**Credit Card Fraud Detection**

**PHASE-4**

**PROGRAM:**

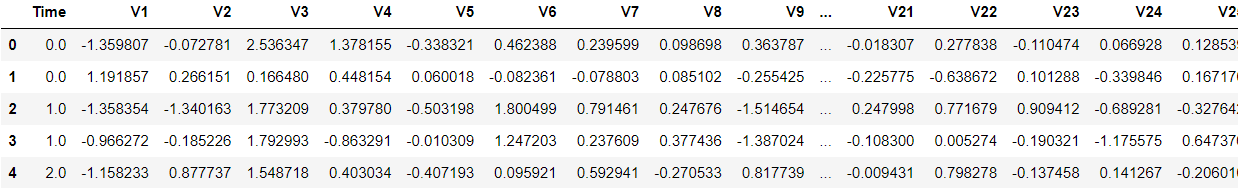
**import pandas as pd**

**from collections import Counter**

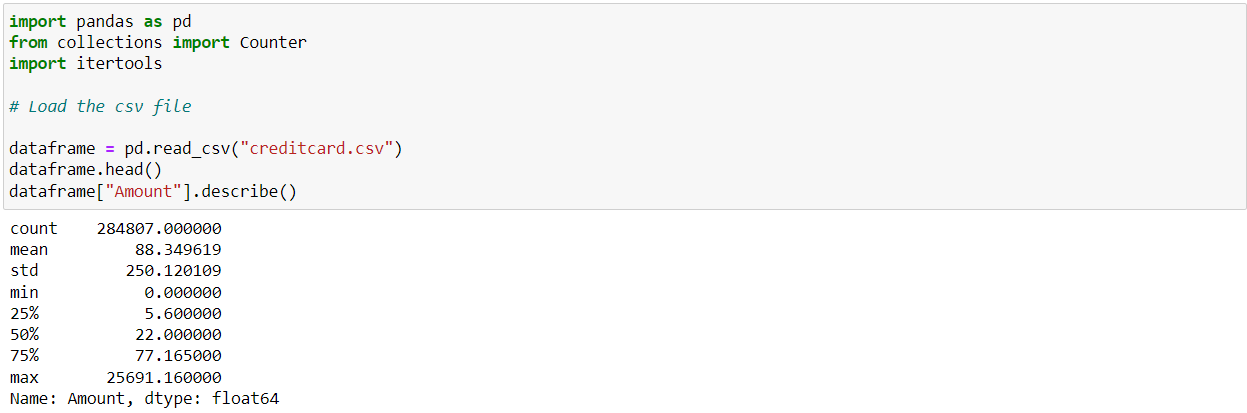
**import itertools**

**dataframe = pd.read\_csv("creditcard.csv")**

**dataframe.head()**

****

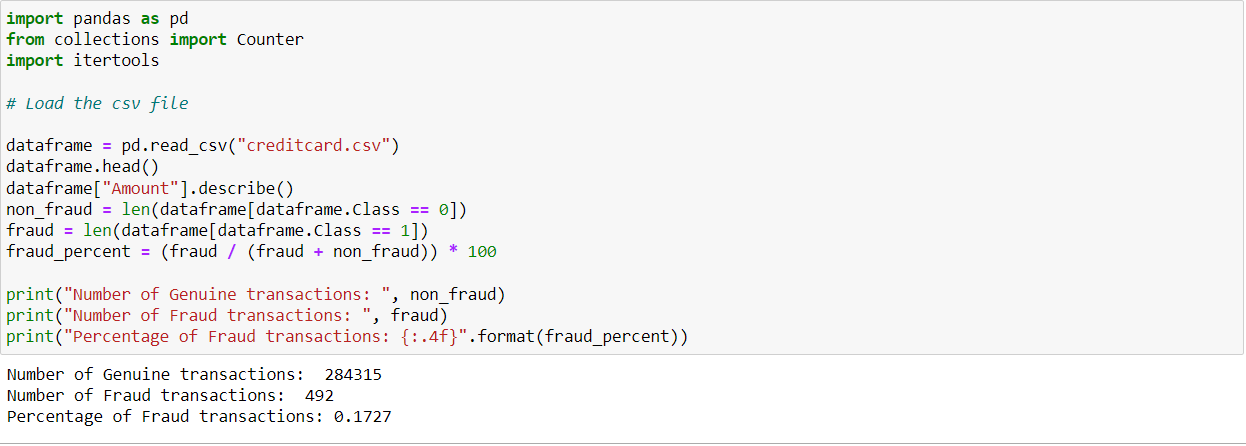
**dataframe["Amount"].describe()**

****

**non\_fraud = len(dataframe[dataframe.Class == 0])**

**fraud = len(dataframe[dataframe.Class == 1])**

**fraud\_percent = (fraud / (fraud + non\_fraud)) \* 100**

****

**import matplotlib.pyplot as plt**

**labels = ["Genuine", "Fraud"]**

**count\_classes = dataframe.value\_counts(dataframe['Class'], sort= True)**

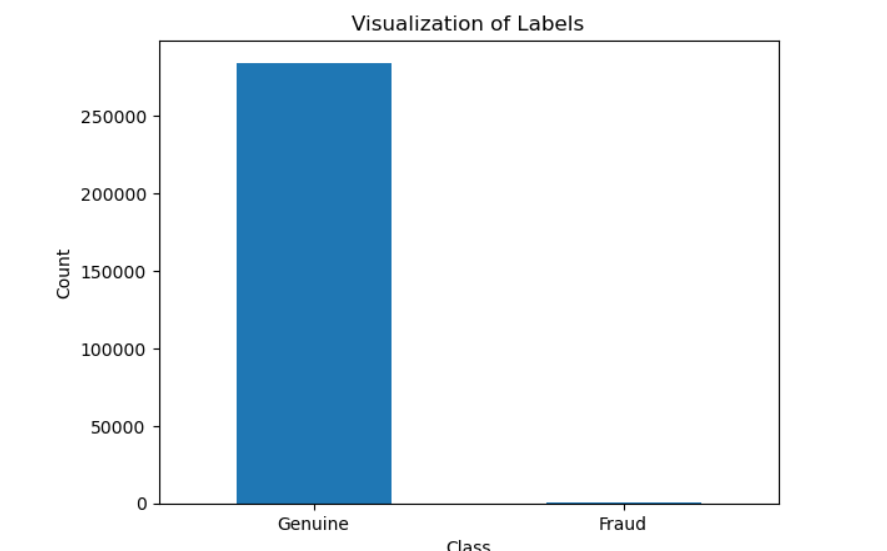
**count\_classes.plot(kind = "bar", rot = 0)**

