

**A Mini -Project Report**

**On**

**“Creditcard Fraud Detection”**

**Machine Learning**

**Submitted by**

**P.Jahnavi sai(19B21A0409)**

**D.Pavan Kalyan(19B21A0434)**

**V.Ganesh (19B21A0437)**

**G.Kalyan Kumar(19B21A0448)**

**V.Rama Supriya(19B21A0404)**

**Department of Electronics and Communication Engineering**

**Submitted to**

**Dr.Krish Naik**

**Chief Innovation Officer**

**Submission Date: November 11,2021**

**INDEX**

ABSTRACT

LIST OF ABBREVATIONS

CHAPTER1: INTRODUCTION

* 1. Problem Definition
  2. Objectives

CHAPTER2: TECHNOLOGIES USED

2.1 Machine Learning

2.2 Python

MODULES

DATASET

DATA FLOW DIAGRAM

SYSTEM REQUIREMENTS

CODE

OBSERVATIONS

ADVANTAGES

SOLUTION

CONCLUSION

**ABSTRACT**

This report represents the mini-project taken by our group of 5 students for the partial fulfilment of “Credit Card Fraud Detection”, Machine Learning, It is vital that credit card companies are able to identify fraudulent credit card transactions so that customers are not charged for items that they did not purchase. Such problems can be tackled with Data Science and its importance, along with Machine Learning, cannot be overstated. This project intends to illustrate the modelling of a data set using machine learning with Credit Card Fraud Detection. The Credit Card Fraud Detection Problem includes modelling past credit card transactions with the data of the ones that turned out to be fraud.

This model is then used to recognize whether a new transaction is fraudulent or not. Our objective here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications. Credit Card Fraud Detection is a typical sample of classification. In this process, we have focused on analysing and pre-processing data sets as well as the deployment of multiple anomaly detection algorithms such as Local Outlier Factor and Isolation Forest algorithm on the PCA transformed Credit Card Transaction data.

Keywords— Credit card fraud, applications of machine

learning, data science, isolation forest algorithm, local outlier

factor.

**LIST OF ABBREVATIONS**

1. IDE: Integrated Development Environment
2. LOF: Local Outlier Factor
3. SVM: Support Vector Machine

**1**. **INTRODUCTION**

Fraud in credit card transactions is unauthorized and unwanted usage of an account by someone other than the owner of that account. Necessary prevention measures can be taken to stop this abuse and the behaviour of such fraudulent practices can be studied to minimize it and protect against similar occurrences in the future. In other words, Credit Card Fraud can be defined as a case where a person uses someone else’s credit card for personal reasons while the owner and the card issuing authorities are unaware of the fact that the card is being used. Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behaviour, which consist of fraud, intrusion, and defaulting.

**1.1 PROBLEM DEFINING**

This problem is particularly challenging from the perspective of learning, as it is characterized by various factors such as class imbalance. The number of valid transactions far outnumber fraudulent ones. Also, the transaction patterns often change their statistical properties over the course of time.

These are not the only challenges in the implementation of a real-world fraud detection system, however. In real world examples, the massive stream of payment requests is quickly scanned by automatic tools that determine which transactions to authorize.

Machine learning algorithms are employed to analyse all the authorized transactions and report the suspicious ones. These reports are investigated by professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent.

## 1.2 Objectives

The main objective of developing this project are:

1. To develop machine learning model to predict frauds.
2. To analyse feature selection methods and understand their working principle.

**2.TECHNOLOGIES USED**

**Machine Learning**

Machine Learning is the scientific study of algorithms and static models that computer system use in order to perform a specific task effectively without using the explicit instruction, relaying on patterns and interface instead.

Machine Learning algorithms build a mathematical models based on sample data.

