|  |
| --- |
| Credit Card Fraud Detection |
|  |

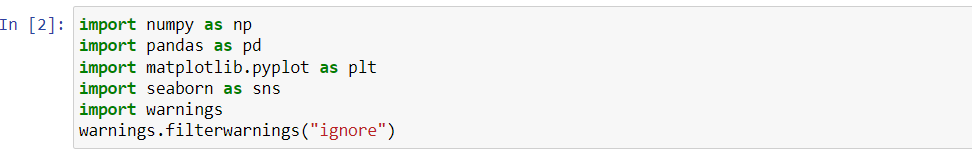
# Context :

It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase.

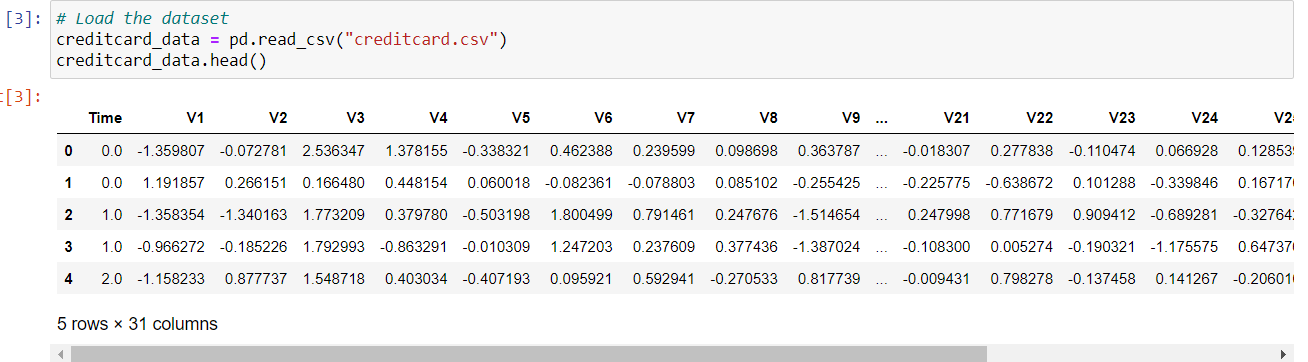
**Description:**

In this analysis, we focused on detecting fraudulent transactions in a credit card dataset. The dataset contains 284,807 transactions with 31 features, including time, transaction amount, and variables V1 to V28. The target variable is 'Class,' where 0 indicates a non-fraudulent transaction and 1 indicates a fraudulent transaction.

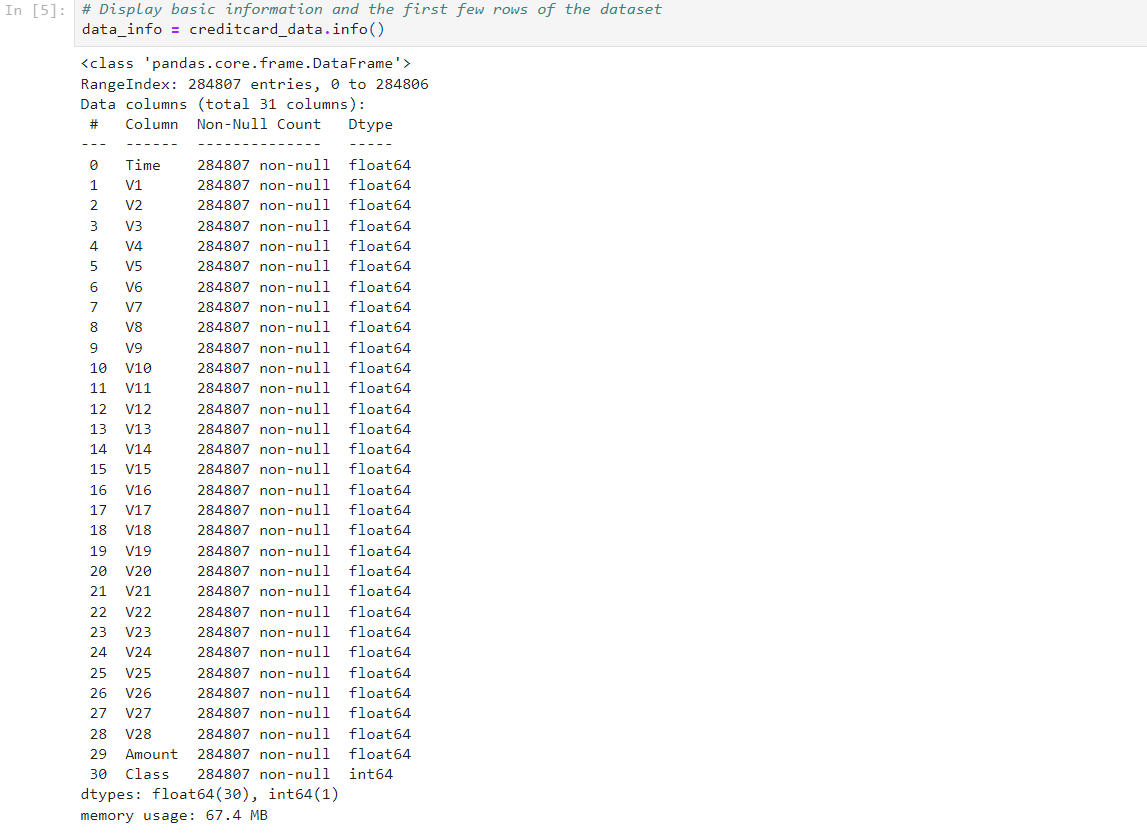
# The first step is to Importing the required Libraries.



