

# 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:

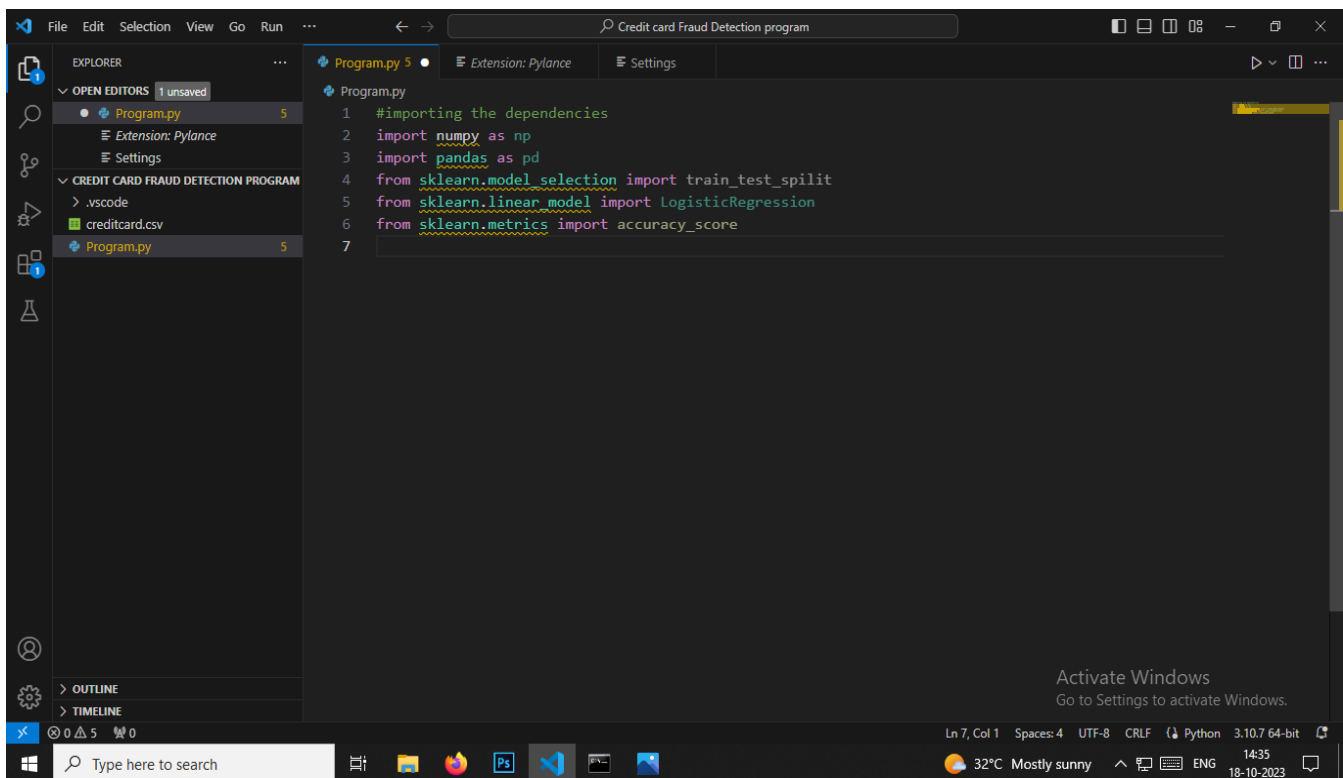
- ❖ Feature Engineering
- ❖ Model Training
- ❖ Evaluation