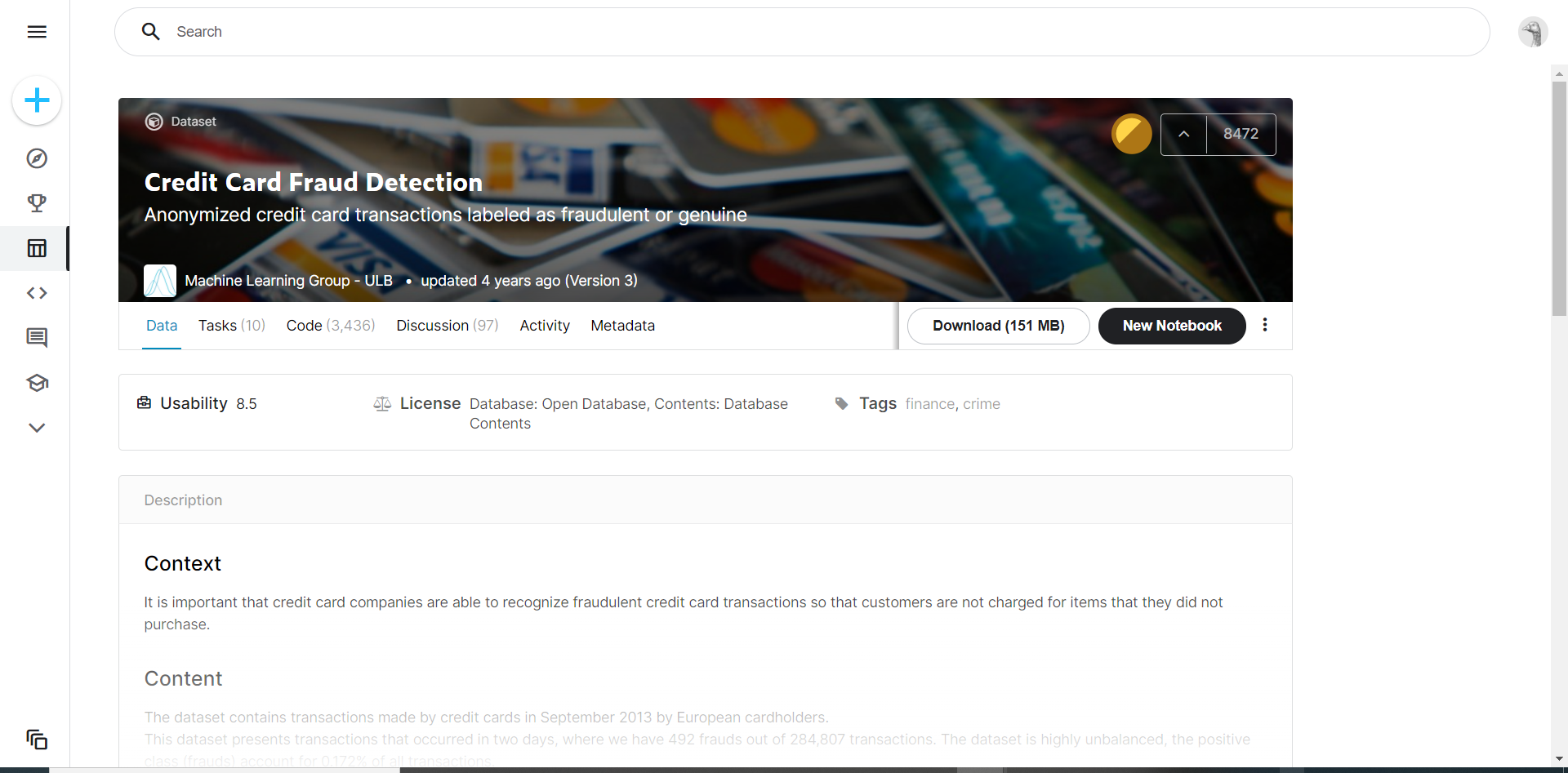
**MODULES**

1. Numpy
2. Pandas
3. Sklearn
4. Scipy
5. Matplotlib.pyplot
6. Seaborn
7. Sklearn
8. pylab

**DATASET**

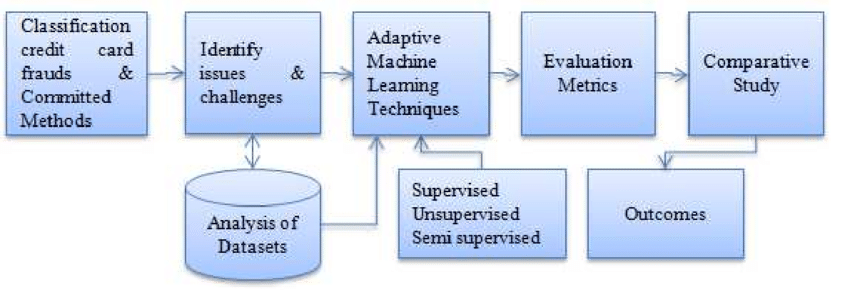
Our required DataSet is creditcard.csv, which is a zip file. From Kaggle platform we have to download dataset first and then unzip it by using extract all.

