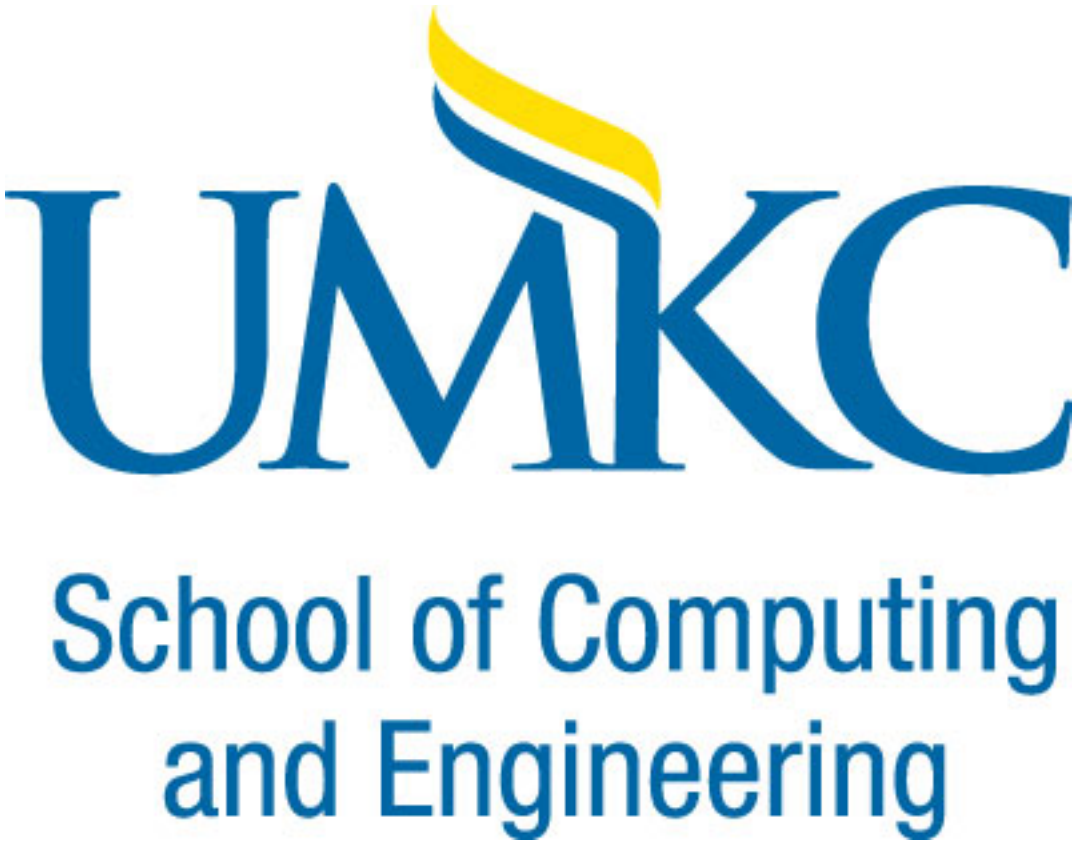


Principles Of Big Data-
US Presidential Candidates

Twitter Data analysis

VIKESH PADARTHI

VAMSHI RAJARIKAM

MANIKANTA MADDULA

PROJECT TOPIC:

United Presidential Candidates.

TECHNOLOGIES:

- Python: To Collect Tweets on Candidates from twitter
- Format Collected: JSON
- Visualization: D3.js
- Eclipse: Dynamic Web based application. Java program for queries

Project Description:

We have started collecting tweets data from twitter site using tweepy API by programming in python language. We have collected 1GB of tweets data which is in json format. Then using eclipse IDE, we have created a java program to run the queries on the tweets. We have created the following queries as stated below. When we run these queries the output of the analyzed data, it is sent to the adjacent CSV file. This CSV file is used in creating the visualization. Using D3.js library files, we have visualized the graph from the data obtained in CSV file. We have graphs for various queries as shown below.

Analytical Queries:

Query 1: Tweet Count based on President Candidates
Query 2: Top 8 Most Frequently Tweeting Users
Query 3: Top 8 Users with highest followers
Query 4: Top Locations with most Tweets
Query 5: Users with Friends greater than 150000
Query 6: Top 8 Most Tweeting Timestamps
Query 7: Sentiment Discovery
Query 8: Tweets from Different Type of Devices
Query 9: Tweet vs Retweet Status

Home Screen: This is the home screen of our project. From here, we can select the queries we want to execute.

