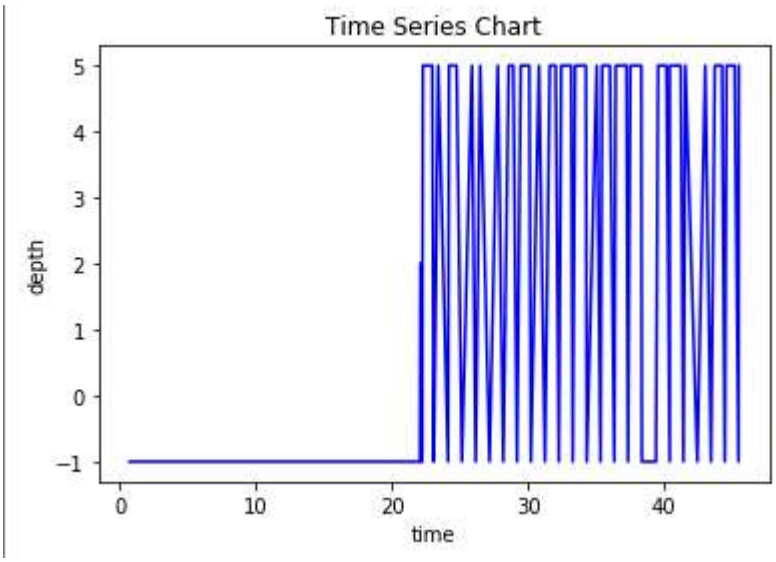
The following time series charts/graphs were created using **python's matplotlib.pyplot** library

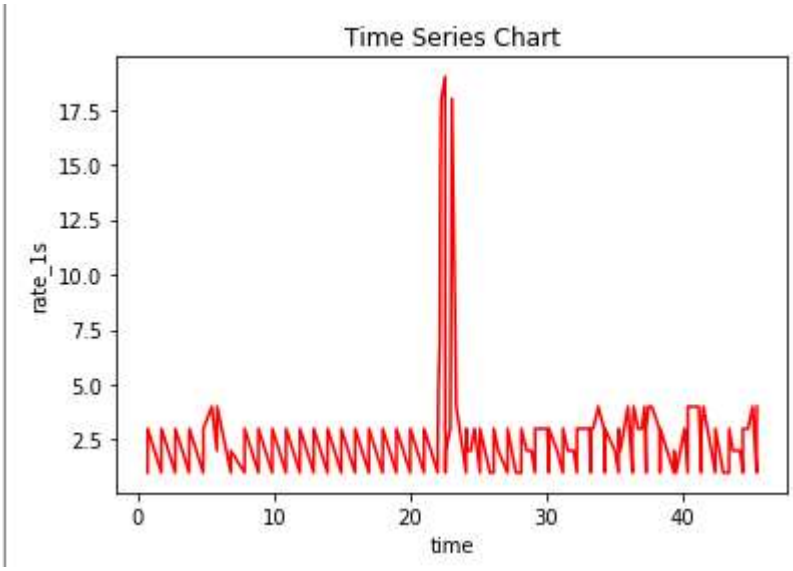
SOLUTION - 1

The values for the graph was retrieved from the exec table of the provided database

