CREDIT CARD FRAUD DETECTION WITH MACHINE LEARNING

Phase 4: Development Part 2

In this part you will continue building your project .Continue conducting the credit card fraud detection project analysis by performing:

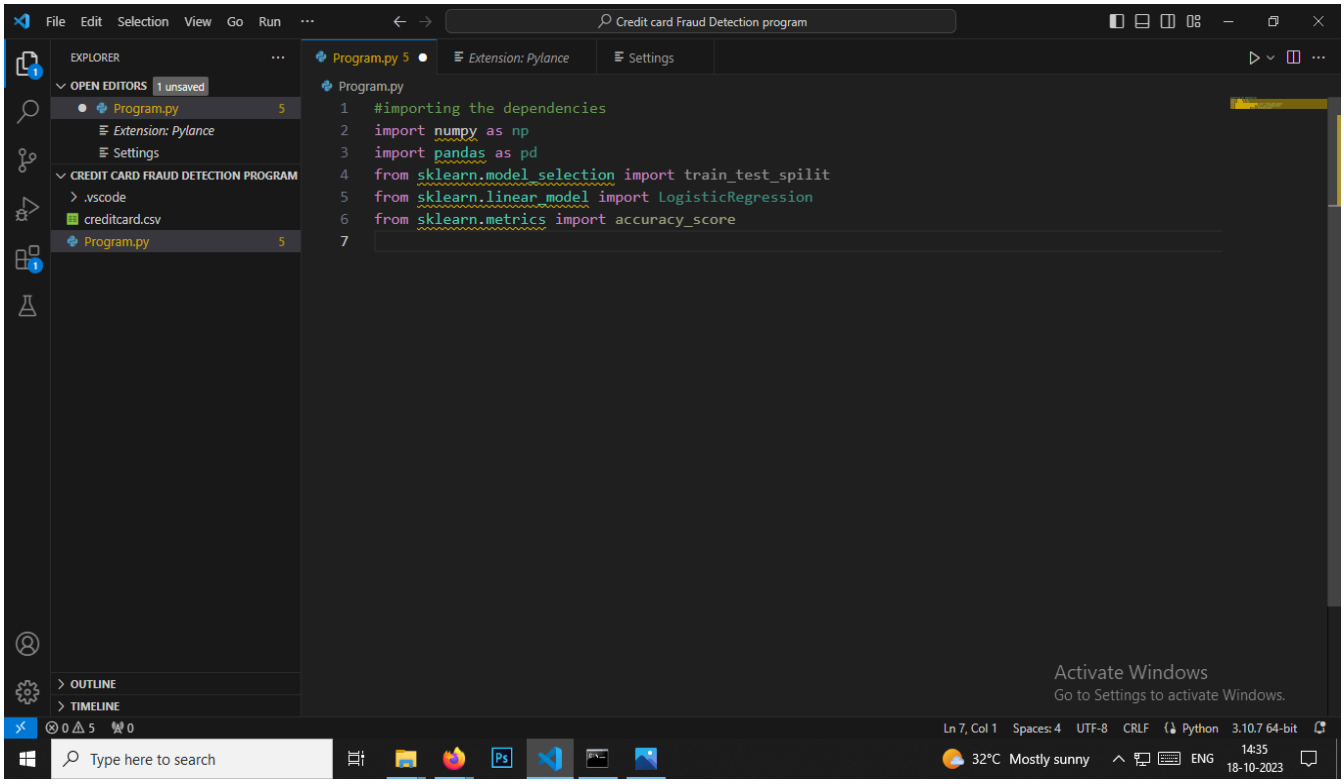
- ❖ Feature Engineering
- ❖ Model Training
- ❖ Evaluation

Dataset Link:

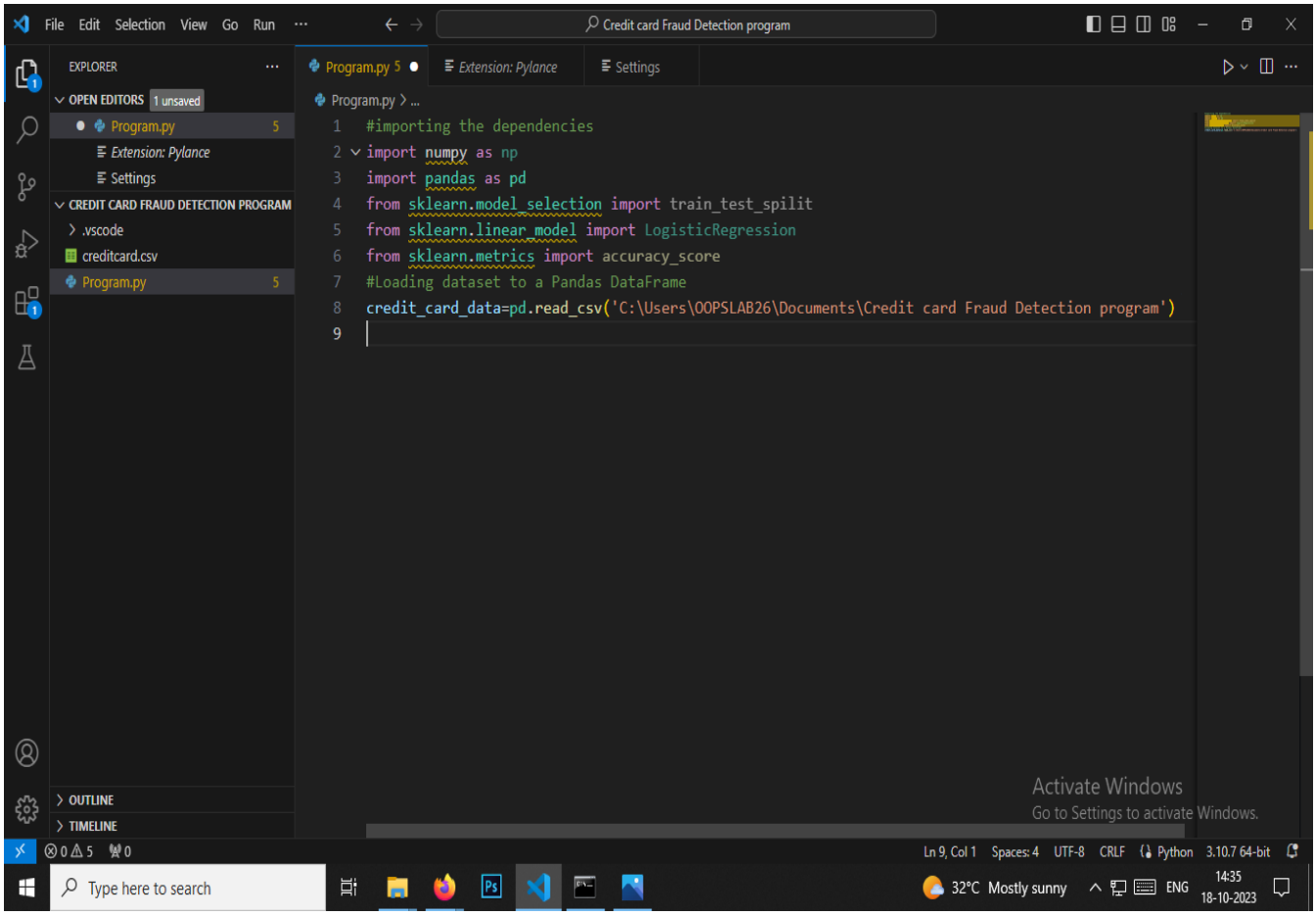
https://www.kaggle.com/datasets/mlgulb/creditcardfraud

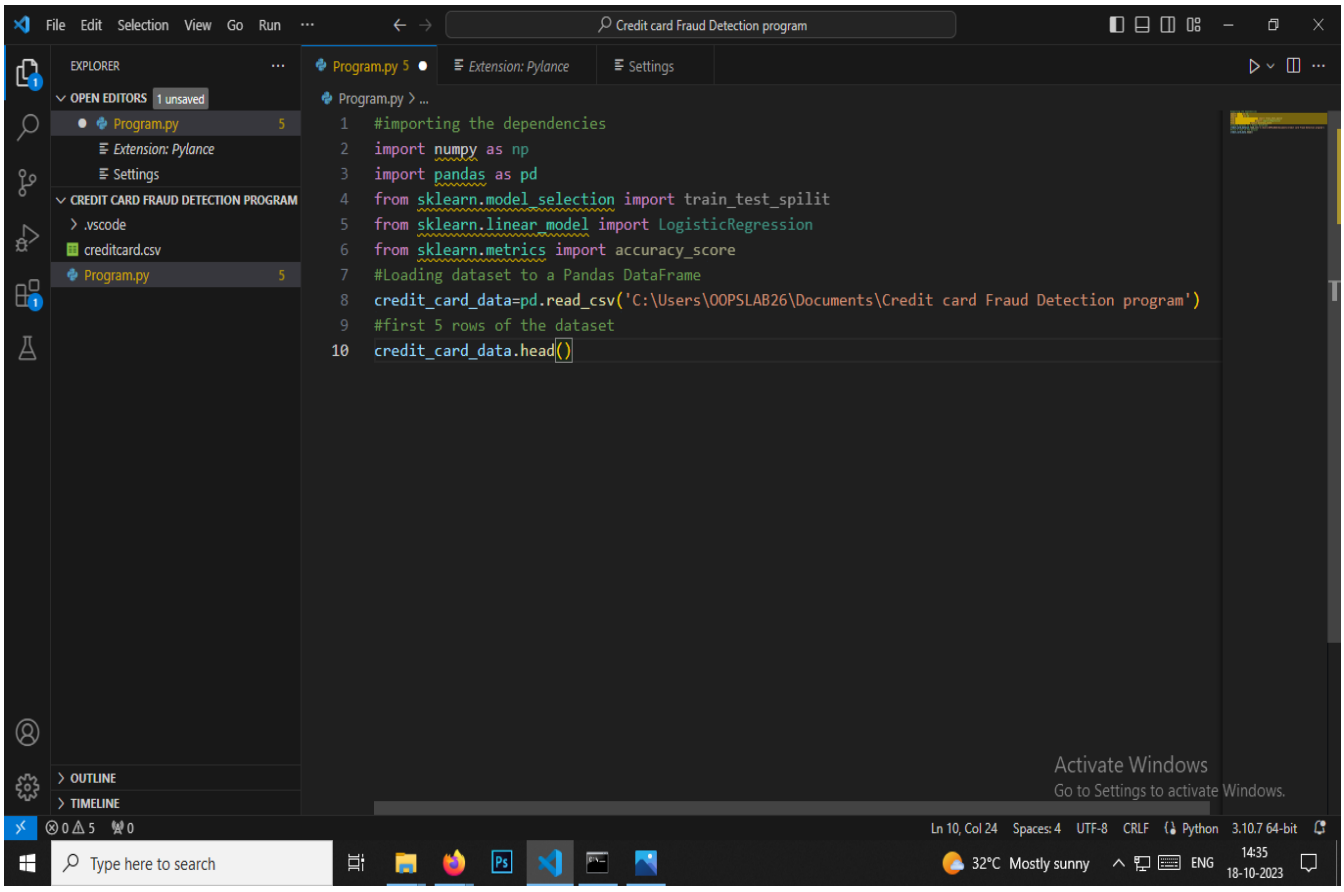
- ❖ Performing exploratory data analysis
- ❖ Statistical analysis