**plt.title("Visualization of Labels")**

**plt.ylabel("Count")**

**plt.xticks(range(2), labels)**

**plt.show()**

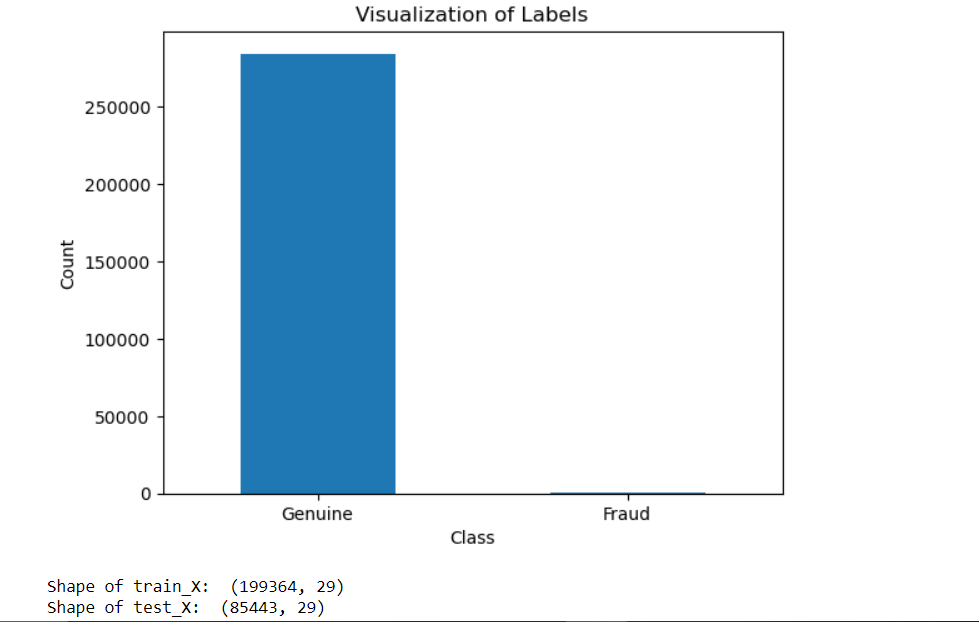
****

**from sklearn.model\_selection import train\_test\_split**