### Dataset Link:

<https://www.kaggle.com/datasets/mlgulg/creditcardfraud>

- ❖ Performing exploratory data analysis
- ❖ Statistical analysis

### #Performing exploratory data analysis:

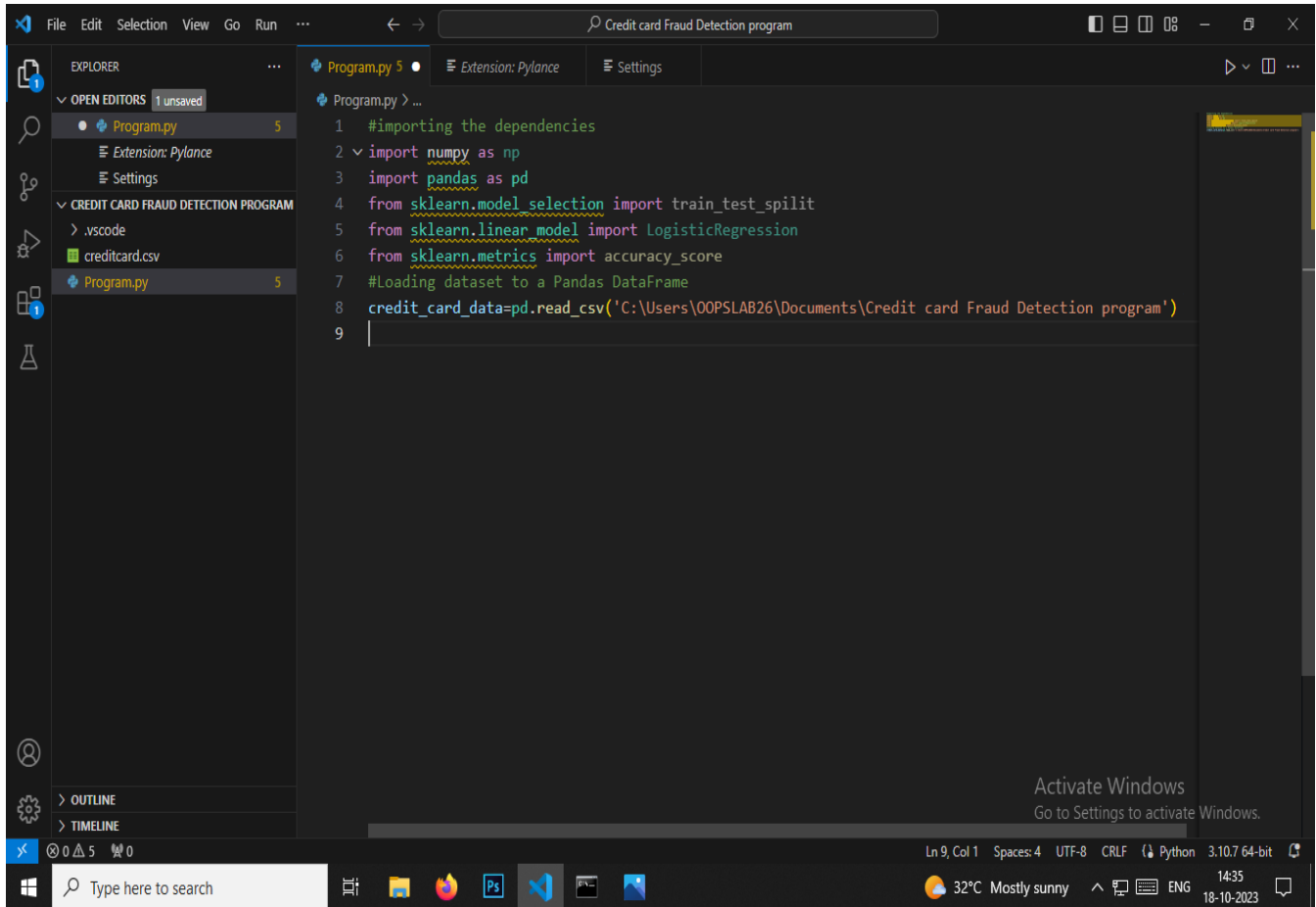


The screenshot shows a Visual Studio Code editor window titled "Credit card Fraud Detection program". The Explorer sidebar on the left shows a project named "CREDIT CARD FRAUD DETECTION PROGRAM" with files ".vscode", "creditcard.csv", and "Program.py". The main editor area displays the content of "Program.py", which includes the following code:

```
1 #importing the dependencies
2 import numpy as np
3 import pandas as pd
4 from sklearn.model_selection import train_test_split
5 from sklearn.linear_model import LogisticRegression
6 from sklearn.metrics import accuracy_score
7
```

The status bar at the bottom indicates the current line and column (Ln 7, Col 1), encoding (UTF-8), line endings (CRLF), Python version (3.10.7 64-bit), and system information (32°C Mostly sunny, 14:35, 18-10-2023).