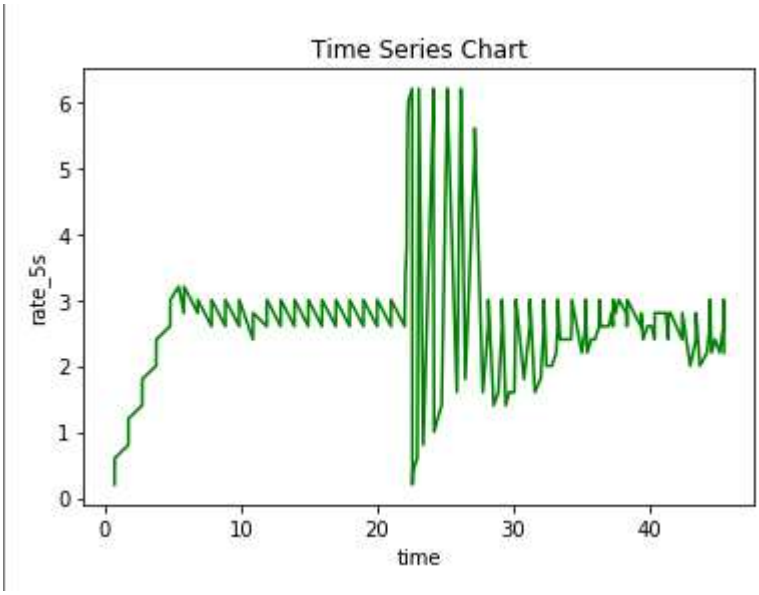
1. Depth vs Time



2. Rate_1s vs Time

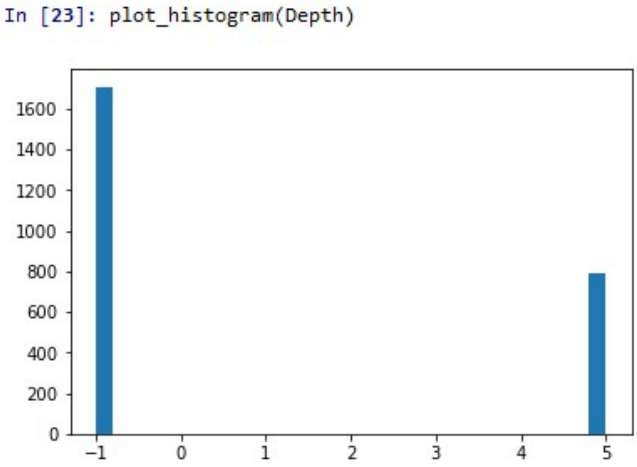


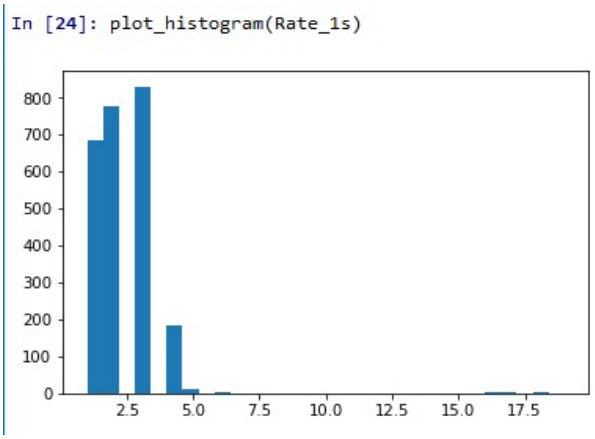
3. Rate_5s vs Time

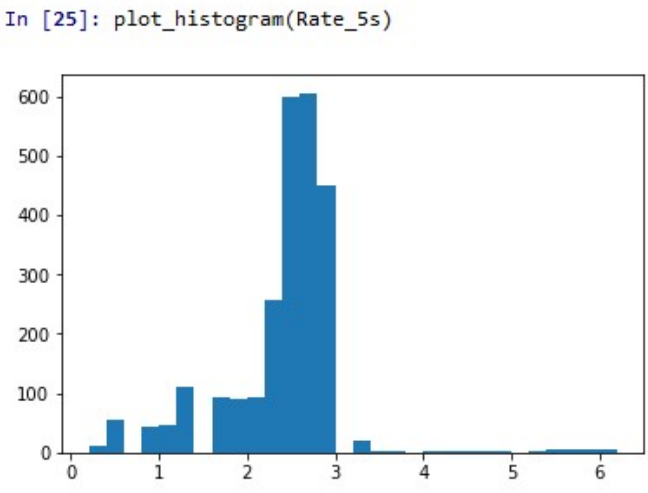


SOLUTION - 2

Histograms or frequency distributions for Depth, Rate_1s and Rate_5s



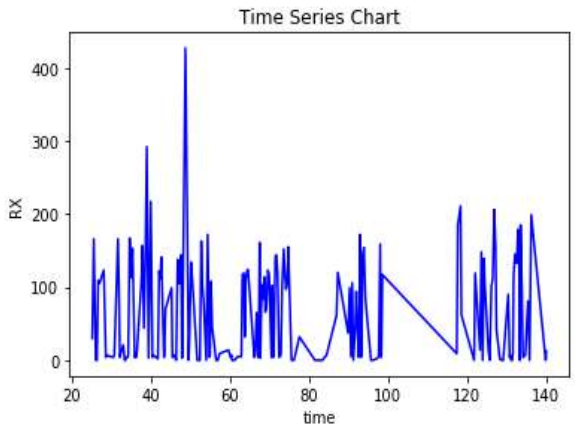




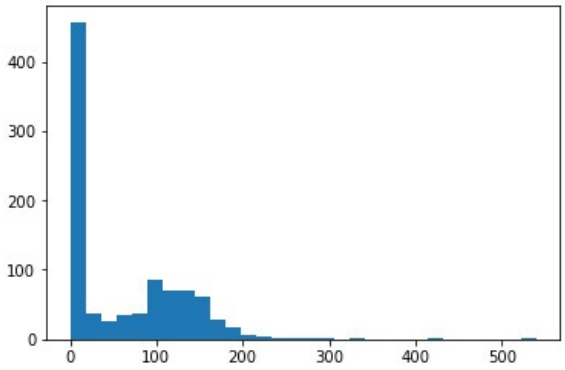
SOLUTION - 3

The values for the following graphs was retrieved from the **tcplife** table of the provided database.

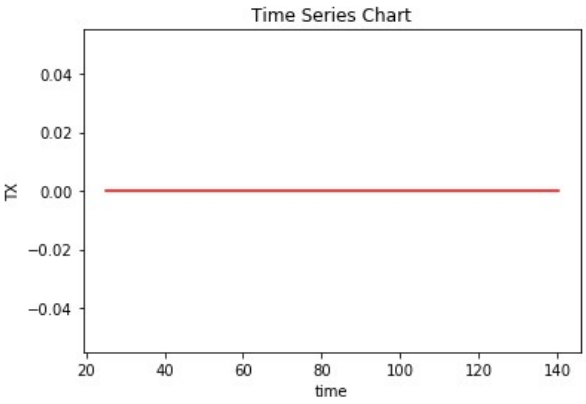
a. RX vs time



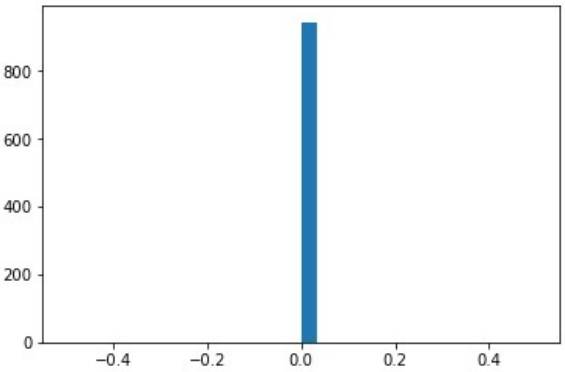

```
In [33]: plot_histogram(Rx)
```



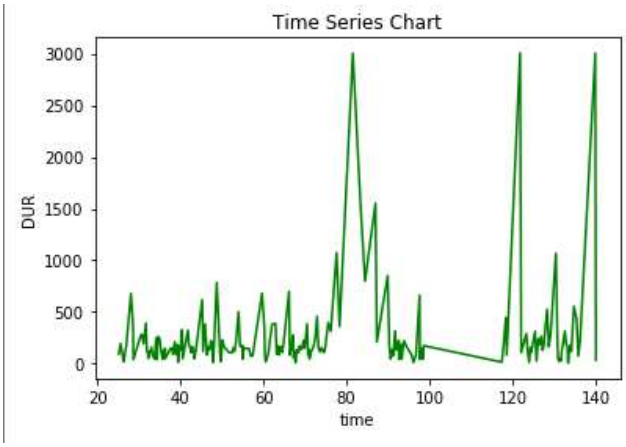
b. TX vs time



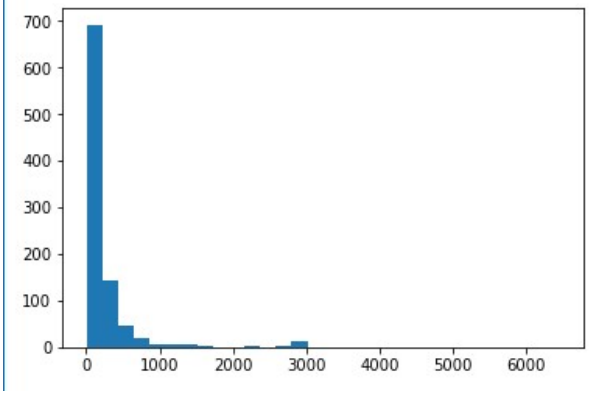

```
In [34]: plot_histogram(Tx)
```



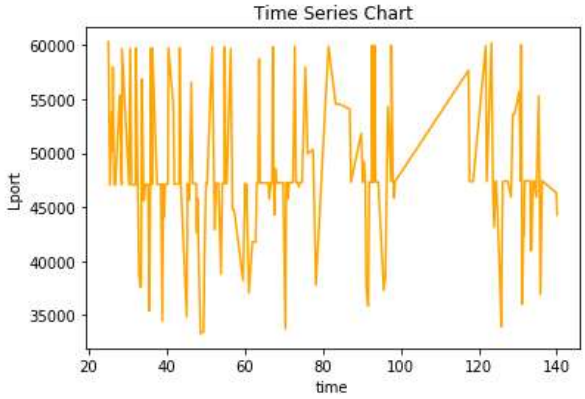
c. Dur vs time



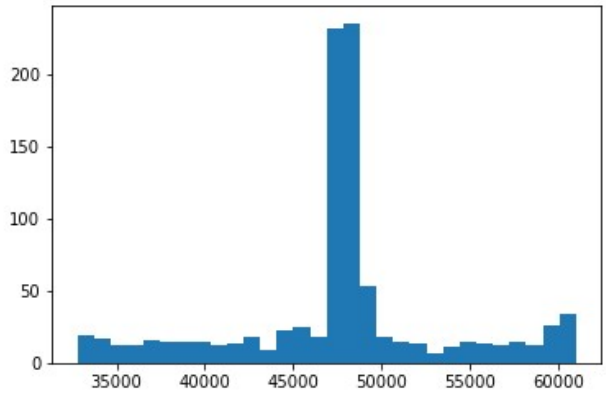

```
