Phase II

Team:

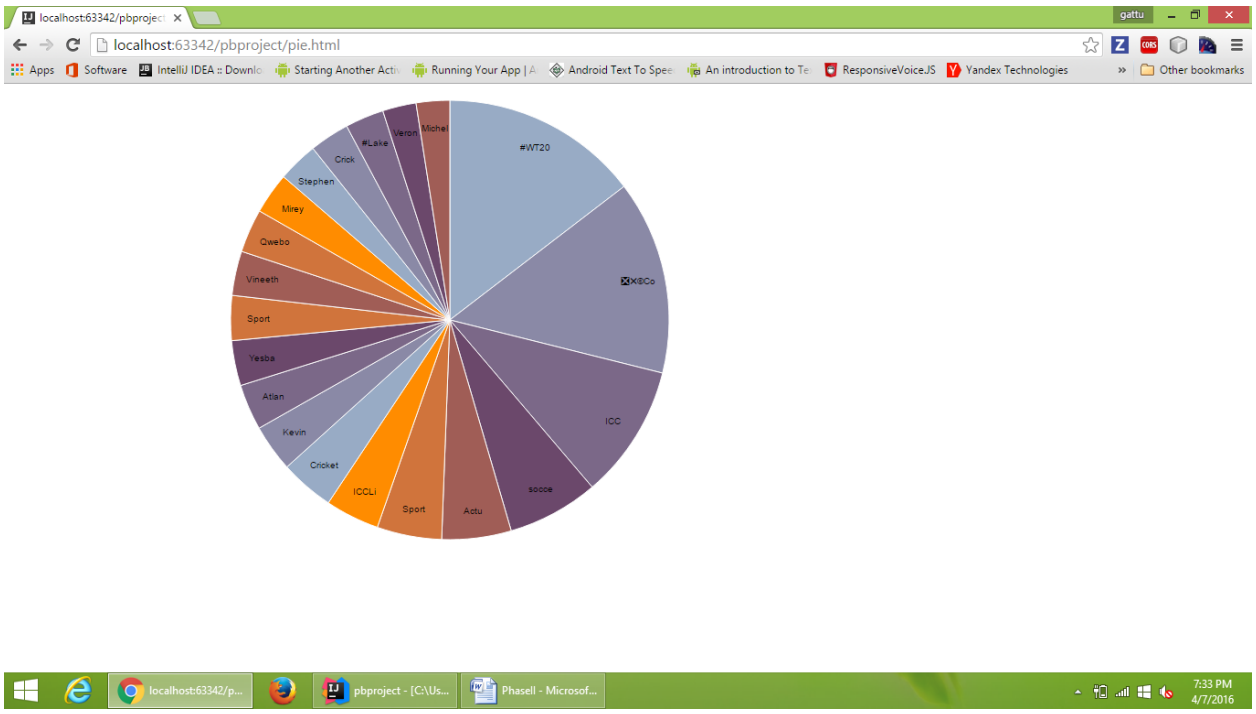
Akhilesh Gattu (16221279)

Nagi Reddy Sollapuram (16225926)

Siva Rama Krishna Linga (16227126)