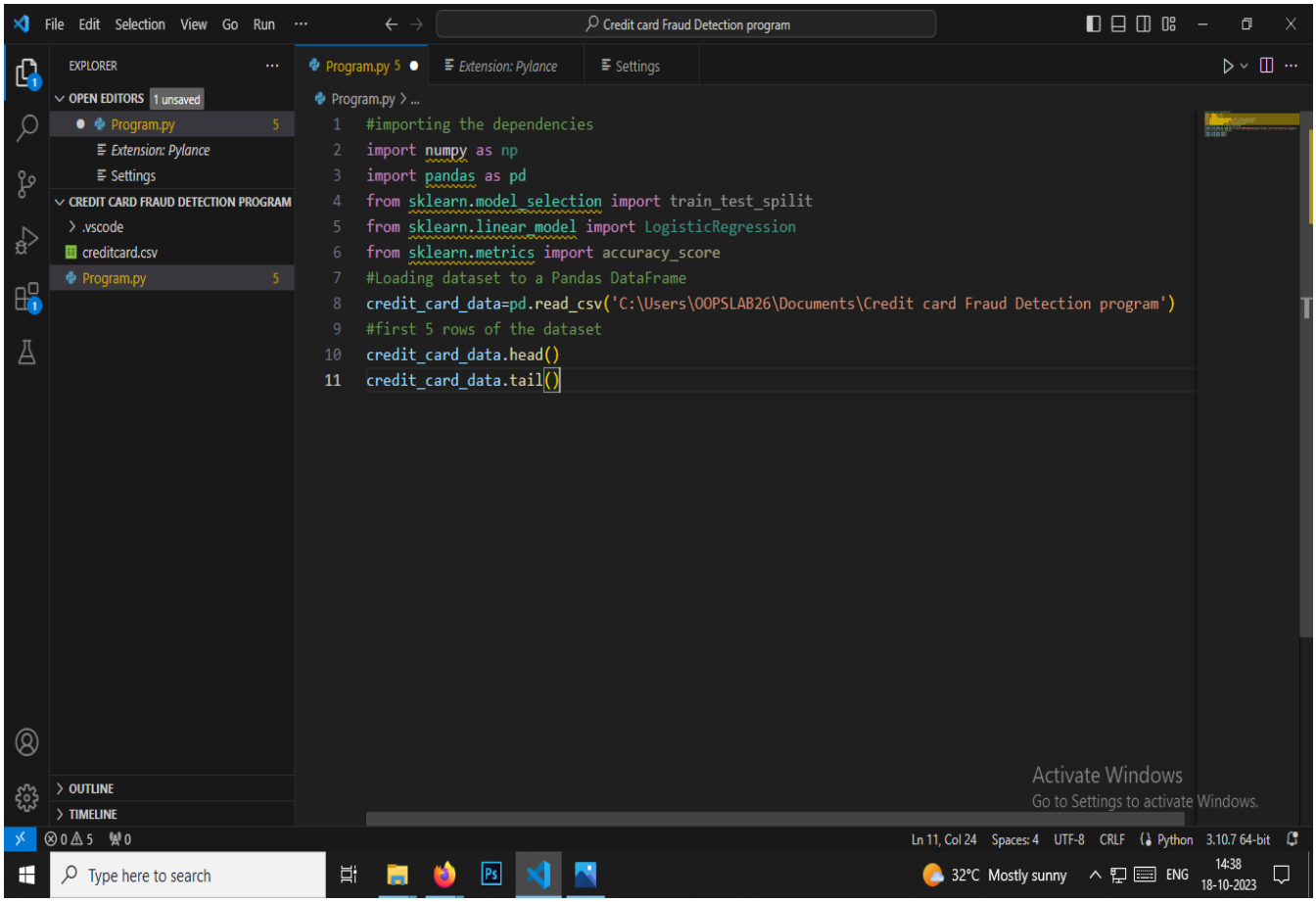
#Performing exploratory data analysis:

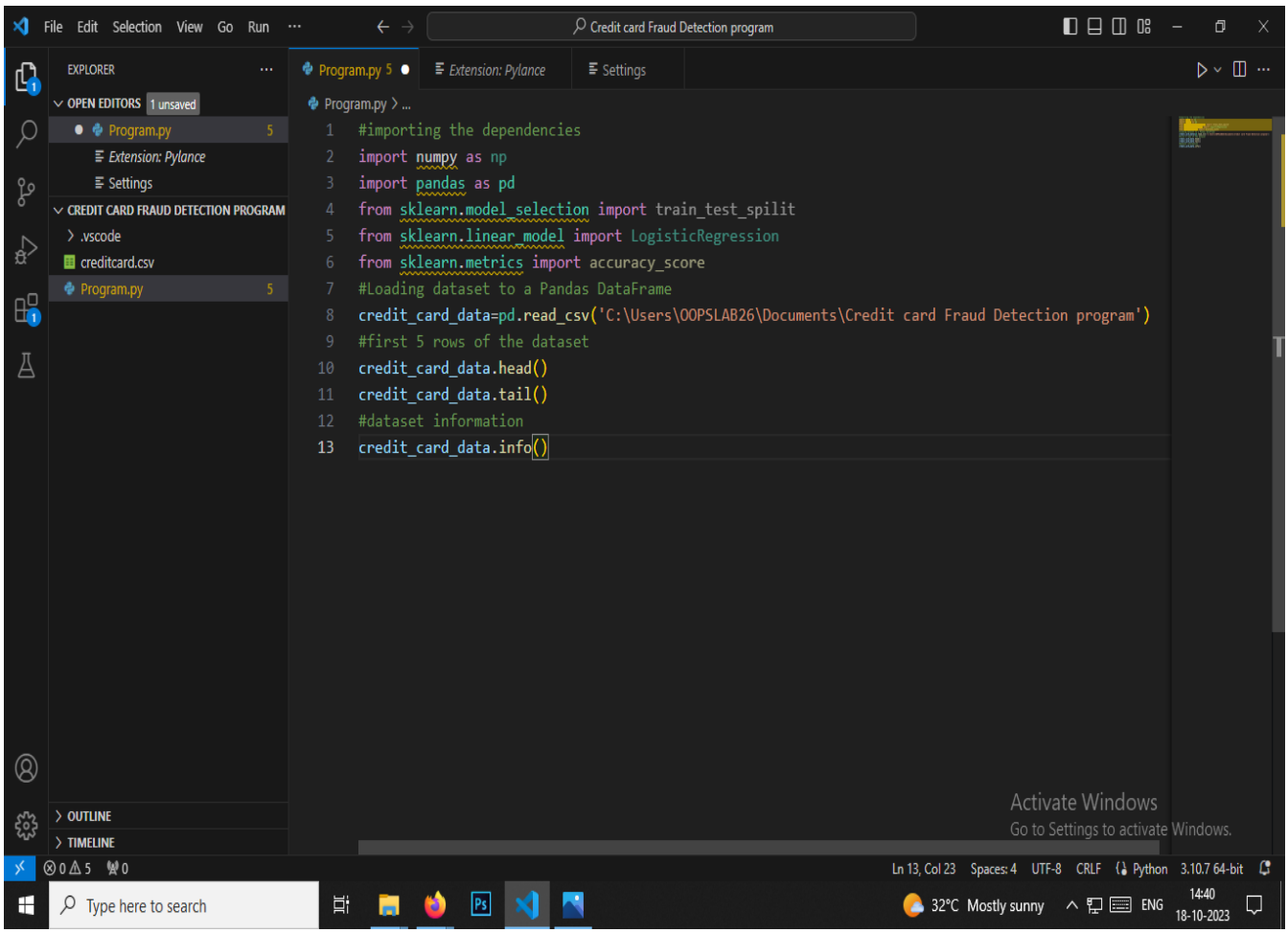


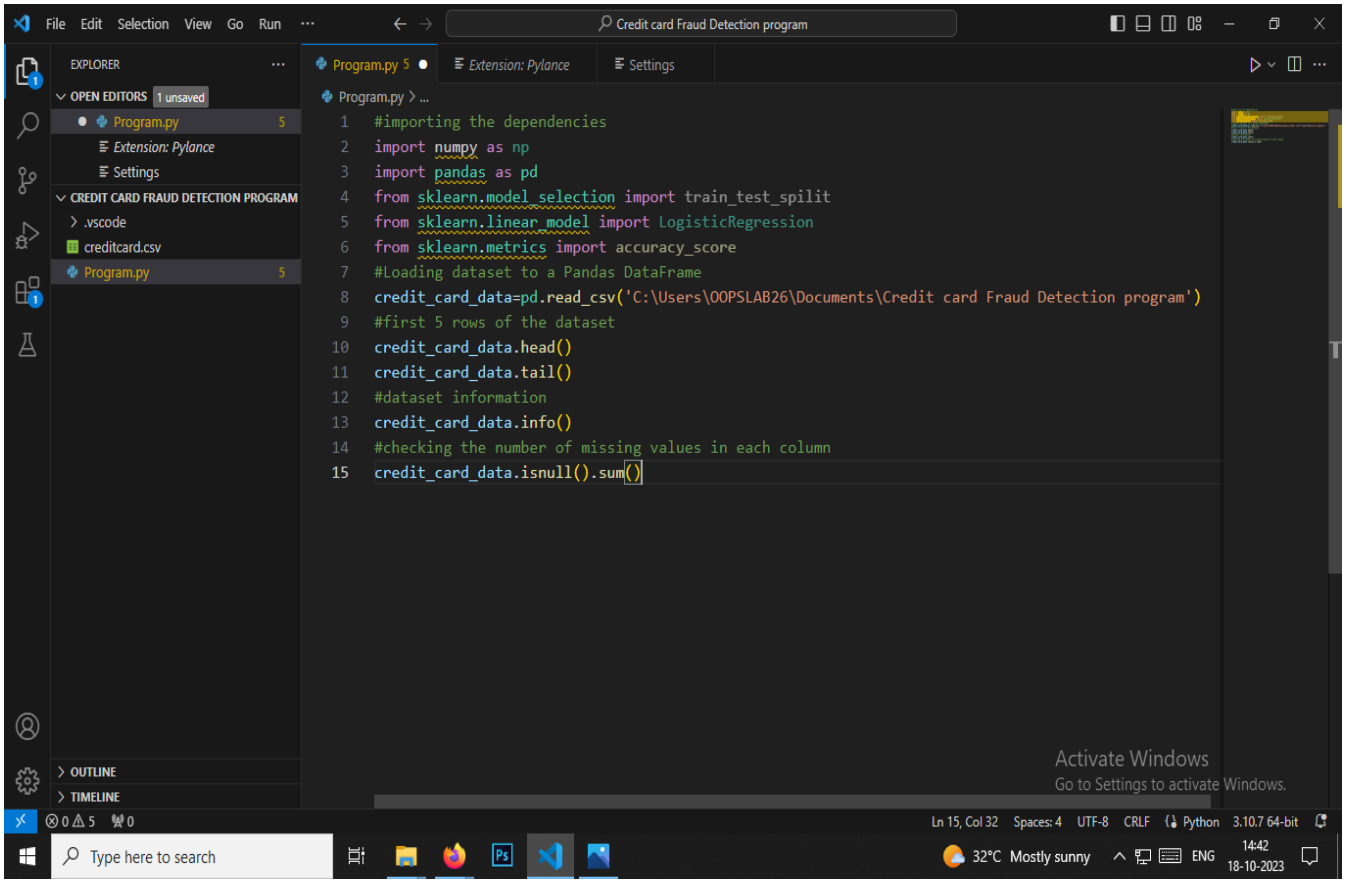
#Data describe :

