|  |
| --- |
| **PRINCIPLES OF BIG DATA MANAGEMENT PROJECT**  **PHASE 2:** |

**TEAM MEMBERS:**

* Kruthika Basvankote Channappa
* Dwarkamoye Mohanty
* Sai Sree Nikitha Pateel

**Objective:**

Main objective of this phase is to download the twitter data in json format and analyze the data based on the topic which we have downloaded and visualize the twitter data, while we query to get various insights from the data.

**Technologies Used:**

1. Spark

2. Zeppelin

3. Hadoop Distributed File System

4. Scala

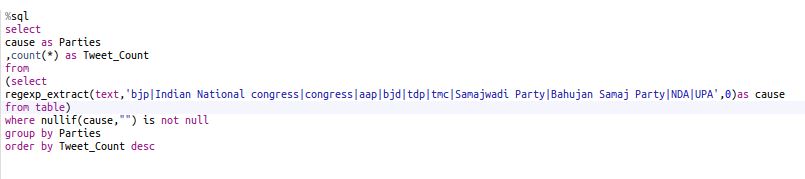
5. Python

Query 1:

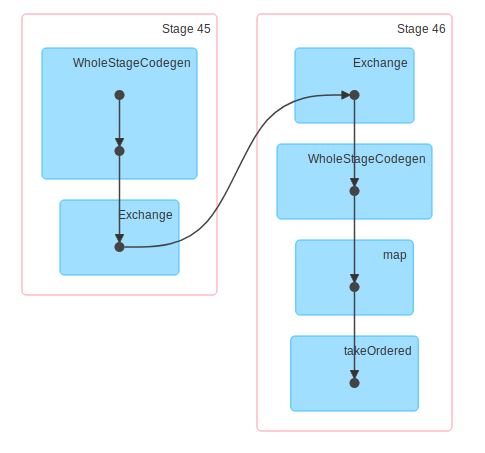
Tweets on different Political Parties

Code:

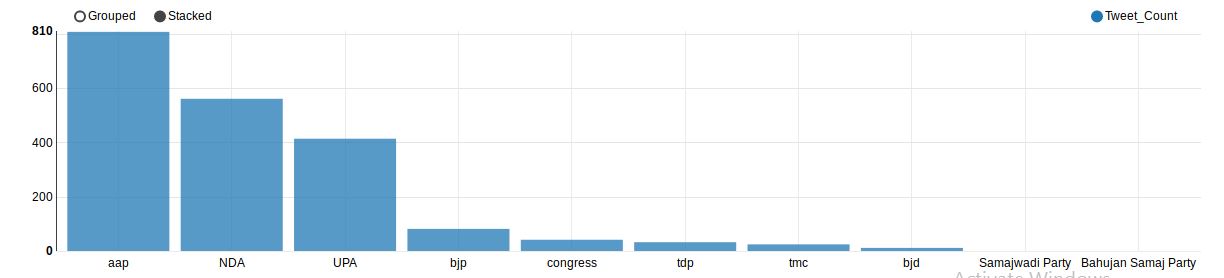
C:\Users\User\Desktop\PB_Project\parti-code1.JPG



DAG Visualization:



Visualization:

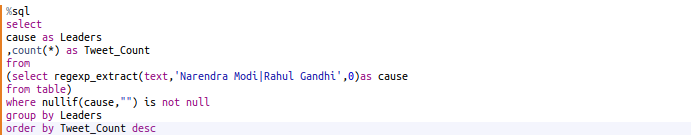


Query 2:

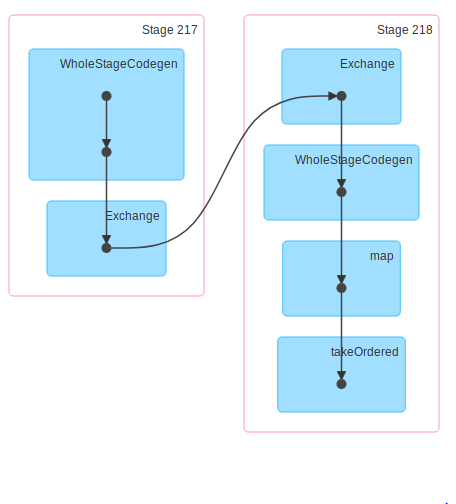
Comparing two politics supreme:

Narendra Modi vs Rahul Gandhi

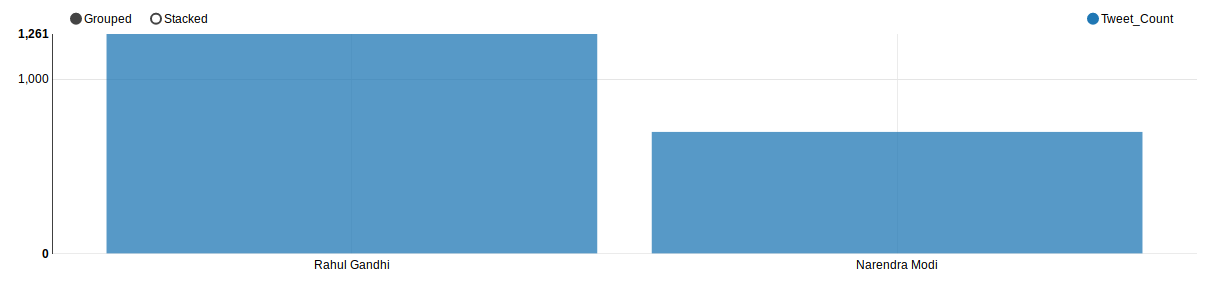
Code:



DAG Visualization



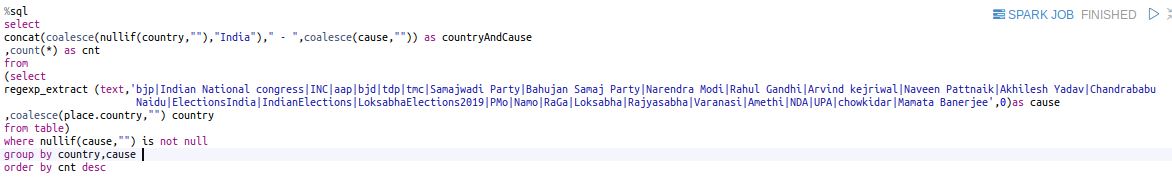
Visualization



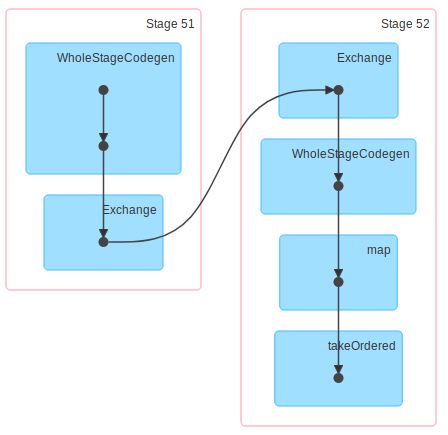
Query 3

Tweets from Countries Based on Different Hashtags

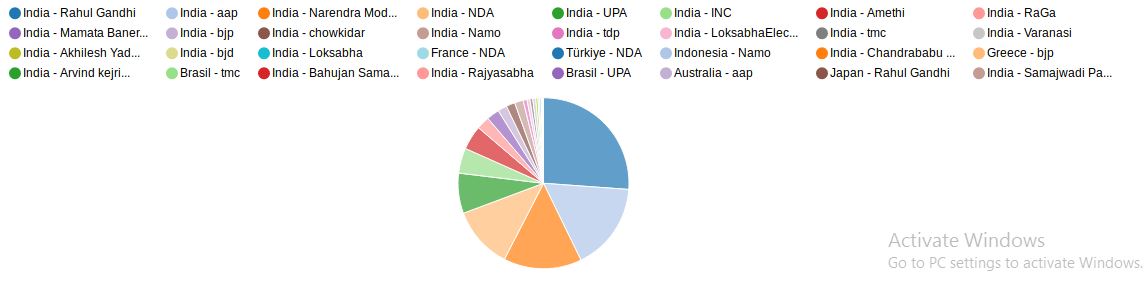
Code:



DAG Visualization:



Visualization:



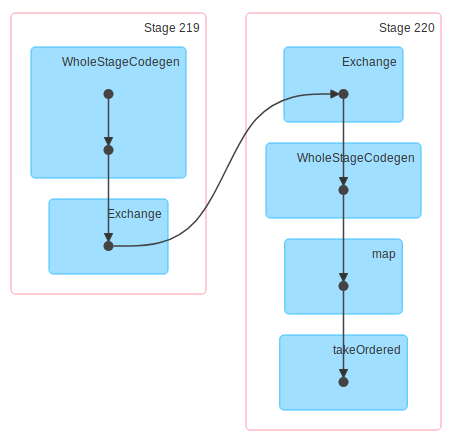
Query 4:

Tweets Based on Languages

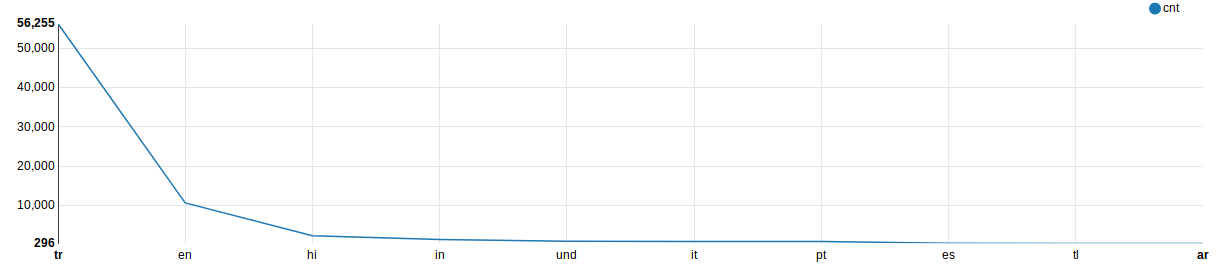
Code:



DAG Visualization:



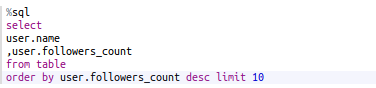
Visualization:



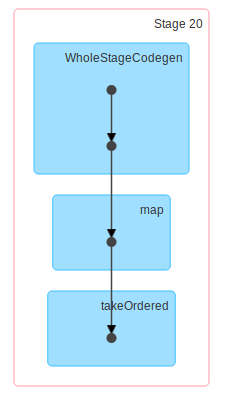
Query 5:

Top 10 users tweeting on basis of follower count

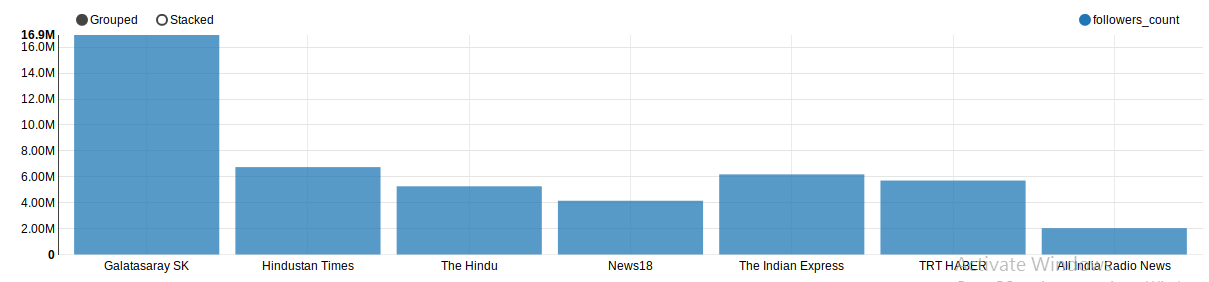
Code:



DAG Visualization:



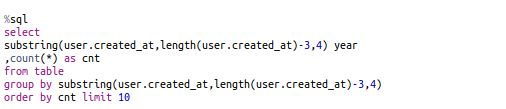
Visualization:



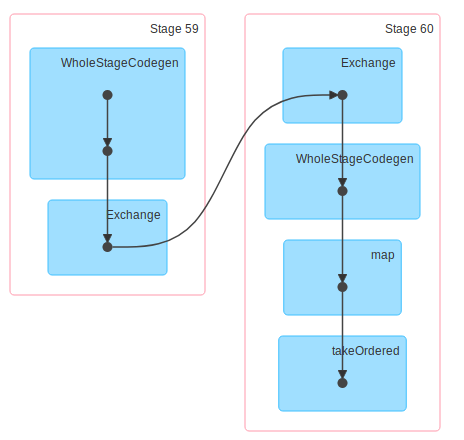
Query 6:

Top 10 years based on tweeted user’s creation details

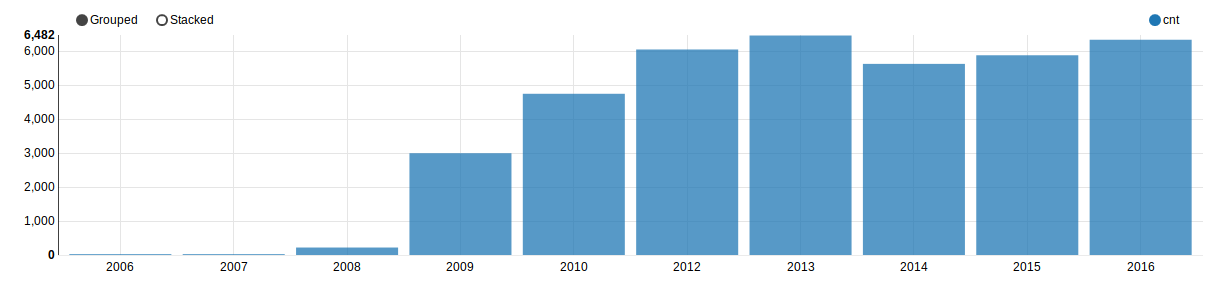
Code:



DAG Visualization:



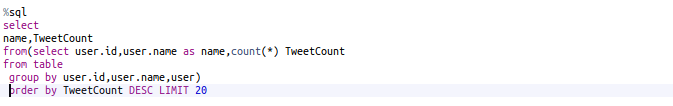
Visualization:



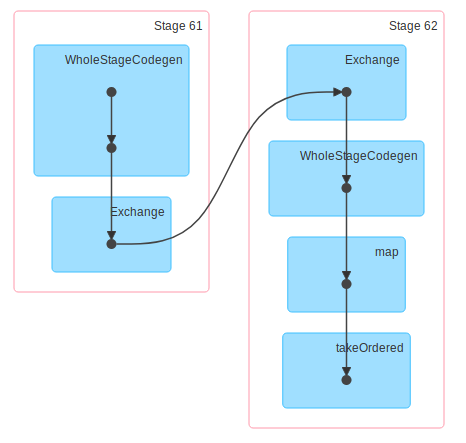
Query 7:

Top 20 users tweeting on Politics based on tweet count

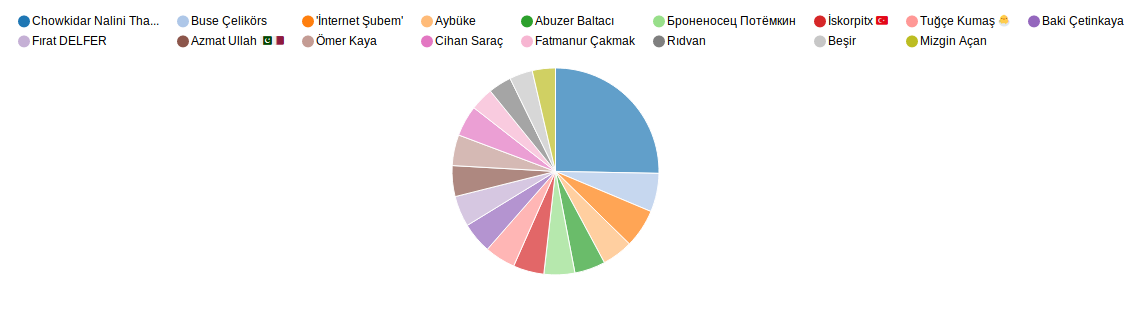
Code:



DAG Visualization:



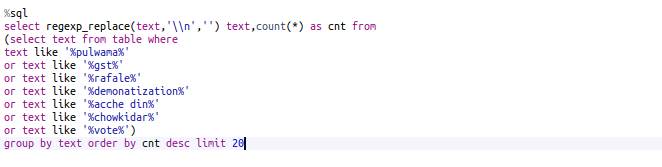
Visualization:



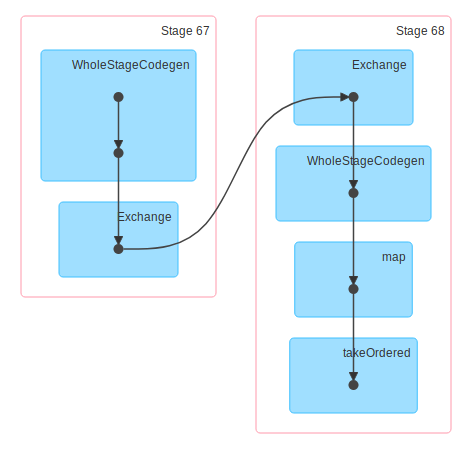
Query 8:

Top 10 tweets related to political issues like Pulwama, GST, Rafael deal etc.