In [35]: plot_histogram(Dur)
```



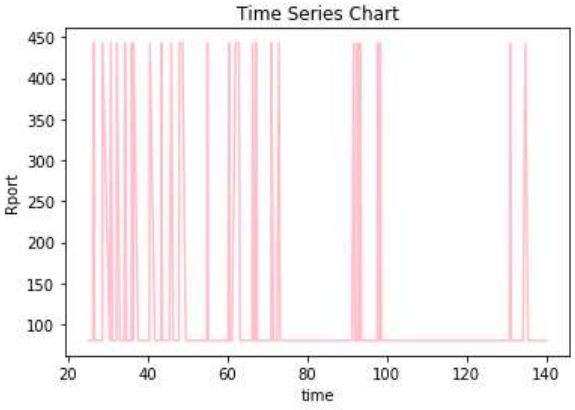
d. Lport vs time



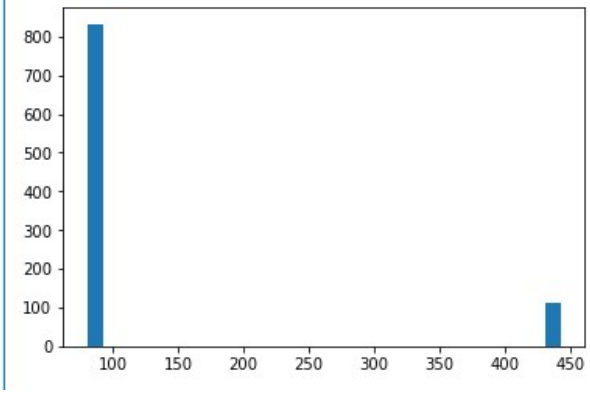

```
In [36]: plot_histogram(Lport)
```



e. Rport vs time




```
In [37]: plot_histogram(Rport)
```



SOLUTION - 4

Basic Descriptive statistics for ‘dur’ values

Mean = 261.219

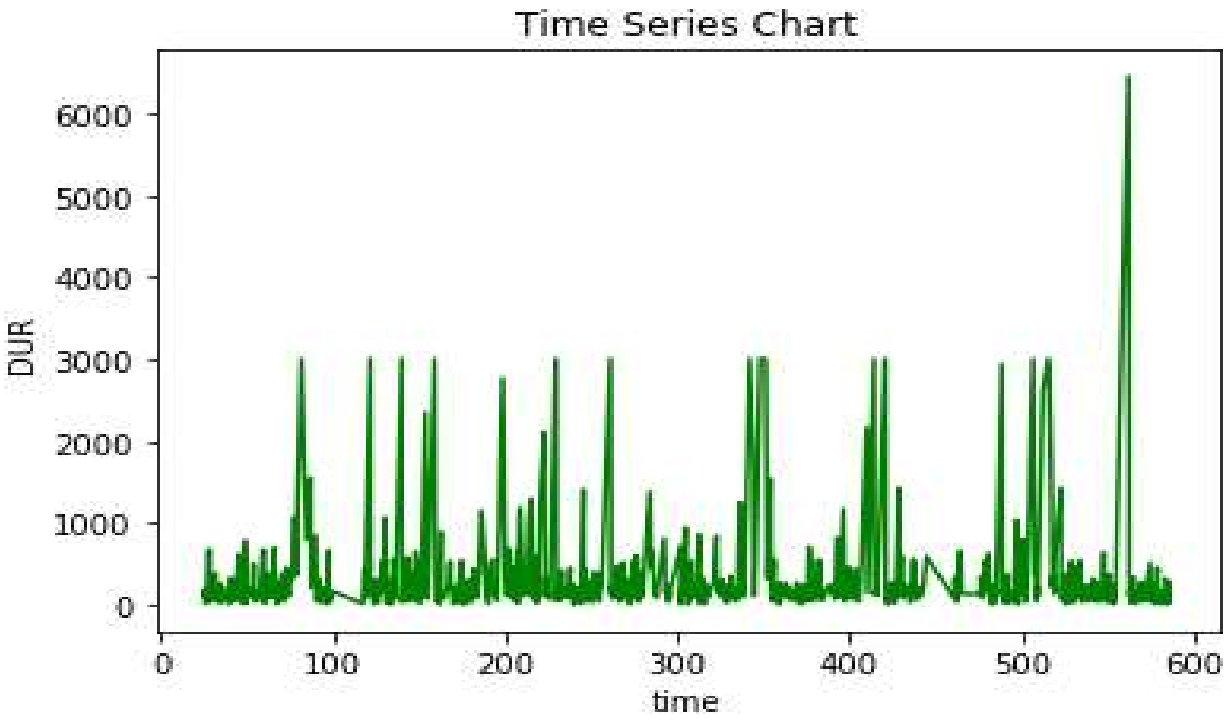
Median = 144.649

Mode = 127.26