or we can download directly from https://www.kaggle.com/mlg-ulb/creditcardfraud



So our dataset is creditcard.csv

**DATA FLOW DIAGRAM**



**SYSTEM REQUIREMENTS**

**Hardware Requirements:**

RAM: 4GB and higher

Processor: Intel I3 and above

Hard Disk: 500GB: Minimum

**Software Requirements**:

OS: Windows or Linux

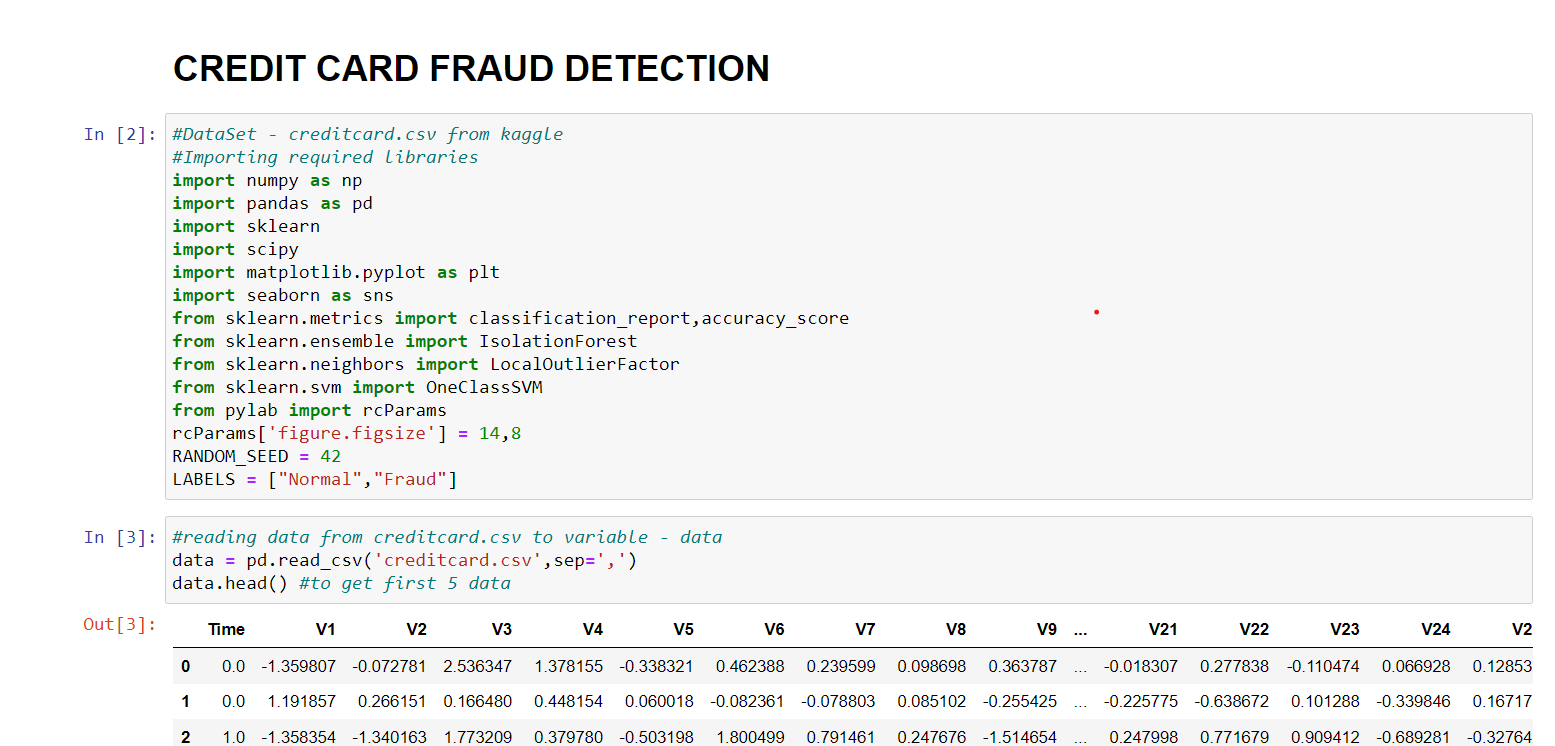
Python IDE: Python 2.7 and above

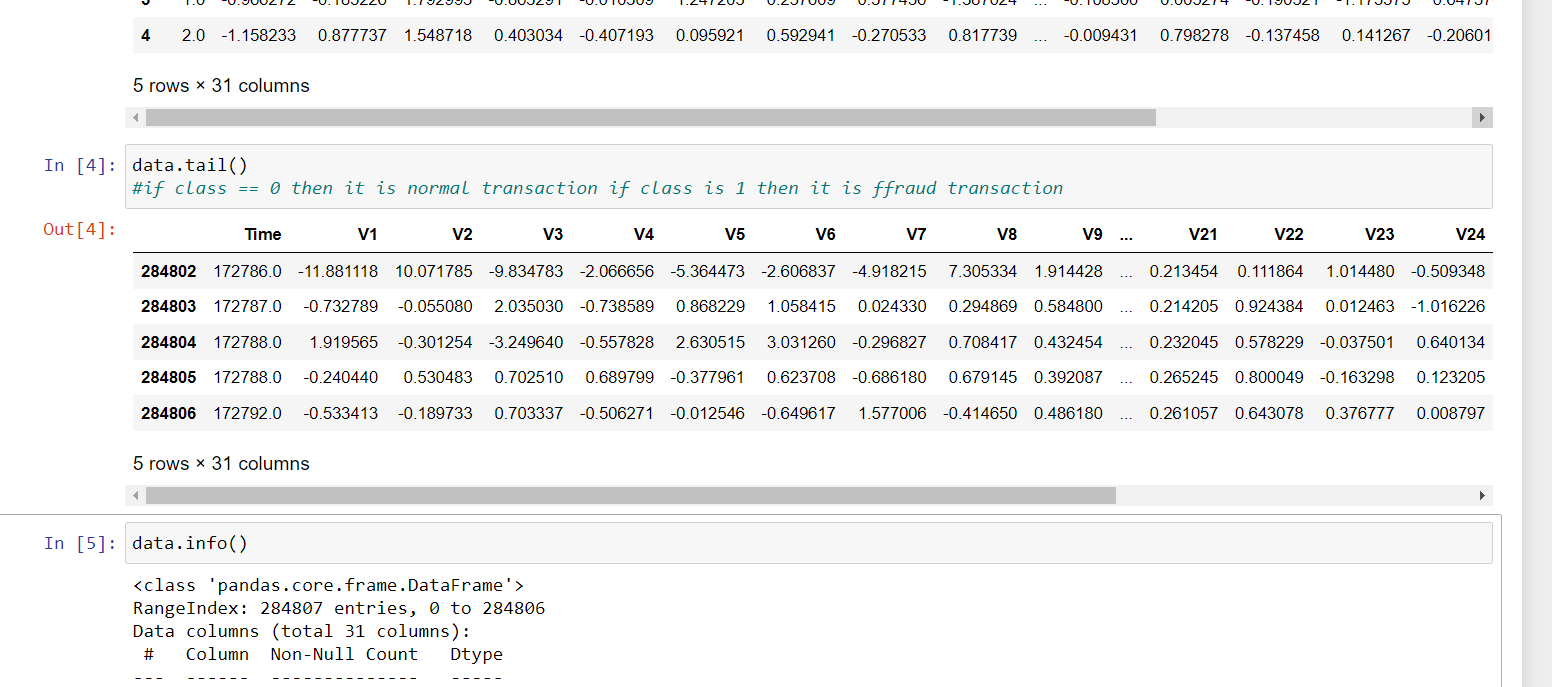
Jupyter notebook

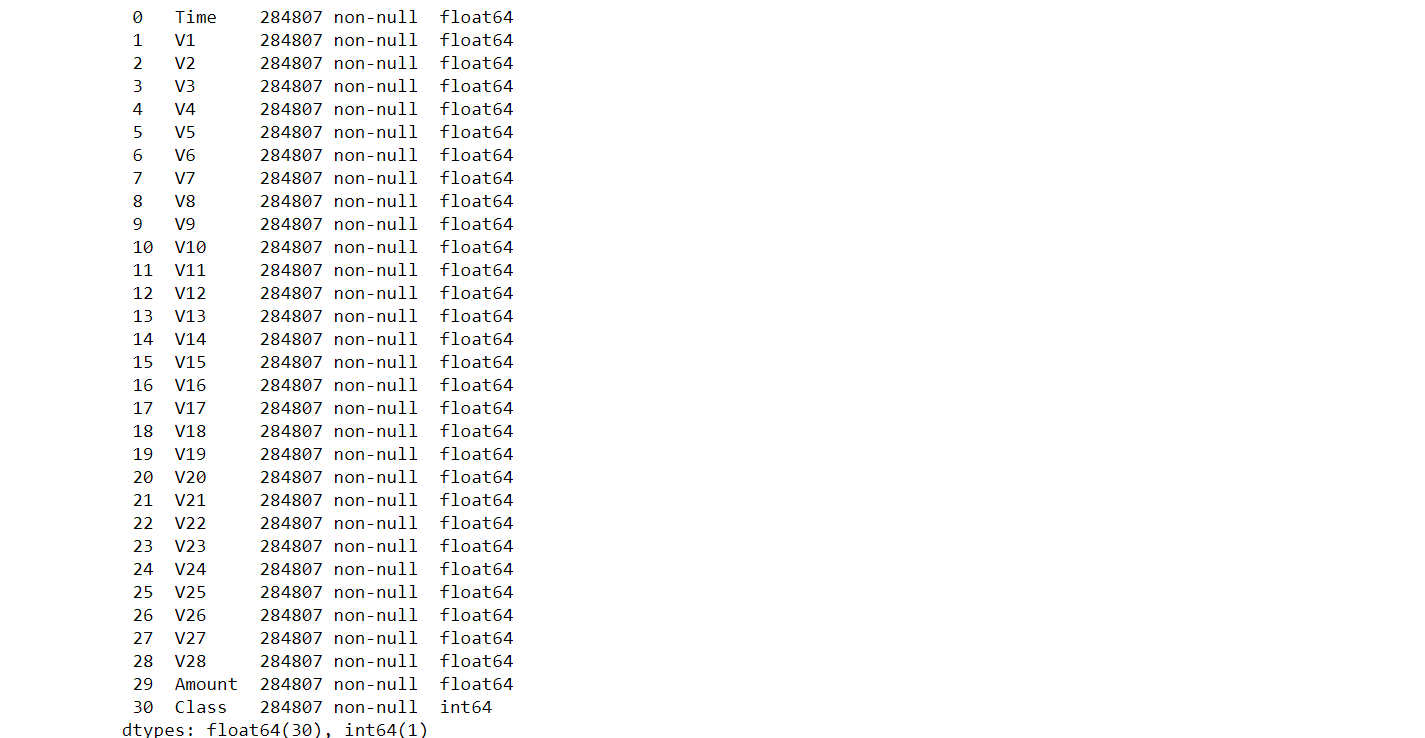
Setup Tools and pip to be installed for 3.6 and above