**(train\_X, test\_X, train\_Y, test\_Y) = train\_test\_split(X, Y, test\_size= 0.3, random\_state= 42)**

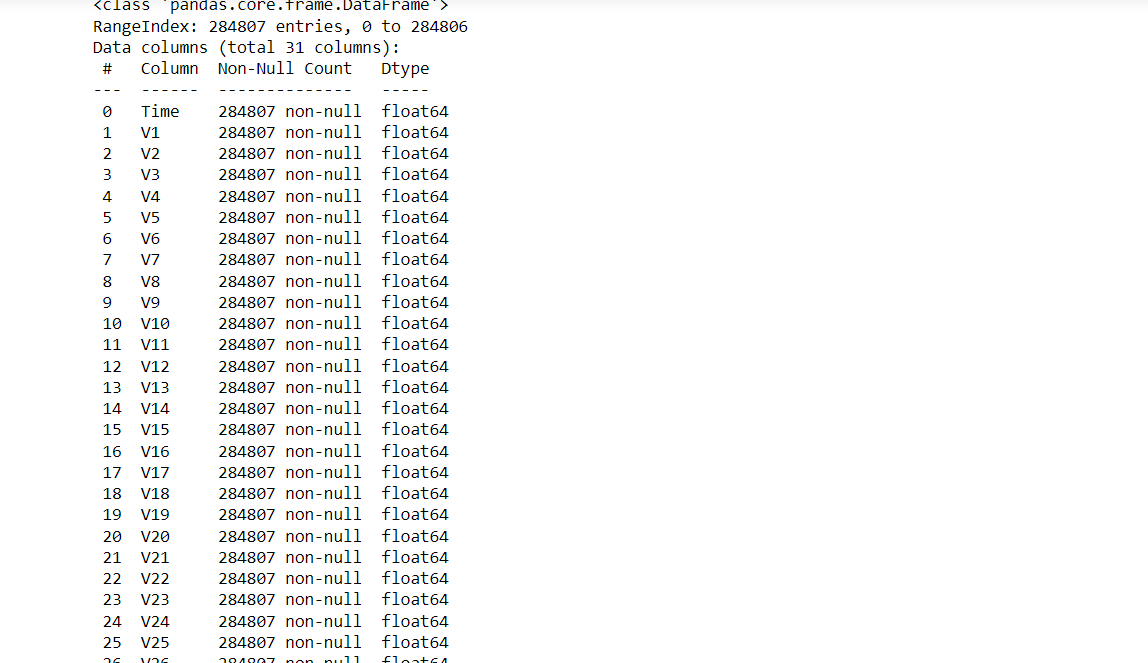
**print("Shape of train\_X: ", train\_X.shape)**

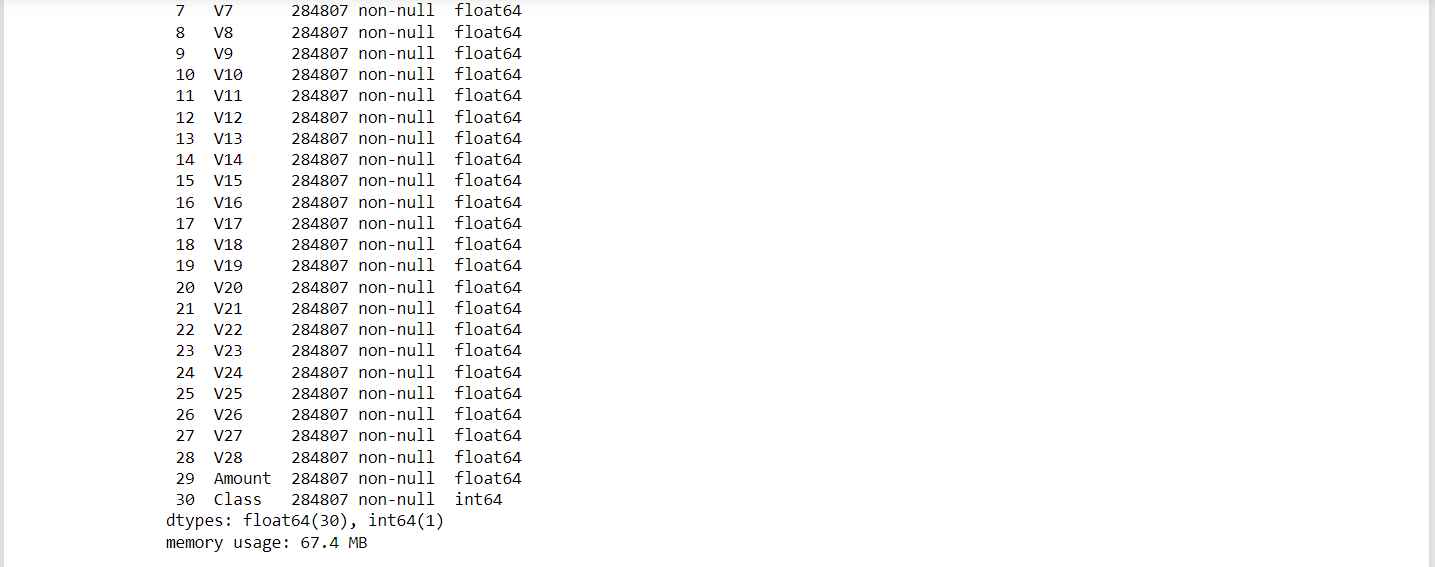
**print("Shape of test\_X: ", test\_X.shape)**