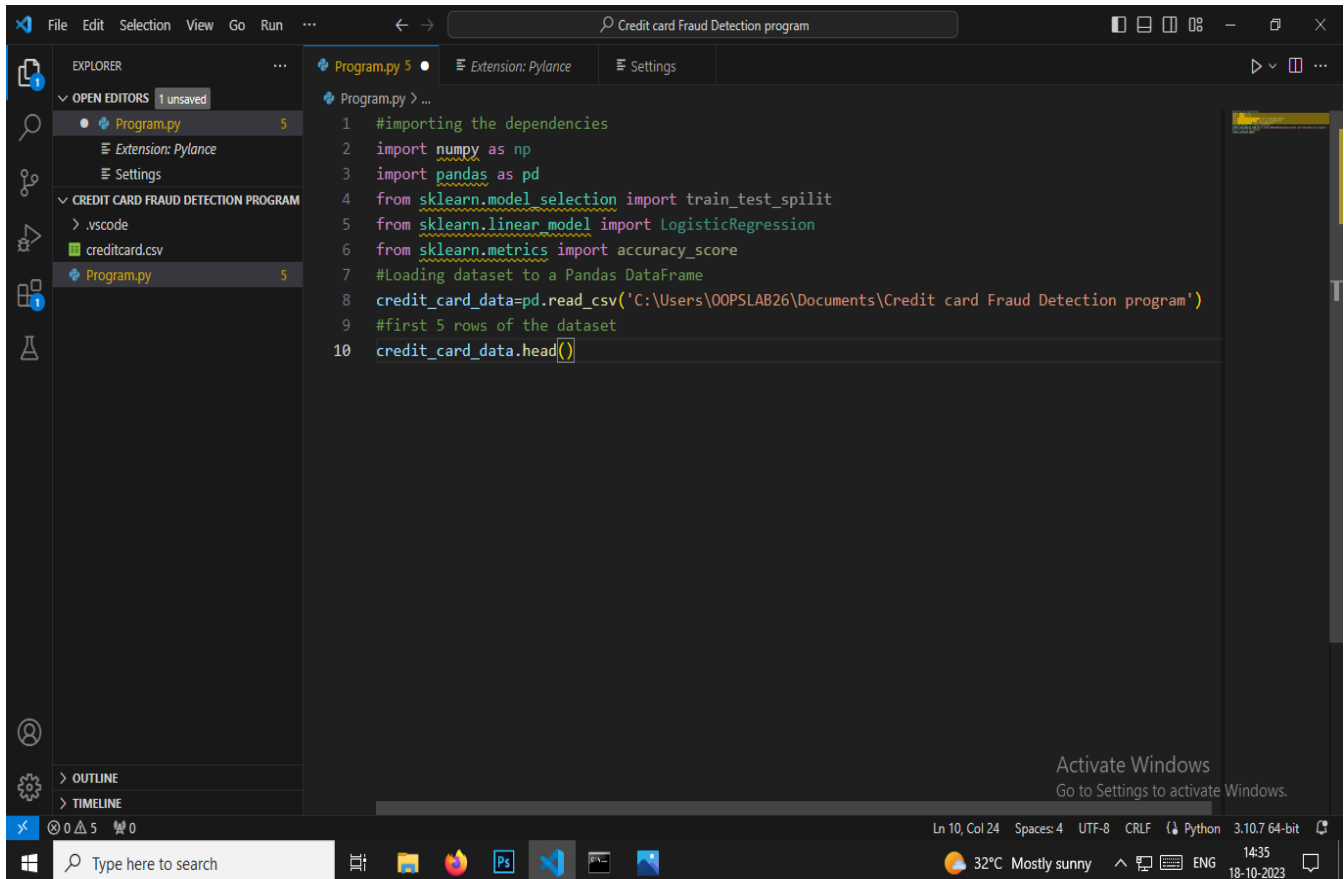
## #Data describe :



VS Code editor interface showing a Python script for credit card fraud detection. The Explorer sidebar on the left shows the project structure: 'OPEN EDITORS' (1 unsaved), 'Program.py' (5 lines), 'Extension: Pylance', 'Settings', 'CREDIT CARD FRAUD DETECTION PROGRAM', '.vscode', 'creditcard.csv', and 'Program.py' (5 lines). The main editor displays the following code:

```
1 #importing the dependencies
2 import numpy as np
3 import pandas as pd
4 from sklearn.model_selection import train_test_split
5 from sklearn.linear_model import LogisticRegression
6 from sklearn.metrics import accuracy_score
7 #Loading dataset to a Pandas DataFrame
8 credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
9
```

The status bar at the bottom indicates: Ln 9, Col 1, Spaces: 4, UTF-8, CRLF, Python, 3.10.7 64-bit. The system tray shows 32°C Mostly sunny, 14:35, 18-10-2023.