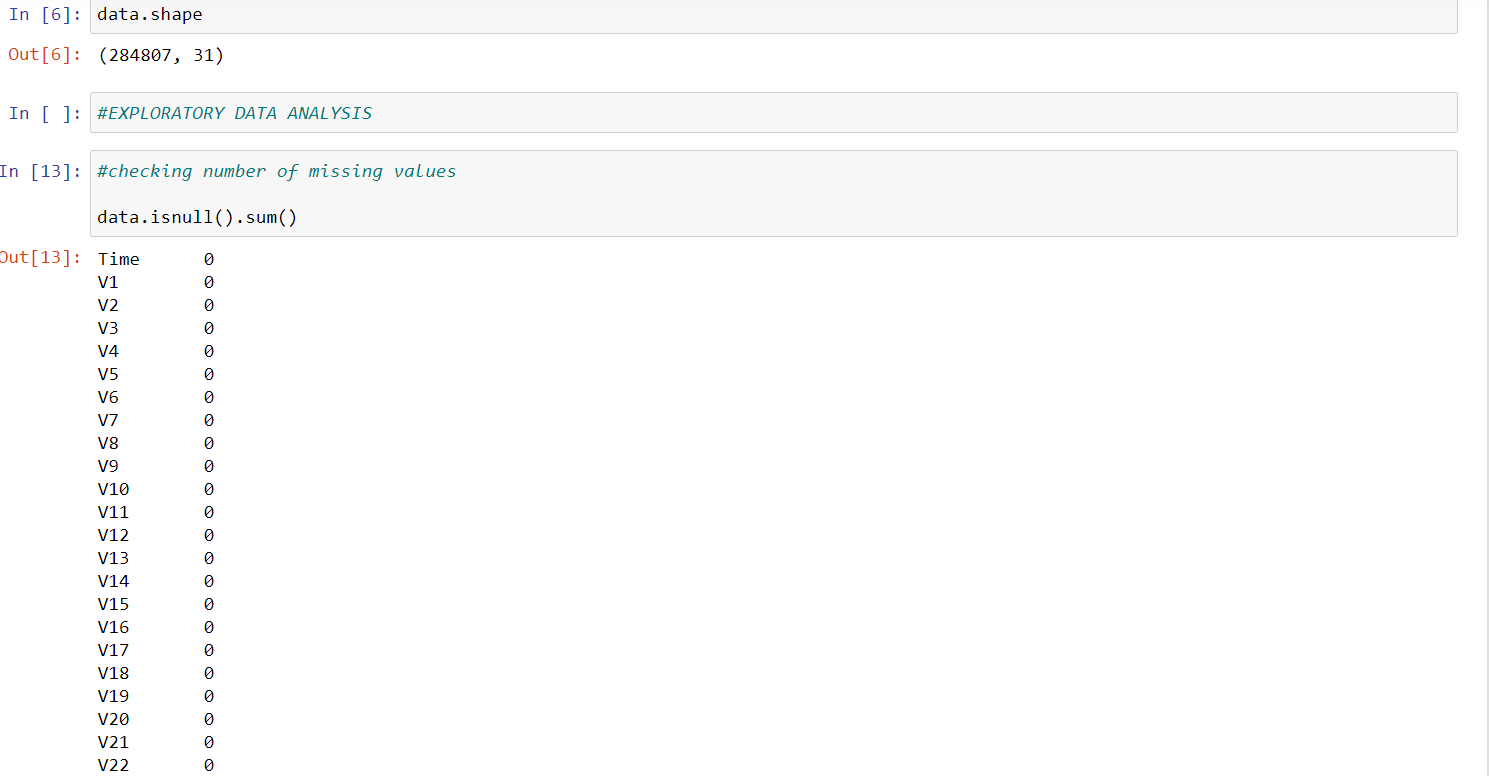
Language: Python

**CODE**

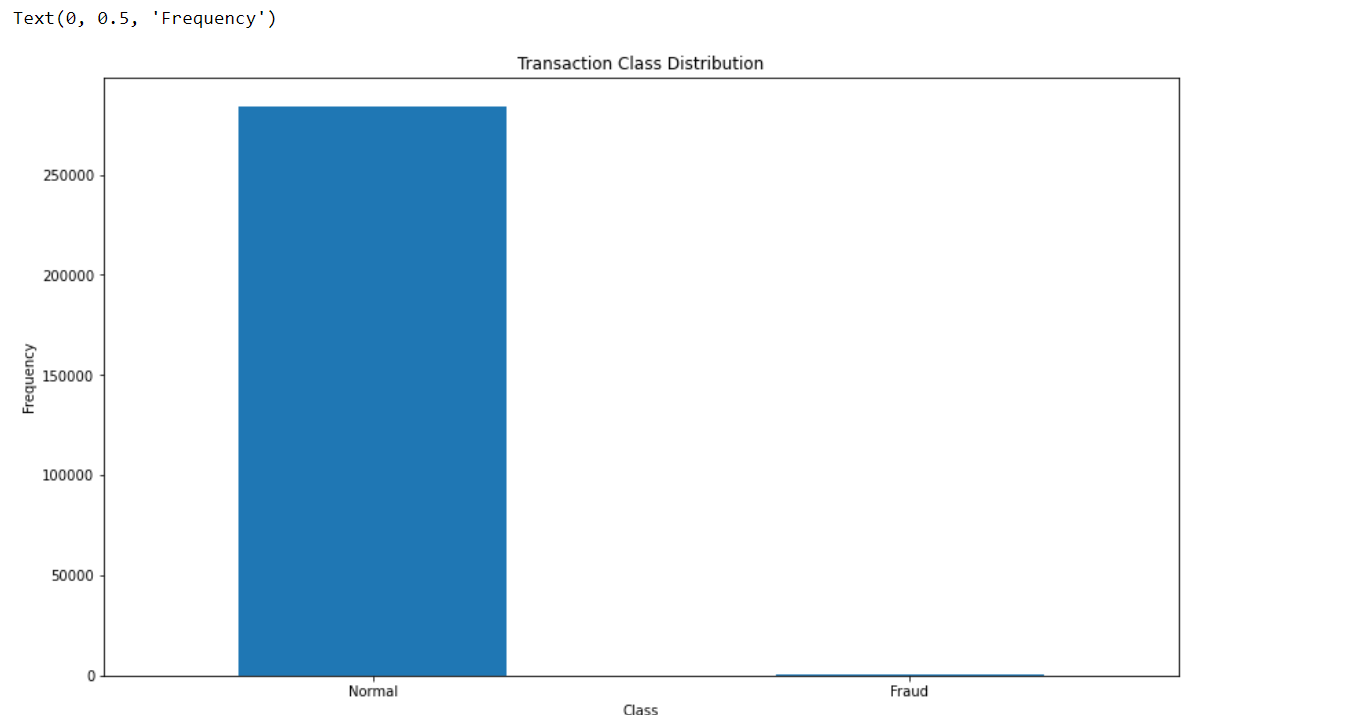


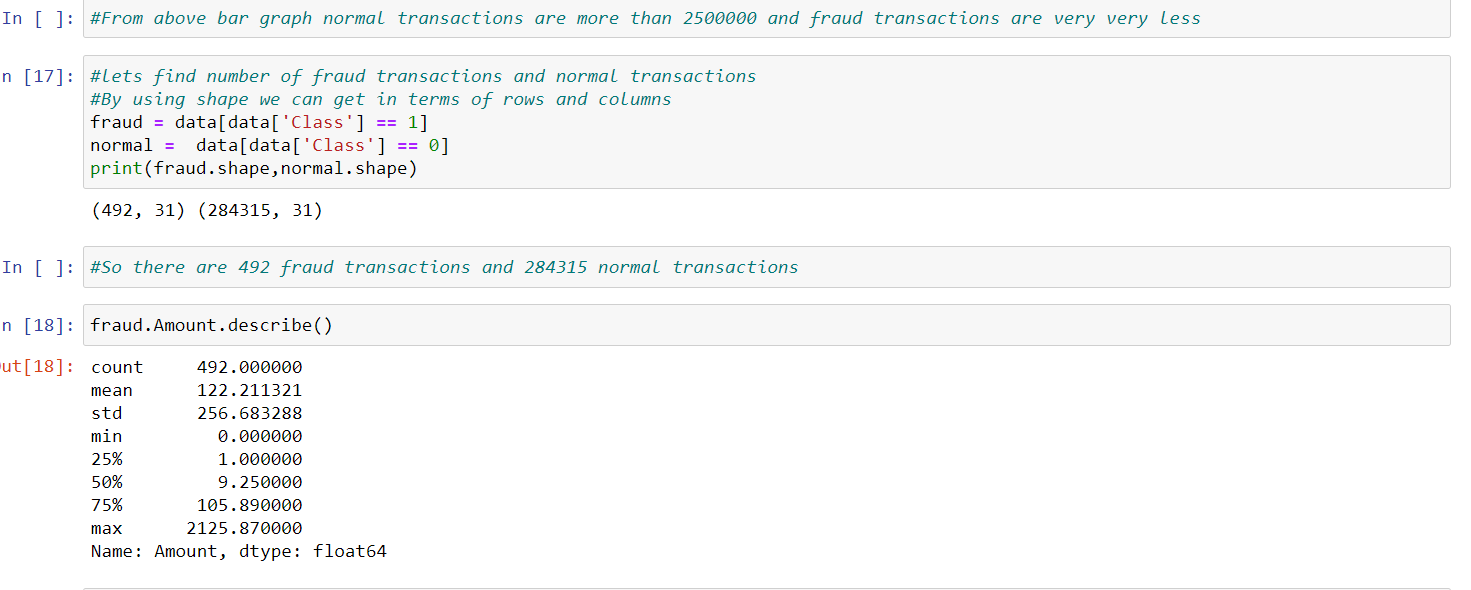


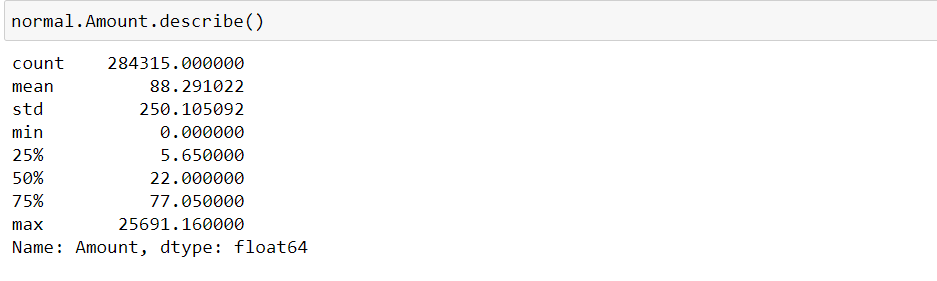


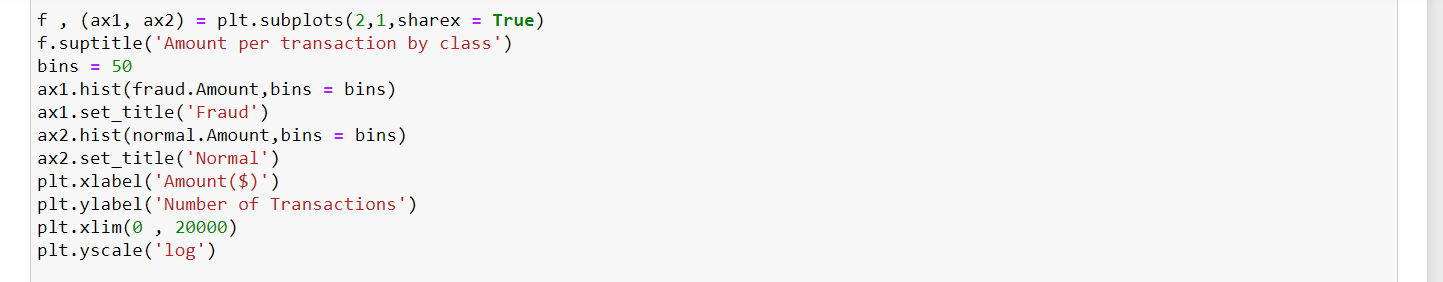


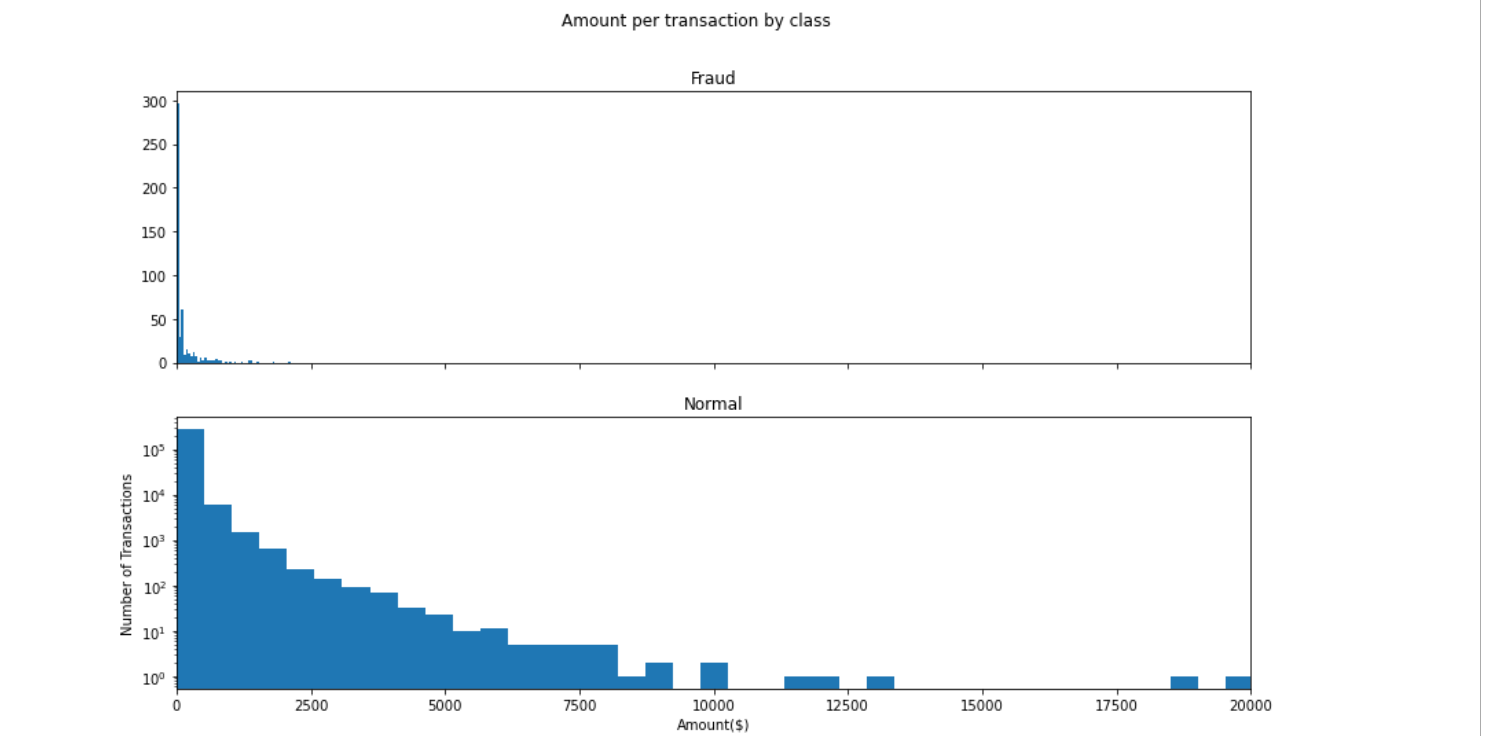


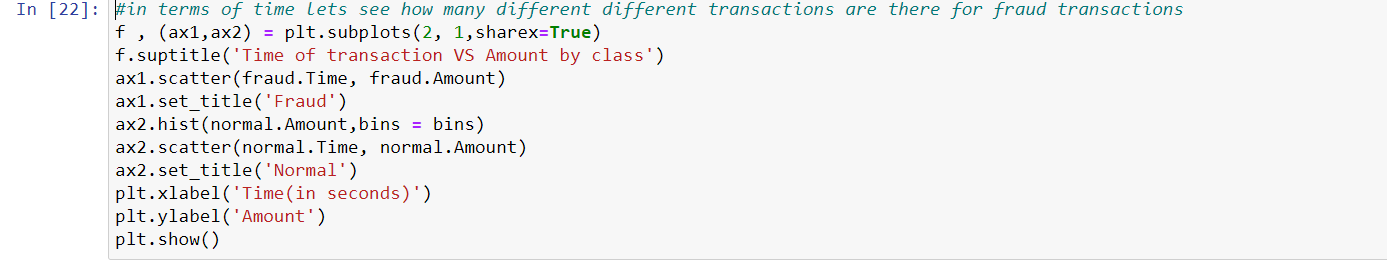