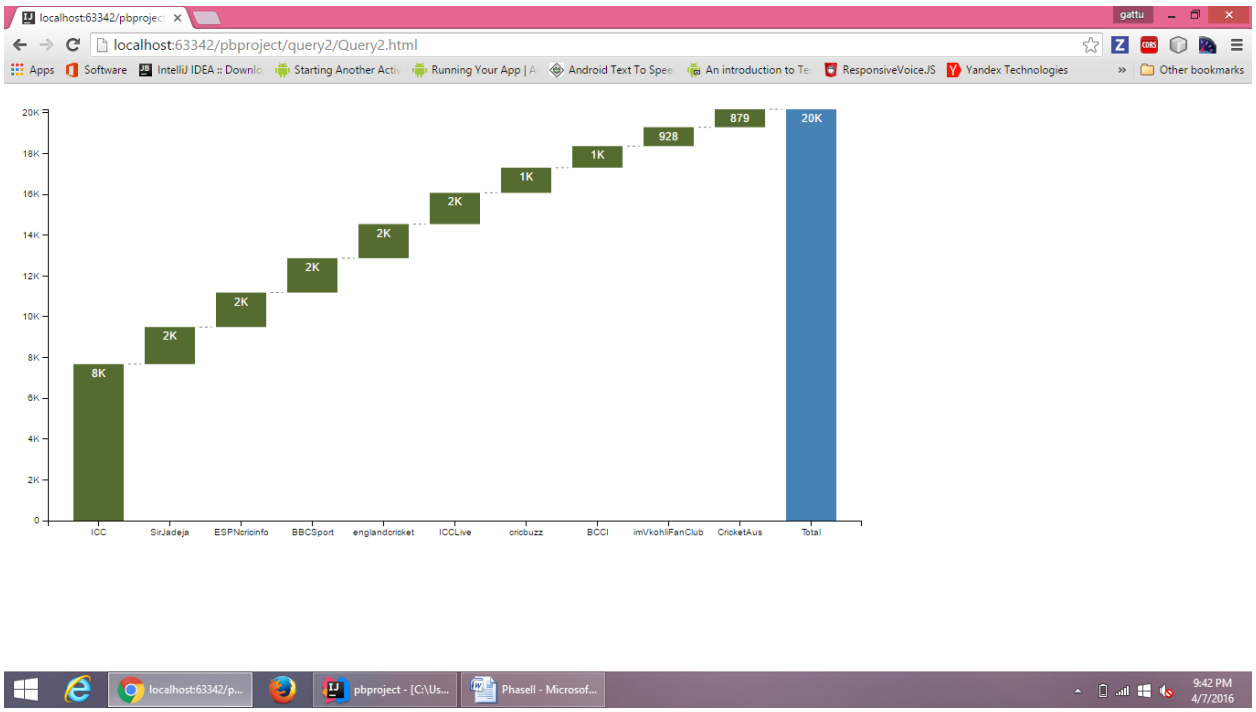
Query-1: This query will display the top 20 user names and their respective followers count in the descending order.

val query1 = sqlContext.sql(**"select user.name, count(user.followers_count) as followersCount from querytable1 group by user.name order by followersCount desc limit 20"**)



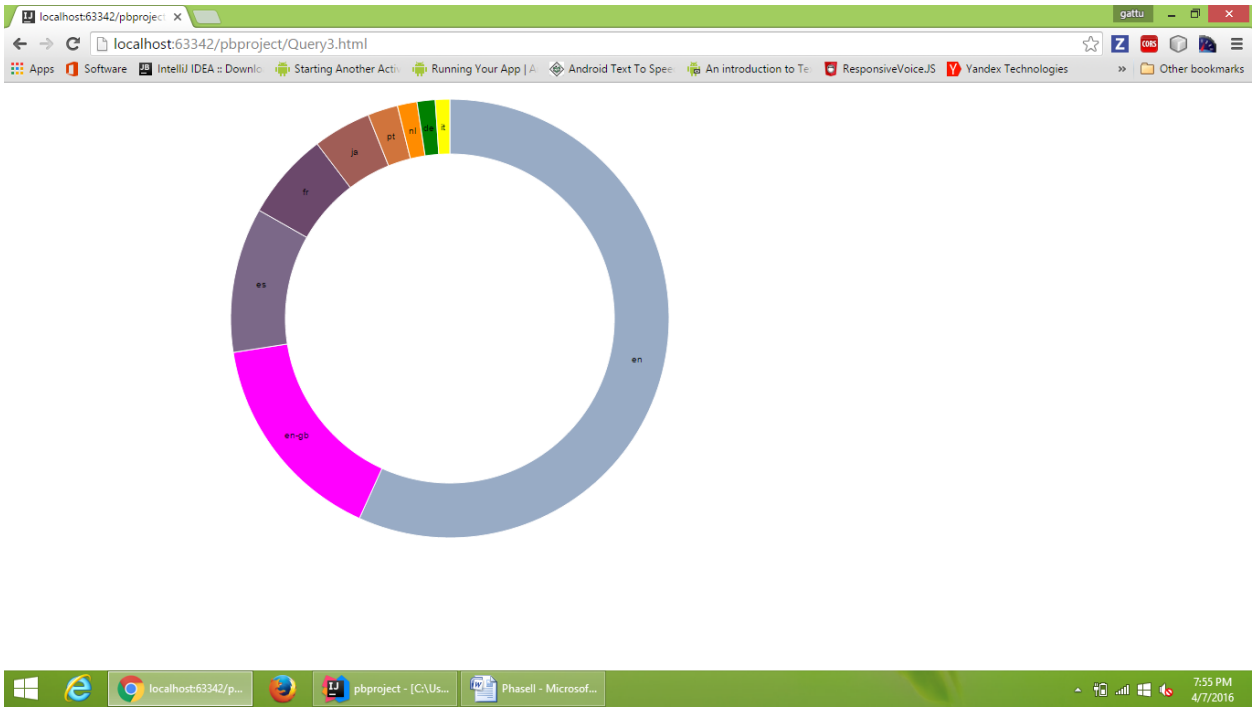
Query-2: This query will display the user on whom there are more number of re-tweets are made. At the same time it will also show the count of the tweets or the number of re-tweets on a particular user. The results are ordered in the descending order of the re-tweet count.