Figure 1: Home page Web UI

Query 2: Tweet count based on President Candidates

This query gives the result based on, how many tweets are tweeted on each Presidential Candidate. The figure below shows, the number of tweets tweeted on each candidate. This does not definitely mean favor for particular candidate as sentiment of a tweet can be different for all tweets.

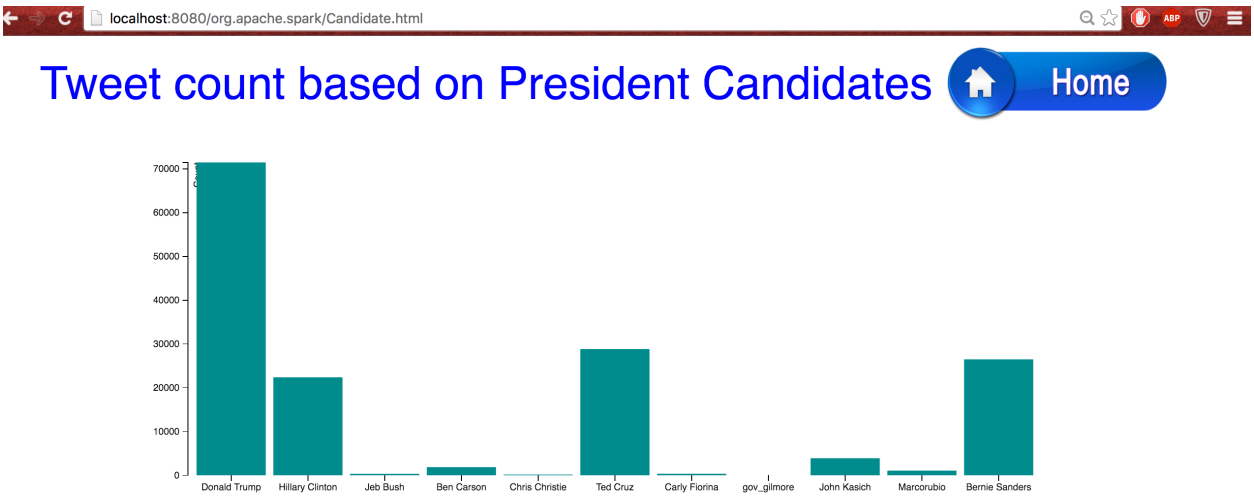


Figure 2: Tweet count based on President Candidates

Query 2: Top 8 Most Frequently Tweeting Users

This query gives the information about the individual who tweets more number of tweets on twitter in the collected data. These are the top 8 users, who tweets the most.

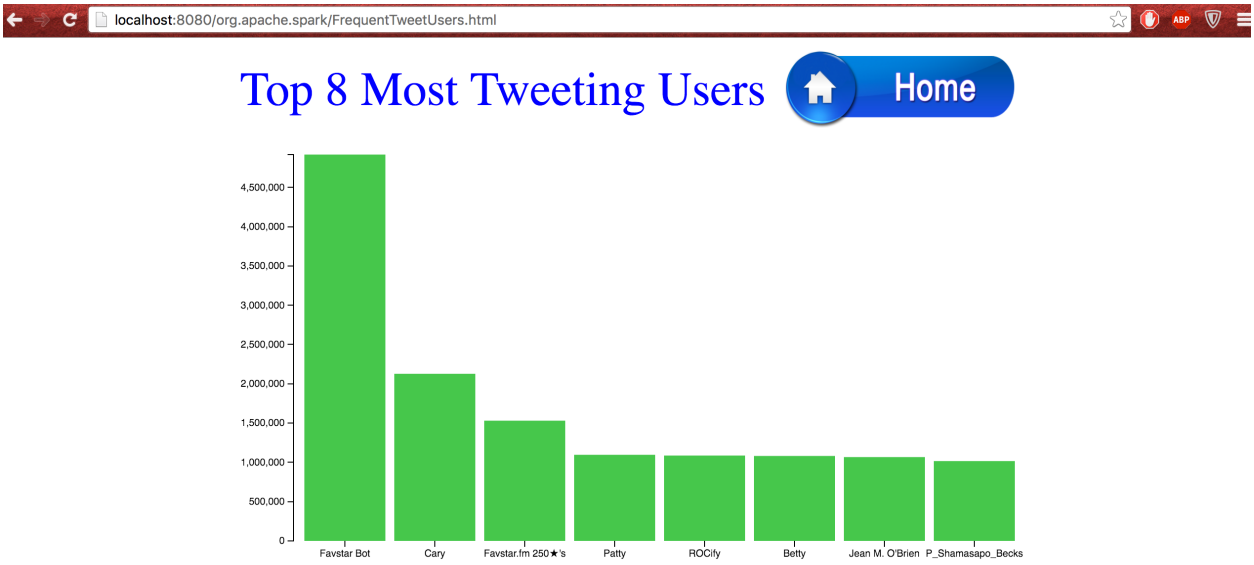


Figure 3: Top 8 Most Frequently Tweeting Users

Query 3: Top 8 Users with highest followers:

This query gives the information about individual user who has highest number of followers. The figure below shows, top 8 user with highest number of followers.

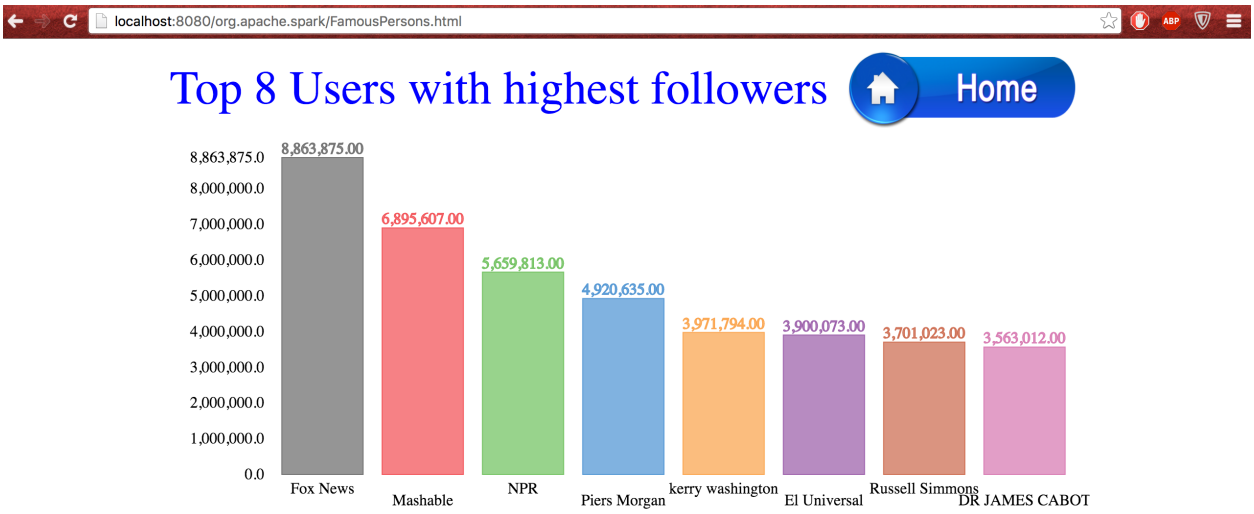


Figure 4: Users with highest followers

Query 4: Top location with highest Tweets

This query gives information about, highest number of tweets posted based on location.