VS Code editor interface showing the same Python script as above, but with an additional line to display the first 5 rows of the dataset using 'credit\_card\_data.head()'. The Explorer sidebar on the left shows the project structure: 'OPEN EDITORS' (1 unsaved), 'Program.py' (5 lines), 'Extension: Pylance', 'Settings', 'CREDIT CARD FRAUD DETECTION PROGRAM', '.vscode', 'creditcard.csv', and 'Program.py' (5 lines). The main editor displays the following code:

```
1 #importing the dependencies
2 import numpy as np
3 import pandas as pd
4 from sklearn.model_selection import train_test_split
5 from sklearn.linear_model import LogisticRegression
6 from sklearn.metrics import accuracy_score
7 #Loading dataset to a Pandas DataFrame
8 credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
9 #first 5 rows of the dataset
10 credit_card_data.head()
```

The status bar at the bottom indicates: Ln 10, Col 24, Spaces: 4, UTF-8, CRLF, Python, 3.10.7 64-bit. The system tray shows 32°C Mostly sunny, 14:35, 18-10-2023.

VS Code interface showing a Python script for credit card fraud detection. The Explorer sidebar shows the project structure with files like `creditcard.csv` and `Program.py`. The main editor displays the following code:

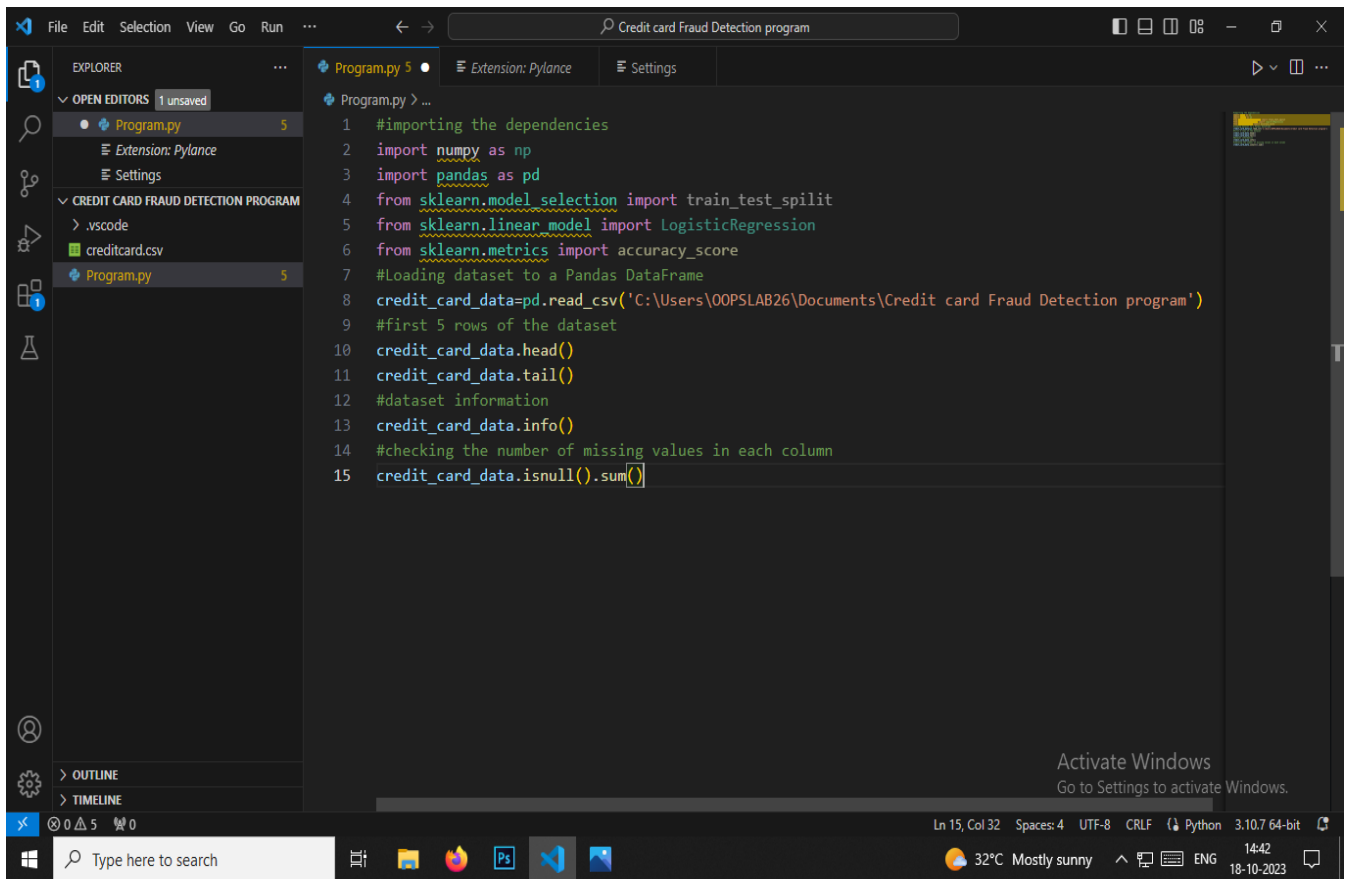
```
1 #importing the dependencies
2 import numpy as np
3 import pandas as pd
4 from sklearn.model_selection import train_test_split
5 from sklearn.linear_model import LogisticRegression
6 from sklearn.metrics import accuracy_score
7 #Loading dataset to a Pandas DataFrame
8 credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
9 #first 5 rows of the dataset
10 credit_card_data.head()
11 credit_card_data.tail()
```

