CREDITCARD FRAUD DETECTION

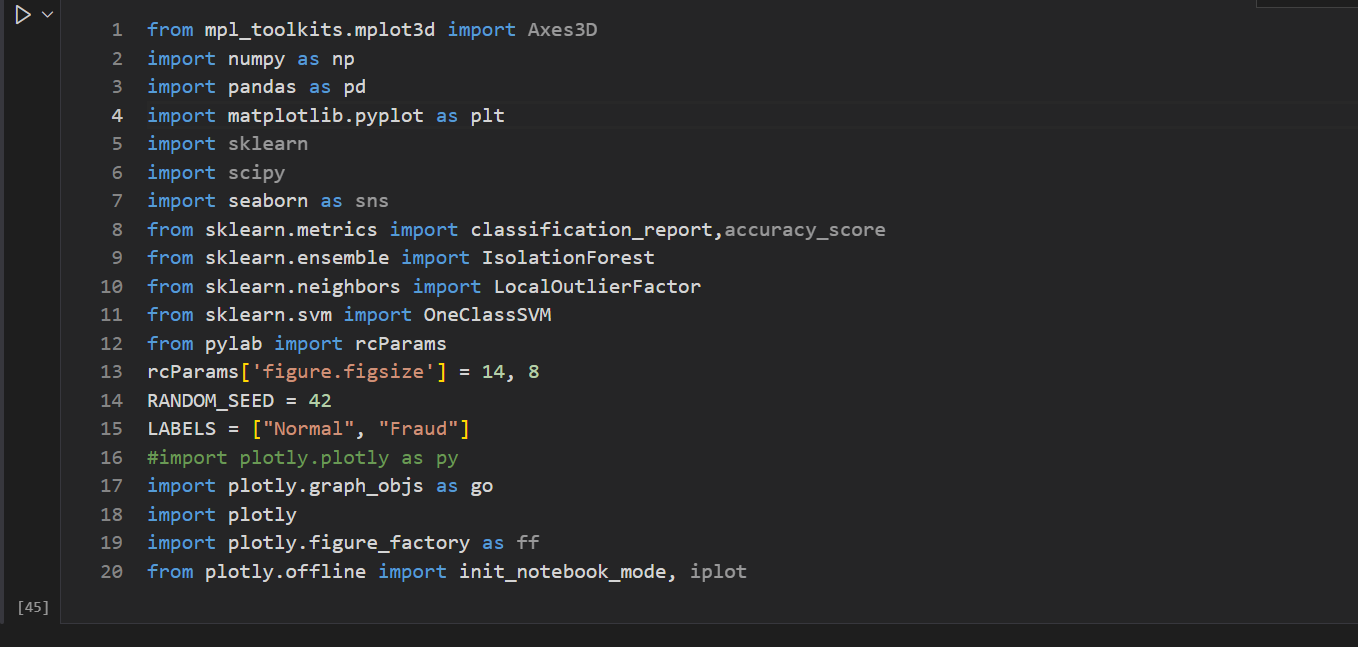
Phase:5

**Problem statement:**

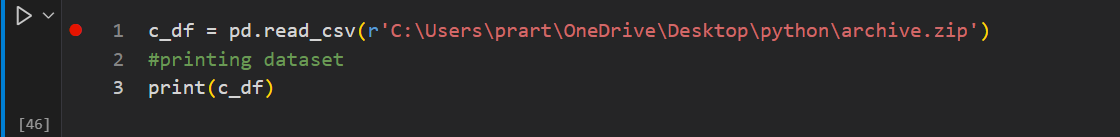
This credit card fraud detection problem statement involves analysing transaction data in real-time and detecting unusual or suspicious patterns to protect cardholders and financial institutions from unauthorized or fraudulent charges.