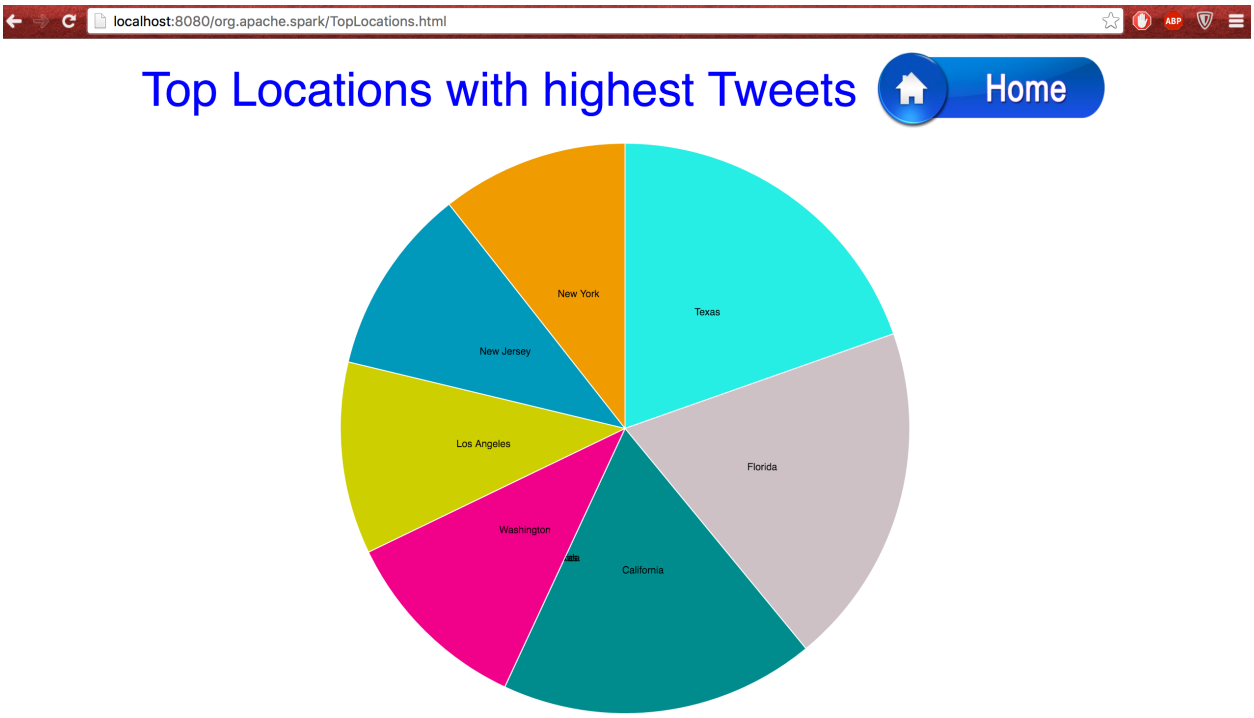


Figure 5: Top Locations with highest Tweets

Query 5: Users with Friends greater than 150000

This query gives the result of individual users with friends greater than 150000.

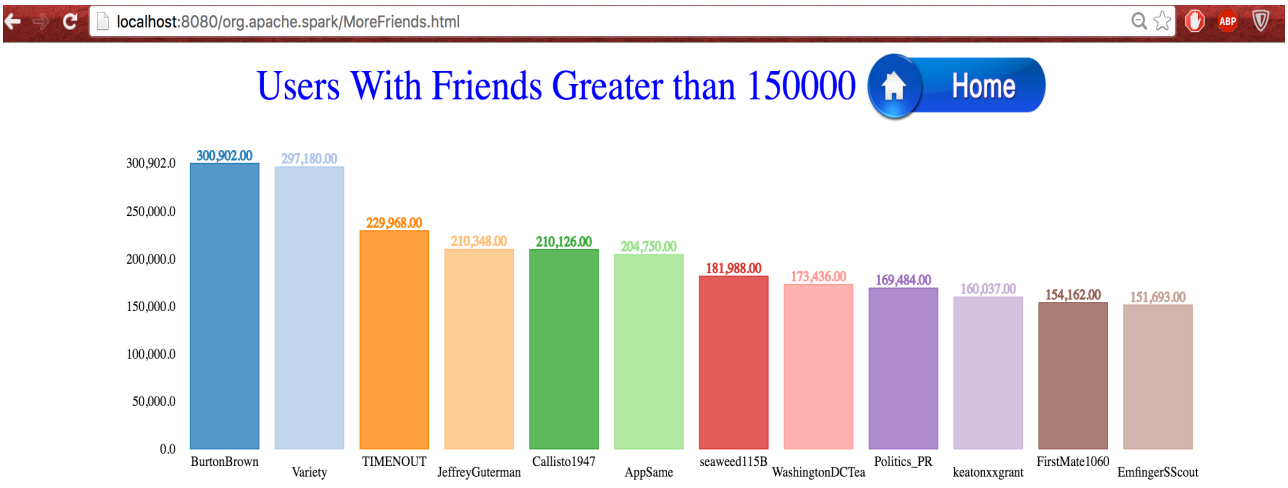


Figure 6: Users with Friends Greater than 150000

Query 6: Top 8 Most Tweeting Timestamps

This will give the information about, most numbers of tweets tweeted at what time.

