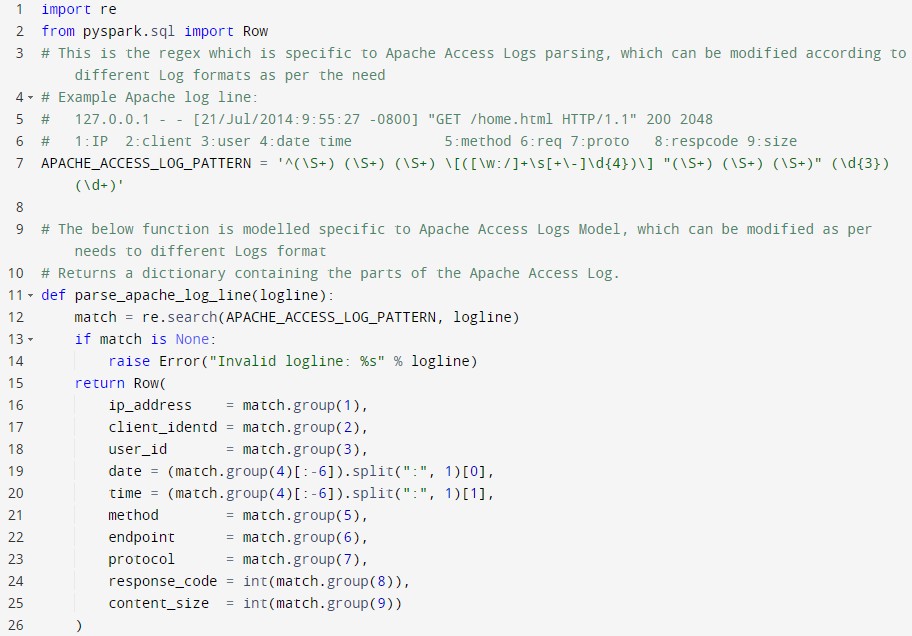
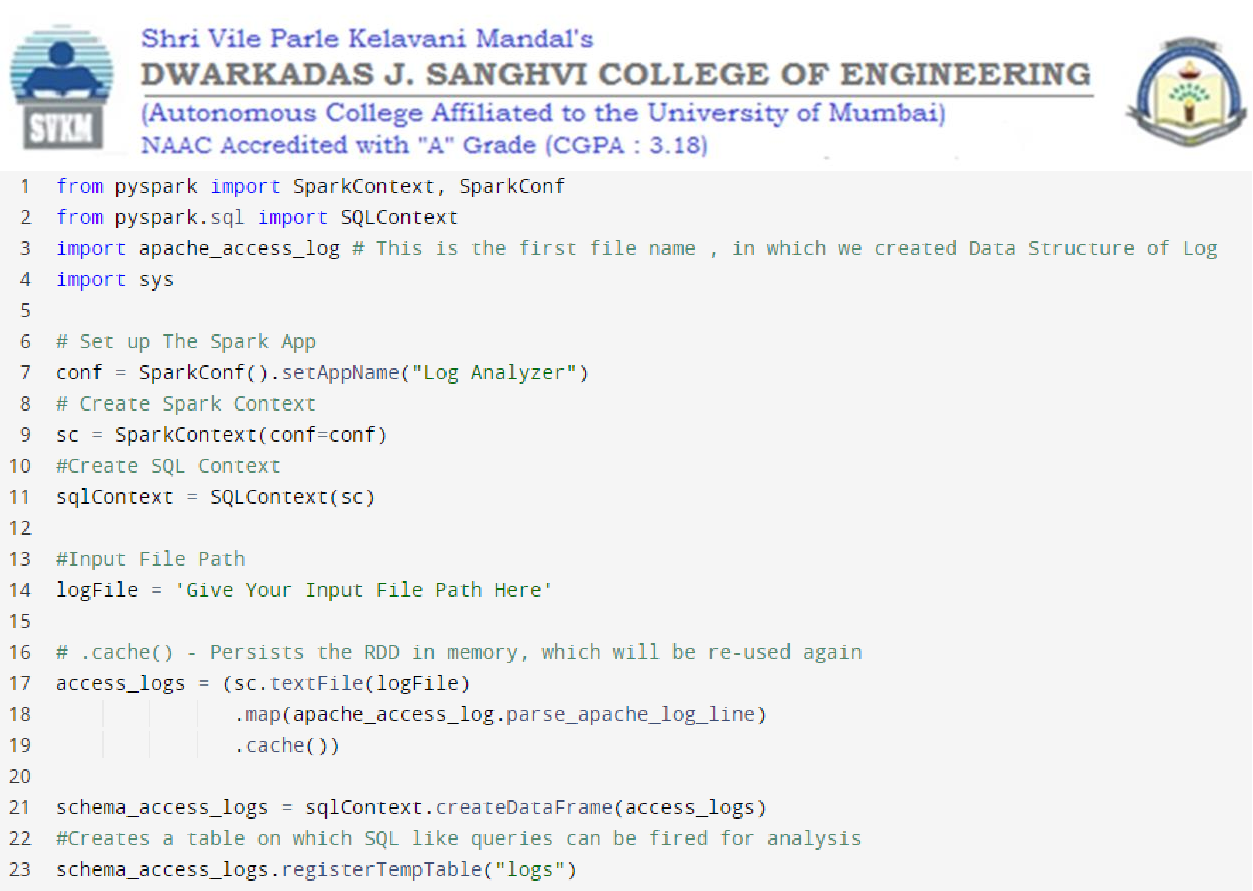
|  |  |
| --- | --- |
| Name: | Prerna Sunil Jadhav |
| Sap Id: | 60004220127 |
| Class: | T. Y. B. Tech (Computer Engineering) |
| Course: | Big Data Infrastructure Laboratory |
| Course Code: | DJ19CEEL6011 |
| Experiment No.: | 10 |