val query2 = sqlContext.sql("**select count(*) as cnt,retweeted_status.user.screen_name as name from querytable1 where retweeted_status.user.screen_name is not NULL group by retweeted_status.user.screen_name order by cnt desc limit 10**")



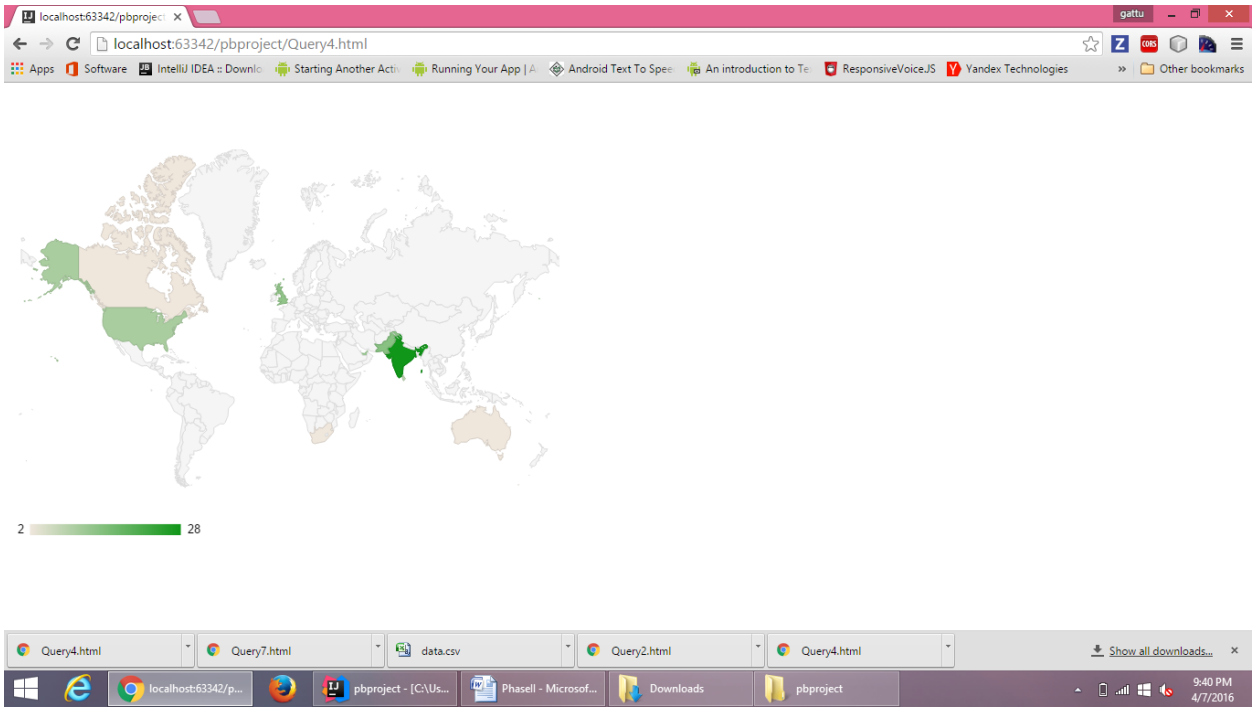
Query-3: This query will list the top 10 languages along with their tweets count in which the tweets are made. The results are ordered in descending order.