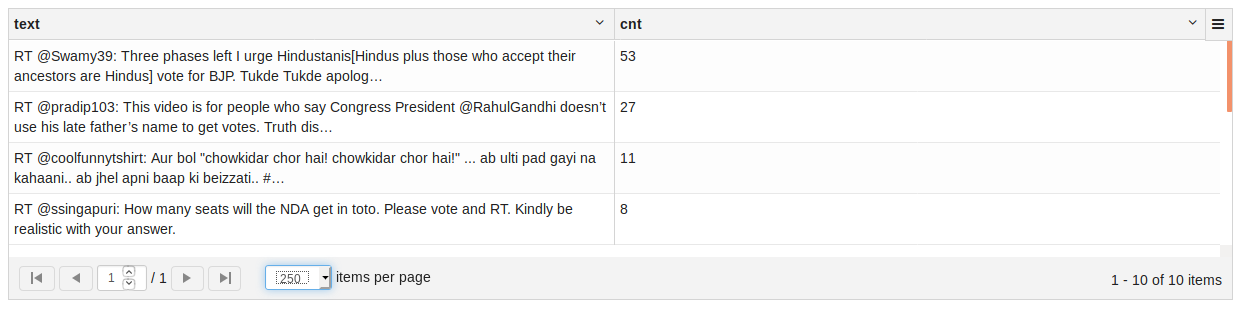
Code:

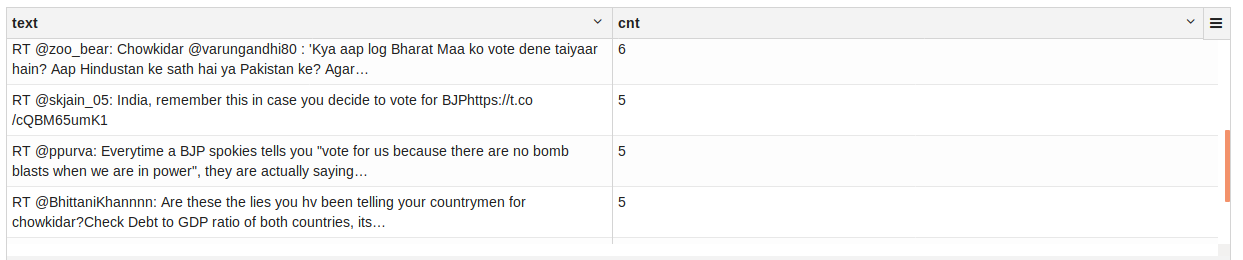


DAG Visualization



Visualization





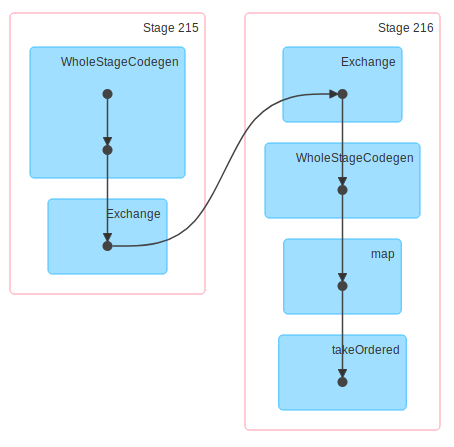
Query 9:

Tweets based on Places

Code



DAG Visualization



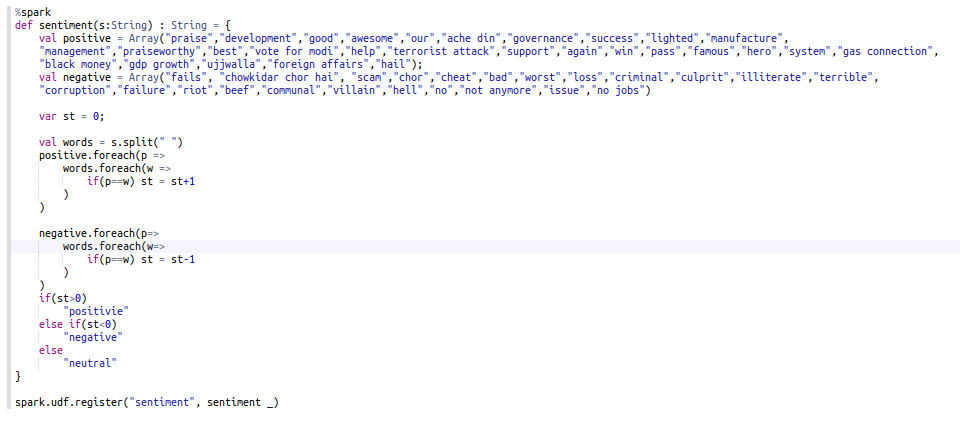
Visualization

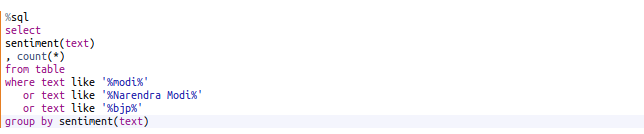


Query 10:

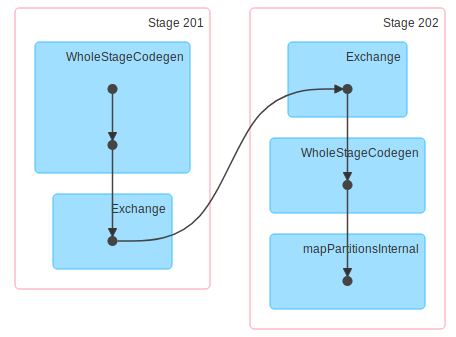
Sentimental Analysis on tweets based on MODI

Code:

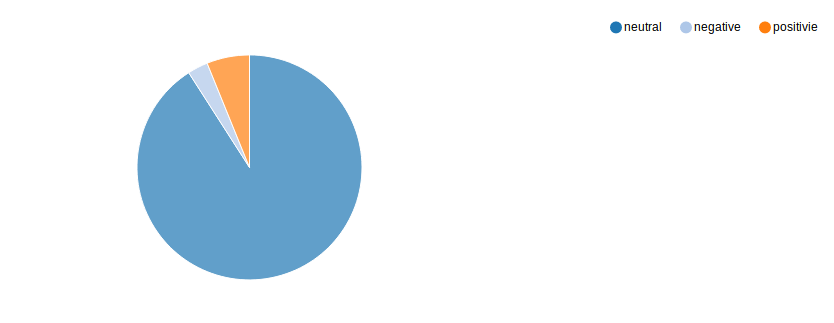