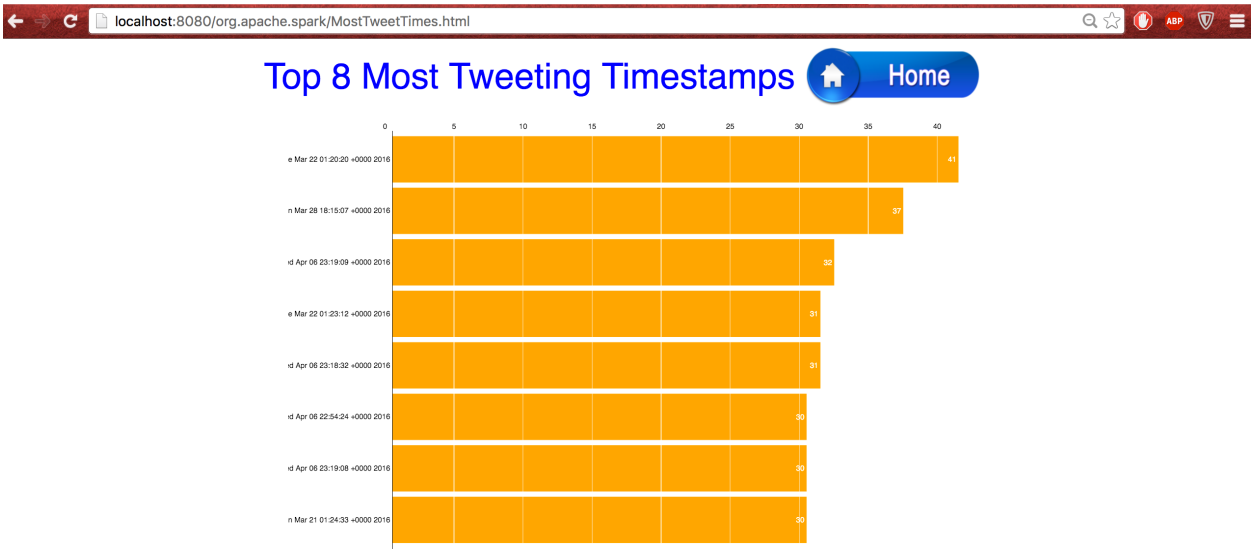


Figure 7: Top 8 Most Tweeting Timestamps

Query 7: Sentiment Discovery

This give the information about sentimental analysis of the tweet text data.

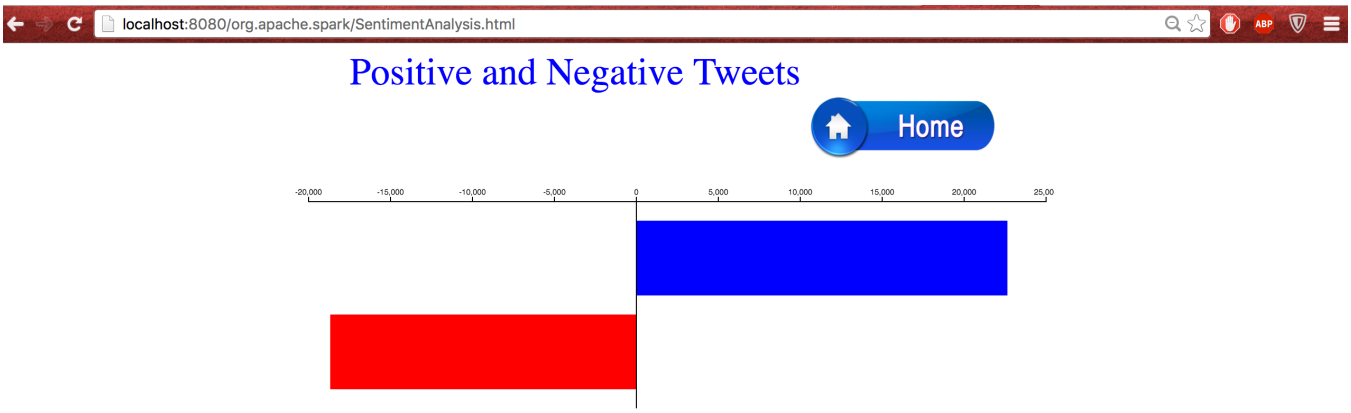


Figure 8: Sentiment Discovery

Query 8: Tweets from Different Type of Devices
This will give the information about the devices on which tweets are posted from.