val query3 = sqlContext.sql("**select user.lang as language, count(*) as cnt from querytable1 where user.lang is not NULL group by user.lang order by cnt desc limit 10**")



Query-4: This query will list the top ten countries from which people have tweeted about **"Virat Kohli"** and also list the count of tweets from that particular country. The results are ordered in the descending order.

val query4 = sqlContext.sql("**select place.country_code as location,count(*) as cnt from querytable1 where place.country_code is not NULL and text like '%virat%' or text like '%kohli%' group by place.country_code order by cnt desc limit 10**")

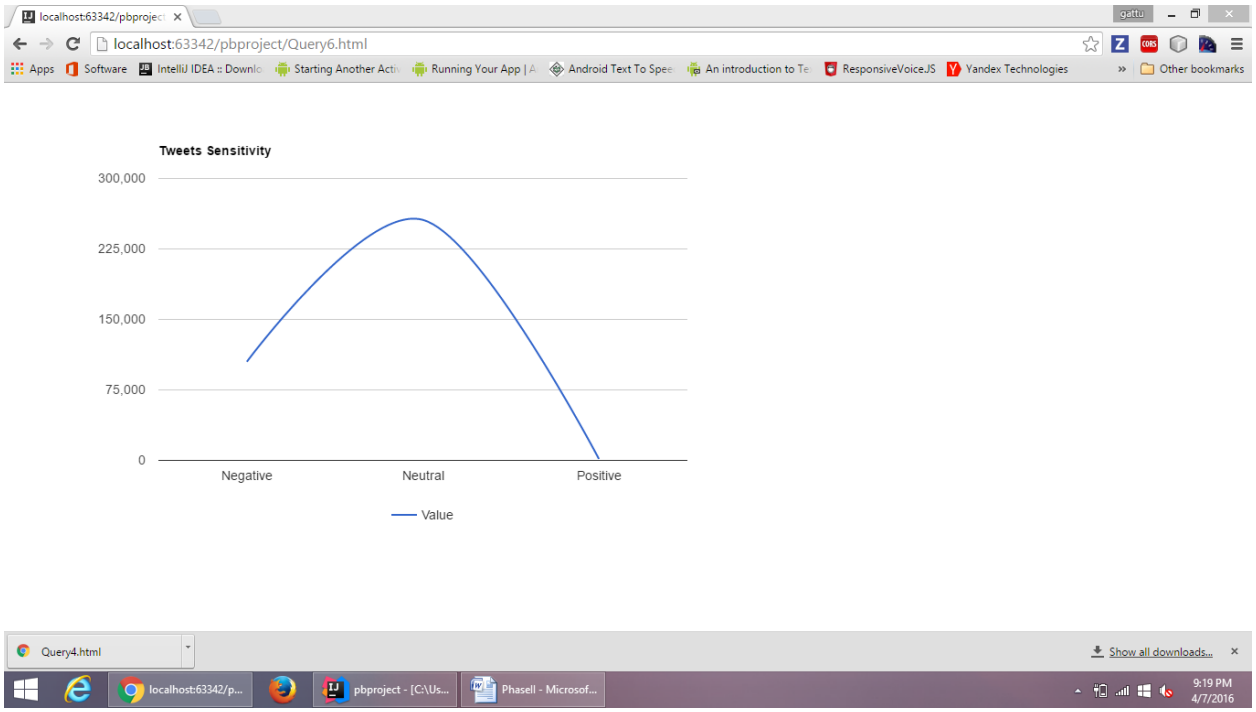


Query-5: This query will display the list of top 10 users along with their re-tweets count, who made more number of re-tweets

