Credit Card Fraud Detection

Credit Card Fraud – Problem definition

- Credit card allows remote online payment transactions.
- A credit card information can be used by hackers to make fraudulent transactions.
- Detection of the fraudulent transactions can help avoiding the completion of the transaction and protect the card holders.

Data Description

The dataset has 31 attributes:

Time, **amount**, **class** and other 28 features.

•**Time** represents the time elapsed between the first transaction in the dataset, and the current transaction. It is then shown as an increasing number of seconds going down the column.

•**Amount** shows the amount of transaction that can be deposited or withdrawn.

•Other features V1-V28 are a result of a dimensionality reduction using principal component analysis (PCA), as they may carry some personal information and cannot be revealed to public users of an open source database and to ensure protection and security of their identity.

•The last column, **Class**, is considered to be our label. It takes values 0 and 1; 1 for a fraudulent transaction and 0 otherwise.

Target Class Distribution

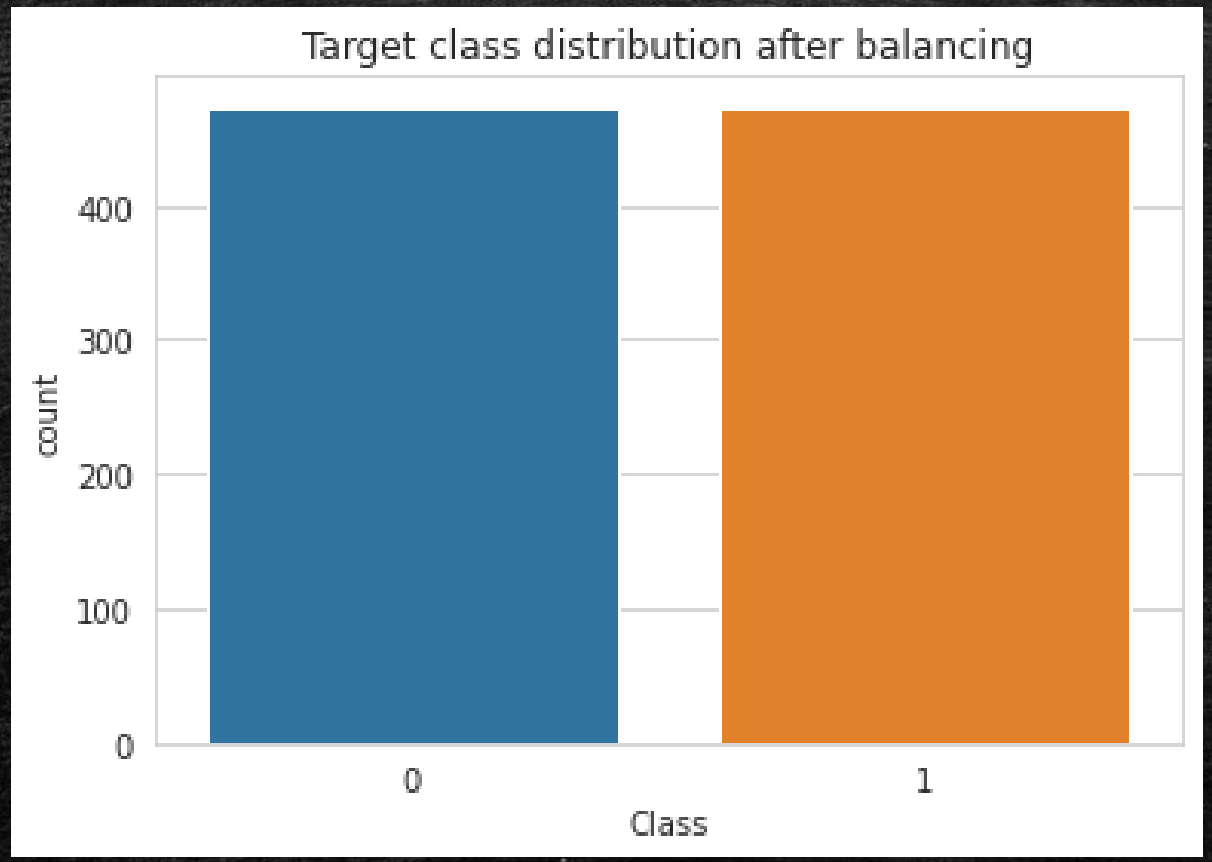
- Huge imbalance in the dataset.
- The number of non-fraudulent transactions is much greater than the number of fraudulent.