**Design thinking process:**

* **Datasource:**  <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud> .
* **Extension imported for processing:**

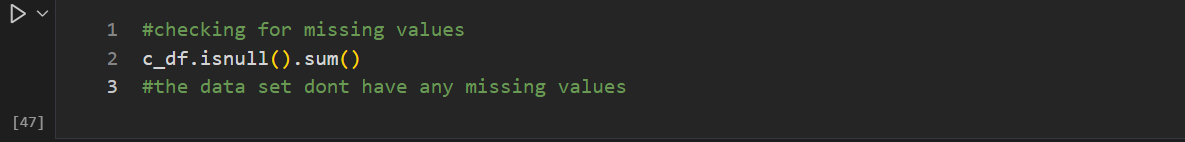


* **Loading datasets:** The pandas library is used to load dataset from the CSV file into a pandas Data frame. Then the dataset is explored with various data frame methods. For example, here “read\_csv” is used with the actual path of the dataset.

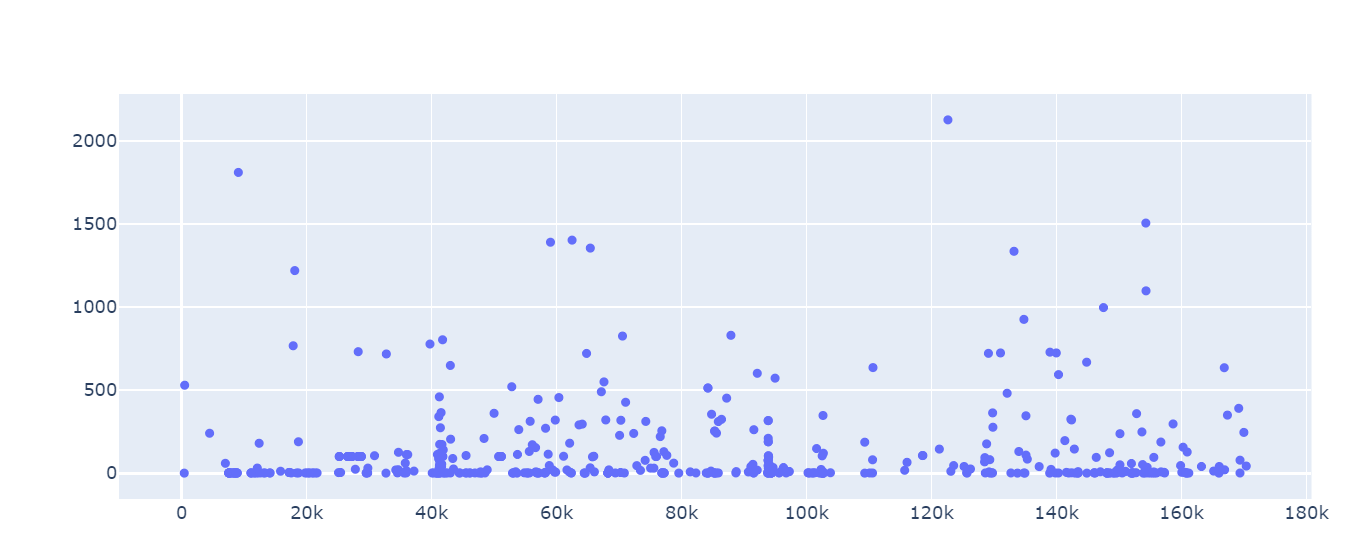


* **Data Cleaning:**

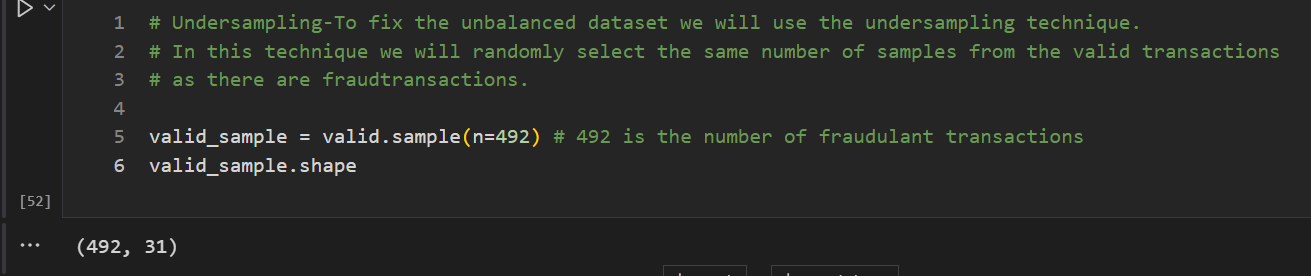
1. Handle missing values: Identify and fill in missing values, possibly using techniques like mean imputation or more advanced.