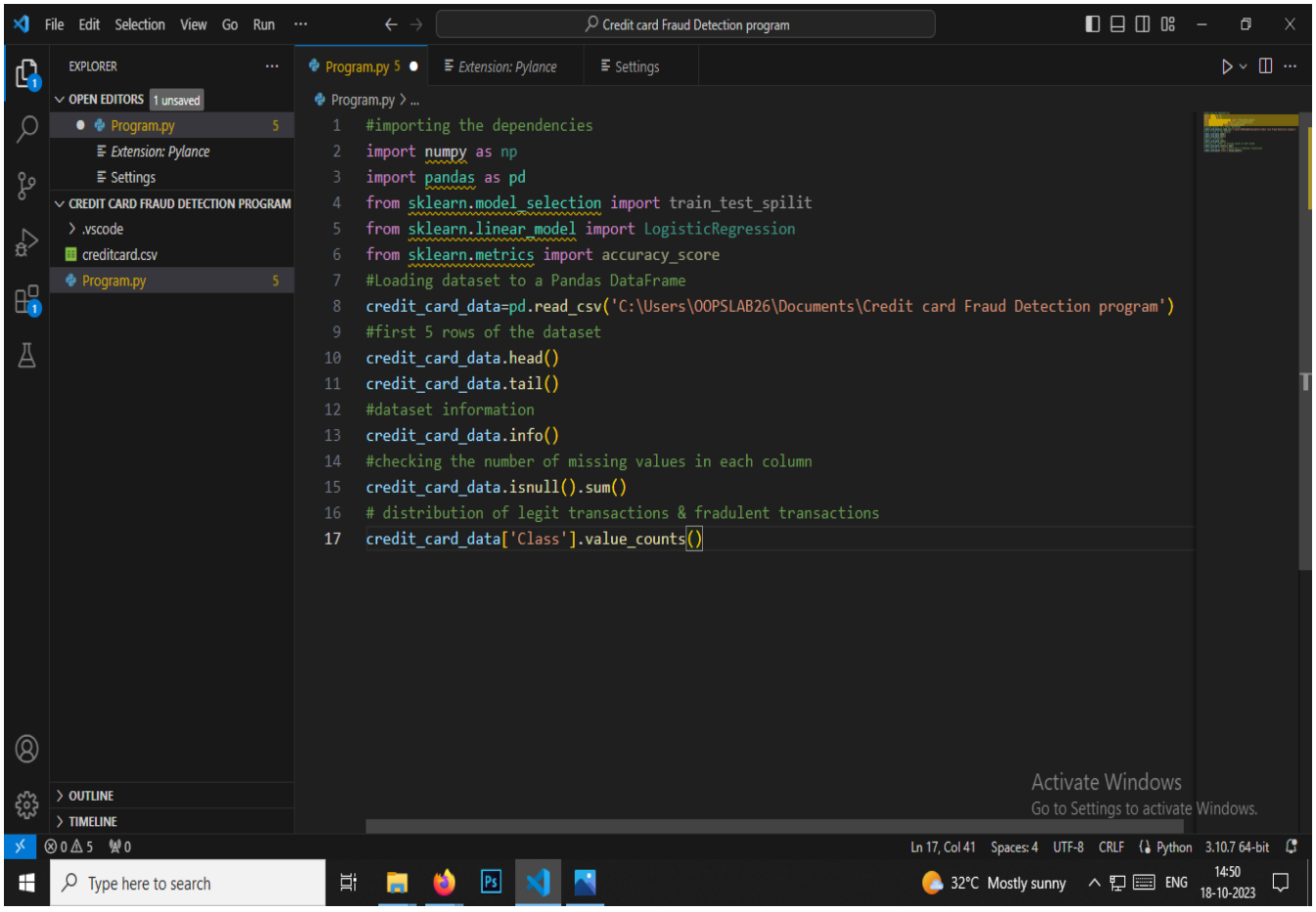


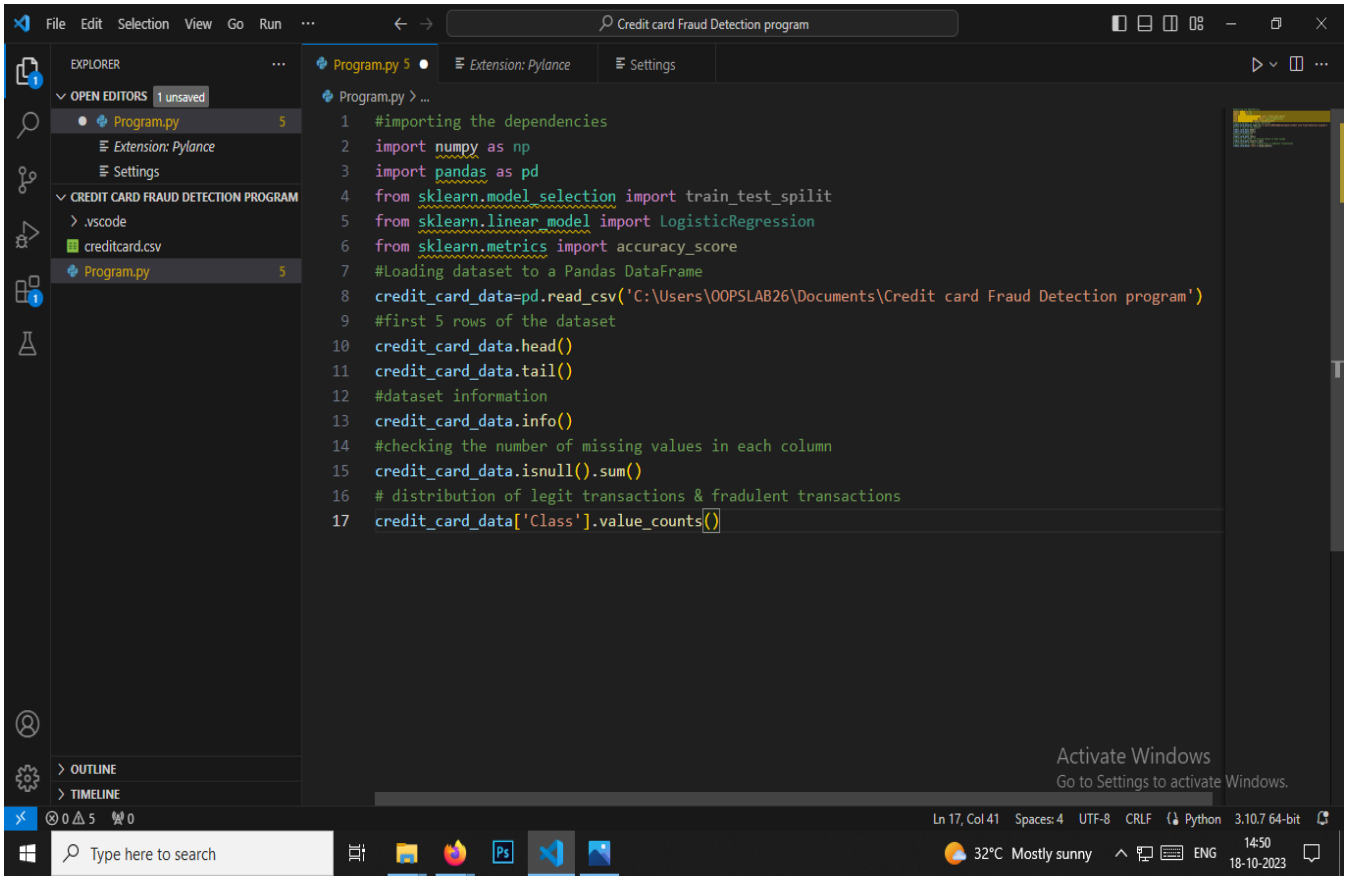


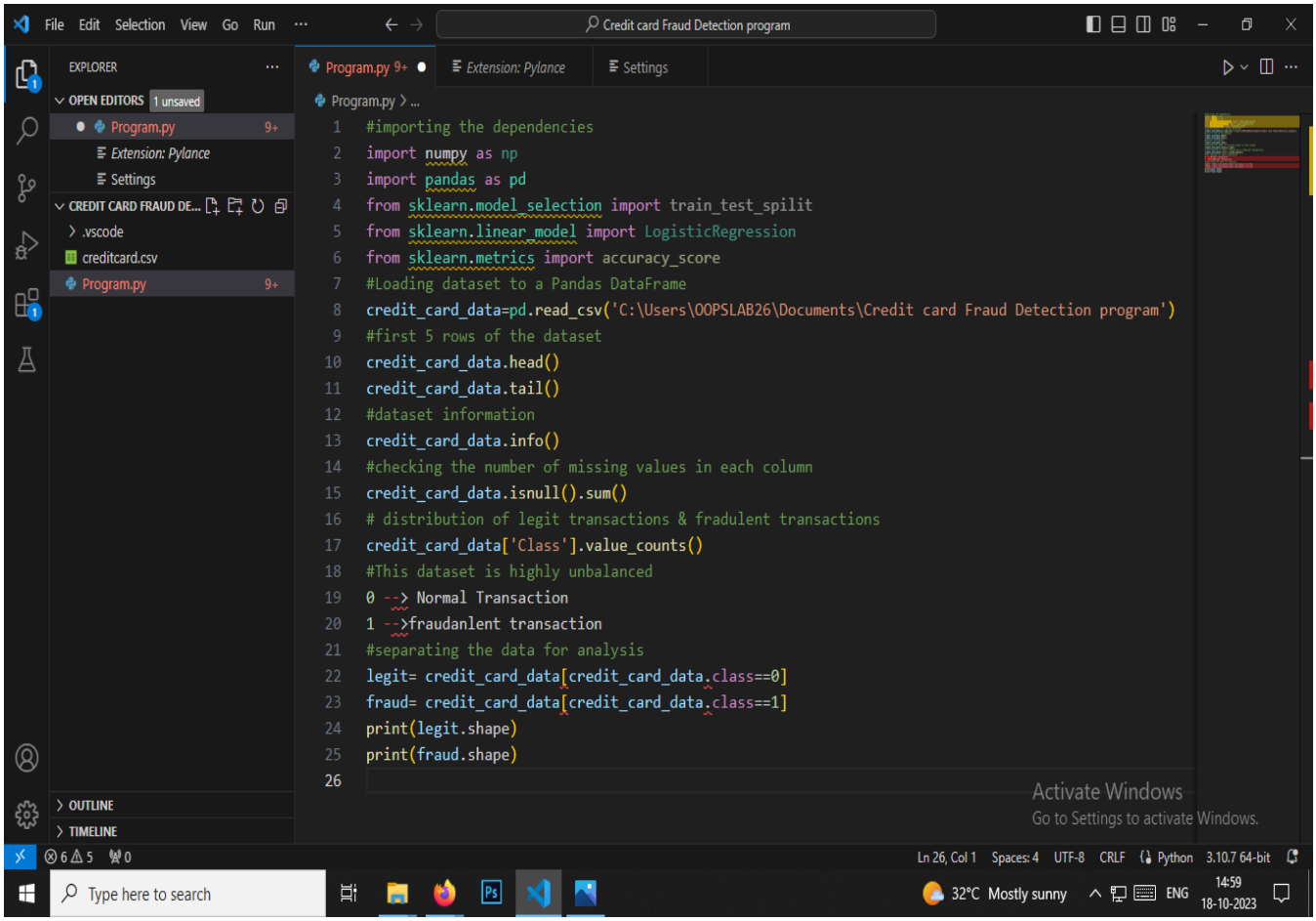

File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 5
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 5
Program.py 5
Extension: Pylance
Settings
Program.py > ...
#importing the dependencies
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_spilit
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
#Loading dataset to a Pandas DataFrame
credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
#first 5 rows of the dataset
credit_card_data.head()
credit_card_data.tail()
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 11, Col 24 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
14:38
18-10-2023

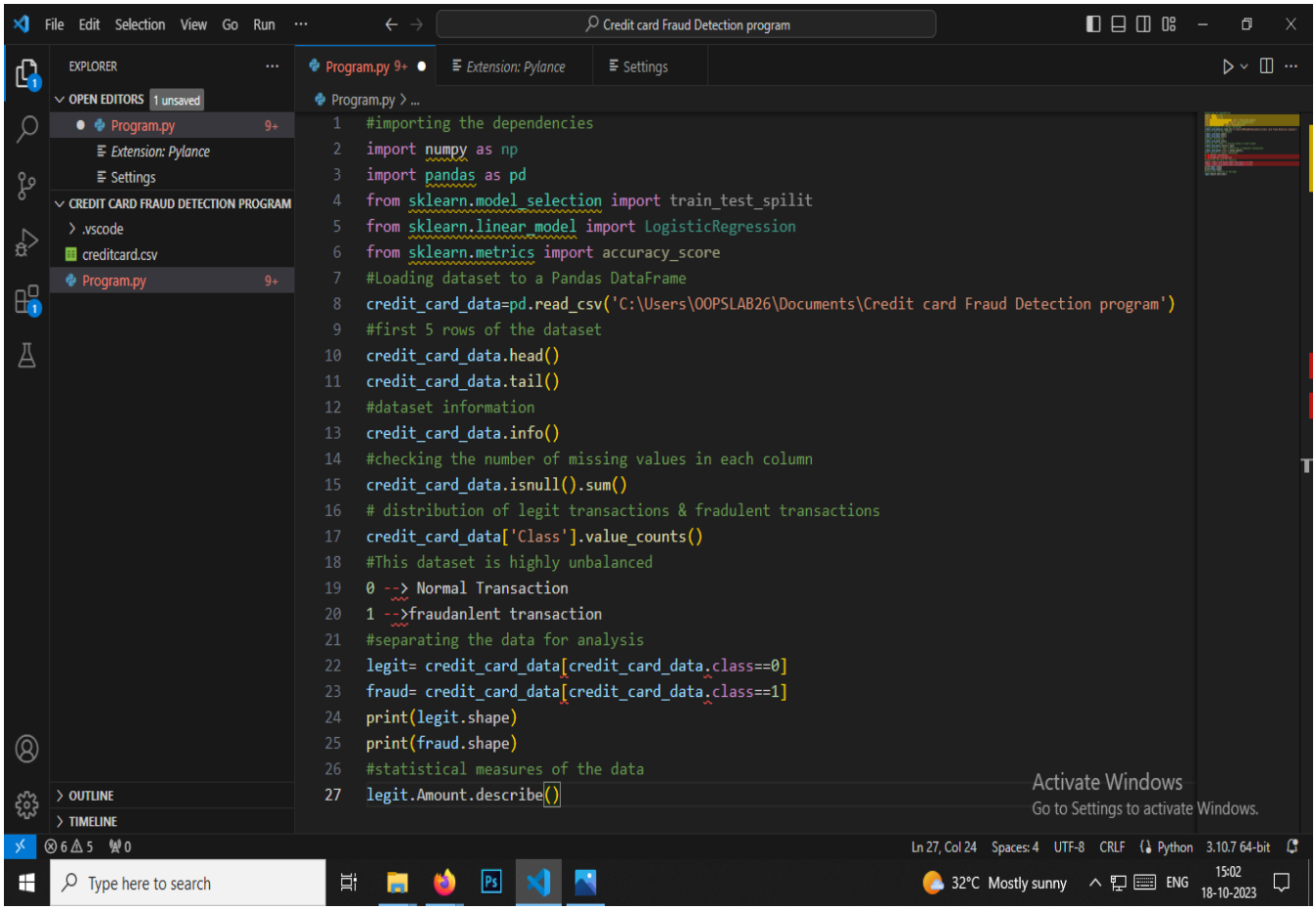

File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 5
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 5
Program.py 5
Extension: Pylance
Settings
Program.py > ...
#importing the dependencies
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_spilit
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
