**Data:**

The dataset for this project can be accessed by clicking the link provided below.

[creditcard.csv](https://kh3-ls-storage.s3.us-east-1.amazonaws.com/Updated%20Project%20guide%20data%20set/creditcard.csv)

The dataset contains transactions made by credit cards in September 2013 by European cardholders.  
This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features **V1, V2, … V28** are the principal components obtained with PCA, the only features which have not been transformed with PCA are **'Time'** and **'Amount**'. Feature **'Time'** contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature **'Amount'** is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning. Feature **'Class'** is the response variable and it takes value 1 in case of fraud and 0 otherwise. So total 31 columns as shown below.



Fig: Raw Data

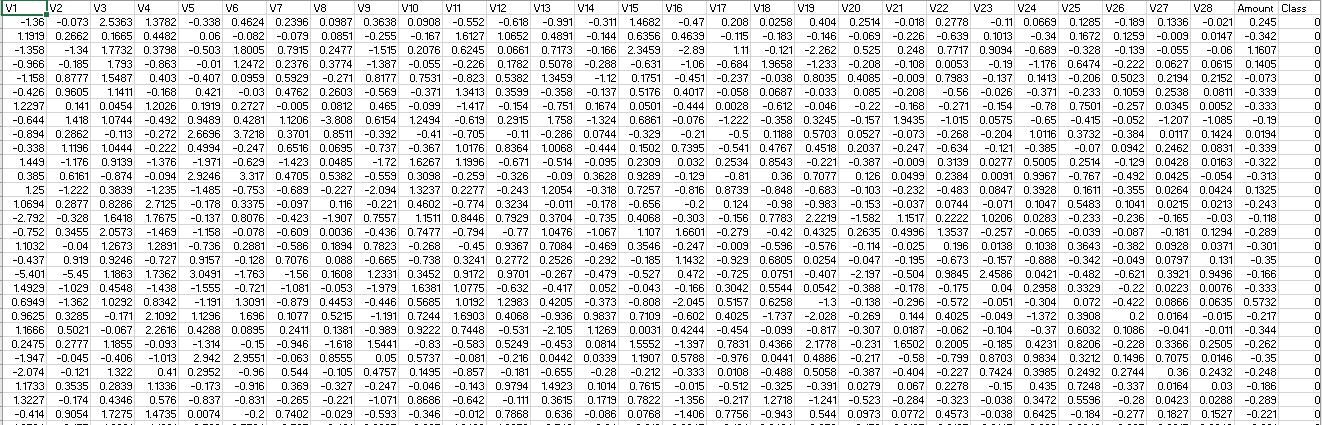


Fig: Processed Data