**FRAUD DETECTION**

**Problem**: To detect fraud transactions in given data. The data has 10000 observations.

**Analysis**

Shape of data: **284807** rows and **31** columns.

data.describe(): min time = 0 max time = 172792 (2 days). Therefore, all transactions have happened within a span of 2 days.

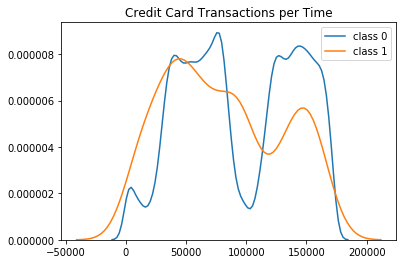
check missing data: No missing data was found.

Check whether the data in balanced or not.



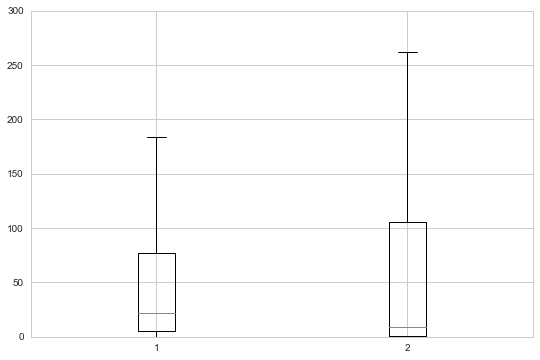
We clearly see that number of non-fraudulent activities outnumbers the number of fraudulent activities by a large number

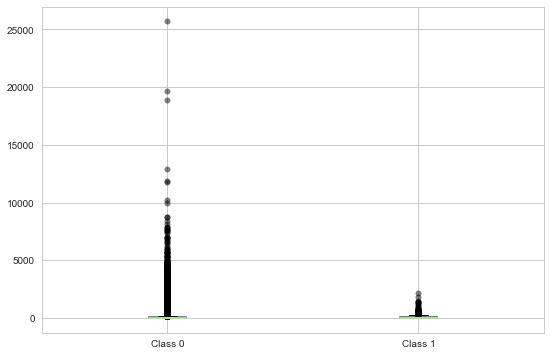
Visualize transactions over time.



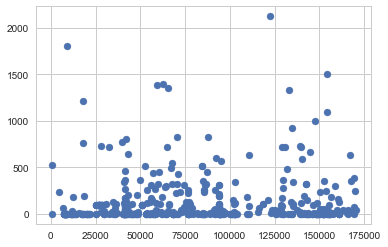
Non-fraudulent class is more uniform

**Check for outliers in amount**

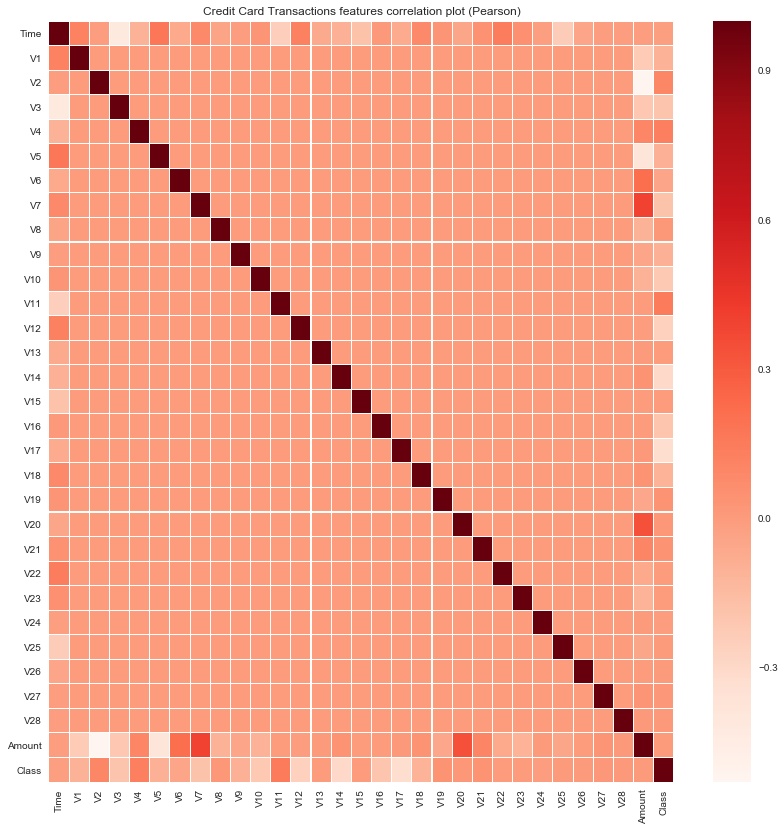




**Scatter plot for fraud class(Amount) vs Time**



**Features co-relation**



**Most of the features are independent except:**

Amount and V2, V5 (-ly correlated)

Amount and V7, V20 (+ly correlated)

Time and V25, V11 (-ly correlated)

Time and V1, V5, V7 and V12 (+ly correlated)

**Split the data into train and test (proportion 0.3)**

**Predicting with Random forest**

Confusion Matrix:

Class Not Fraud Class Fraud

85288 8

39 108

ROC: 86 %

Accuracy: 99.4%

**Predicting with SVM**

Confusion Matrix:

Class Not Fraud Class Fraud

85280 16

94 53

ROC: 86.7 %

Accuracy: 99.9%