DAG Visualization



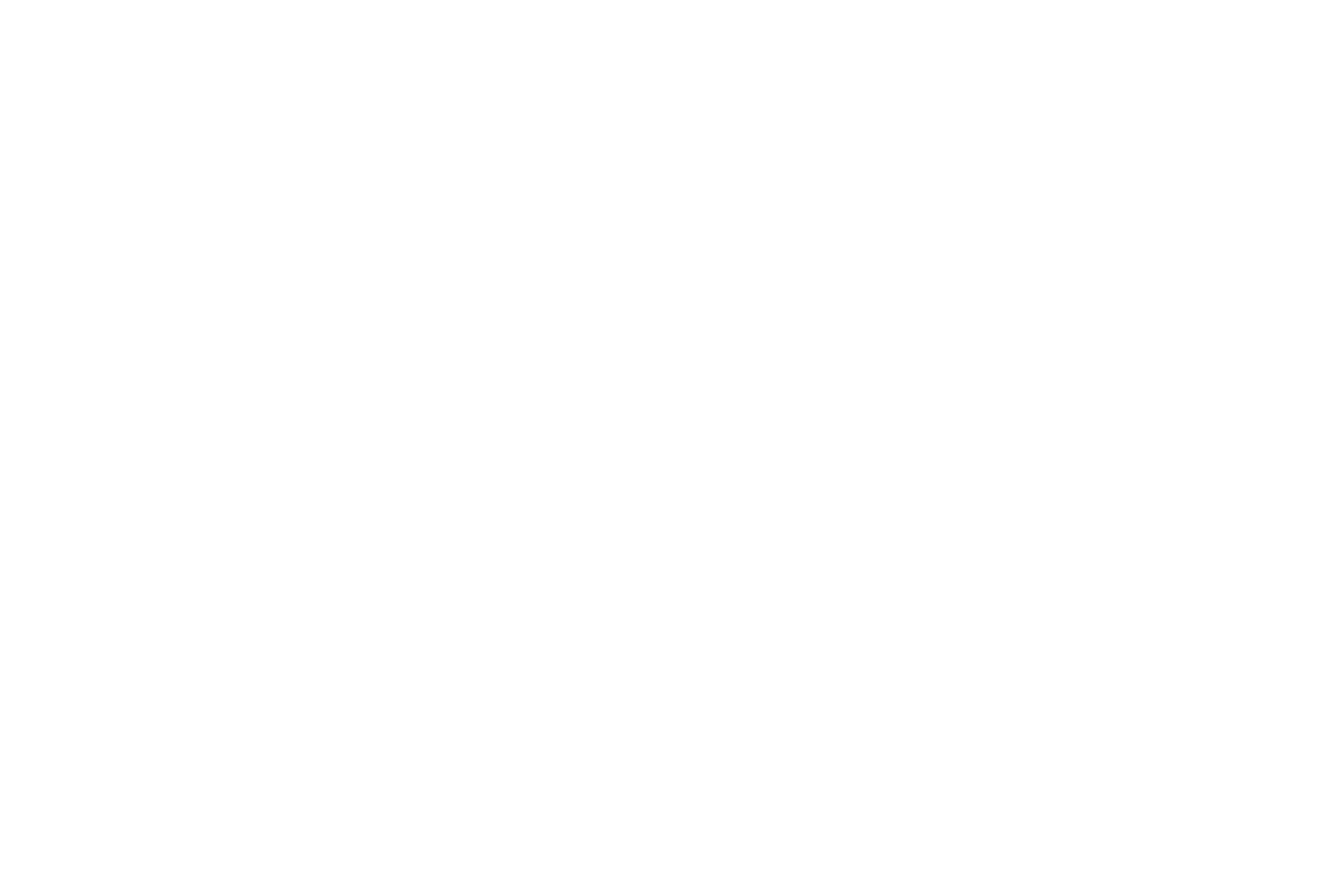
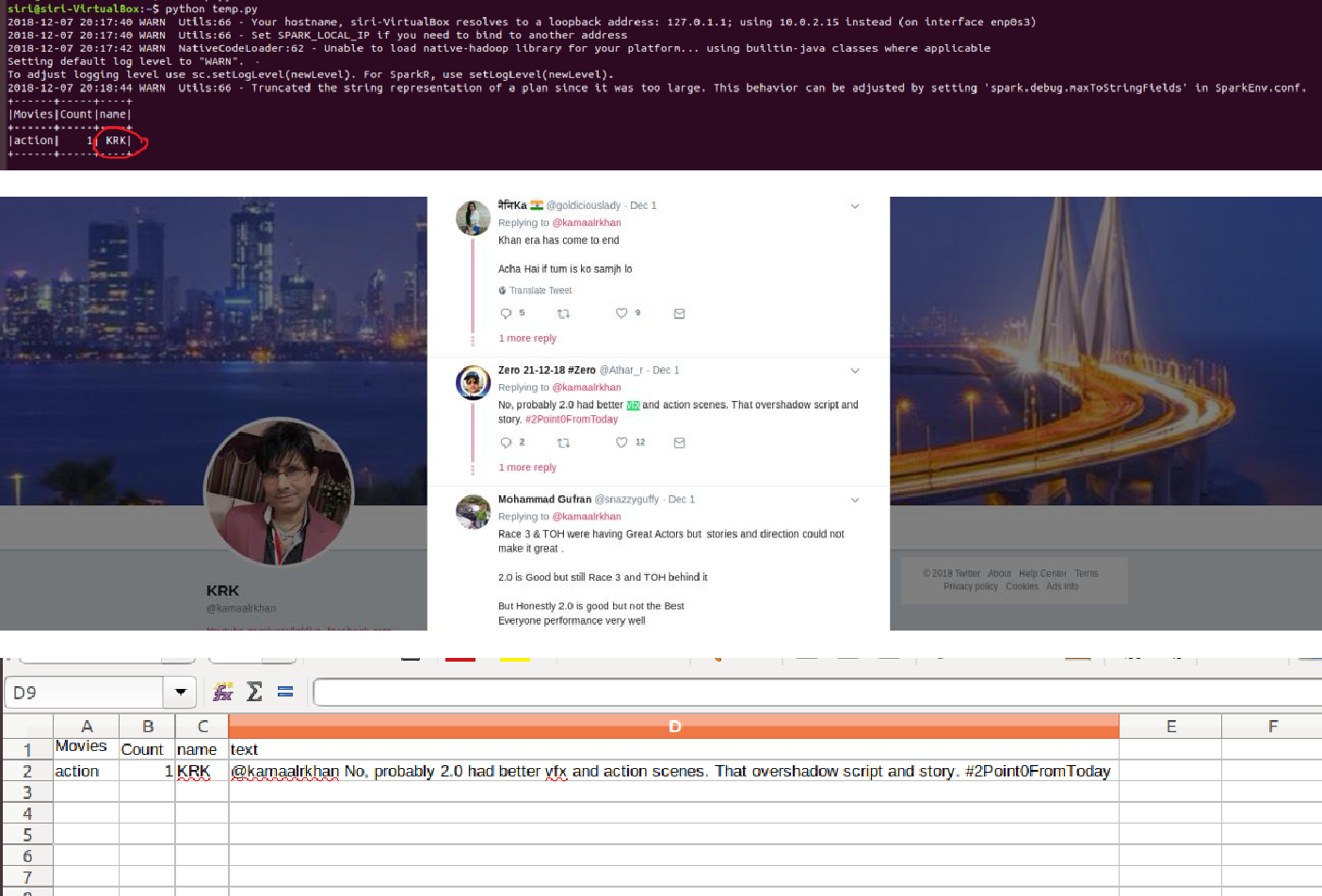
Visualization