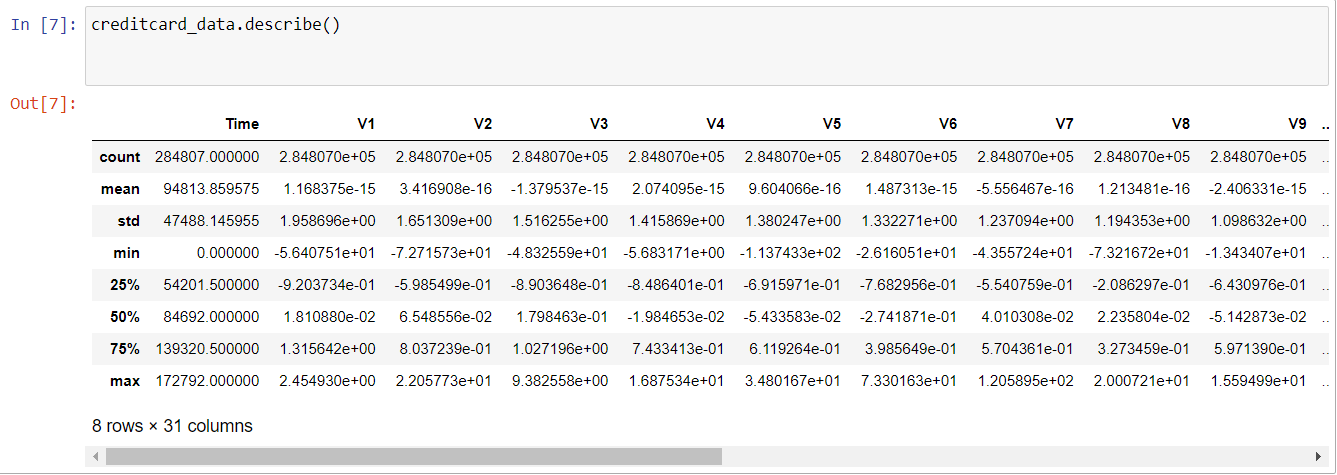
Upload the csv file named (credir.csv) in the Python folder saved on Jupyter Notebook. Load the Dataset from the csv file name (credit.csv) using the pandas function.



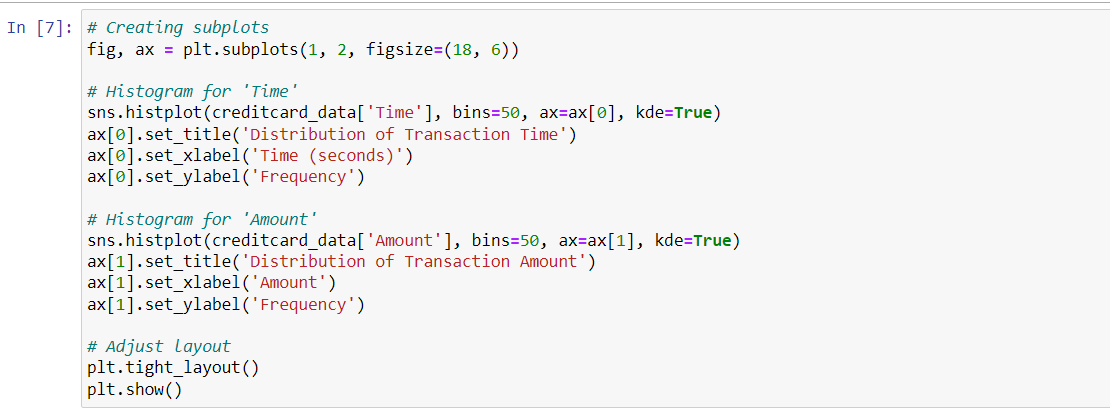
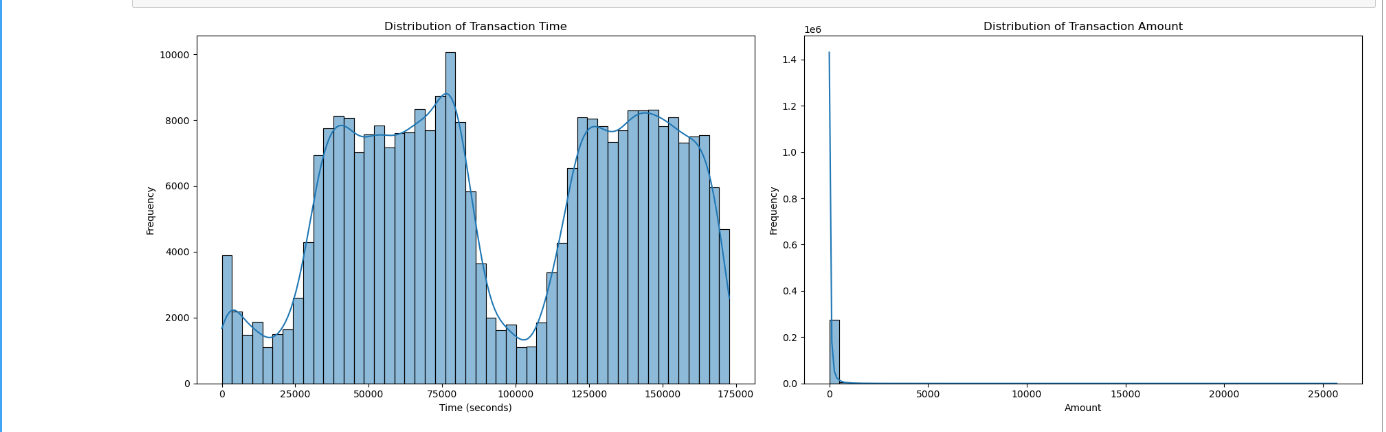
Now, we loaded the dataset and examined its basic information using info(). The dataset has no missing values, and all features are numerical.

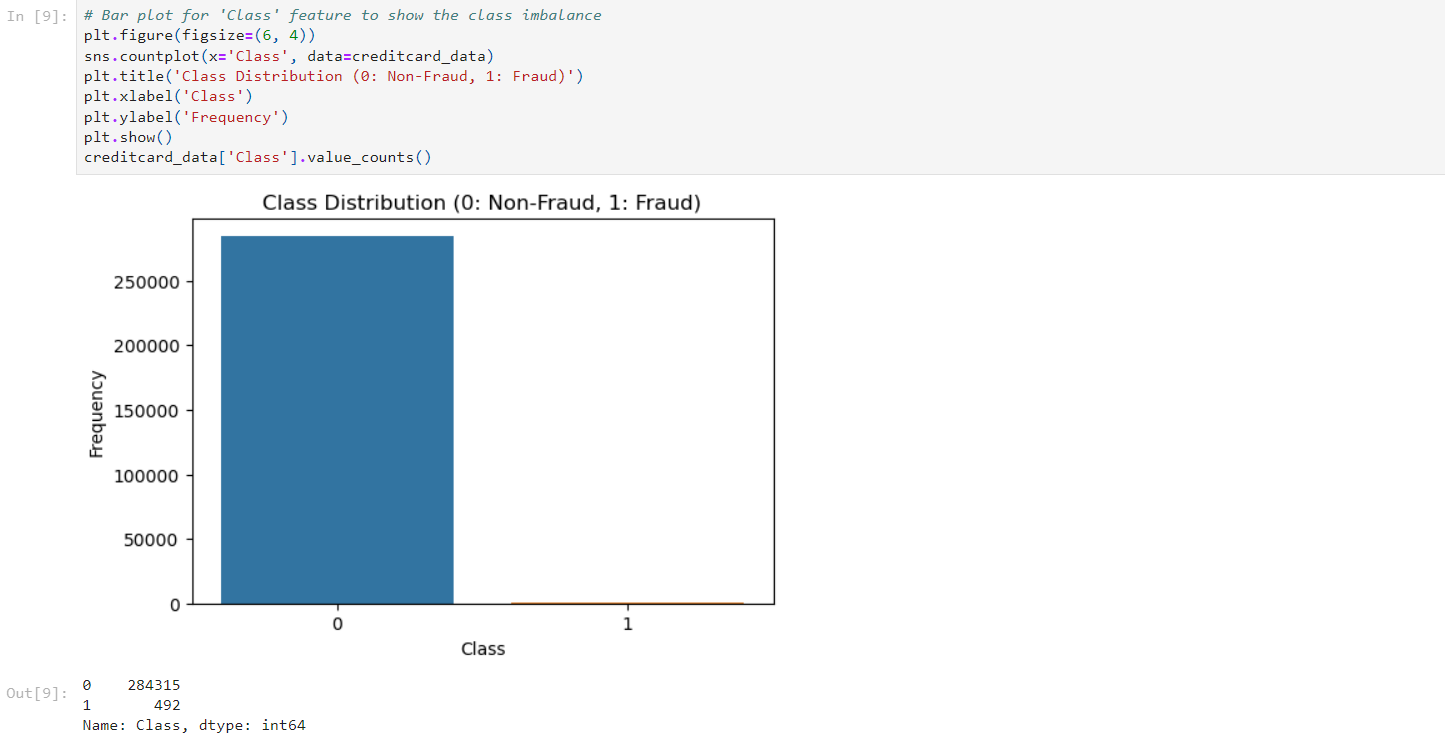


Descriptive statistics revealed that the mean of the 'Class' variable is close to 0, indicating a class imbalance.