****

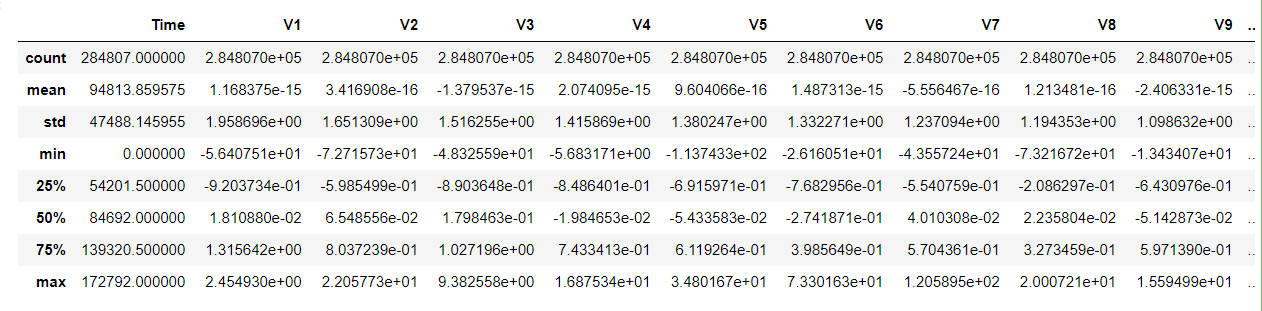
**dataframes.head()**

**dataframes.info()**

****

****

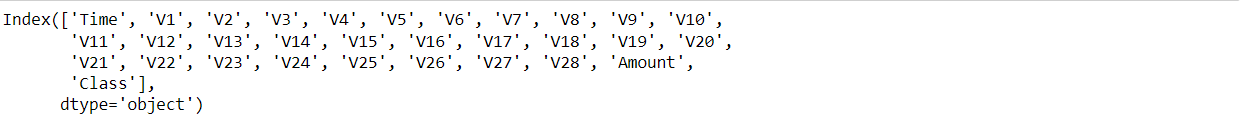
**dataframe.describe()**

****

**dataframe.isnull().sum()**

****

**dataframe.columns**

****

**dataframe.shape**

(284807, 31)

**dataframe.duplicated().value\_counts()**

False 283726

True 1081

Name: count, dtype: int64

**dataframe.Class.unique()**

array([0, 1], dtype=int64)

**print(legitimate)**

**print(fraud)**

284315

492

# ****Handling Imbalance of the data using SMOTE:****

x=data.drop("Class",axis=1)

y=data["Class"]

In [18]:

linkcode

smote = SMOTE(random\_state=42)

x\_smote, y\_smote = smote.fit\_resample(x, y)

## Data Visualization:

corr\_balanced = x\_smote.corr()

plt.figure(figsize=(8, 6))

sns.heatmap(corr\_balanced, annot=False, cmap="plasma", linewidths=0.5)

plt.title("Correlation Matrix for Balanced Data")

plt.show()