**Testing:**

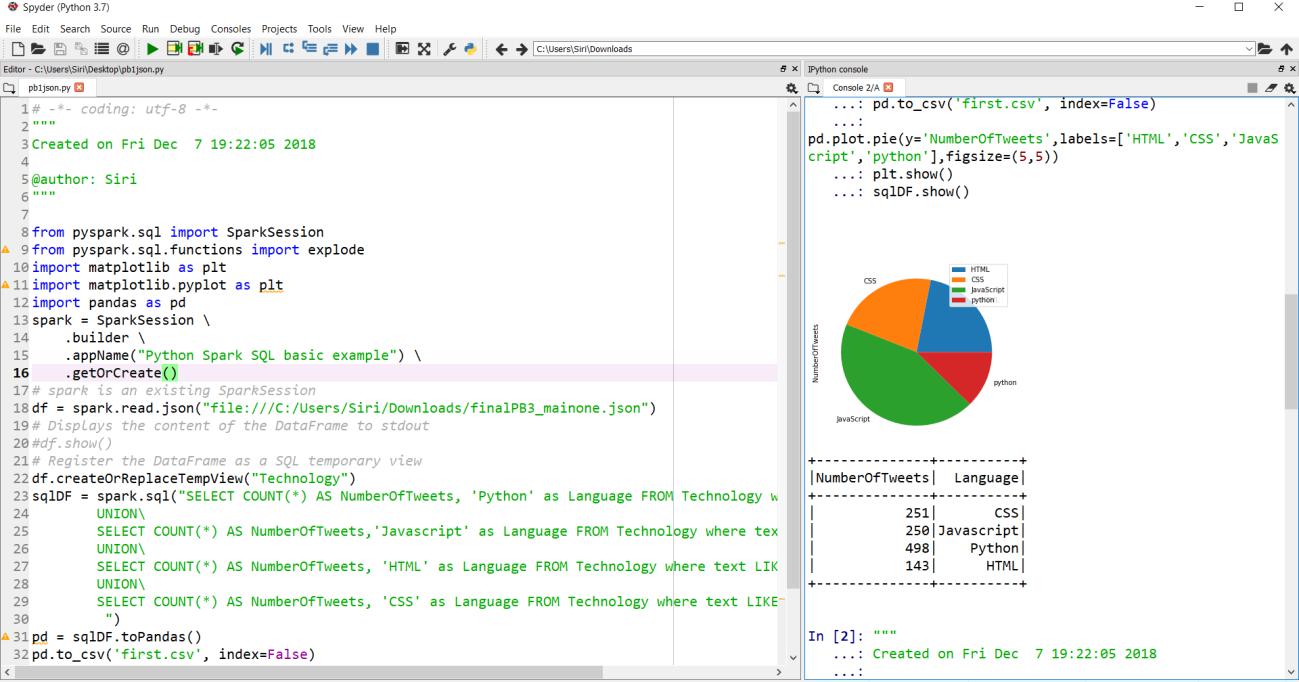
**Manual testing:**

We tried to test the results manually by taking the keywords used for the query and finding the tweets for the same keywords using twitter search engine in google and tried to match the tweets with the twitter data we retrieved.

For instance, consider the VFX query for which we retrieved the profile user name of the particular tweet and searched it manually in twitter to find the tweet and chec

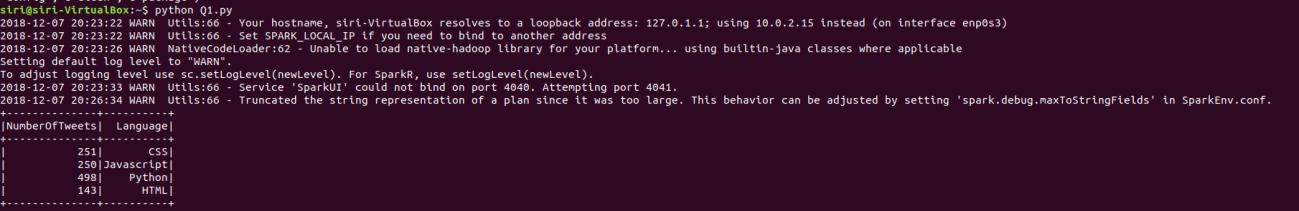
**Testing in tools:**

**Running in SPYDER:**



**Running in LINUX Terminal:**





As we can observe from the above two outputs both the outputs match each other therefore we confirm that the result is correct.

GITHUB URL: <https://github.com/NikithaPateel/PB_Project/tree/master/Phase_2_new>