val query5 = sqlContext.sql("**select user.name as name, retweeted_status.retweet_count as cnt from querytable1 where user.name is not NULL order by cnt desc limit 10**")

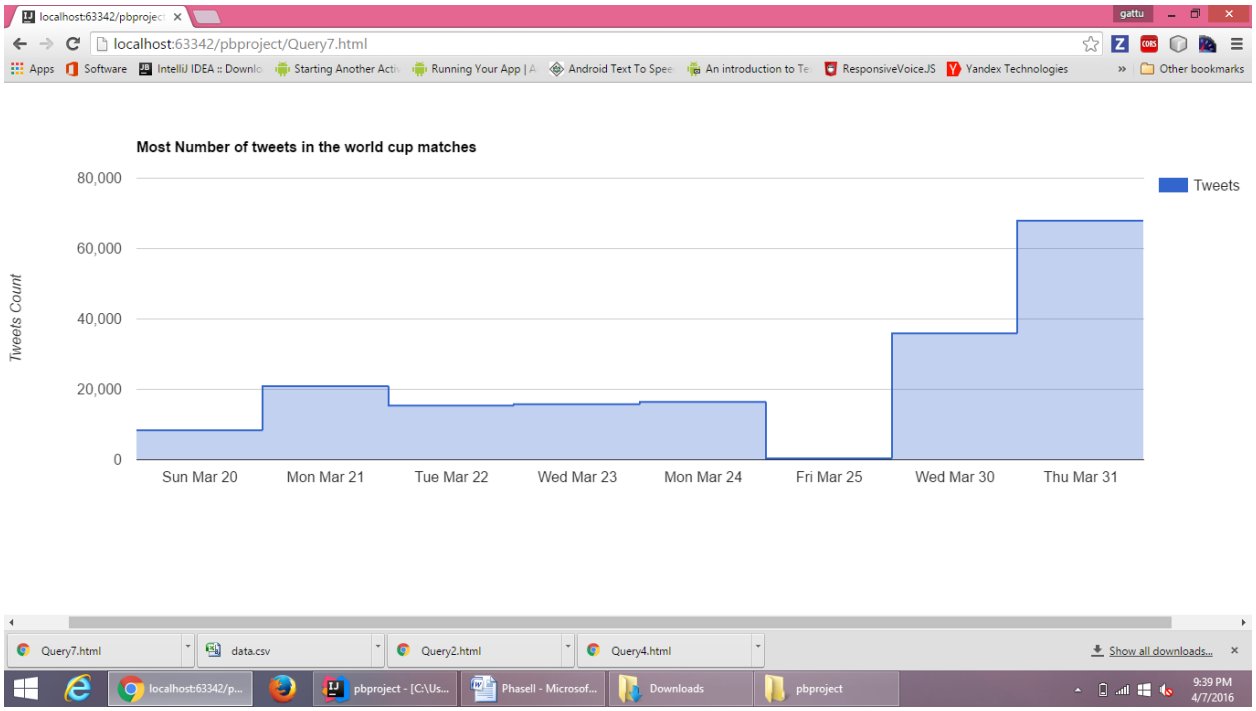
Query-6: This query will list the count of tweets that are possibly sensitive or not and also the count of tweets for the respective category.

val query6 = sqlContext.sql("**select possibly_sensitive as sensitive,count(*) from querytable1 group by possibly_sensitive limit 3**")



