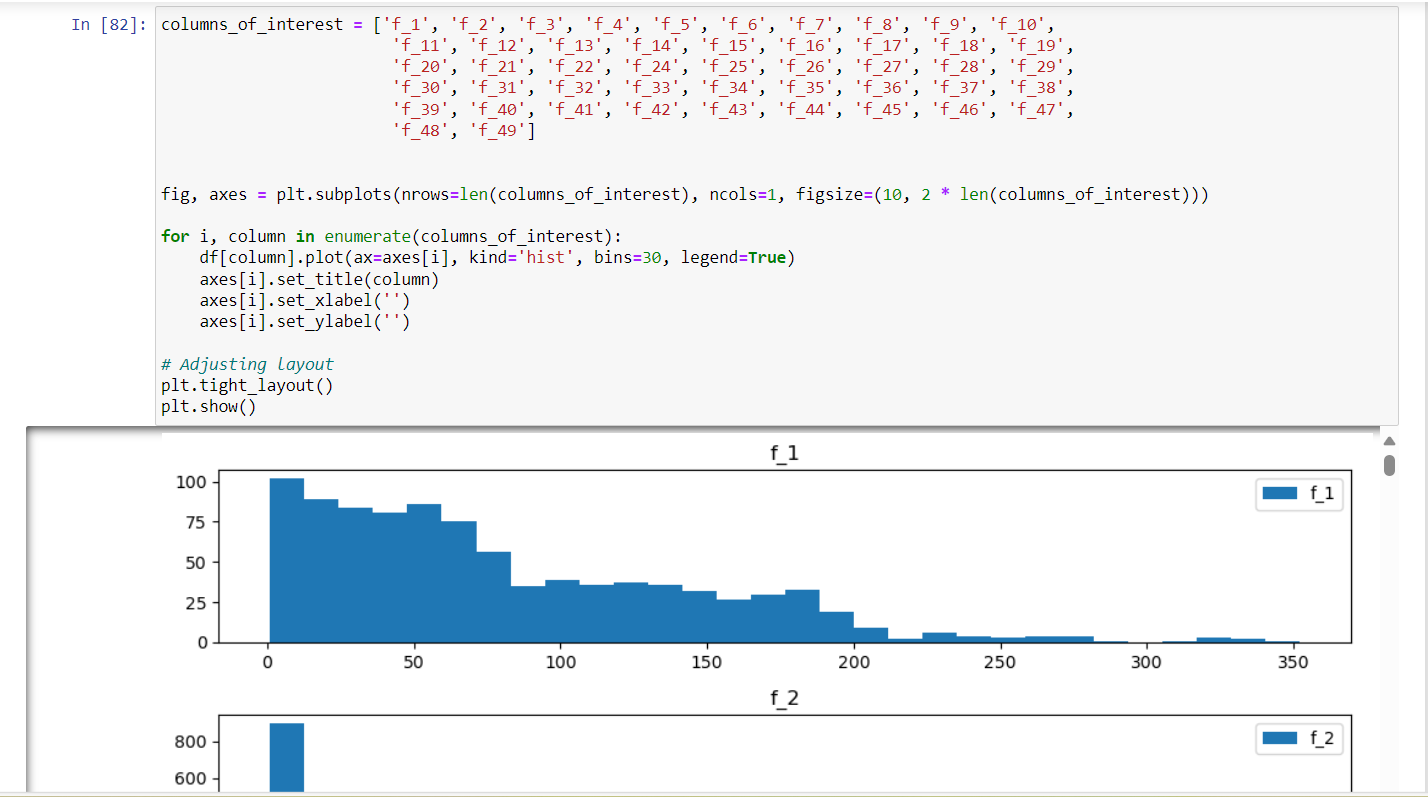
****

**Q3)Derive some insights from the dataset.**

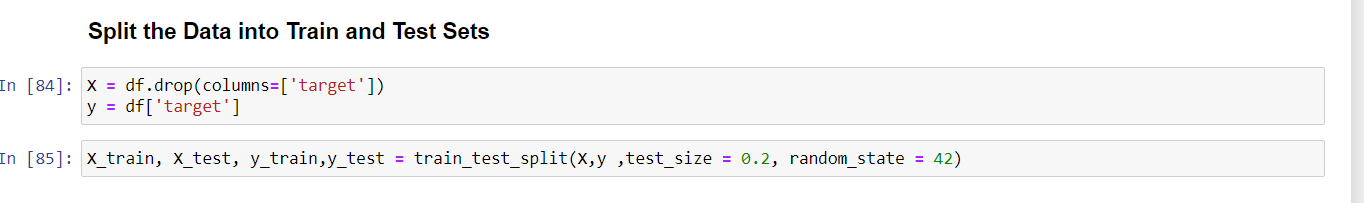
****

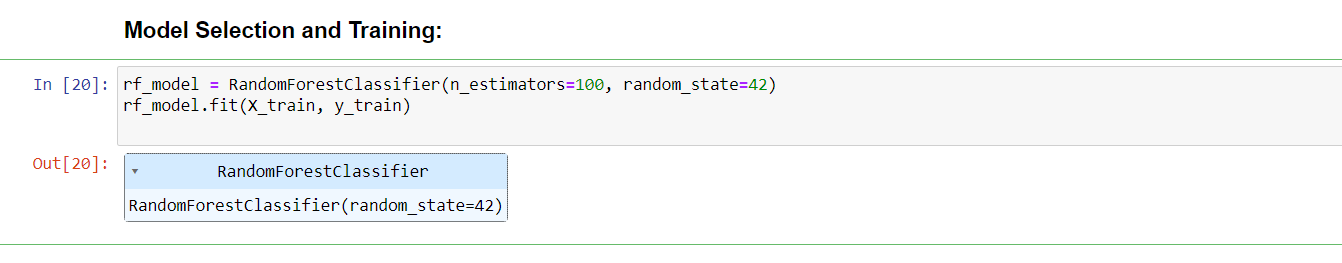
****

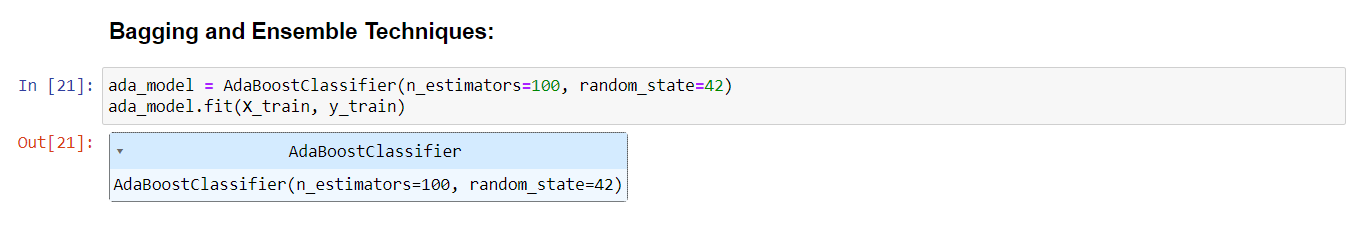
****

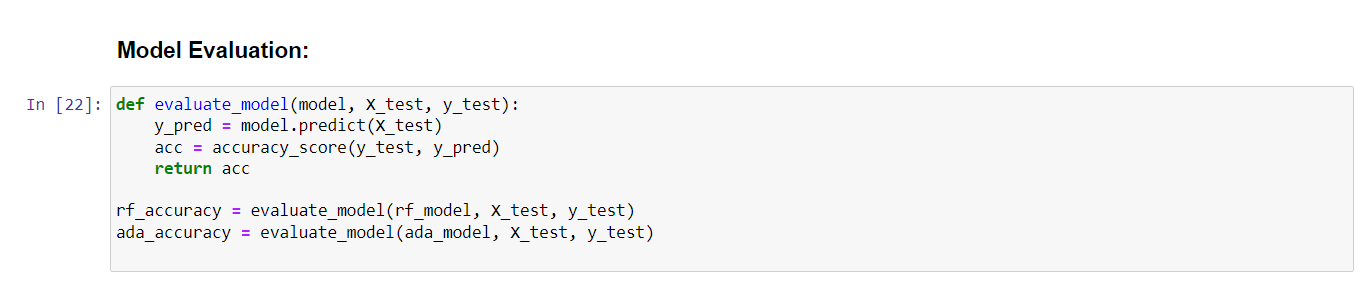
****

**04) Apply various Machine Learning techniques to predict the output in the target column, make use of Bagging and Ensemble as required, and find the best model by evaluating the model using Model evaluation techniques.**