#Loading dataset to a Pandas DataFrame
credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
#first 5 rows of the dataset
credit_card_data.head()
credit_card_data.tail()
#dataset information
credit_card_data.info()
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 13, Col 23 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
14:40
18-10-2023

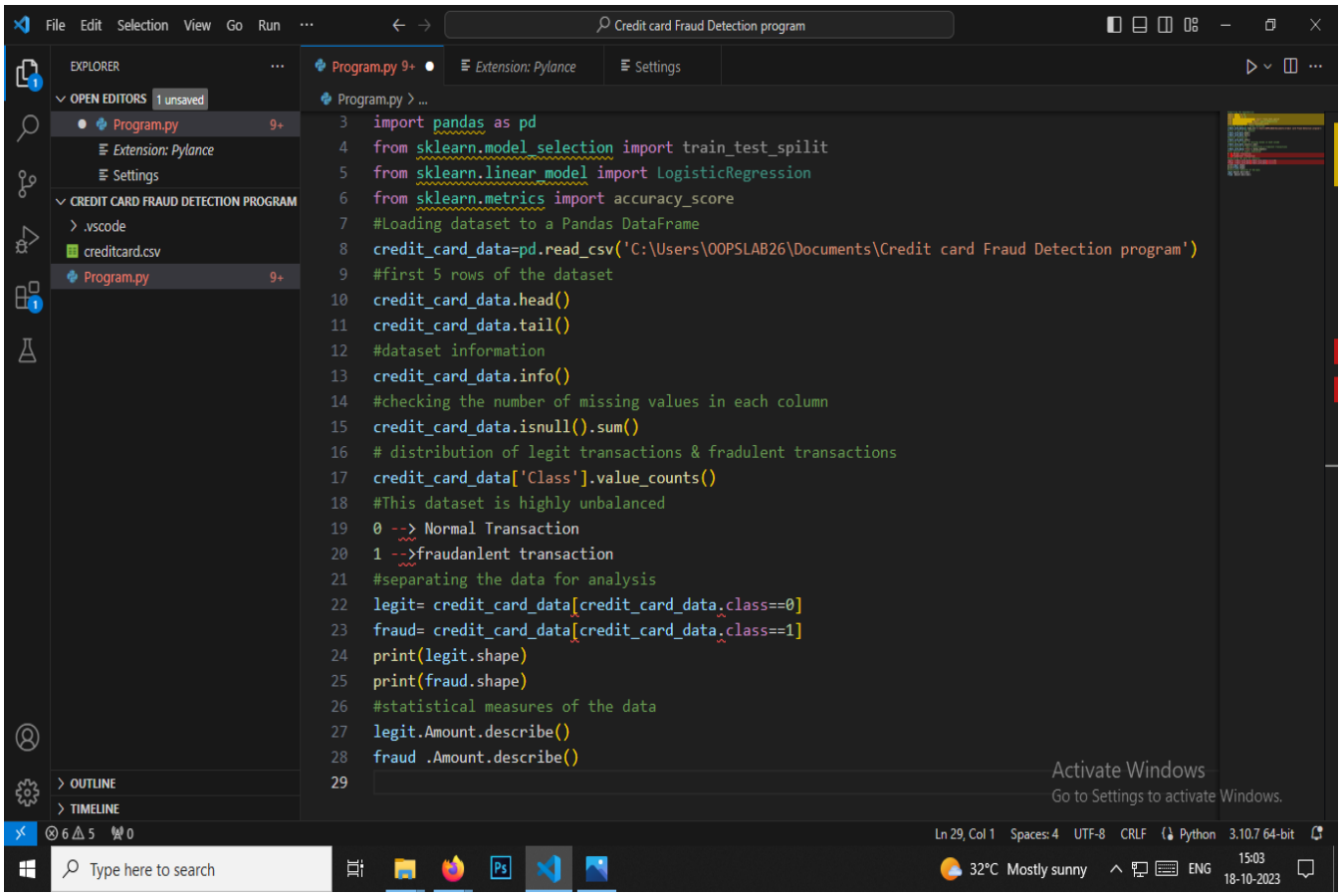

File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 5
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 5
Program.py > ...
#importing the dependencies
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_spilit
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
#Loading dataset to a Pandas DataFrame
credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
#first 5 rows of the dataset
credit_card_data.head()
credit_card_data.tail()
#dataset information
credit_card_data.info()
#checking the number of missing values in each column
credit_card_data.isnull().sum()
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 15, Col 32 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
14:42
18-10-2023