Fairness and Bias

- A dataset of the credit card transactions was collected through 2 days of recorded transactions done by European credit card holders.
- A selected timing (certain two days) for the transactions of a some category of people (Europe) cannot be generalized and used for any fraud detection. The data is biased.

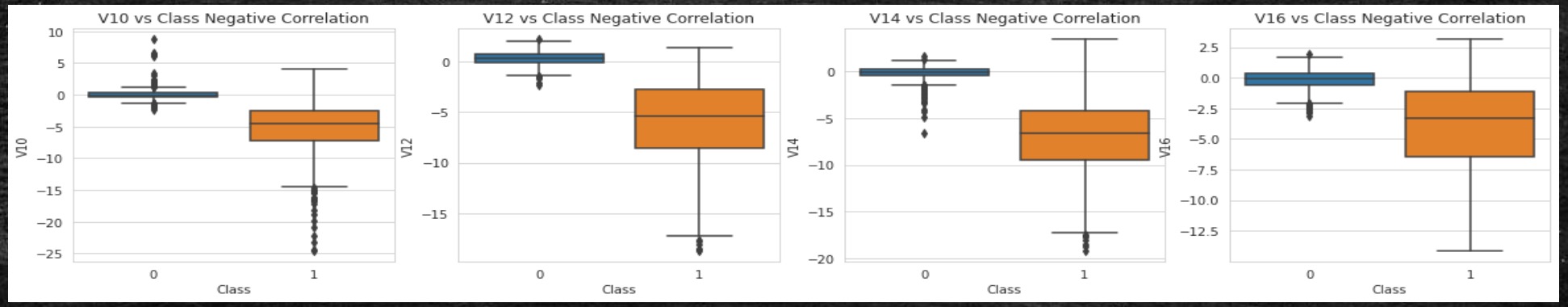
- The dataset is unbalanced (skewed towards non-fraud transactions), which results in biased decisions towards Legitimate transactions over the - more important – fraudulent transactions. Thus, balancing the data is recommended to have better results.

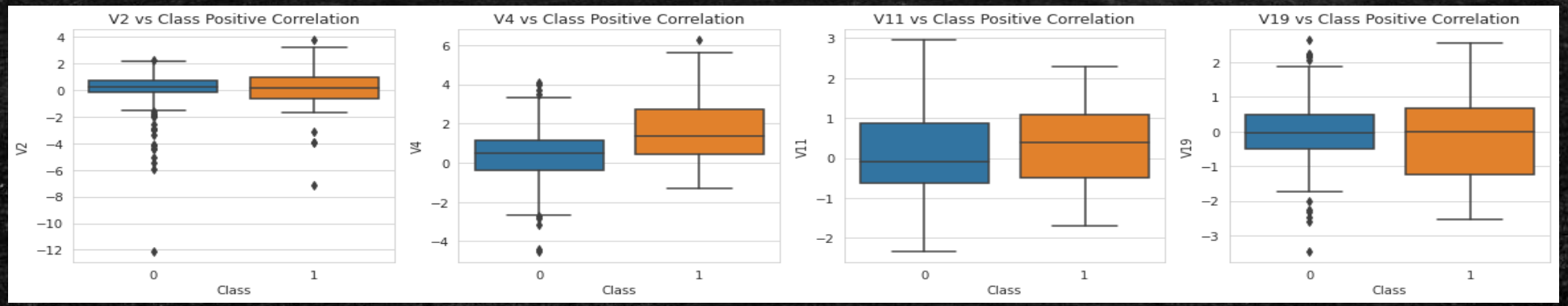
Solution: Balancing the dataset



- The number of fraudulent transactions is 492 so to balance the data consider a random 492 legitimate transactions

Outliers


V10 vs Class Negative Correlation
V10
Class
V12 vs Class Negative Correlation
V12
Class
V14 vs Class Negative Correlation
V14
Class
V16 vs Class Negative Correlation
V16
Class


V2 vs Class Positive Correlation
V2
Class
V4 vs Class Positive Correlation
V4
Class
V11 vs Class Positive Correlation
V11
Class
V19 vs Class Positive Correlation
V19
Class

What we are going to do

Table of contents