CS 6350

ASSIGNMENT \_\_\_\_\_3\_\_\_\_\_\_\_

Names of students in your group:

Abhijeetsinh Vaghela: abv210000

Yash Niraj Majmudar: ynm210000

Number of free late days used: \_\_\_\_\_\_\_\_1\_\_\_\_\_\_\_\_\_\_\_\_   
Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

<https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html>

<https://kafka.apache.org/quickstart>

PART 1:

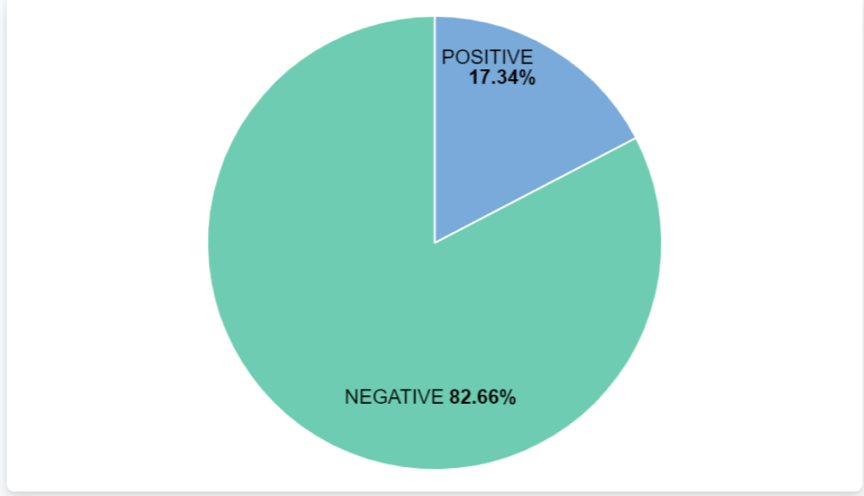
The config.ini file can be used to modify parameters to connect to kafka. The search keyword for the tweets is also defined in it.

We have used docker-compose to spin up kafka and ELK Stack containers.

Tweets are collected using the filter #FifaWorldCup.

The results and images are down below:

As we can see, majority of the sentiment are Negative which is 82.66% , given the topic we have chosen.



Chart, bar chart, histogram

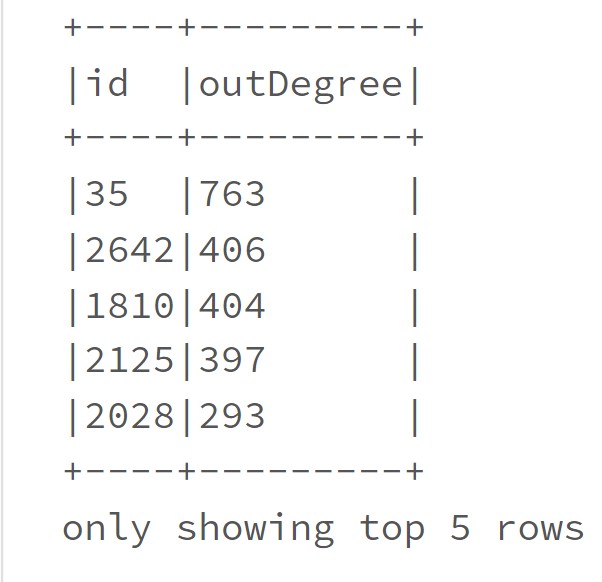
Description automatically generated

PART 2:

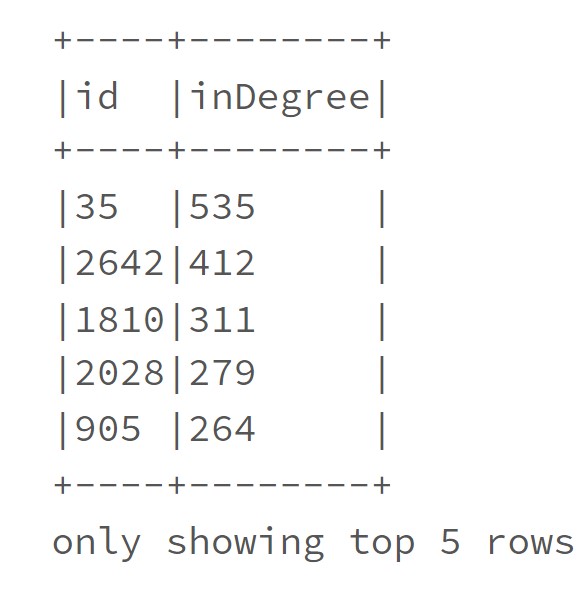
Databricks link: <https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/3441969184380383/2209160058816177/4638214704103479/latest.html>