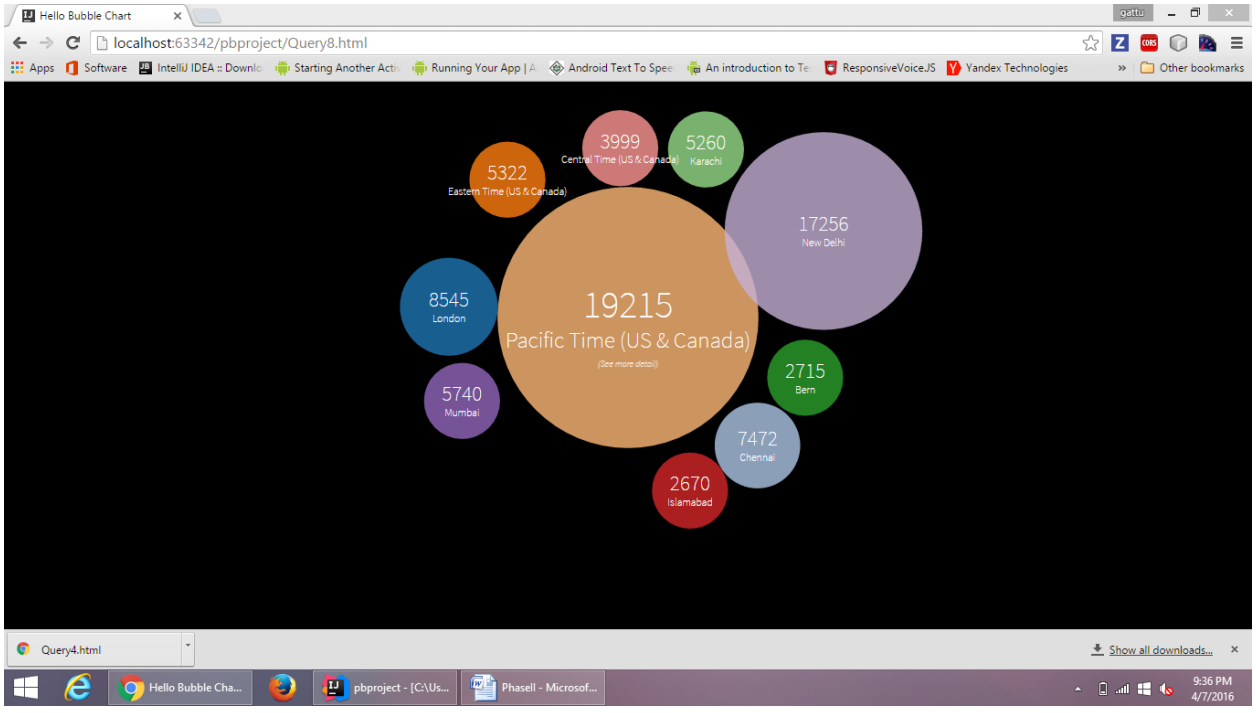
Query-7: This query will display the list of top ten days on which the number of tweets are more along with the number of tweets on that particular day. The results are displayed in the descending order.

val query7 = sqlContext.sql("**select substr(created_at,0,10) as time, count(*) as cnt from querytable1 where created_at is not NULL group by substr(created_at,0,10) order by cnt desc limit 10**")



Query-8: This query will display the list of top ten countries from which there are maximum number of tweets along with their count of tweets. The results are ordered in descending order.

val query8 = sqlContext.sql("**select user.time_zone as time, count(*) as cnt from querytable1 where user.time_zone is not NULL group by user.time_zone order by cnt desc limit 10**")



Source Code: Source code is available at the following link

https://github.com/AkhileshGattu/PBProject

Tweets are available at

https://drive.google.com/open?id=0B8QQeMrq6yxTbWlHMWdxVVZZM28

References:

- https://d3js.org/
- https://developers.google.com/chart/
- http://marcobonzanini.com/2015/03/02/mining-twitter-data-with-python-part-1/