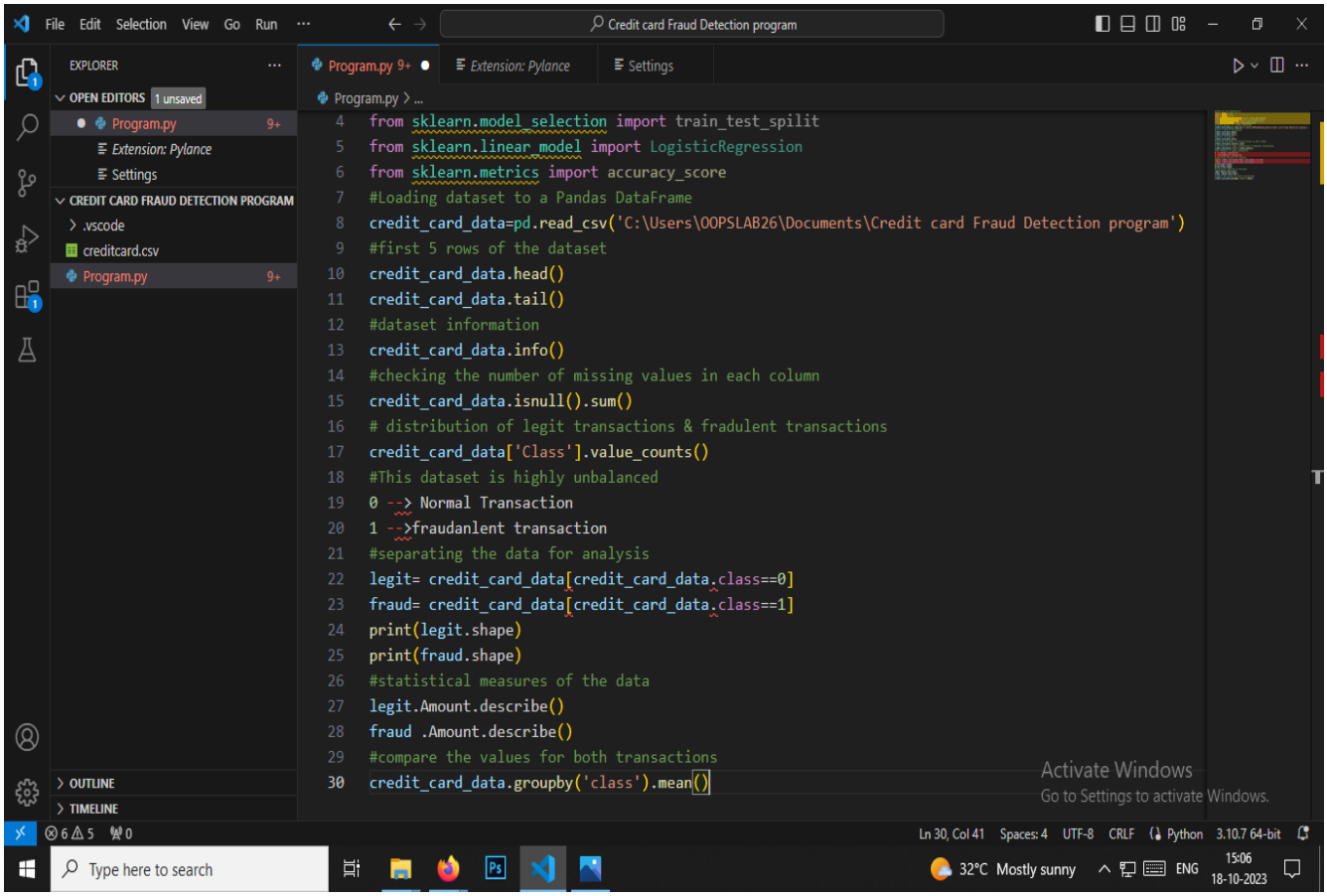

File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 5
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 5
Program.py > ...
#importing the dependencies
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_spilit
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
#Loading dataset to a Pandas DataFrame
credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
#first 5 rows of the dataset
credit_card_data.head()
credit_card_data.tail()
#dataset information
credit_card_data.info()
#checking the number of missing values in each column
credit_card_data.isnull().sum()
# distribution of legit transactions & fradulent transactions
credit_card_data['Class'].value_counts()
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 17, Col 41 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
14:50
18-10-2023

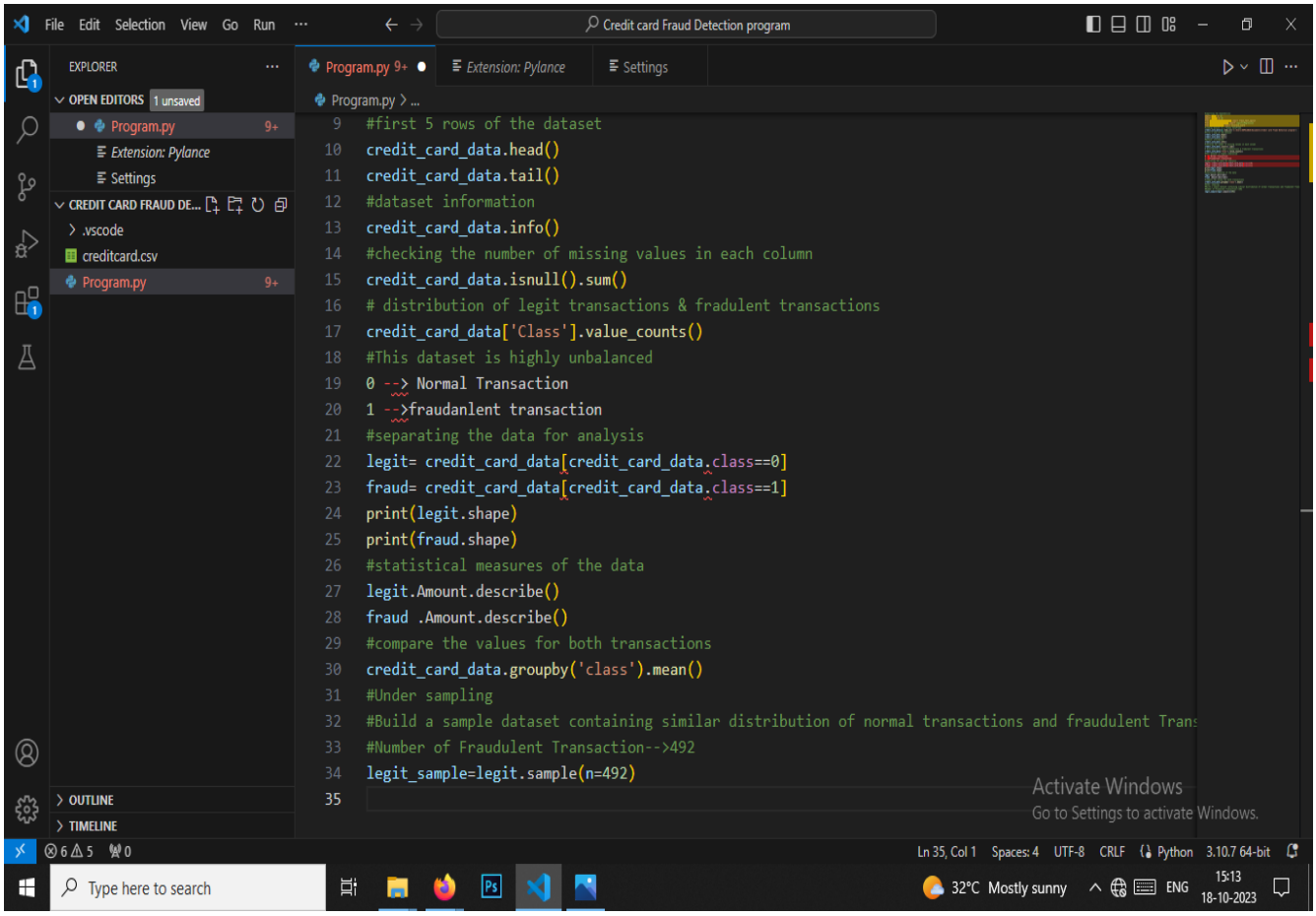

File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 5
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 5
Program.py > ...
#importing the dependencies
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_spilit
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
#Loading dataset to a Pandas DataFrame
credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
#first 5 rows of the dataset
credit_card_data.head()
credit_card_data.tail()
#dataset information
credit_card_data.info()
#checking the number of missing values in each column
credit_card_data.isnull().sum()
# distribution of legit transactions & fradulent transactions
credit_card_data['Class'].value_counts()
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 17, Col 41 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
14:50
18-10-2023

