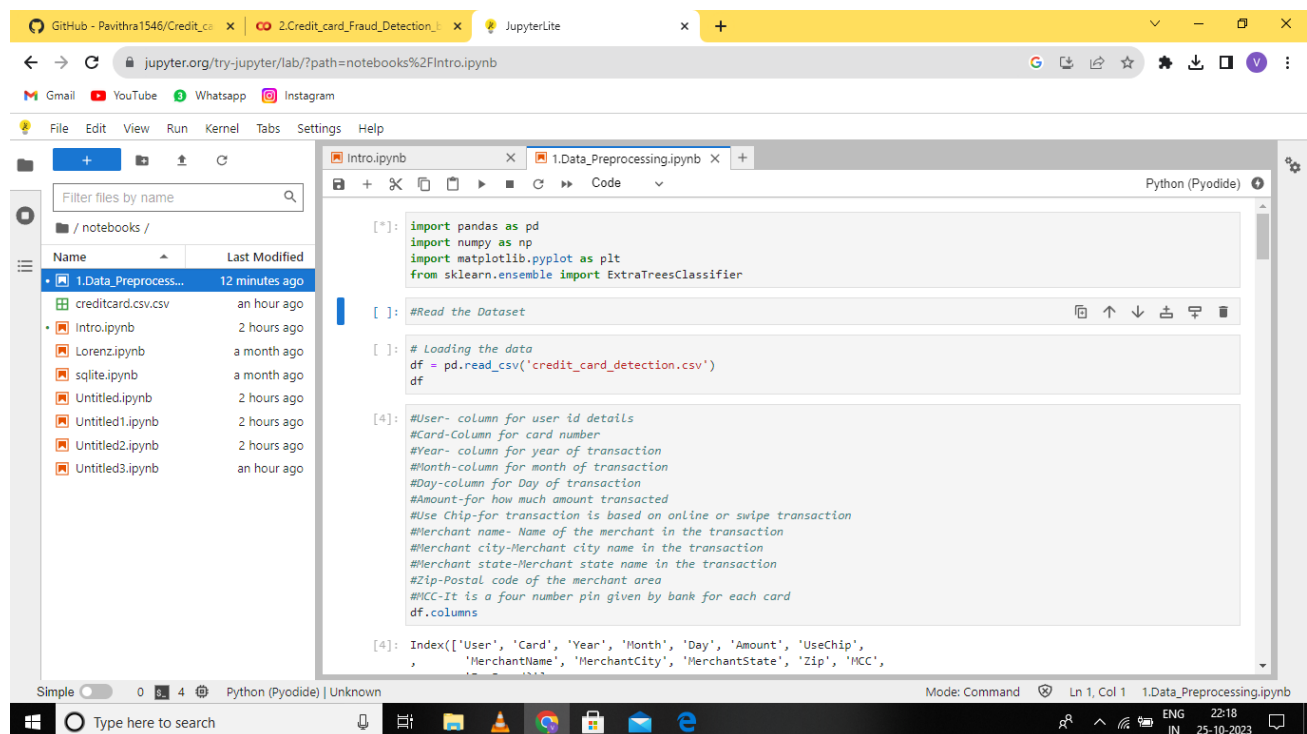


CREDIT CARD FRAUD DETECTION

The dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions. It contains only numeric input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data.