**AIM:** Perform Sentiment Analysis using Kafka.

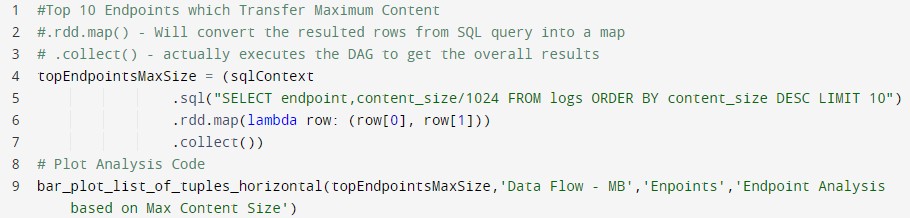
Step 1 : As the Log Data is unstructured, we parse and create a structure from each line, which will in turn become each row while analysis.

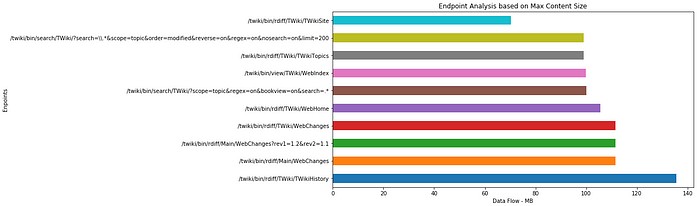


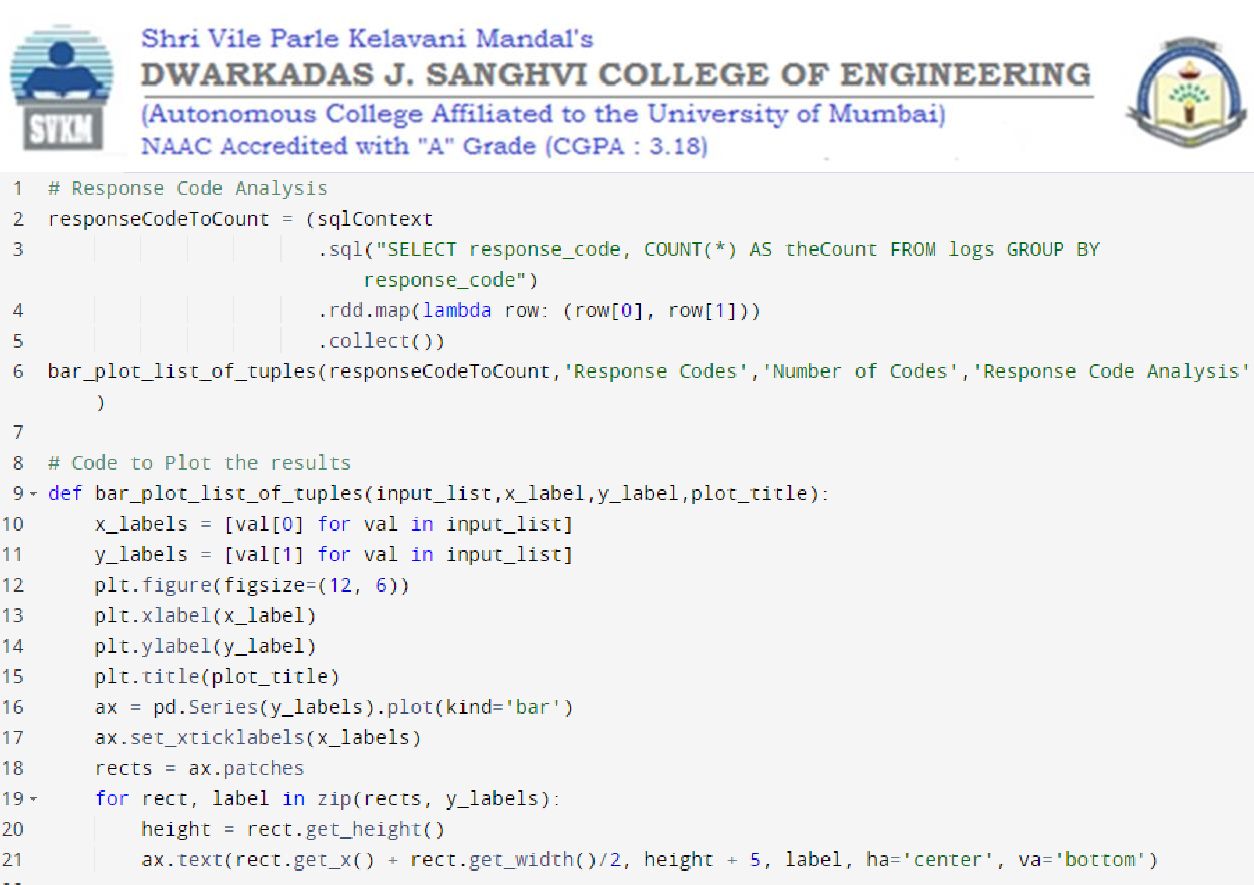
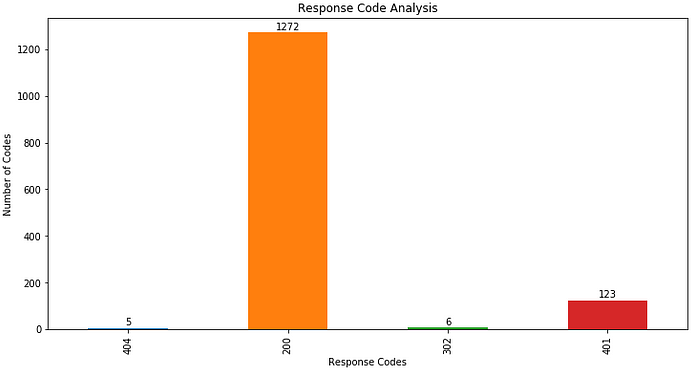
Step 2: Create Spark Context, SQL Context, DataFrame (is a distributed collection of data organized into named columns. It is conceptually equivalent to a table in a relational database)