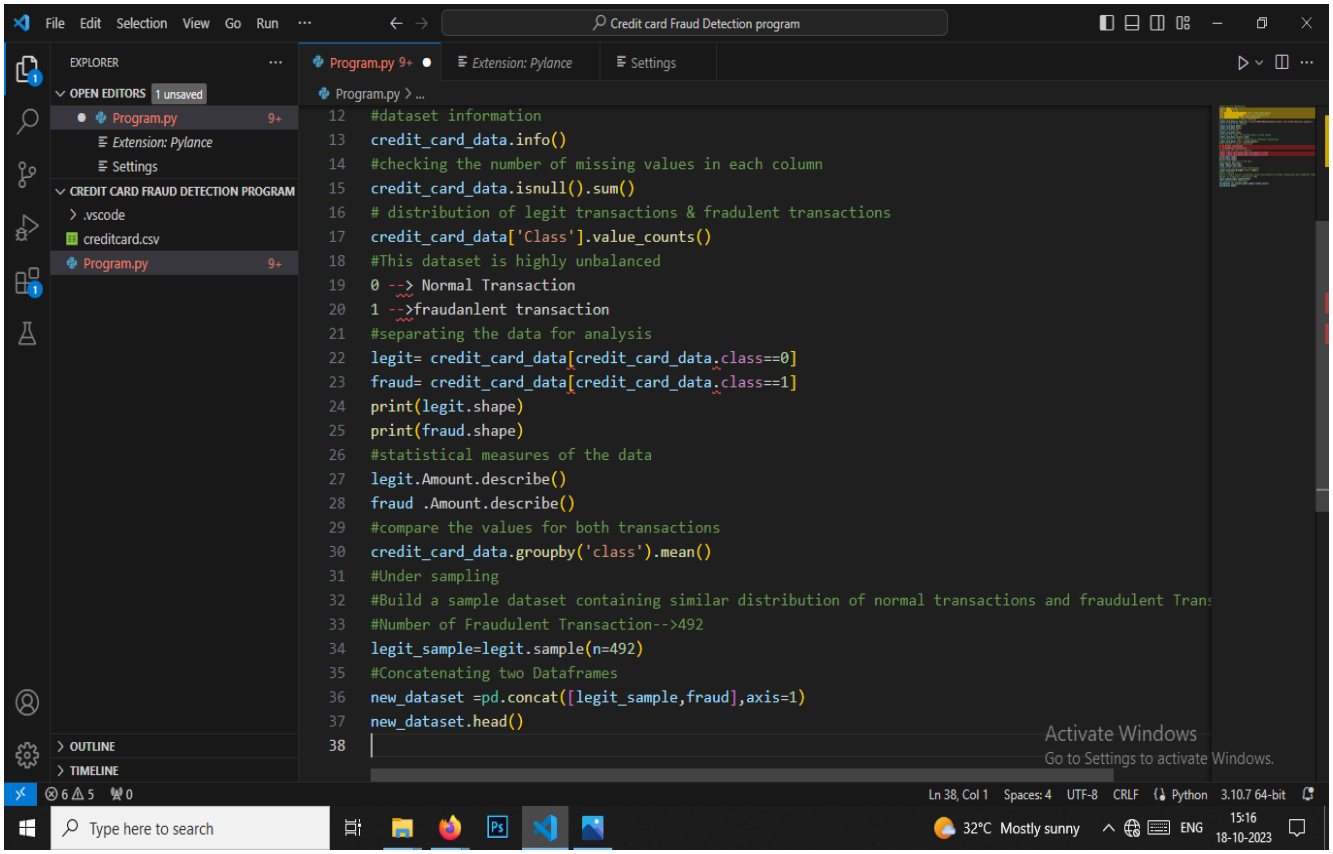

File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 9+
Extension: Pylance
Settings
CREDIT CARD FRAUD DE...
.vscode
creditcard.csv
Program.py 9+
Program.py > ...
#importing the dependencies
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_spilit
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
#Loading dataset to a Pandas DataFrame
credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
#first 5 rows of the dataset
credit_card_data.head()
credit_card_data.tail()
#dataset information
credit_card_data.info()
#checking the number of missing values in each column
credit_card_data.isnull().sum()
# distribution of legit transactions & fradulent transactions
credit_card_data['Class'].value_counts()
#This dataset is highly unbalanced
0 --> Normal Transaction
1 -->fraudanlent transaction
#separating the data for analysis
legit= credit_card_data[credit_card_data.class==0]
fraud= credit_card_data[credit_card_data.class==1]
print(legit.shape)
print(fraud.shape)
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 26, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
14:59
18-10-2023

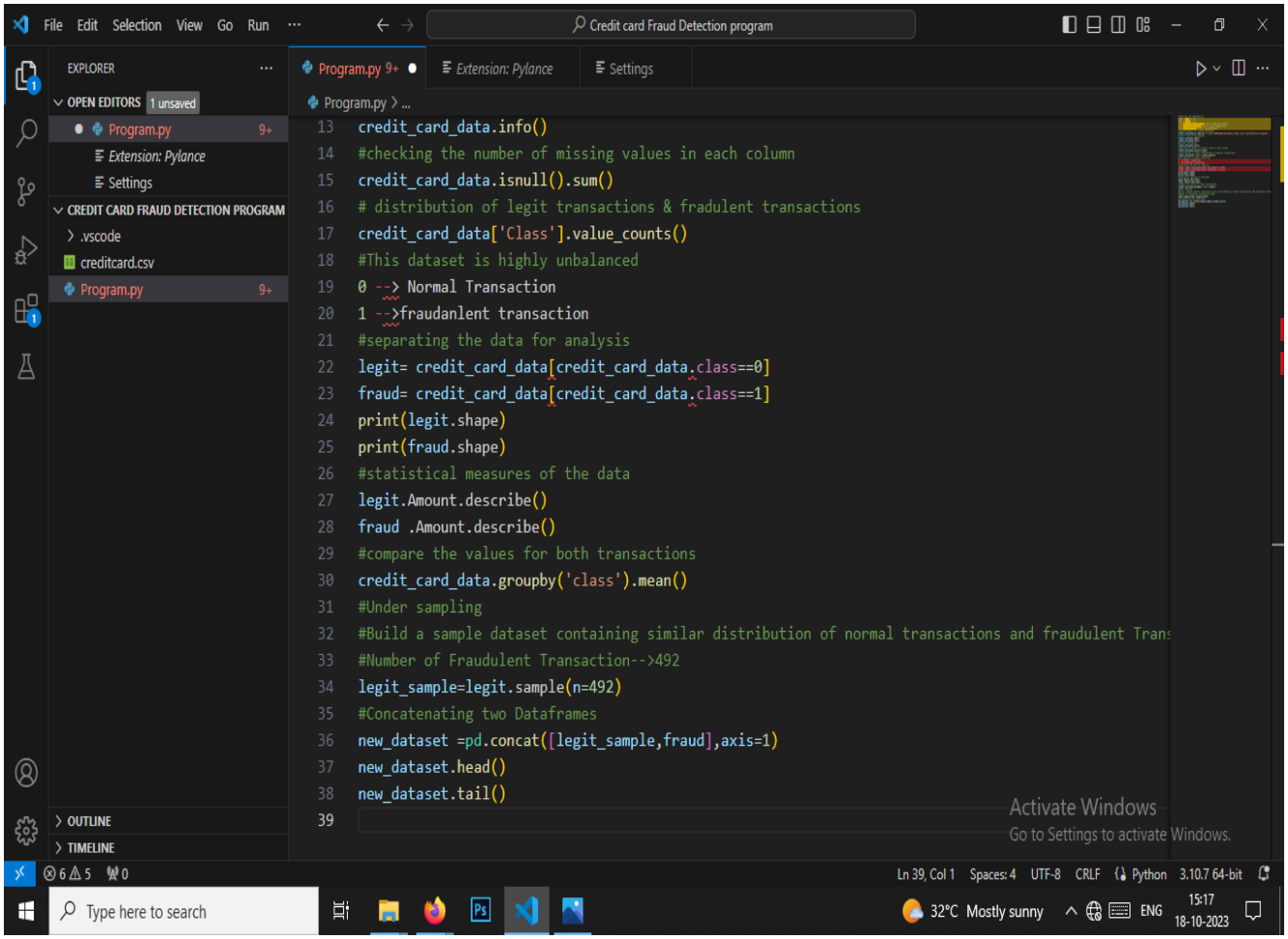

File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 9+
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 9+
Program.py 9+
Extension: Pylance
Settings
Program.py > ...
#importing the dependencies
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_spilit
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
#Loading dataset to a Pandas DataFrame
credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
#first 5 rows of the dataset
credit_card_data.head()
credit_card_data.tail()
#dataset information
credit_card_data.info()
#checking the number of missing values in each column
credit_card_data.isnull().sum()
# distribution of legit transactions & fradulent transactions
credit_card_data['Class'].value_counts()
#This dataset is highly unbalanced
0 --> Normal Transaction
1 -->fraudanlent transaction
#separating the data for analysis
legit= credit_card_data[credit_card_data.class==0]
fraud= credit_card_data[credit_card_data.class==1]
print(legit.shape)
print(fraud.shape)
#statistical measures of the data
legit.Amount.describe()
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 27, Col 24 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
15:02
18-10-2023

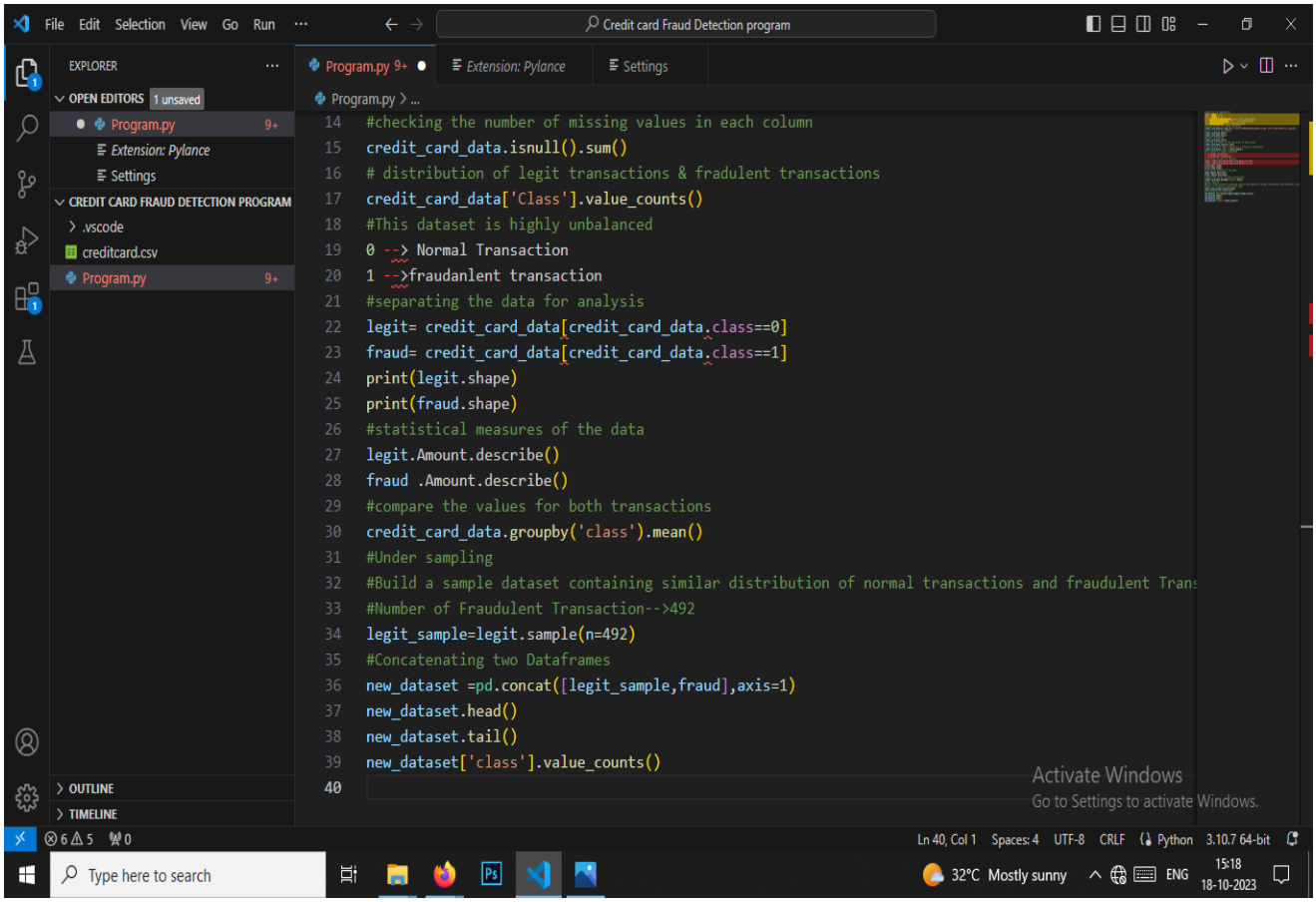

File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 9+
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 9+
Program.py 9+
Extension: Pylance
Settings
Program.py > ...
import pandas as pd
from sklearn.model_selection import train_test_spilit
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
#Loading dataset to a Pandas DataFrame
credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
#first 5 rows of the dataset
credit_card_data.head()
credit_card_data.tail()
#dataset information
credit_card_data.info()
#checking the number of missing values in each column
credit_card_data.isnull().sum()
# distribution of legit transactions & fradulent transactions
credit_card_data['Class'].value_counts()
#This dataset is highly unbalanced
0 --> Normal Transaction
1 -->fraudanlent transaction
#separating the data for analysis
legit= credit_card_data[credit_card_data.class==0]
fraud= credit_card_data[credit_card_data.class==1]
print(legit.shape)
print(fraud.shape)
#statistical measures of the data
legit.Amount.describe()
fraud .Amount.describe()
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 29, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
15:03
18-10-2023

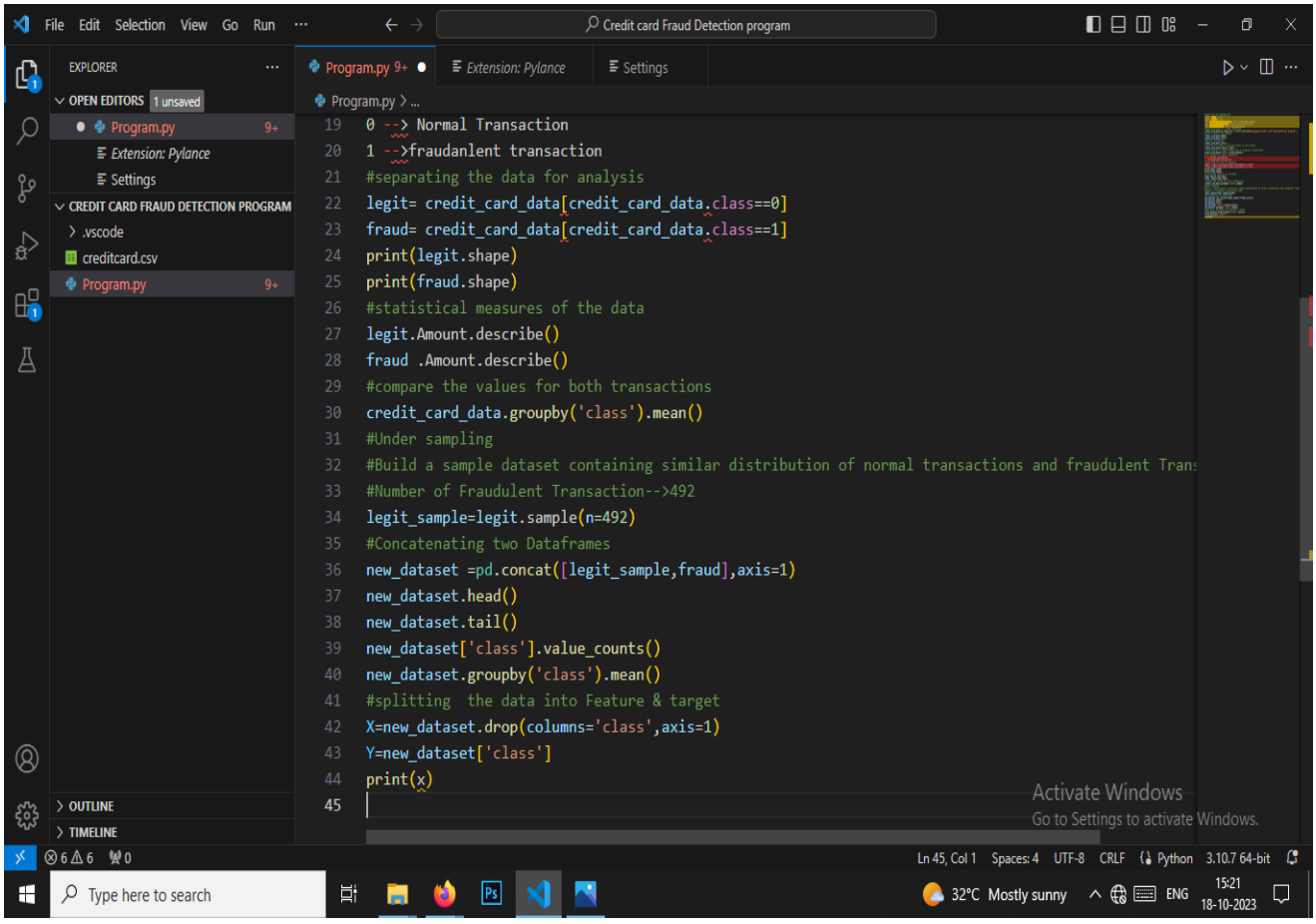

File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 9+
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 9+
Program.py 9+
Extension: Pylance
Settings
Program.py > ...
from sklearn.model_selection import train_test_spilit
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
#Loading dataset to a Pandas DataFrame
credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
#first 5 rows of the dataset
credit_card_data.head()
credit_card_data.tail()
#dataset information
credit_card_data.info()
#checking the number of missing values in each column
credit_card_data.isnull().sum()
# distribution of legit transactions & fradulent transactions
credit_card_data['Class'].value_counts()
#This dataset is highly unbalanced
0 --> Normal Transaction
1 -->fraudanlent transaction
#separating the data for analysis
legit= credit_card_data[credit_card_data.class==0]
fraud= credit_card_data[credit_card_data.class==1]
print(legit.shape)
print(fraud.shape)
#statistical measures of the data
legit.Amount.describe()
fraud .Amount.describe()
#compare the values for both transactions
credit_card_data.groupby('class').mean()
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 30, Col 41 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
15:06
18-10-2023

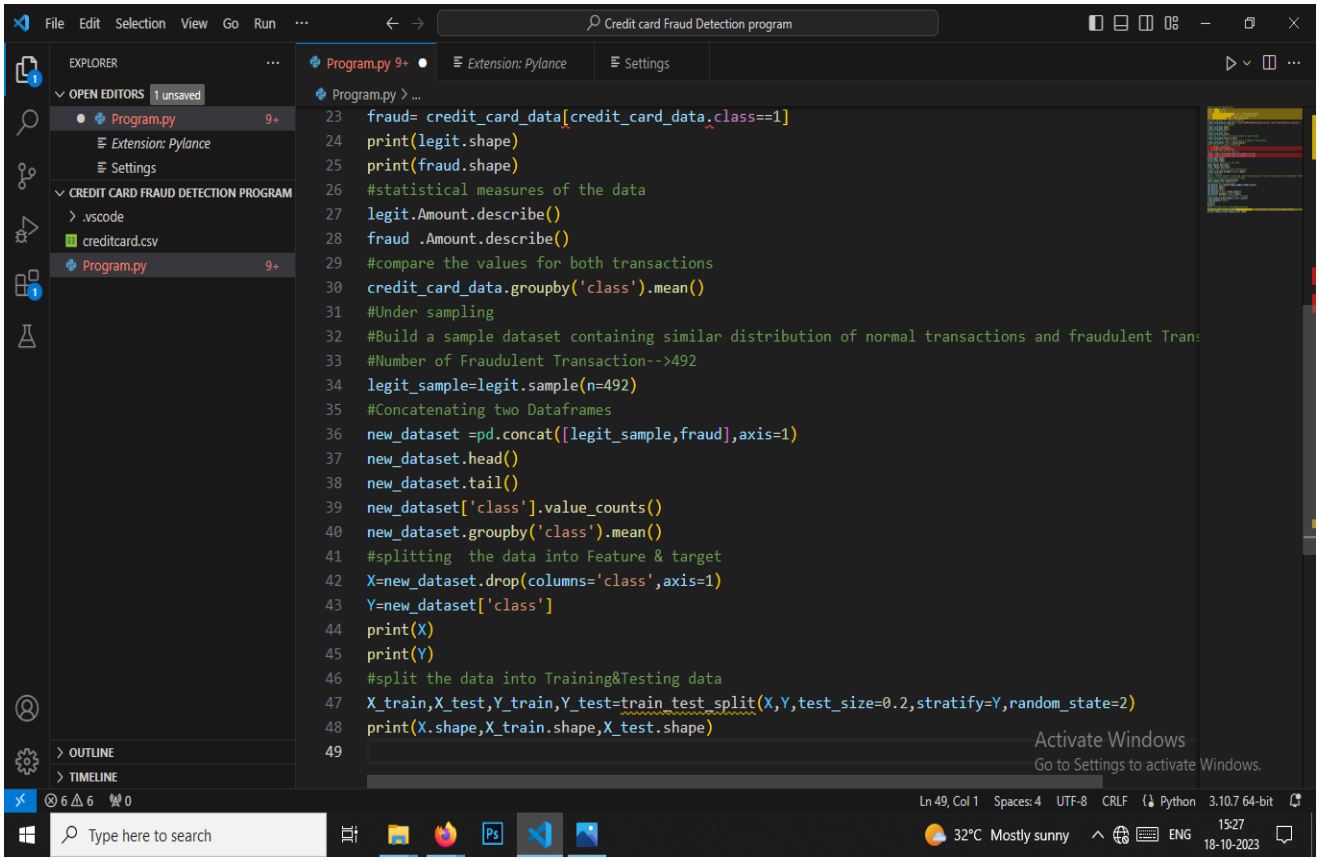

File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 9+
Extension: Pylance
Settings
CREDIT CARD FRAUD DE...
.vscode
creditcard.csv
Program.py 9+
Program.py 9+
Extension: Pylance
Settings
Program.py > ...
#first 5 rows of the dataset
credit_card_data.head()
credit_card_data.tail()
#dataset information
credit_card_data.info()
#checking the number of missing values in each column
credit_card_data.isnull().sum()
# distribution of legit transactions & fradulent transactions
credit_card_data['Class'].value_counts()
#This dataset is highly unbalanced
0 --> Normal Transaction
1 -->fraudanlent transaction
#separating the data for analysis
legit= credit_card_data[credit_card_data.class==0]
fraud= credit_card_data[credit_card_data.class==1]
print(legit.shape)
print(fraud.shape)
#statistical measures of the data
legit.Amount.describe()
fraud .Amount.describe()
#compare the values for both transactions
credit_card_data.groupby('class').mean()
#Under sampling
#Build a sample dataset containing similar distribution of normal transactions and fraudulent Trans
#Number of Fraudulent Transaction-->492
legit_sample=legit.sample(n=492)
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 35, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
15:13
18-10-2023

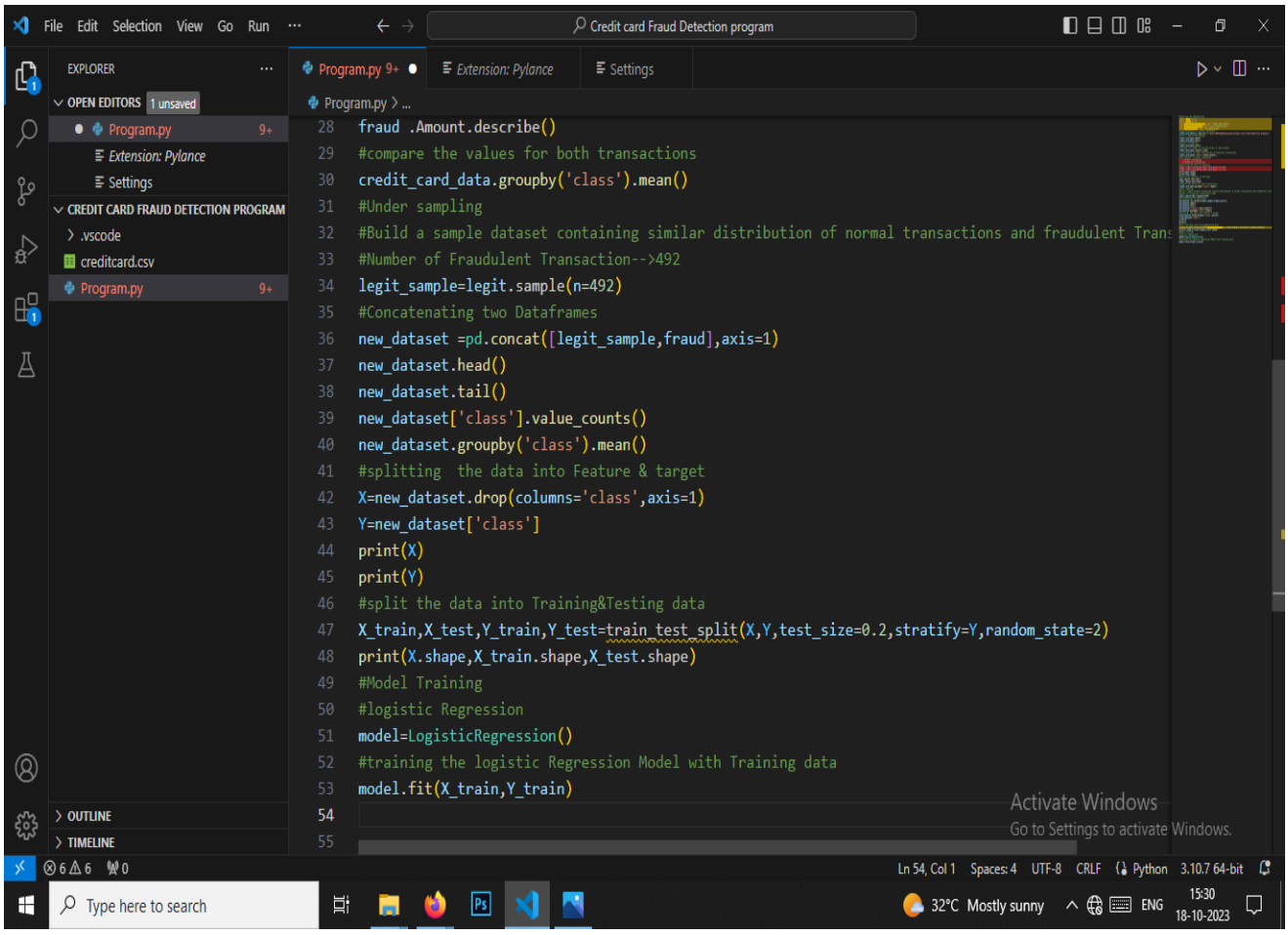

File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 9+
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 9+
Program.py > ...
#dataset information
credit_card_data.info()
#checking the number of missing values in each column
credit_card_data.isnull().sum()
# distribution of legit transactions & fradulent transactions
credit_card_data['Class'].value_counts()
#This dataset is highly unbalanced
0 --> Normal Transaction
1 -->fraudanlent transaction
#separating the data for analysis
legit= credit_card_data[credit_card_data.class==0]
fraud= credit_card_data[credit_card_data.class==1]
print(legit.shape)
print(fraud.shape)
#statistical measures of the data
legit.Amount.describe()
fraud .Amount.describe()
#compare the values for both transactions
credit_card_data.groupby('class').mean()
#Under sampling
#Build a sample dataset containing similar distribution of normal transactions and fraudulent Trans
#Number of Fraudulent Transaction-->492
legit_sample=legit.sample(n=492)
#Concatenating two Dataframes
new_dataset =pd.concat([legit_sample,fraud],axis=1)
new_dataset.head()
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 38, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
15:16
18-10-2023


File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 9+
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 9+
Program.py > ...
credit_card_data.info()
#checking the number of missing values in each column
credit_card_data.isnull().sum()
# distribution of legit transactions & fradulent transactions
credit_card_data['Class'].value_counts()
#This dataset is highly unbalanced
0 --> Normal Transaction
1 -->fraudanlent transaction
#separating the data for analysis
legit= credit_card_data[credit_card_data.class==0]
fraud= credit_card_data[credit_card_data.class==1]
print(legit.shape)
print(fraud.shape)
#statistical measures of the data
legit.Amount.describe()
fraud .Amount.describe()
#compare the values for both transactions
credit_card_data.groupby('class').mean()
#Under sampling
#Build a sample dataset containing similar distribution of normal transactions and fraudulent Trans
#Number of Fraudulent Transaction-->492
legit_sample=legit.sample(n=492)
#Concatenating two Dataframes
new_dataset =pd.concat([legit_sample,fraud],axis=1)
new_dataset.head()
new_dataset.tail()
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 39, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
15:17
18-10-2023


File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 9+
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 9+
Program.py 9+
Extension: Pylance
Settings
Program.py > ...
#checking the number of missing values in each column
credit_card_data.isnull().sum()
# distribution of legit transactions & fradulent transactions
credit_card_data['Class'].value_counts()
#This dataset is highly unbalanced
0 --> Normal Transaction
1 -->fraudanlent transaction
#separating the data for analysis
legit= credit_card_data[credit_card_data.class==0]
fraud= credit_card_data[credit_card_data.class==1]
print(legit.shape)
print(fraud.shape)
#statistical measures of the data
legit.Amount.describe()
fraud .Amount.describe()
#compare the values for both transactions
credit_card_data.groupby('class').mean()
#Under sampling
#Build a sample dataset containing similar distribution of normal transactions and fraudulent Trans
#Number of Fraudulent Transaction-->492
legit_sample=legit.sample(n=492)
#Concatenating two Dataframes
new_dataset =pd.concat([legit_sample,fraud],axis=1)
new_dataset.head()
new_dataset.tail()
new_dataset['class'].value_counts()
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 40, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
15:18
18-10-2023


File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 9+
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 9+
Program.py 9+
Extension: Pylance
Settings
Program.py > ...
0 --> Normal Transaction
1 -->fraudanlent transaction
#separating the data for analysis
legit= credit_card_data[credit_card_data.class==0]
fraud= credit_card_data[credit_card_data.class==1]
print(legit.shape)
print(fraud.shape)
#statistical measures of the data
legit.Amount.describe()
fraud .Amount.describe()
#compare the values for both transactions
credit_card_data.groupby('class').mean()
#Under sampling
#Build a sample dataset containing similar distribution of normal transactions and fraudulent Trans
#Number of Fraudulent Transaction-->492
legit_sample=legit.sample(n=492)
#Concatenating two Dataframes
new_dataset =pd.concat([legit_sample,fraud],axis=1)
new_dataset.head()
new_dataset.tail()
new_dataset['class'].value_counts()
new_dataset.groupby('class').mean()
#splitting the data into Feature & target
X=new_dataset.drop(columns='class',axis=1)
Y=new_dataset['class']
print(x)
OUTLINE
TIMELINE
Activate Windows
Go to Settings to activate Windows.
Ln 45, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
15:21
18-10-2023


File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 9+
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 9+
Program.py 9+
Extension: Pylance
Settings
Program.py > ...
23 fraud= credit_card_data[credit_card_data.class==1]
24 print(legit.shape)
25 print(fraud.shape)
26 #statistical measures of the data
27 legit.Amount.describe()
28 fraud .Amount.describe()
29 #compare the values for both transactions
30 credit_card_data.groupby('class').mean()
31 #Under sampling
32 #Build a sample dataset containing similar distribution of normal transactions and fraudulent Trans
33 #Number of Fraudulent Transaction-->492
34 legit_sample=legit.sample(n=492)
35 #Concatenating two Dataframes
36 new_dataset =pd.concat([legit_sample,fraud],axis=1)
37 new_dataset.head()
38 new_dataset.tail()
39 new_dataset['class'].value_counts()
40 new_dataset.groupby('class').mean()
41 #splitting the data into Feature & target
42 X=new_dataset.drop(columns='class',axis=1)
43 Y=new_dataset['class']
44 print(X)
45 print(Y)
46 #split the data into Training&Testing data
47 X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.2,stratify=Y,random_state=2)
48 print(X.shape,X_train.shape,X_test.shape)
49
Activate Windows
Go to Settings to activate Windows.
OUTLINE
TIMELINE
Ln 49, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
15:27
18-10-2023


File Edit Selection View Go Run
Credit card Fraud Detection program
EXPLORER
OPEN EDITORS 1 unsaved
Program.py 9+
Extension: Pylance
Settings
CREDIT CARD FRAUD DETECTION PROGRAM
.vscode
creditcard.csv
Program.py 9+
Program.py 9+
Extension: Pylance
Settings
Program.py > ...
28 fraud .Amount.describe()
29 #compare the values for both transactions
30 credit_card_data.groupby('class').mean()
31 #Under sampling
32 #Build a sample dataset containing similar distribution of normal transactions and fraudulent Trans
33 #Number of Fraudulent Transaction-->492
34 legit_sample=legit.sample(n=492)
35 #Concatenating two Dataframes
36 new_dataset =pd.concat([legit_sample,fraud],axis=1)
37 new_dataset.head()
38 new_dataset.tail()
39 new_dataset['class'].value_counts()
40 new_dataset.groupby('class').mean()
41 #splitting the data into Feature & target
42 X=new_dataset.drop(columns='class',axis=1)
43 Y=new_dataset['class']
44 print(X)
45 print(Y)
46 #split the data into Training&Testing data
47 X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.2,stratify=Y,random_state=2)
48 print(X.shape,X_train.shape,X_test.shape)
49 #Model Training
50 #logistic Regression
51 model=LogisticRegression()
52 #training the logistic Regression Model with Training data
53 model.fit(X_train,Y_train)
54
55
Activate Windows
Go to Settings to activate Windows.
OUTLINE
TIMELINE
Ln 54, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit
Type here to search
32°C Mostly sunny
ENG
15:30
18-10-2023

#Statistical Analysis:

- ❖ Performing exploratory data analysis
- ❖ Statistical analysis

✓ So these all are credit card fraud detection is analyzed.