SOLUTION – 5

Anomalous values as evident from the time series graph are the most evident outliers. There seems to be a peak in the value of dur for some 560 <= time <= 575.

At time T = 561.61435, dur = 6454.02 => outlier



SOLUTION - 6

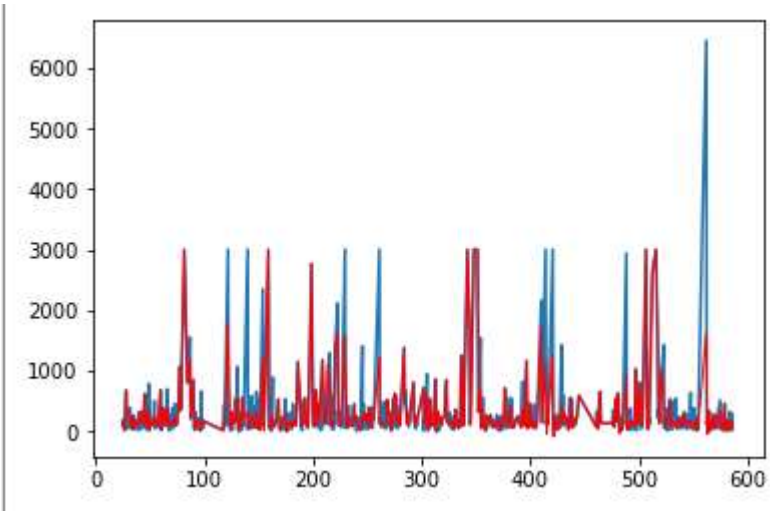
Used the svm regression algorithm implemented by scikit-learn(sklearn) to predict the next 25 values for dur.