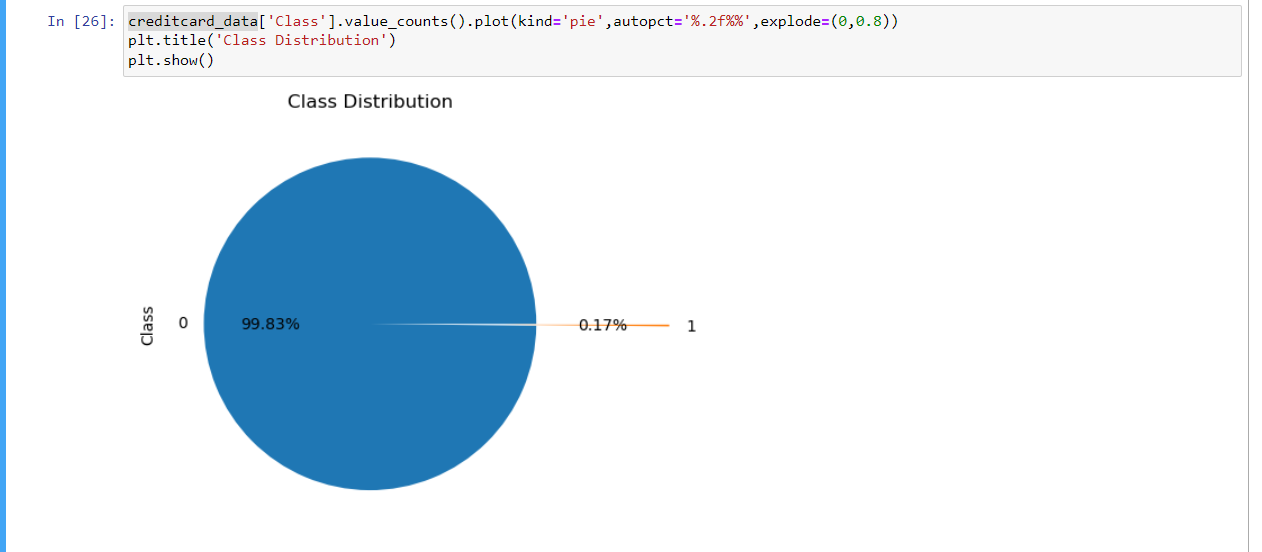


Now, we created histograms to visualize the distribution of transaction time and amount. The histograms revealed that transaction time has a relatively uniform distribution, while transaction amounts are right-skewed, with a few extremely high transactions.



Next, we used a bar plot to show the class distribution, highlighting the significant class imbalance. There are 284,315 non-fraudulent transactions ‘Class 0’ and only 492 fraudulent transactions ‘Class 1’.

Created a pie chart to visualize the percentage distribution of normal and fraudulent transactions.



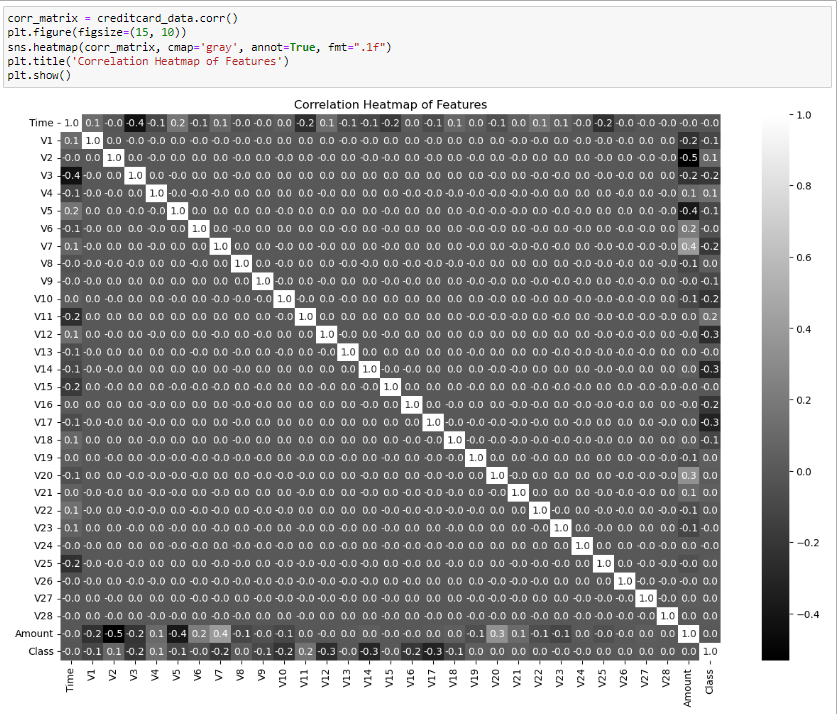
We created box plots to compare the distribution of transaction amounts and times between non-fraudulent and fraudulent transactions. The log scale is applied to the amount axis due to the wide range of transaction amounts.

The plots suggest that fraudulent transactions tend to have higher amounts and occur more evenly across different times.



We calculated the correlation matrix and visualized it using a heatmap.

Most features show low correlation, indicating that the variables (V1 to V28) are relatively independent of each other.



|  |  |
| --- | --- |
| **Name** | Rushikesh.S.Sakhare |
| **Course** | P.G. in Data Analytics |
| **Module** | Data Analytics Essentials |