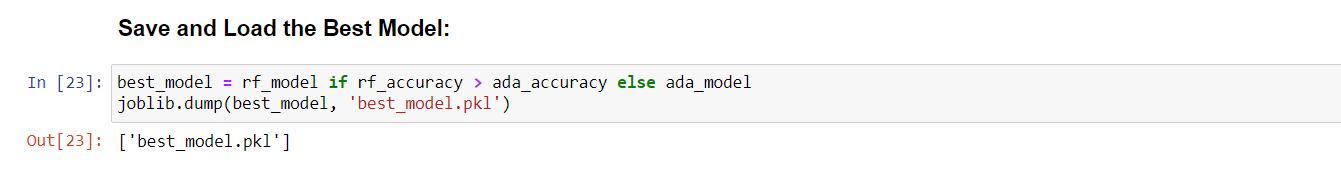
****

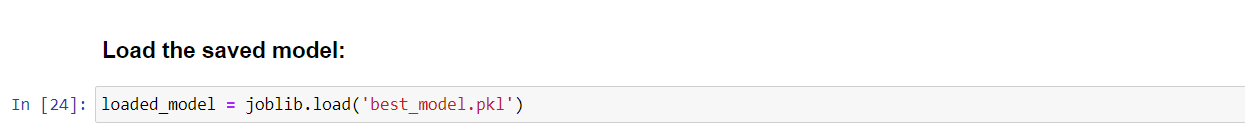
****

****

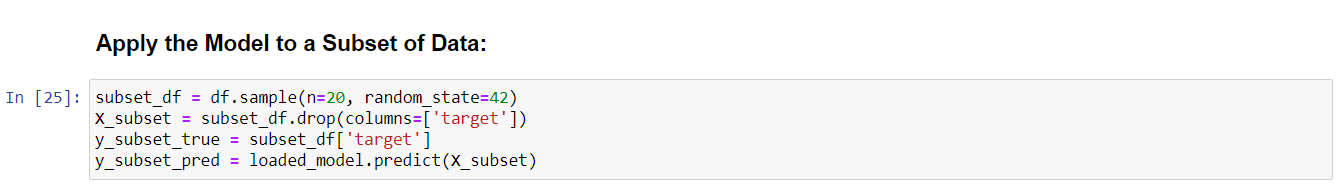
****

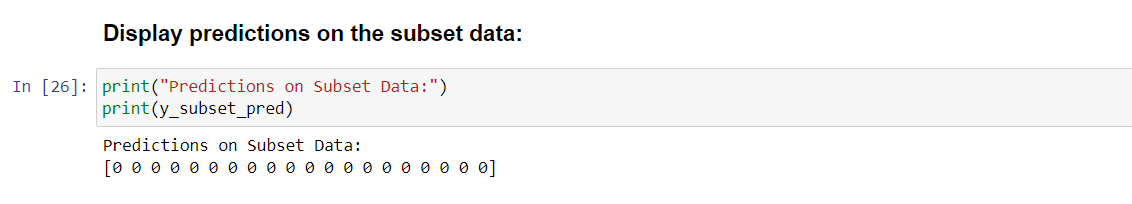
**Q5) Save the best model and Load the model.**