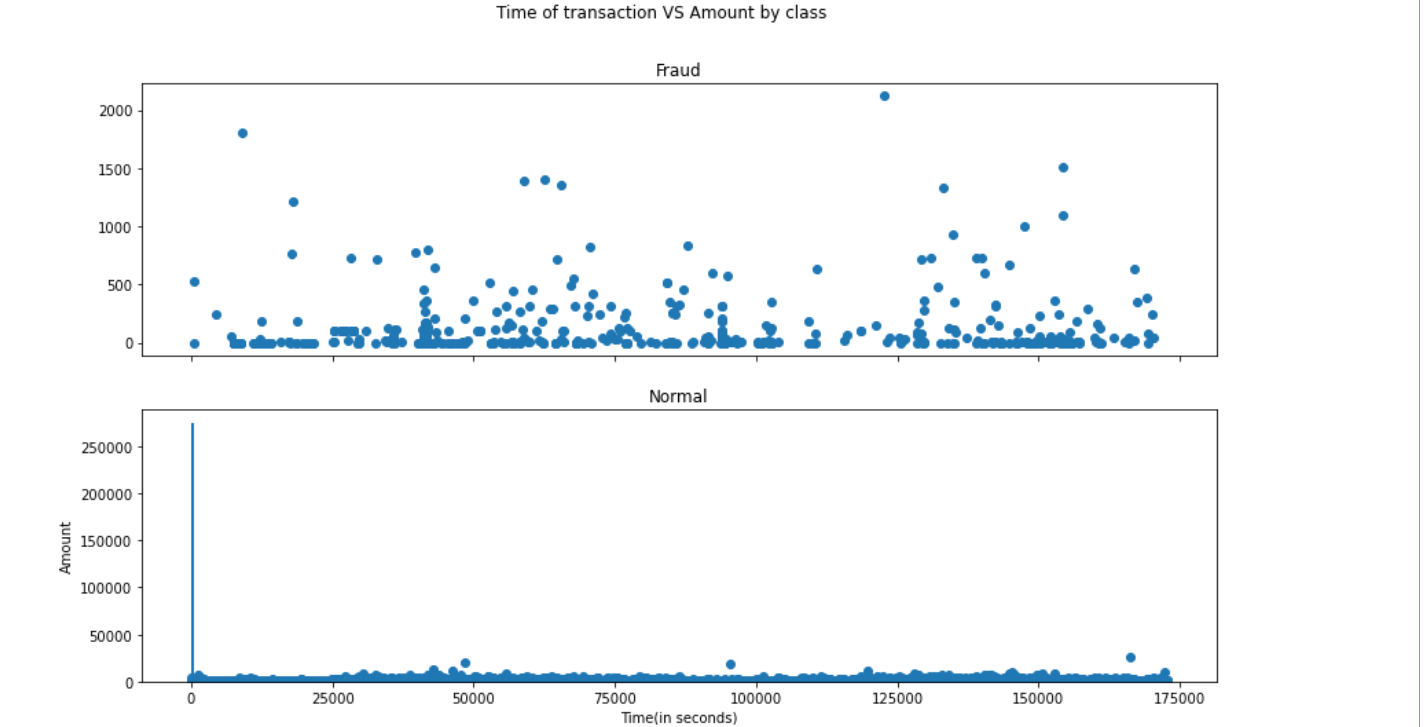


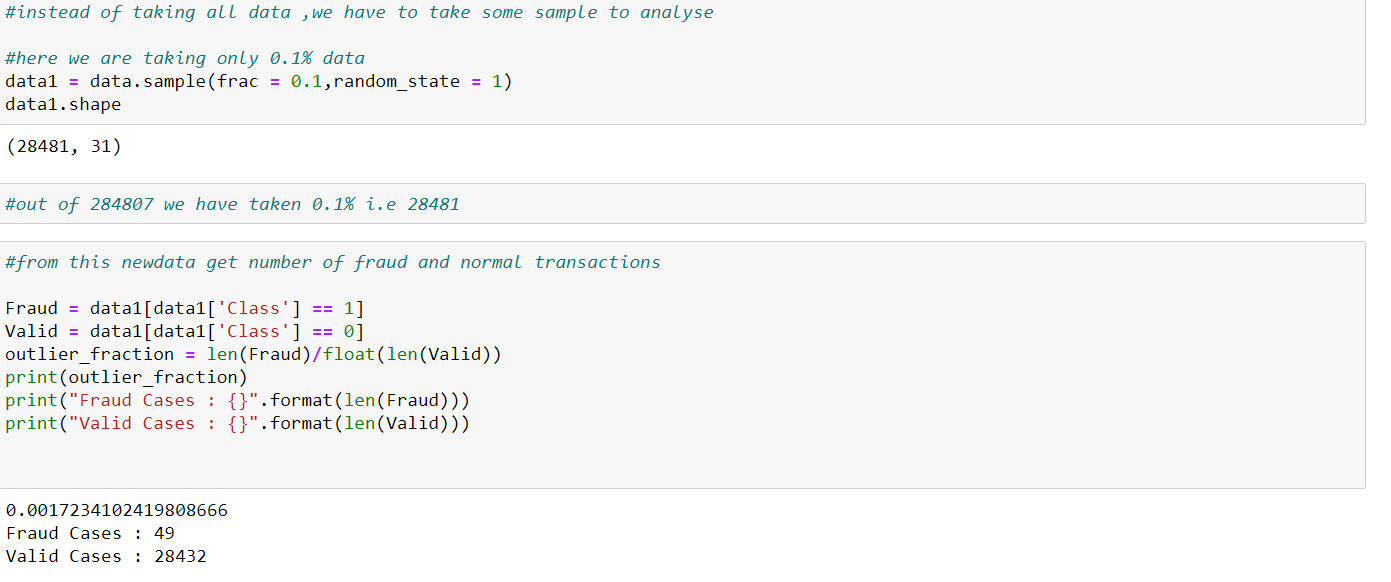


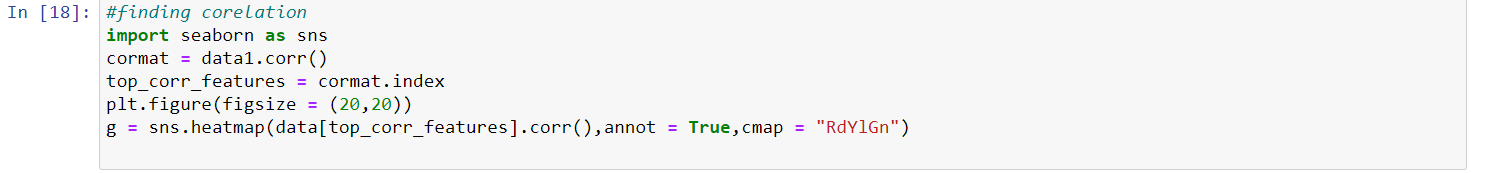


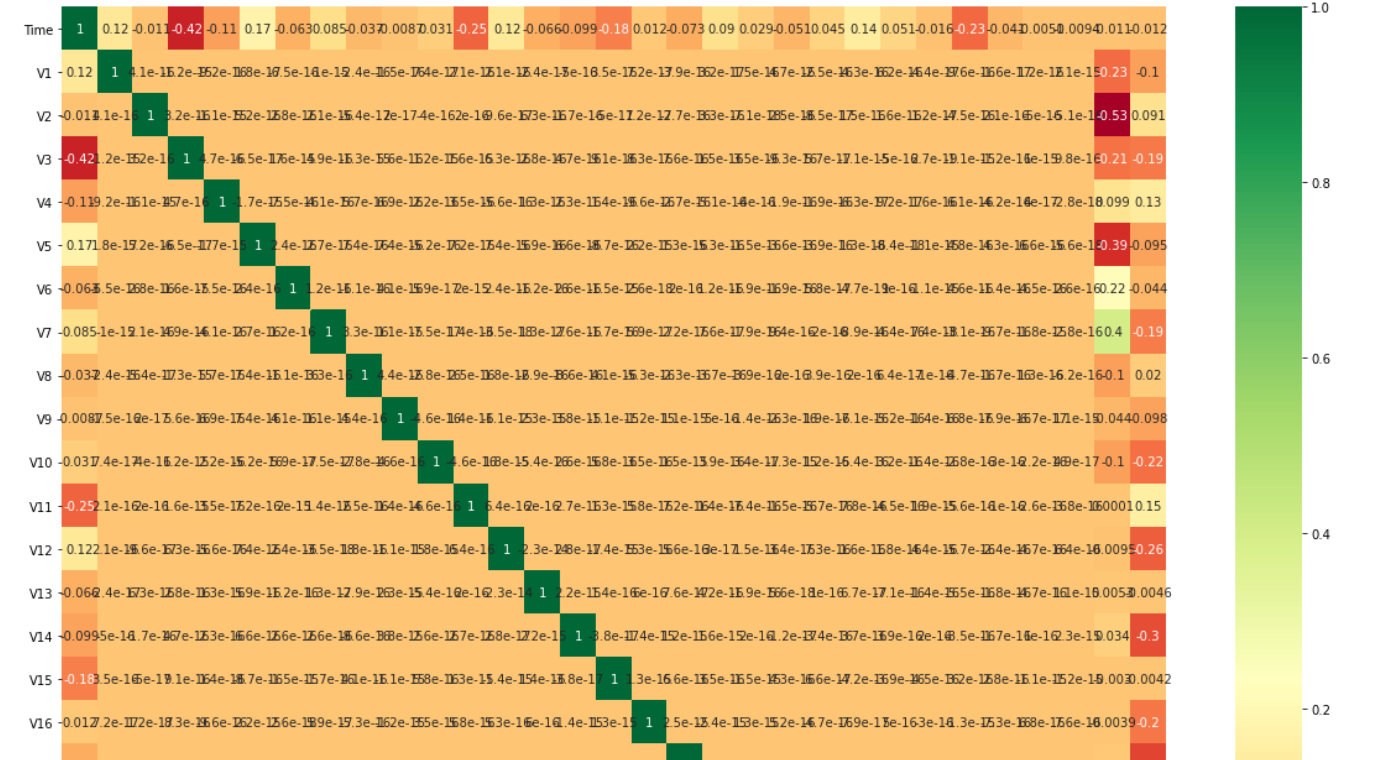


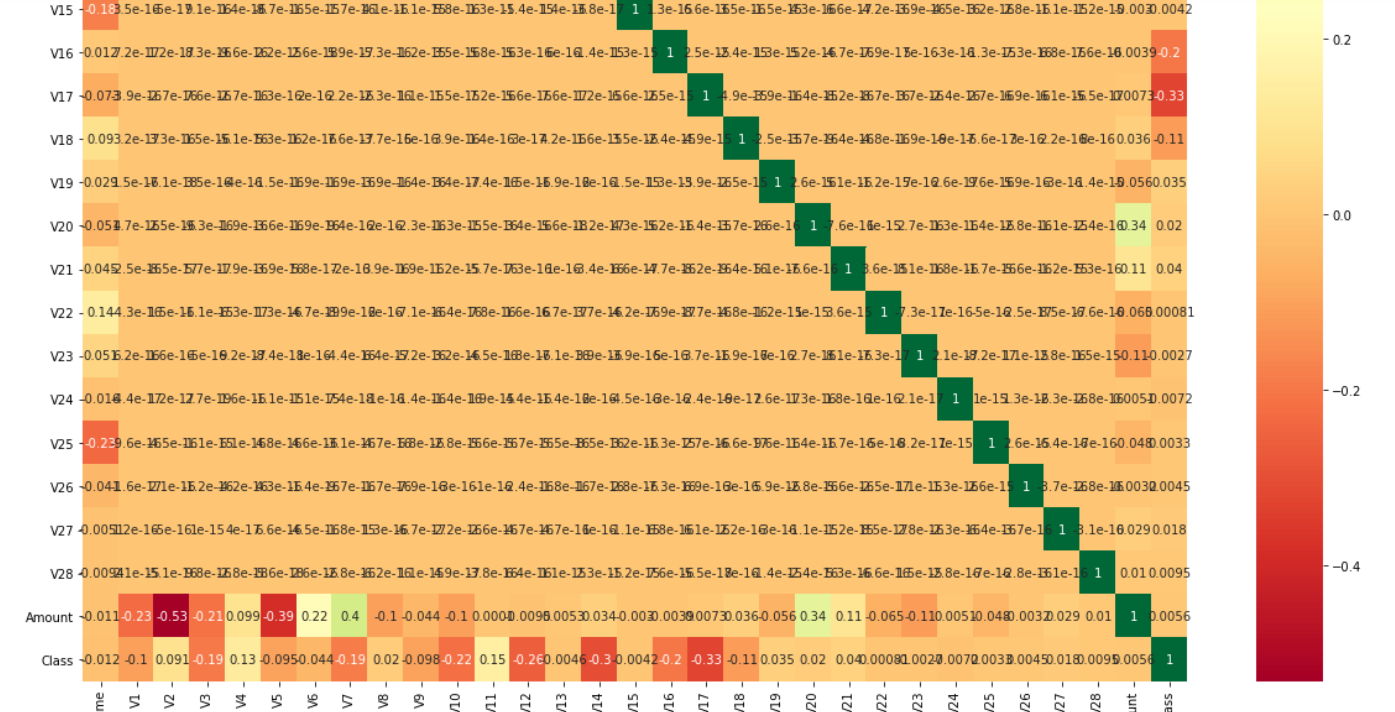


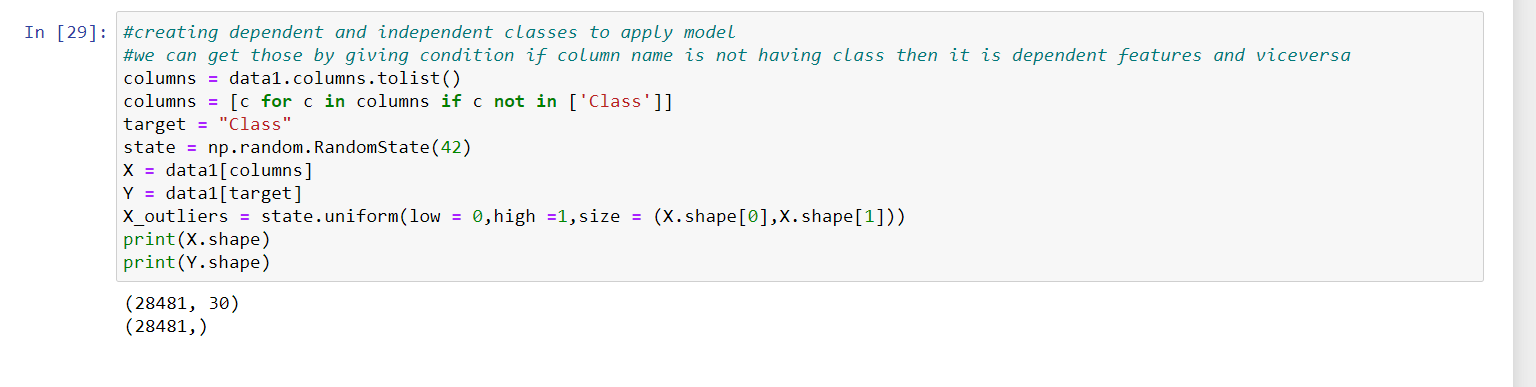




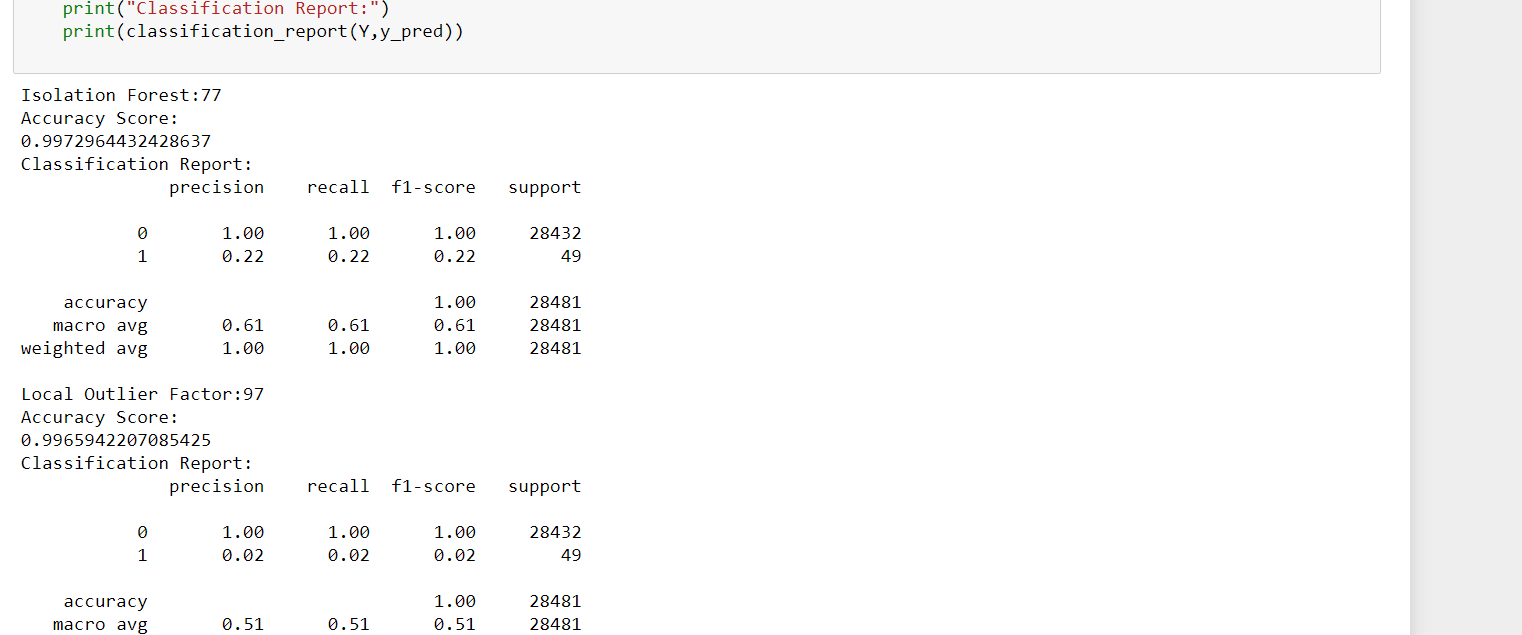


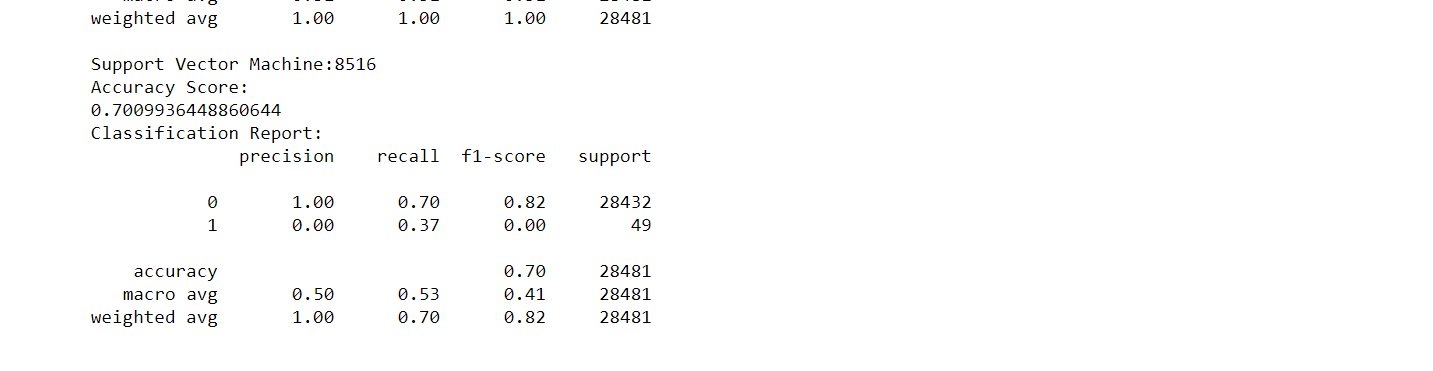