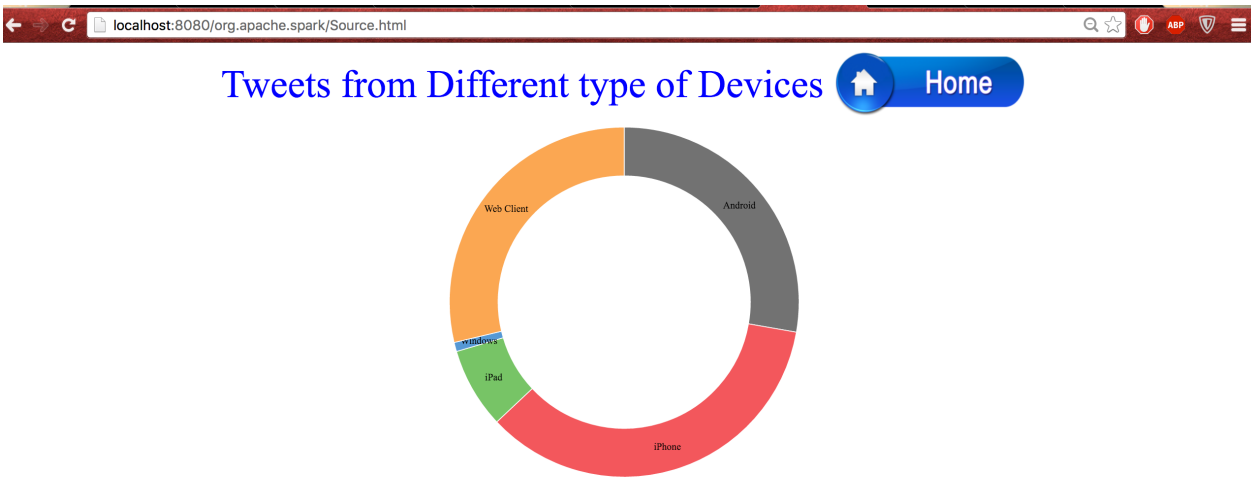
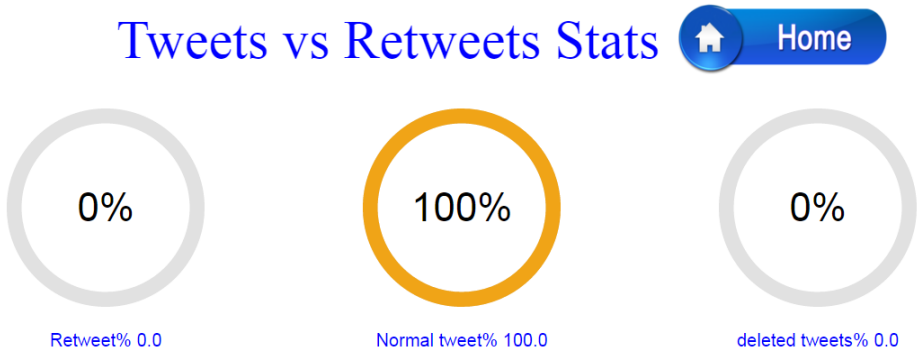


Figure 9: Tweets from Different Type of Devices

Query 9: Tweet vs Retweet Status

This gives the information about the type of tweets i.e. tweets and retweets, that are tweeted on twitter.



SOURCE CODE: https://github.com/manikantamaddula/PBProject/tree/master/Phase2

References: 1. https://github.com/AshokYaganti/PBBlumixCode/graphs/contributors

2. https://github.com/stefani75/workspace/blob/master/projet1/Hands-On-Spark-java-solution/src/main/java/com/duchessfr/spark/part3/sparksql/FunWithSparkSQL.java

3. https://www.youtube.com/watch?v=d-Et9uD463A
4. https://www.youtube.com/watch?v=pUUxmvvl2FE&feature=youtu.be