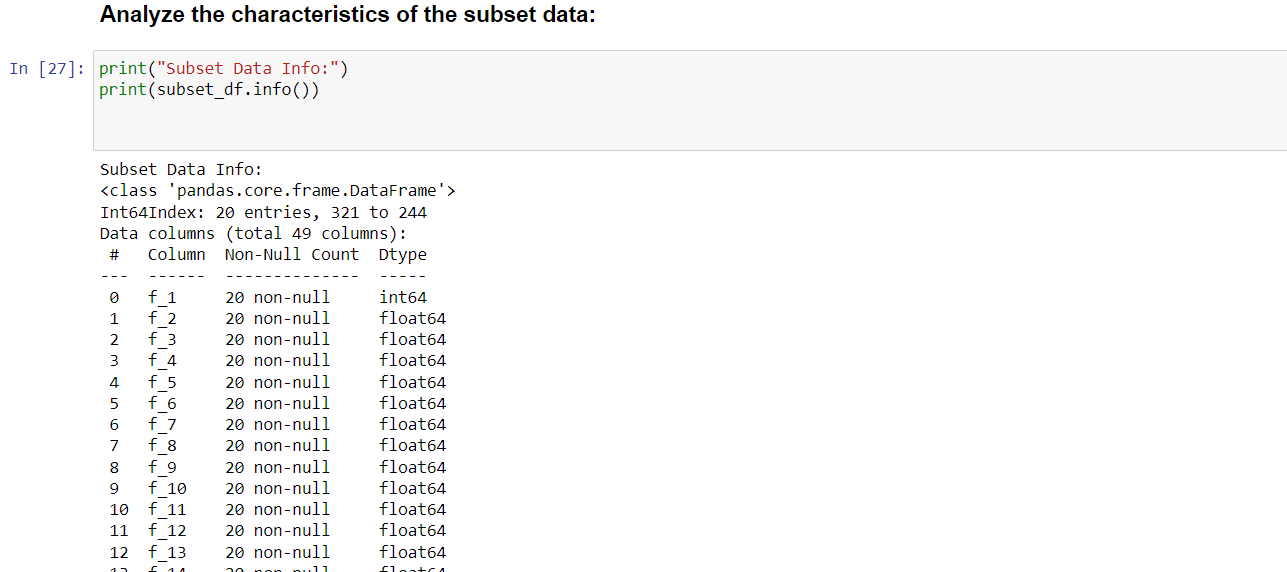
****

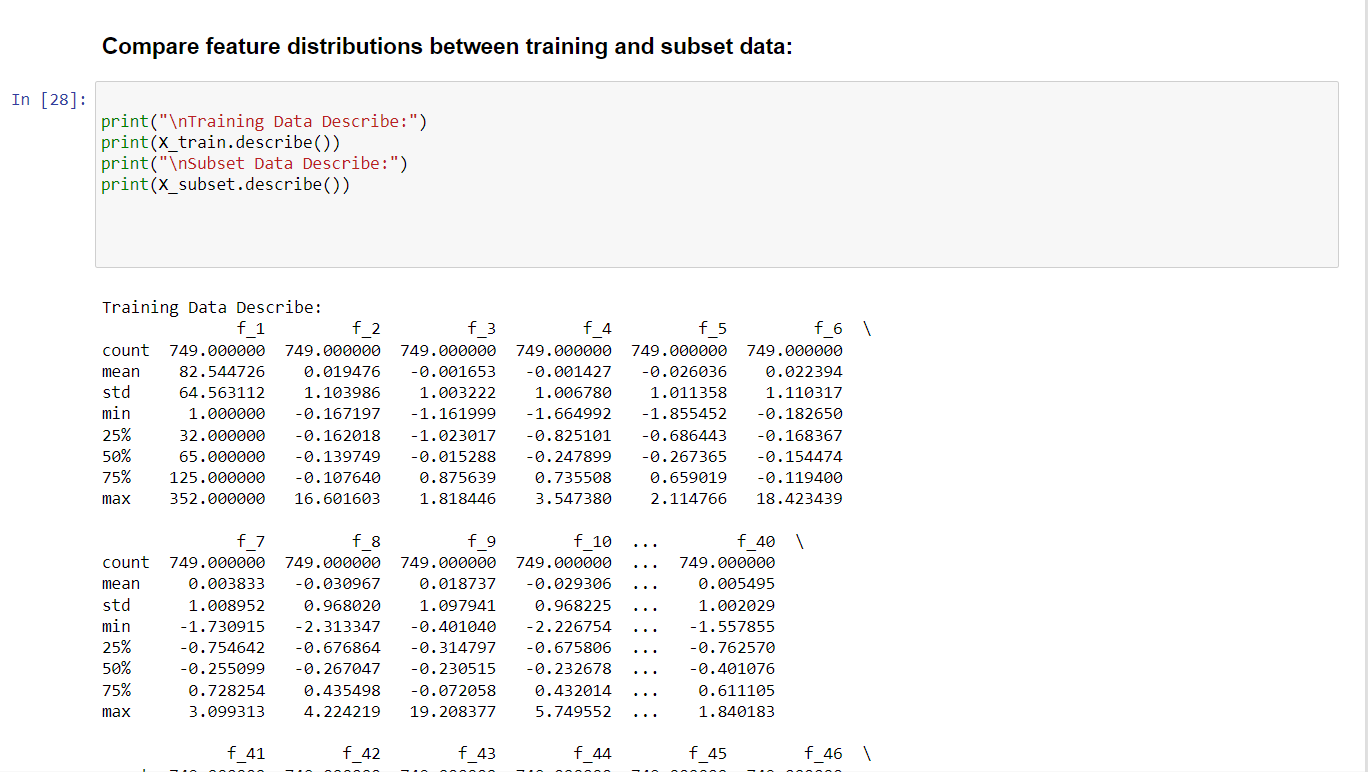
****

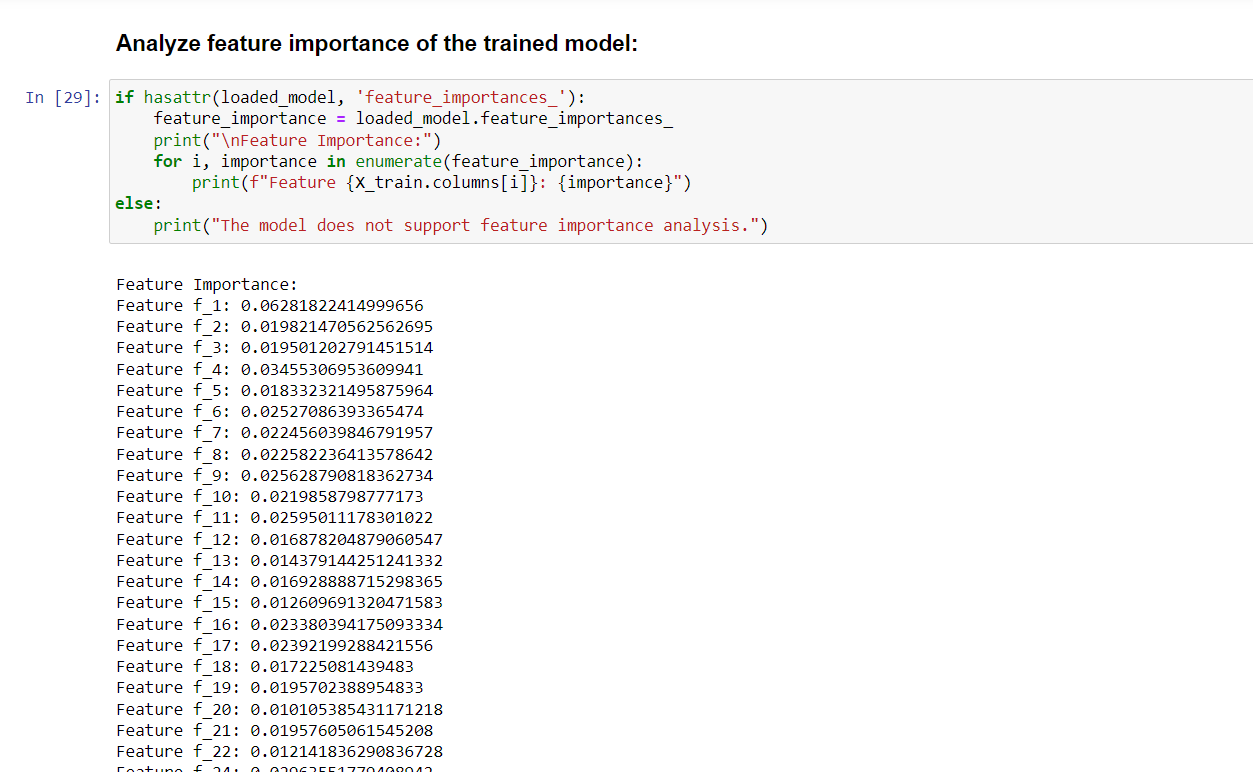
**Q6) Take the original data set and make another dataset by randomly picking 20 data points from the oil spill dataset and applying the saved model to the same.**

****

****

****

****

****