2. Outlier detection: it detects and deals with outlier that could skew the model. Outliers in transaction amounts for example, could indicate potential fraud.



3. Data Sampling: If the dataset is highly imbalanced employ techniques like oversampling or under sampling to balance the dataset.



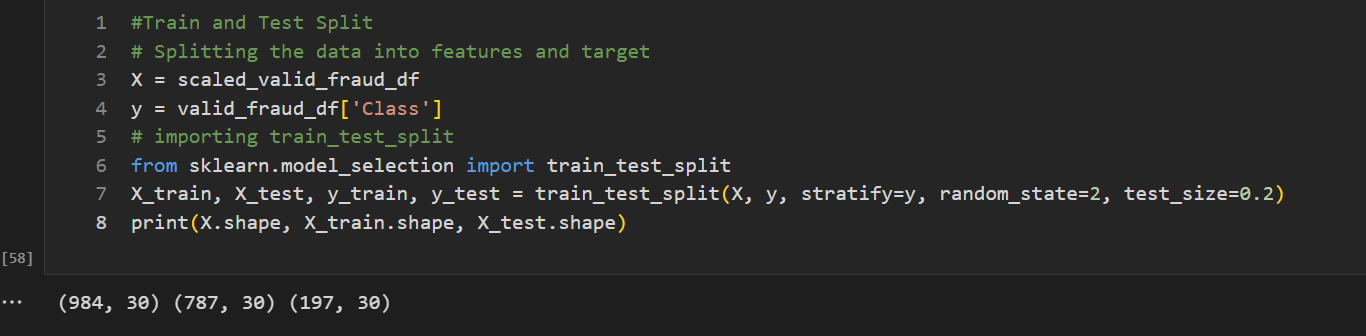
4. Data Splitting: Split the dataset into training, validation, and test sets to evaluate the model’s performance effectively.

* **Model training:**

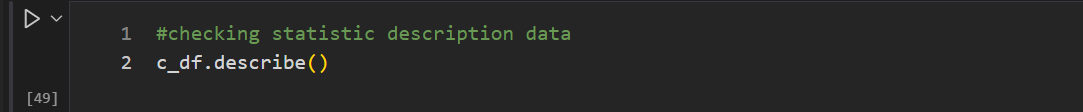
1.Data Splitting: Split your pre-processed dataset into three subsets: training, Split your pre-processed dataset into three subsets: training, validation, and 15% for testing.

2. Model Persistence: Once you are satisfied with the model’s performance, save the trained model so that it can be deployed and used for real-time fraud detection.

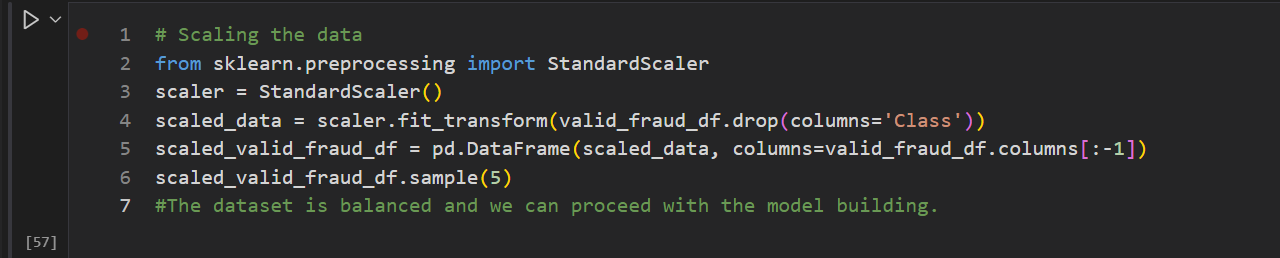
3. Validation: Continuously monitor the model’s performance on the validation set during training. This allows you to detect issues like overfitting or under fitting and adjust hyper parameters accordingly.



4.Summary statistics:



5. 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.