The status bar at the bottom indicates the current position is Line 11, Column 24, with 4 spaces, UTF-8 encoding, CRLF line endings, and Python 3.10.7 64-bit. The system tray shows a temperature of 32°C and the date 18-10-2023.

VS Code interface showing the same Python script as the first image, but with an additional line to display dataset information. The Explorer sidebar shows the project structure with files like `creditcard.csv` and `Program.py`. The main editor displays the following code:

```
1 #importing the dependencies
2 import numpy as np
3 import pandas as pd
4 from sklearn.model_selection import train_test_split
5 from sklearn.linear_model import LogisticRegression
6 from sklearn.metrics import accuracy_score
7 #Loading dataset to a Pandas DataFrame
8 credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
9 #first 5 rows of the dataset
10 credit_card_data.head()
11 credit_card_data.tail()
12 #dataset information
13 credit_card_data.info()
```

The status bar at the bottom indicates the current position is Line 13, Column 23, with 4 spaces, UTF-8 encoding, CRLF line endings, and Python 3.10.7 64-bit. The system tray shows a temperature of 32°C and the date 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

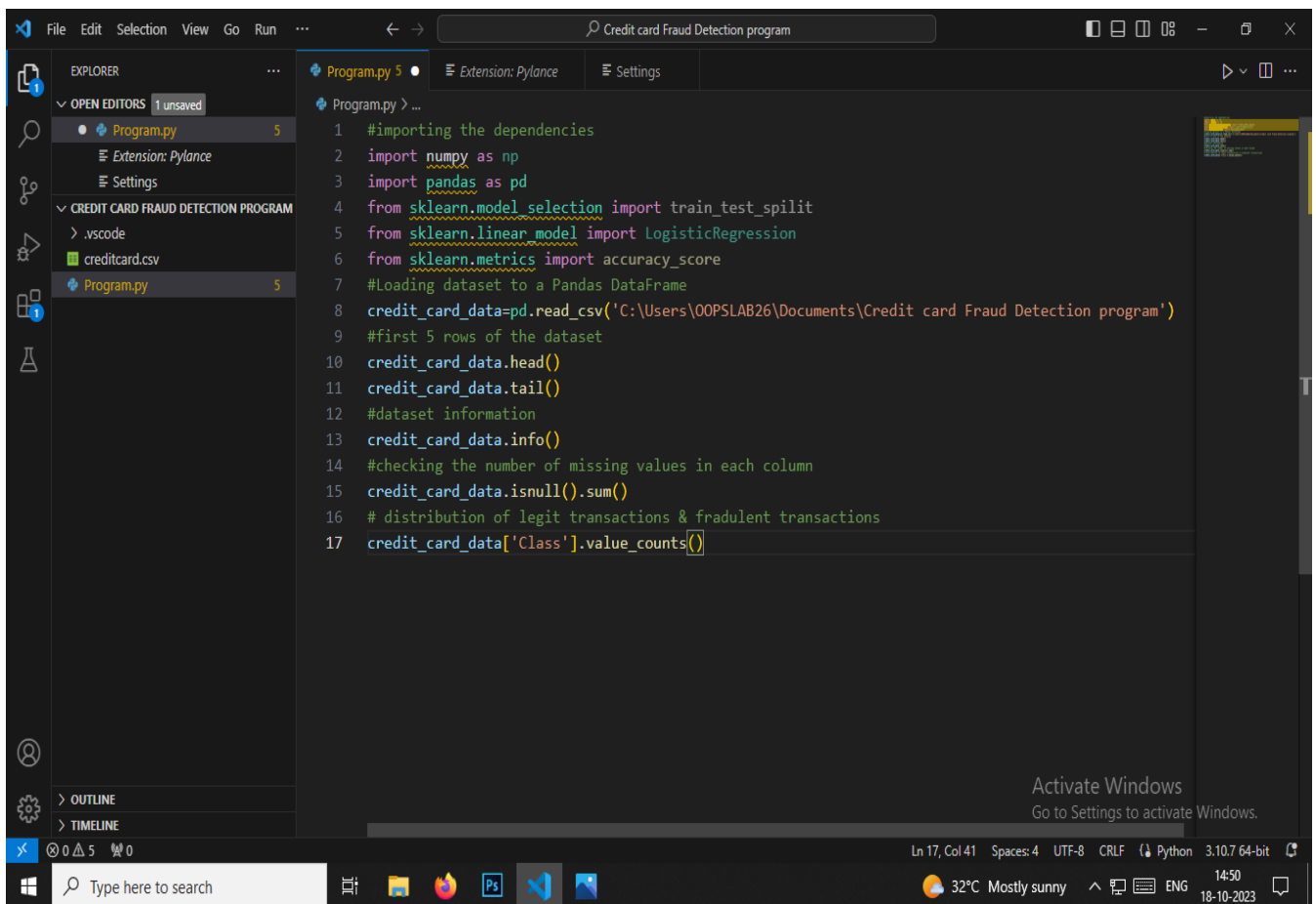
Program.py > ...

```
1 #importing the dependencies
2 import numpy as np
3 import pandas as pd
4 from sklearn.model_selection import train_test_split
5 from sklearn.linear_model import LogisticRegression
6 from sklearn.metrics import accuracy_score
7 #Loading dataset to a Pandas DataFrame
8 credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
9 #first 5 rows of the dataset
10 credit_card_data.head()
11 credit_card_data.tail()
12 #dataset information
13 credit_card_data.info()
14 #checking the number of missing values in each column
15 credit_card_data.isnull().sum()
```

Activate Windows  
Go to Settings to activate Windows.

Ln 15, Col 32 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit

32°C Mostly sunny 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 > ...

```
1 #importing the dependencies
2 import numpy as np
3 import pandas as pd
4 from sklearn.model_selection import train_test_split
5 from sklearn.linear_model import LogisticRegression
6 from sklearn.metrics import accuracy_score
7 #Loading dataset to a Pandas DataFrame
8 credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
9 #first 5 rows of the dataset
10 credit_card_data.head()
11 credit_card_data.tail()
12 #dataset information
13 credit_card_data.info()
14 #checking the number of missing values in each column
15 credit_card_data.isnull().sum()
16 # distribution of legit transactions & fraudulent transactions
17 credit_card_data['Class'].value_counts()
```

Activate Windows  
Go to Settings to activate Windows.

Ln 17, Col 41 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit

32°C Mostly sunny 14:50 18-10-2023

```
1 #importing the dependencies
2 import numpy as np
3 import pandas as pd
4 from sklearn.model_selection import train_test_split
5 from sklearn.linear_model import LogisticRegression
6 from sklearn.metrics import accuracy_score
7 #Loading dataset to a Pandas DataFrame
8 credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
9 #first 5 rows of the dataset
10 credit_card_data.head()
11 credit_card_data.tail()
12 #dataset information
13 credit_card_data.info()
14 #checking the number of missing values in each column
15 credit_card_data.isnull().sum()
16 # distribution of legit transactions & fraudulent transactions
17 credit_card_data['Class'].value_counts()
```

Activate Windows  
Go to Settings to activate Windows.

Ln 17, Col 41 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit

32°C Mostly sunny 14:50 18-10-2023

```
18 #This dataset is highly unbalanced
19 0 --> Normal Transaction
20 1 --> fraudulent transaction
21 #separating the data for analysis
22 legit= credit_card_data[credit_card_data.class==0]
23 fraud= credit_card_data[credit_card_data.class==1]
24 print(legit.shape)
25 print(fraud.shape)
26
```

Activate Windows  
Go to Settings to activate Windows.