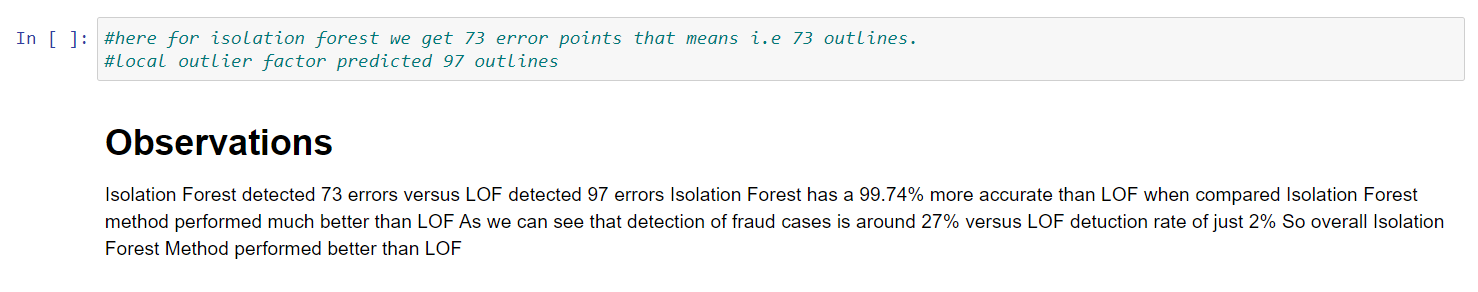












**ADVANTAGES**

The results obtained by those Algorithms are best compared to any other Algorithms.

The Accuracy obtained was almost equal to cent percent which proves using of Forest best analysis.

**SOLUTIONS**

Some proposed solutions by our team are:

1. A mechanism is developed to determine whether the given transaction is fraud or not.
2. The mechanism uses Hidden Markov Model to detect fraud detection.
3. Hidden Markov Model works on the basis of spending habit of user.
4. Classifies User into Low, Medium or High category.

CONCLUSION

Fraud Detection system have become essential for banks and financial institution, to minimize their losses.

However, there is a lack of published literature on credit card fraud detection techniques , due to unavailable credit card transaction dataset for researches.

The dataset available on day to day processing may become outdated , it is necessary to have updated data for effective fraud behaviour identification.