Used the average time difference between all the time stamps

Dur prediction

The line in blue represents the time series plot of ‘dur’ values from the tcplife table

The line in red represents th values of ‘dur’ predicted by the svm regression model.



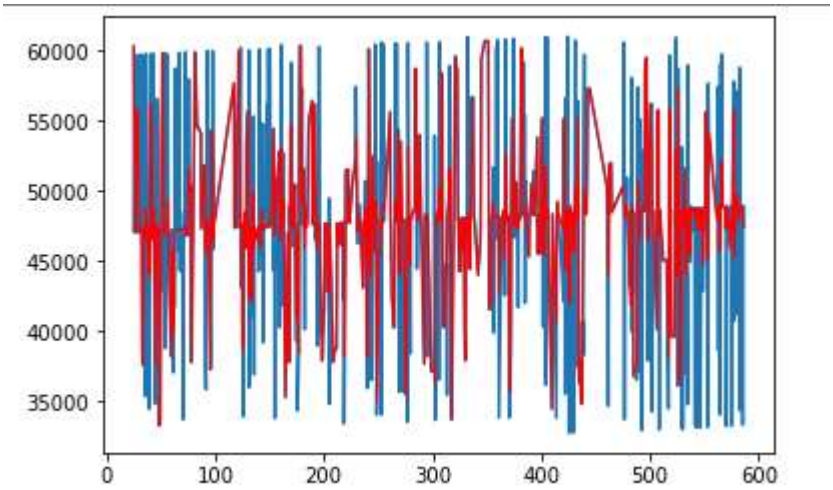
SOLUTION - 7

Used the svm regression algorithm implemented by scikit-learn(sklearn) to predict the next 25 values for 'lport'.

Used the average time difference between all the time stamps

The line in blue represents the time series plot of 'lport' values from the tcplife table

The line in red represents the values of 'lport' predicted by the svm regression model.



SOLUTION – 8

The data in the tables tcplife and exec represent the properties of unix system processes over a network.

```
In [12]: headers = retrieve_schema(Conn, 'tcplife')
    ...: for h in headers:
    ...:     print(h)
(0, 'ts', 'REAL', 0, None, 0)
(1, 'pid', 'INTEGER', 0, None, 0)
(2, 'lport', 'TEXT', 0, None, 0)
(3, 'rport', 'TEXT', 0, None, 0)
(4, 'rx', 'INTEGER', 0, None, 0)
(5, 'tx', 'INTEGER', 0, None, 0)
(6, 'dur', 'REAL', 0, None, 0)
(7, 'histotimes', 'TEXT', 0, None, 0)
(8, 'histosizes', 'TEXT', 0, None, 0)
(9, 'histoports', 'TEXT', 0, None, 0)
(10, 'docker', 'TEXT', 0, None, 0)
(11, 'prediction', 'INTEGER', 0, None, 0)
(12, 'predquality', 'INTEGER', 0, None, 0)
```

```
In [13]: headers = retrieve_schema(Conn, 'exec')
    ...: for h in headers:
    ...:     print(h)
(0, 'ts', 'REAL', 0, None, 0)
(1, 'exe', 'TEXT', 0, None, 0)
(2, 'pid', 'INTEGER', 0, None, 0)
(3, 'ppid', 'INTEGER', 0, None, 0)
(4, 'path', 'TEXT', 0, None, 0)
(5, 'docker', 'TEXT', 0, None, 0)
(6, 'depth', 'INTEGER', 0, None, 0)
(7, 'rates', 'TEXT', 0, None, 0)
```