Dataset used: <https://snap.stanford.edu/data/soc-sign-bitcoinotc.csv.gz>

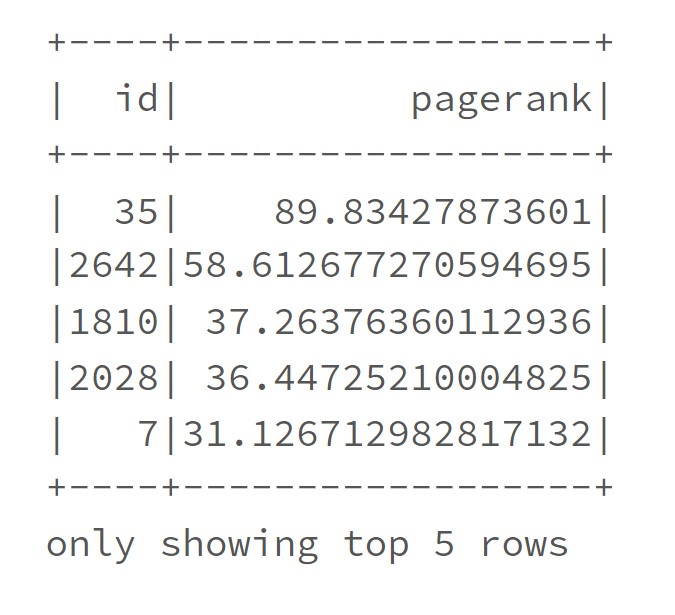
1. Find the top 5 nodes with the highest outdegree and find the count of the number of outgoing edges in each



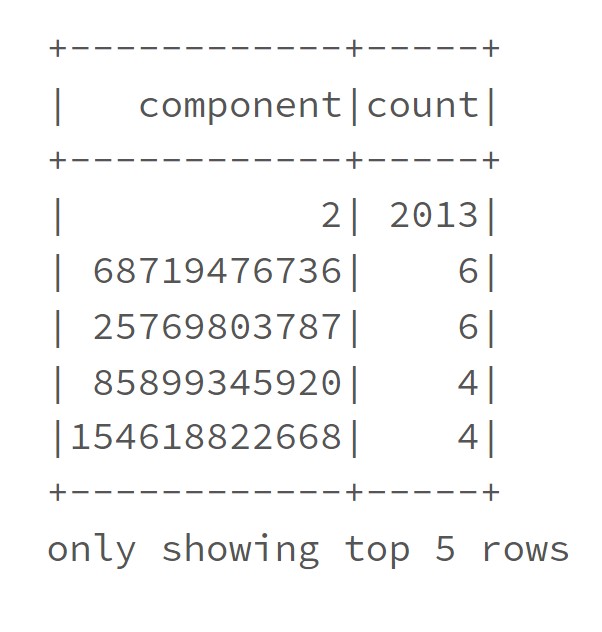
1. Find the top 5 nodes with the highest indegree and find the count of the number of incoming edges in each



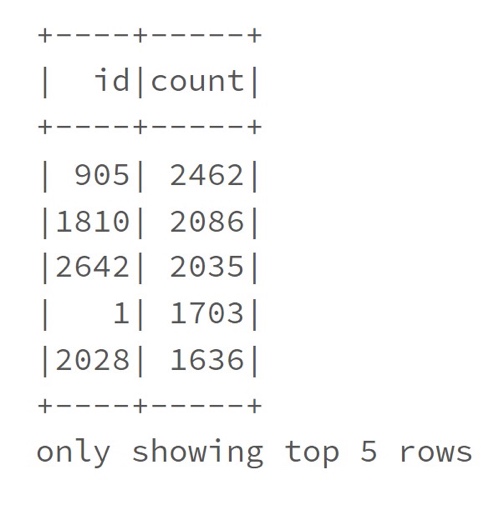
1. Calculate PageRank for each of the nodes and output the top 5 nodes with the highest PageRank values. You are free to define any suitable parameters.



1. Run the connected components algorithm on it and find the top 5 components with the largest number of nodes



1. Run the triangle counts algorithm on each of the vertices and output the top 5 vertices with the largest triangle count. In case of ties, you can randomly select the top 5 vertices.



Summary of Results:

Node-35 has highest outDegree.

Node-35 has highest indegree.

PageRank for Node-35 is highest amount all other nodes.

Since PageRank nodes are in accordance with the max indegree nodes, we can say that our results are valid ones.

Node-905 has largest triangle count of "2462”.