So, first we are importing pandas, numpy, matplotlib, and certain other libraries to make the access given to the data analysis and make the performance better.



```
[*]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.ensemble import ExtraTreesClassifier

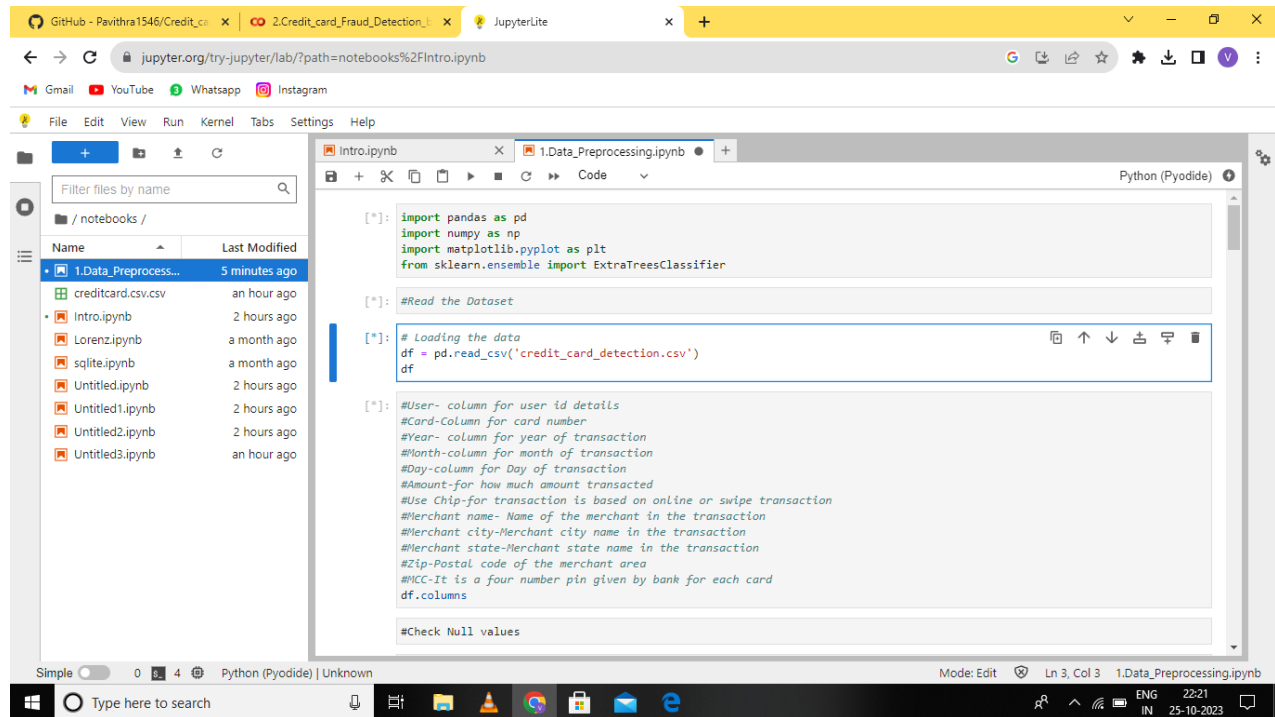
[ ]: #Read the Dataset

[ ]: # Loading the data
df = pd.read_csv('credit_card_detection.csv')
df

[4]: #User- column for user id details
#Card-Column for card number
#Year- column for year of transaction
#Month-column for month of transaction
#Day-column for Day of transaction
#Amount-for how much amount transacted
#Use Chip-for transaction is based on online or swipe transaction
#Merchant name- Name of the merchant in the transaction
#Merchant city-Merchant city name in the transaction
#Merchant state-Merchant state name in the transaction
#Zip-Postal code of the merchant area
#ICC-It is a four number pin given by bank for each card
df.columns

[4]: Index(['User', 'Card', 'Year', 'Month', 'Day', 'Amount', 'UseChip',
         'MerchantName', 'MerchantCity', 'MerchantState', 'Zip', 'ICC',
```

After installing the important libraries that are needed for the program code to succeed, we are reading the CSV file in which the data set is given. Therefore, we run the code for reading the CSV file with the path or the name of the file in which the data set is given.



The screenshot displays a JupyterLab environment. The left sidebar shows a file explorer with a list of notebooks and files, including '1.Data_Preprocess...', 'creditcard.csv', and 'Intro.ipynb'. The main area shows a notebook titled '1.Data_Preprocessing.ipynb' with the following code:

```
[*]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.ensemble import ExtraTreesClassifier

[*]: #Read the Dataset

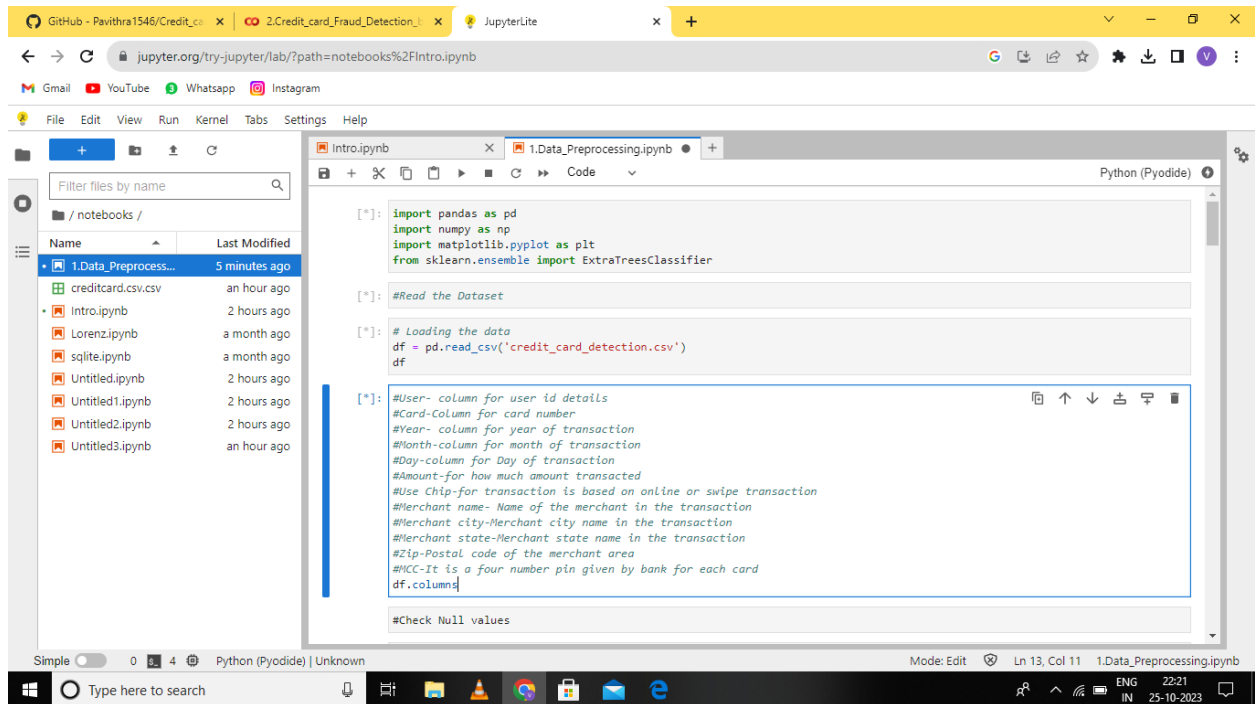
[*]: # Loading the data
df = pd.read_csv('credit_card_detection.csv')
df

[*]: #User- column for user id details
#Card-Column for card number
#Year- column for year of transaction
#Month-column for month of transaction
#Day-column for Day of transaction
#Amount-for how much amount transacted
#Use Chip-for transaction is based on online or swipe transaction
#Merchant name- Name of the merchant in the transaction
#Merchant city-Merchant city name in the transaction
#Merchant state-Merchant state name in the transaction
#Zip-Postal code of the merchant area
#MCC-It is a four number pin given by bank for each card
df.columns

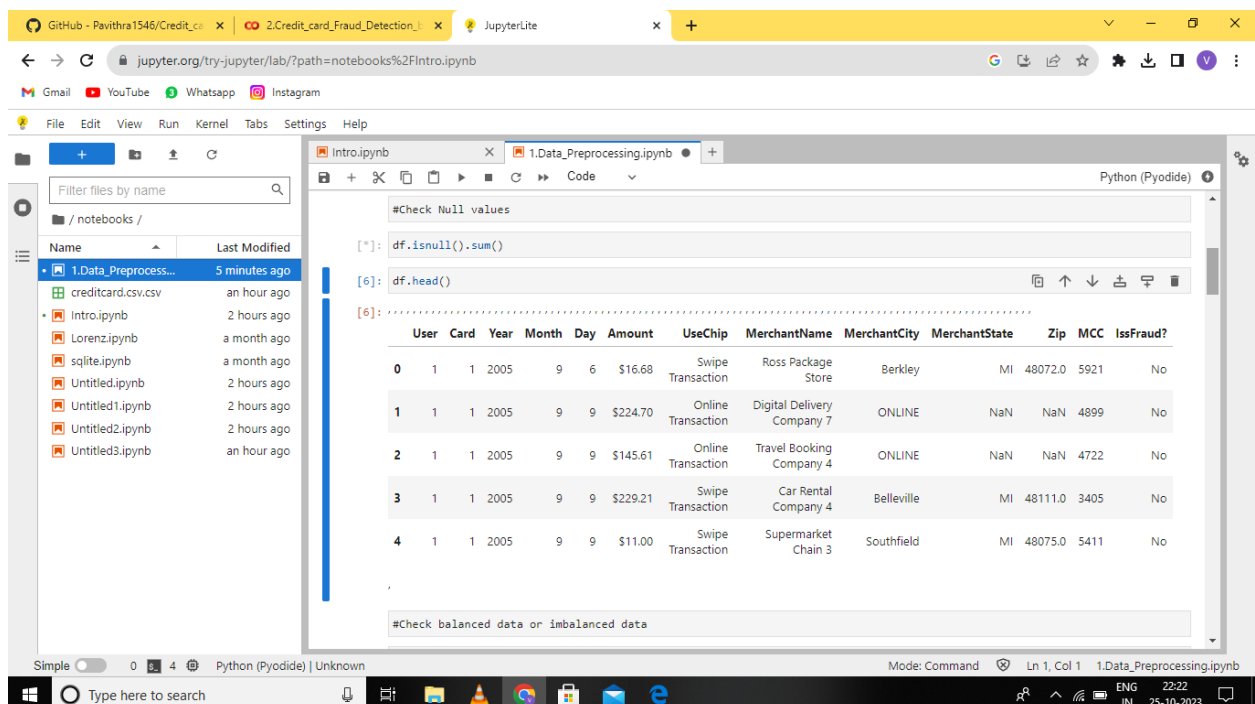
#Check Null values
```

The bottom status bar indicates the current mode is 'Edit', the cursor is at 'Ln 3, Col 3', and the file is '1.Data_Preprocessing.ipynb'. The system tray at the bottom shows the date and time as '25-10-2023' and '22:21'.

After reading the CSV file in the format of CSV, we run the code for columns to be read. This is to check for the dataset to be arranged in columns.



Then after, the data set null values are printed with the sum values and for the performance of the data set to be performed this code is run. And then, the head method is used to return a specified number of rows, string from the top.



This is how the code is runned for each of the following section for credit card fraud detection and each part of the code is used for checking of any fraud is happened.