Ln 26, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit

32°C Mostly sunny 14:59 18-10-2023

```
1 #importing the dependencies
2 import numpy as np
3 import pandas as pd
4 from sklearn.model_selection import train_test_split
5 from sklearn.linear_model import LogisticRegression
6 from sklearn.metrics import accuracy_score
7 #Loading dataset to a Pandas DataFrame
8 credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
9 #first 5 rows of the dataset
10 credit_card_data.head()
11 credit_card_data.tail()
12 #dataset information
13 credit_card_data.info()
14 #checking the number of missing values in each column
15 credit_card_data.isnull().sum()
16 # distribution of legit transactions & fraudulent transactions
17 credit_card_data['Class'].value_counts()
18 #This dataset is highly unbalanced
19 0 --> Normal Transaction
20 1 -->fraudulent transaction
21 #separating the data for analysis
22 legit= credit_card_data[credit_card_data.class==0]
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()
```

Ln 27, Col 24 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit

```
3 import pandas as pd
4 from sklearn.model_selection import train_test_split
5 from sklearn.linear_model import LogisticRegression
6 from sklearn.metrics import accuracy_score
7 #Loading dataset to a Pandas DataFrame
8 credit_card_data=pd.read_csv('C:\Users\OOPSLAB26\Documents\Credit card Fraud Detection program')
9 #first 5 rows of the dataset
10 credit_card_data.head()
11 credit_card_data.tail()
12 #dataset information
13 credit_card_data.info()
14 #checking the number of missing values in each column
15 credit_card_data.isnull().sum()
16 # distribution of legit transactions & fraudulent transactions
17 credit_card_data['Class'].value_counts()
18 #This dataset is highly unbalanced
19 0 --> Normal Transaction
20 1 -->fraudulent transaction
21 #separating the data for analysis
22 legit= credit_card_data[credit_card_data.class==0]
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
```

Ln 29, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit

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+

OUTLINE

TIMELINE

```
4 from sklearn.model_selection import train_test_split
5 from sklearn.linear_model import LogisticRegression
6 from sklearn.metrics import accuracy_score
7 #Loading dataset to a Pandas DataFrame
8 credit_card_data=pd.read_csv('C:\Users\00PSLAB26\Documents\Credit card Fraud Detection program')
9 #first 5 rows of the dataset
10 credit_card_data.head()
11 credit_card_data.tail()
12 #dataset information
13 credit_card_data.info()
14 #checking the number of missing values in each column
15 credit_card_data.isnull().sum()
16 # distribution of legit transactions & fraudulent transactions
17 credit_card_data['Class'].value_counts()
18 #This dataset is highly unbalanced
19 0 --> Normal Transaction
20 1 -->fraudulent transaction
21 #separating the data for analysis
22 legit= credit_card_data[credit_card_data.class==0]
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()
```

Activate Windows  
Go to Settings to activate Windows.

Ln 30, Col 41 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit

32°C Mostly sunny 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+

OUTLINE

TIMELINE

```
9 #first 5 rows of the dataset
10 credit_card_data.head()
11 credit_card_data.tail()
12 #dataset information
13 credit_card_data.info()
14 #checking the number of missing values in each column
15 credit_card_data.isnull().sum()
16 # distribution of legit transactions & fraudulent transactions
17 credit_card_data['Class'].value_counts()
18 #This dataset is highly unbalanced
19 0 --> Normal Transaction
20 1 -->fraudulent transaction
21 #separating the data for analysis
22 legit= credit_card_data[credit_card_data.class==0]
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
```

Activate Windows  
Go to Settings to activate Windows.

Ln 35, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit

32°C Mostly sunny 15:13 18-10-2023

The screenshot shows the Visual Studio Code editor with a Python file named `Program.py` open. The Explorer sidebar on the left shows the project structure with `creditcard.csv` and `Program.py`. The main editor area displays the following code:

```
12 #dataset information
13 credit_card_data.info()
14 #checking the number of missing values in each column
15 credit_card_data.isnull().sum()
16 # distribution of legit transactions & fraudulent transactions
17 credit_card_data['Class'].value_counts()
18 #This dataset is highly unbalanced
19 0 --> Normal Transaction
20 1 --> fraudulent transaction
21 #separating the data for analysis
22 legit= credit_card_data[credit_card_data.class==0]
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
```

The status bar at the bottom indicates the cursor is at line 38, column 1, with 4 spaces, UTF-8 encoding, CRLF line endings, and Python 3.10.7 64-bit. The system tray shows a temperature of 32°C and the date 18-10-2023.

