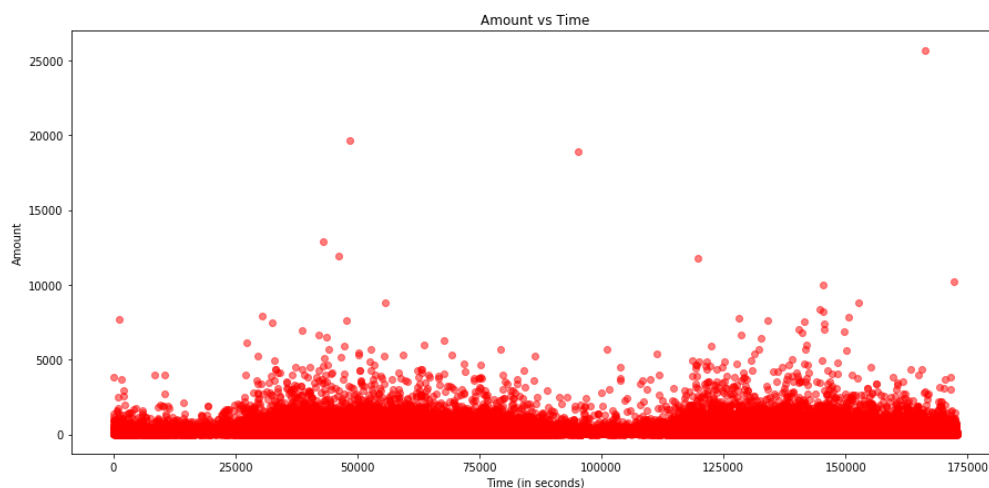


Credit Card Fraud

1. Amount vs. Time Graph:

I plotted a scatter plot to visualise the distribution of transaction amounts over time. Each dot on the graph represents a transaction. This visualisation helps in understanding if there's any trend or pattern related to transaction amounts as time progresses.



2. Fraudulent and Valid Transactions:

The data contains a total of 492 fraudulent transactions and 284,315 valid transactions. This shows a highly imbalanced distribution between the two classes, with fraudulent transactions being much less frequent.

```
Fraudulent Transactions: 492
Valid Transactions: 284315
```

3. Outlier Fraction:

The outlier fraction, representing the proportion of fraudulent transactions in the data, was calculated to be approximately 0.001730.00173 or 0.173%0.173%. This fraction is crucial when considering techniques like anomaly detection, as it indicates the rarity of fraudulent transactions.

```
Outlier Fraction: 0.00173
```

4. Correlation Between Features:

The correlation coefficients between the features and the Class column were computed. Some parts positively correlate with fraudulence, meaning they tend to increase when a transaction is fraudulent. Conversely, some features have a negative correlation, meaning their values tend to decrease for fraudulent transactions. Understanding these correlations can be essential for building predictive models.

```
Class      1.000000
V11        0.154876
```

V4	0.133447
V2	0.091289
V21	0.040413
V19	0.034783
V20	0.020090
V8	0.019875
V27	0.017580
V28	0.009536
Amount	0.005632
V26	0.004455
V25	0.003308
V22	0.000805
V23	-0.002685
V15	-0.004223
V13	-0.004570
V24	-0.007221
Time	-0.012323
V6	-0.043643
V5	-0.094974
V9	-0.097733
V1	-0.101347
V18	-0.111485
V7	-0.187257
V3	-0.192961
V16	-0.196539
V10	-0.216883
V12	-0.260593
V14	-0.302544
V17	-0.326481

5. Heatmap of Correlations:

A heatmap was created to visualise the correlation coefficients between all the features in the data. This heatmap offers a visual representation of how different parts relate. It can be helpful for feature selection and understanding multicollinearity between elements.

