CREDIT CARD FRAUD DETECTION

Phase-3: Development

**Objective:**

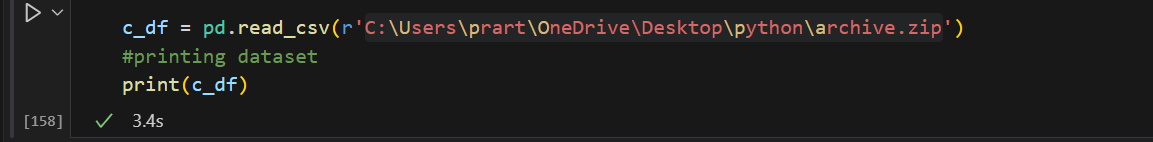
The main objective for Creditcard fraud detection project in applied data science is to implement and develop a system that identify and prevent credit card fraud transaction. The aim is to reduce the number of fraudulent transactions to improve user experience.

**Dataset:**

<https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud> .

**Loading datasets:**

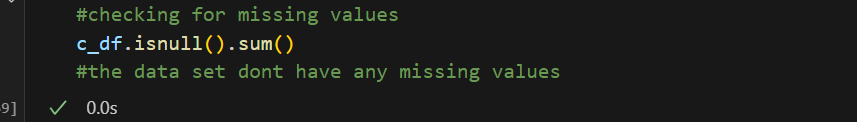
The pandas library is used to load dataset from the CSV file into a pandas Data frame. Then the dataset is explored with various dataframe methods. For example, here “read\_csv” is used with the actual path of the dataset.



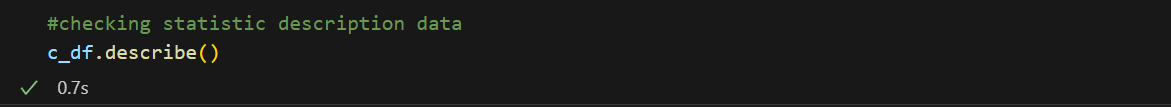
**Data preprocessing:**

Explore dataset to understand its structure and characteristics.it includes checking of missing values, understanding the data distribution and identifying potential issues

* Checking of missing values:



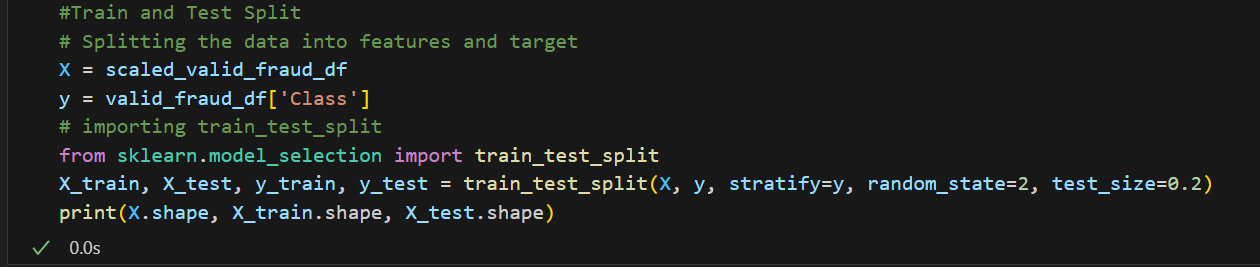
* Summary statistics:



**Training and test split:**

* Splitting the dataset:

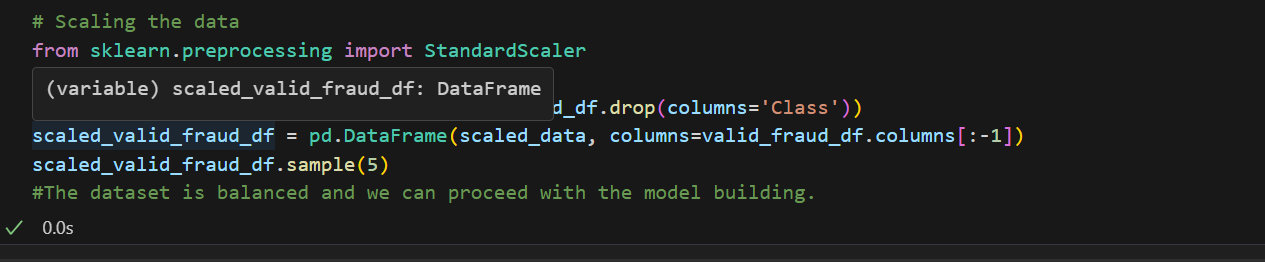
The dataset is divided into training and testing set to evaluate the performance of the model



* Scaling and standardizing:

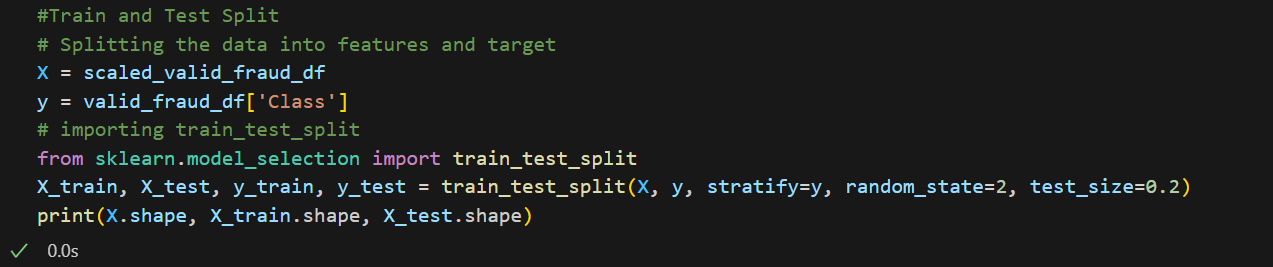
It is mandatory to scale and standardize to have zero mean and unit variance.

This can improve the performance of some machine learning algorithm.



**Model building:**

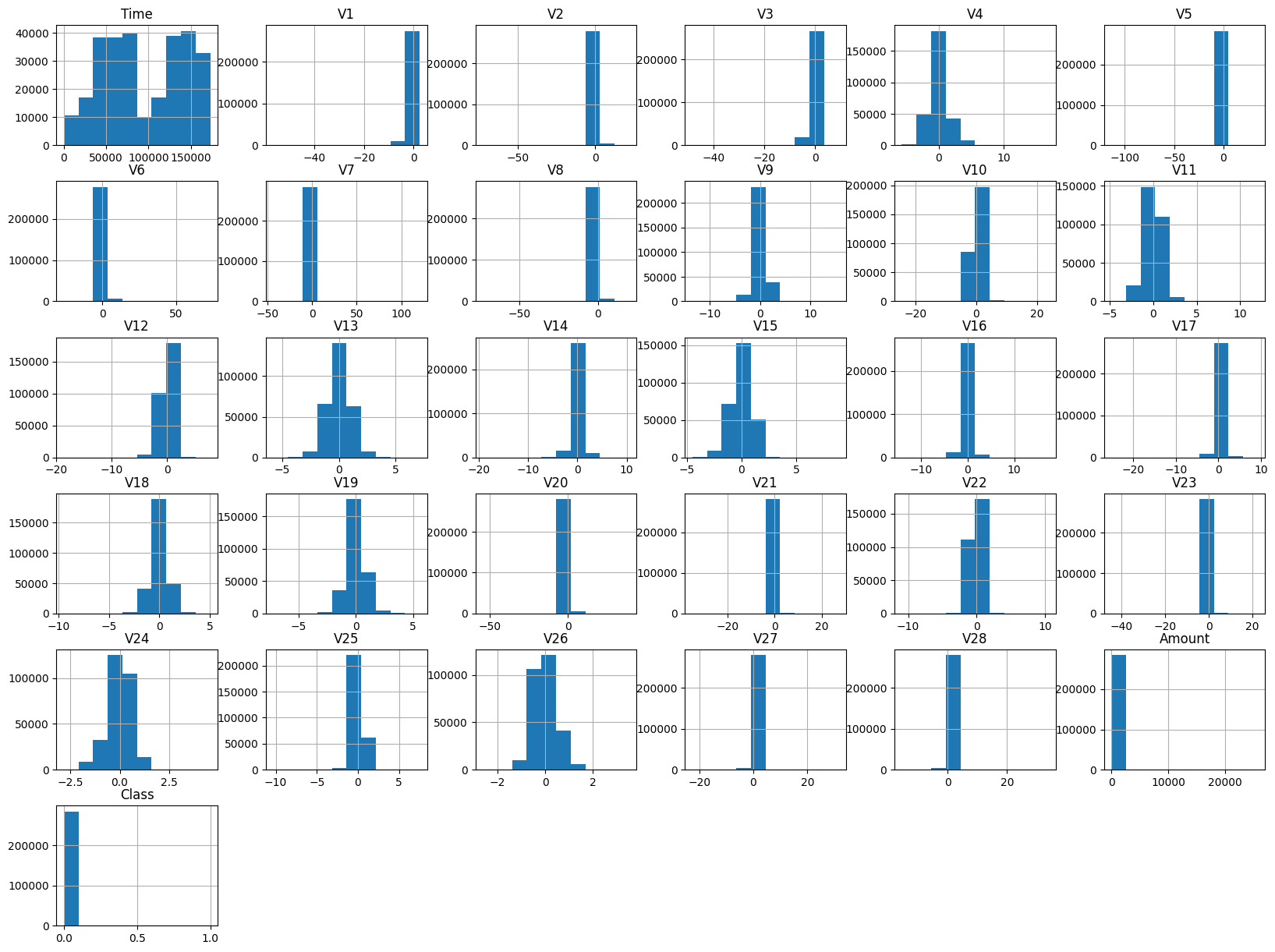
After preprocessing, the training and building of machine learning model is done. the common algorithm for credit card fraud detection is logistic regression.



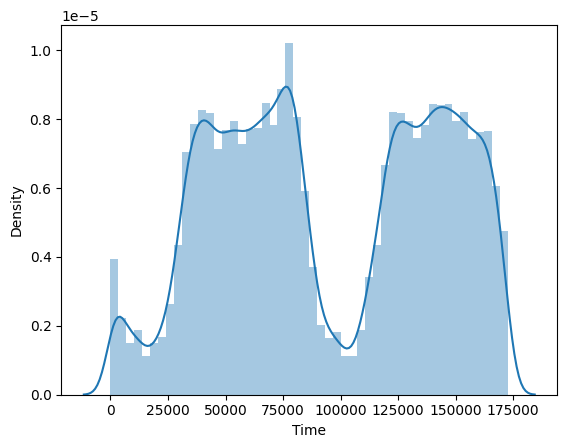
**Model evaluation:**

Evaluating the performance of a credit card fraud detection model in data science is crucial to pivotal to ensure its effectiveness in identifying fraudulent transactions while minimizing false positives. The accuracy measures the overall correctness of the model's prediction. It is the ratio of correctly predicted transactions to the total number of transactions. However, in highly imbalanced datasets the accuracy can be misleading. The precision measures the proportion of correctly identified fraudulent transaction among all transactions predicted as fraud

**Credit card fraud detection analysis of data in histogram visualization:**



**Credit card fraud detection analysis of data in histogram visualization:**



**Correlation matrix heatmap:**

A correlation matrix heatmap is a valuable visualization tool in credit card fraud detection for understanding the relationships between various features in the dataset.it can help identify potential correlations between features and highlight those that might be useful in fraud detection in feature engineering.