This screenshot shows the same Visual Studio Code editor with the `Program.py` file. The code is identical to the first screenshot, but with an additional line at the end:

```
39
```

The status bar at the bottom now indicates the cursor is at line 39, column 1, with 4 spaces, UTF-8 encoding, CRLF line endings, and Python 3.10.7 64-bit. The system tray shows a temperature of 32°C and the date 18-10-2023.



```
14 #checking the number of missing values in each column
15 credit_card_data.isnull().sum()
16 # distribution of legit transactions & fraudulent transactions
17 credit_card_data['Class'].value_counts()
18 #This dataset is highly unbalanced
19 0 --> Normal Transaction
20 1 --> fraudulent transaction
21 #separating the data for analysis
22 legit= credit_card_data[credit_card_data.class==0]
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
```

Activate Windows  
Go to Settings to activate Windows.

```
19 0 --> Normal Transaction
20 1 --> fraudulent transaction
21 #separating the data for analysis
22 legit= credit_card_data[credit_card_data.class==0]
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
```

Activate Windows  
Go to Settings to activate Windows.

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 > ...

```
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.

Ln 49, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit

32°C Mostly sunny 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 > ...

```
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
```

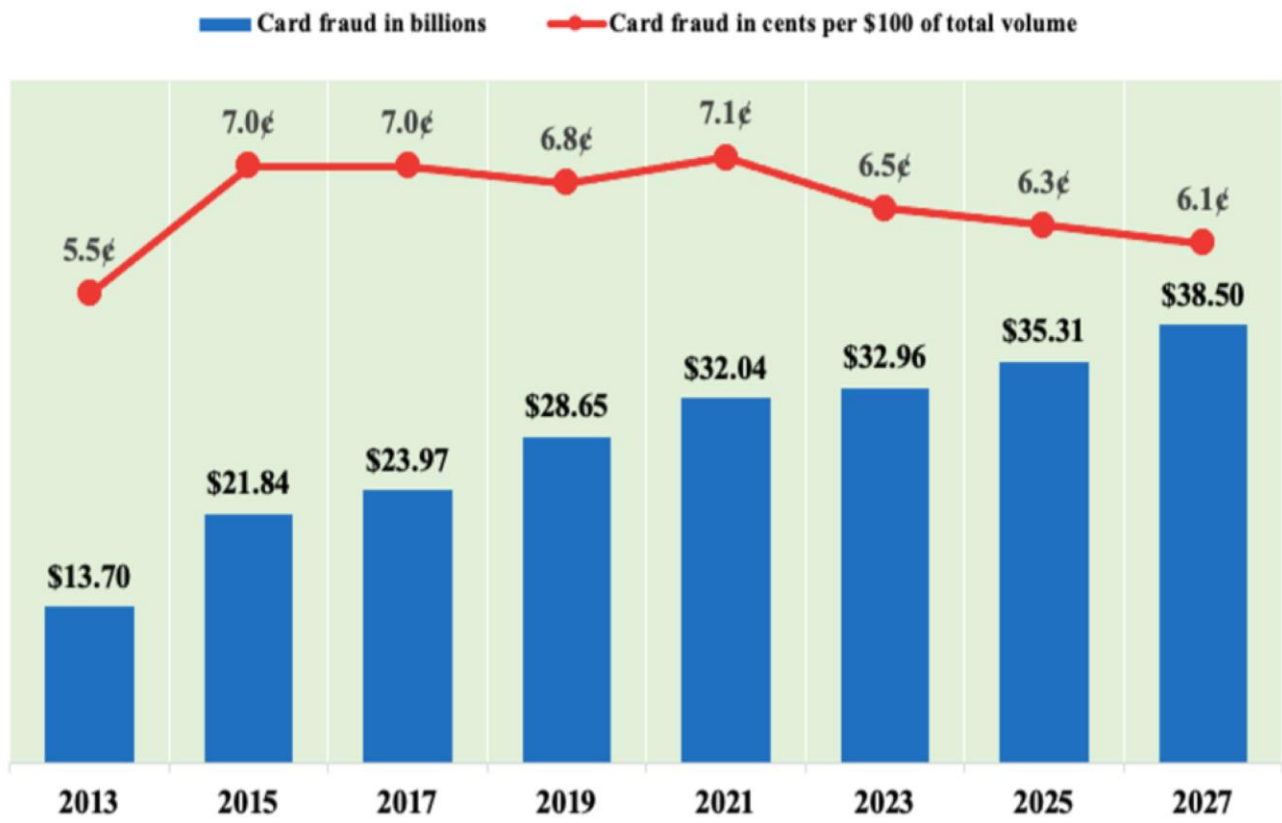
Activate Windows  
Go to Settings to activate Windows.

Ln 54, Col 1 Spaces: 4 UTF-8 CRLF Python 3.10.7 64-bit

32°C Mostly sunny 15:30 18-10-2023

## #Statistical Analysis:

### Card Fraud Worldwide



❖ Performing exploratory data analysis

❖ Statistical analysis

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