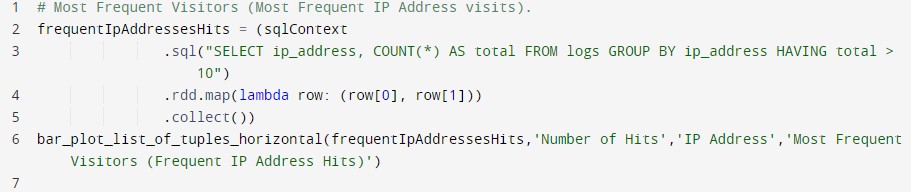


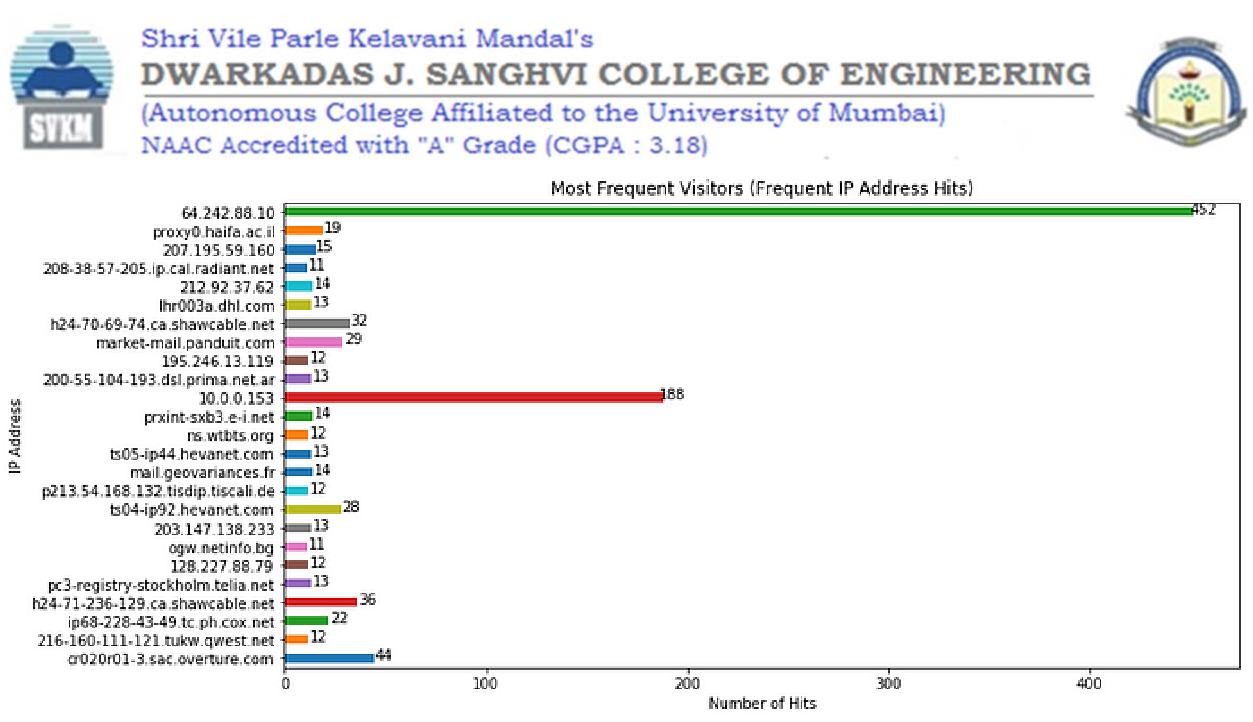
Step 3 : Analyze Top 10 Endpoints which Transfer Maximum Content in MB

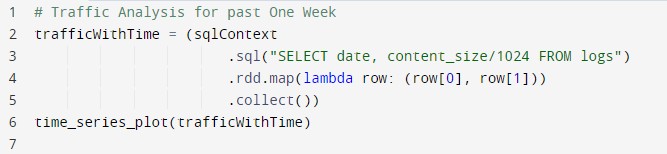


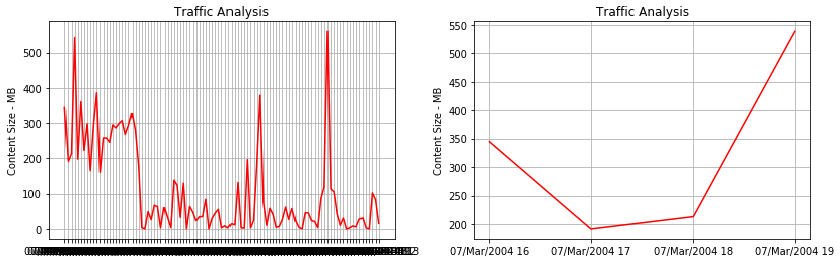


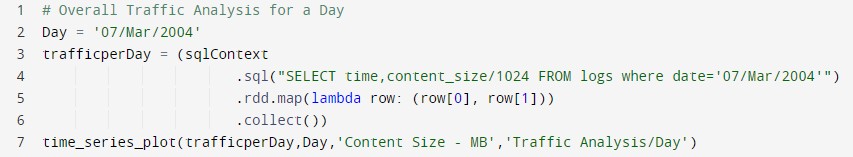


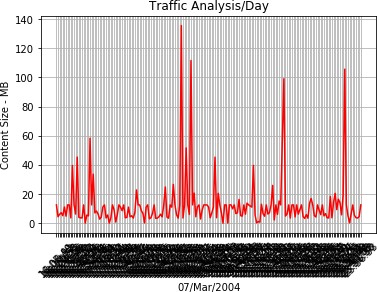




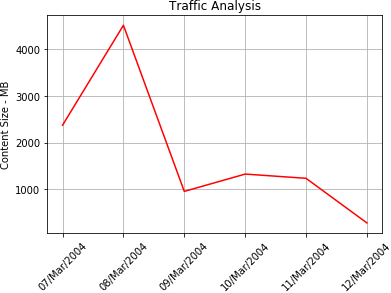








Outliers can be clearly detected by analysis the spikes and which end points were been hit at time by what IP Addresses.



Here, we can see an unusual spike on 8th March, which can be analyzed further for identifying discrepancy.

Code for Plot Analysis:

