**CS5540**

**PRINCIPLES OF BIG DATA MANAGEMENT**

**(Spring 2019)**

**PROJECT REPORT**

A picture containing object

Description generated with very high confidence

BY

**Bindu Sravya Karumuru(16246180)**

**Y.Anusha(16271202)**

**K.Bhavana Deepthi(16275870)**

**OBJECTIVE:**

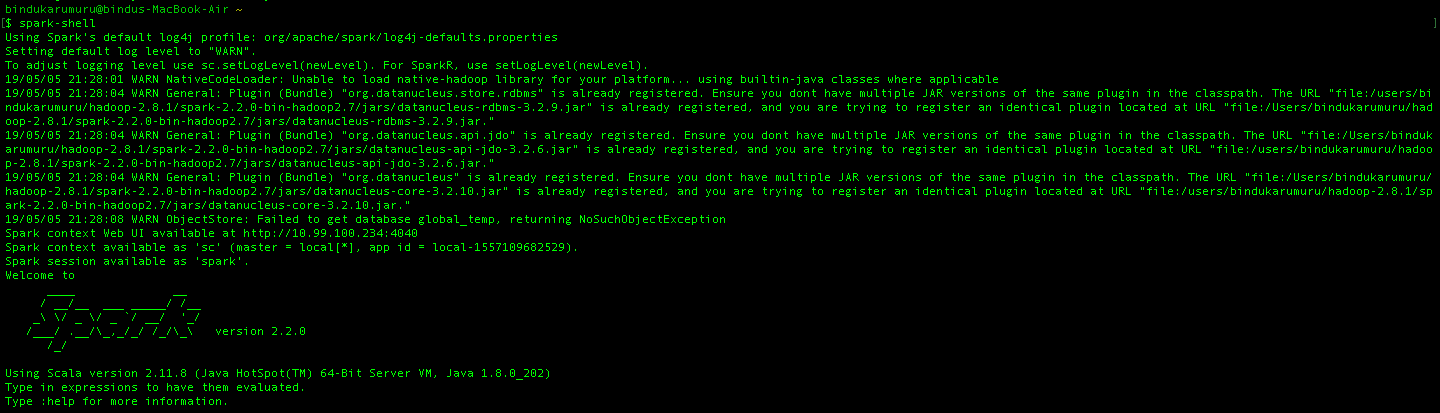
* To work on the tweets related to gadgets and to figure out how to store them in Spark SQL.
* To write interesting analytical queries to explore and understand the data collected.
* To develop interesting visualizations of the above written queries.

**VISUALIZATION :** To perform the visualization on the result sets and to view it in some pictorial representations like the pie charts, bar graphs etc.

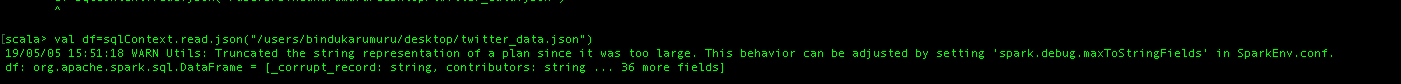
**SOFTWARES PLANNING TO USE:**

* Apache Spark
* Scala (Execution of Queries)
* High Charts for Visualizations, Tableau

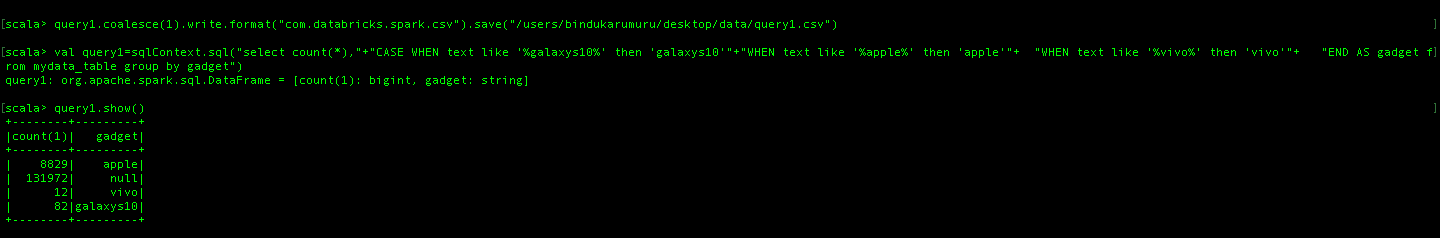
Implementation in spark:



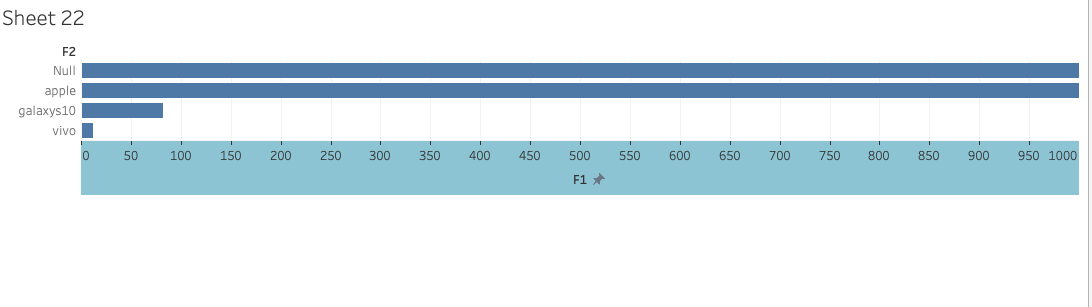
Loaded data set into spark-shell:



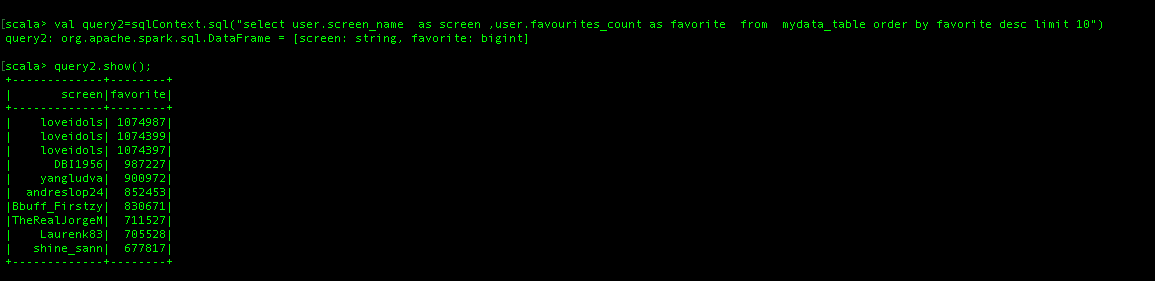
**1.Query to fetch information number of people who talks about galaxys10,apple,vivo**



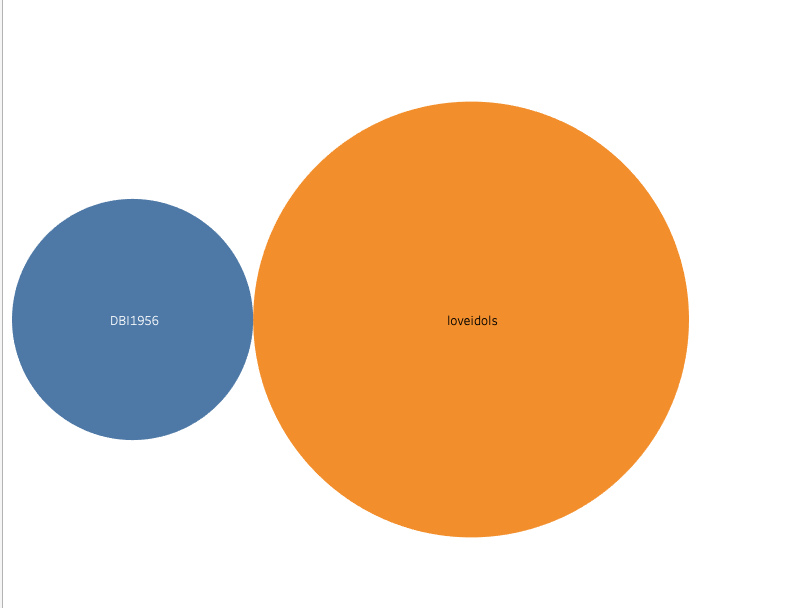
**Graph for Query 1:**



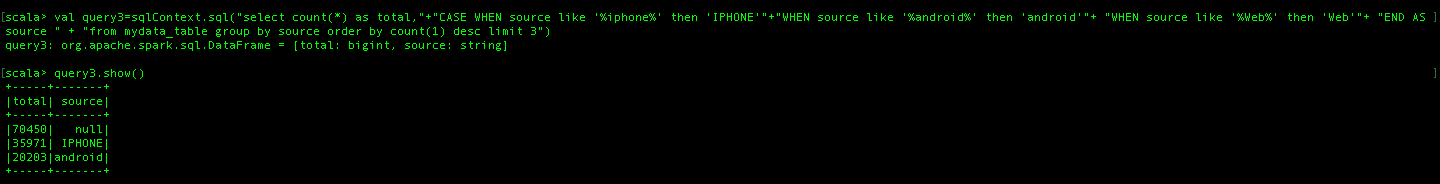
**2.Query collects favorite count of all tweeters**



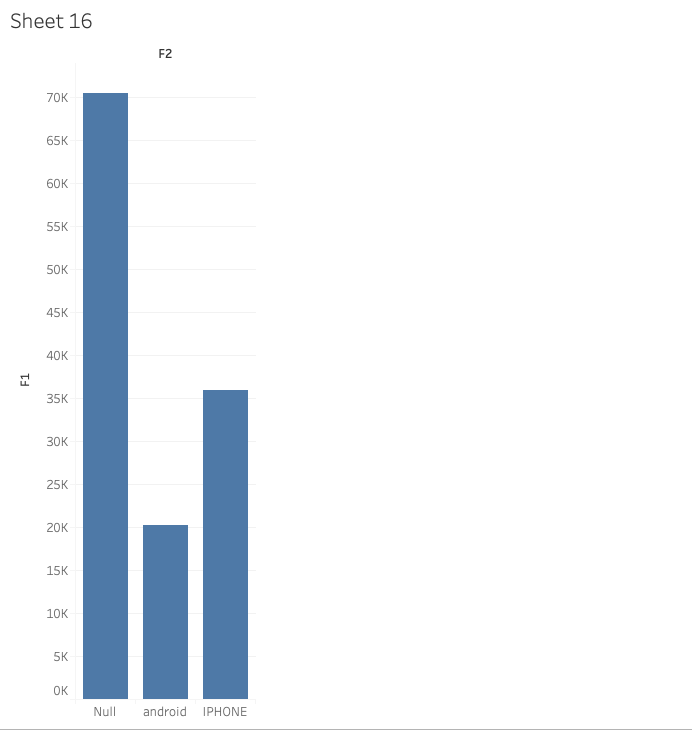
**Graph for Query 2:-**



3.Query to extract the mostly talked mobile categories(iphone, android) on twitter



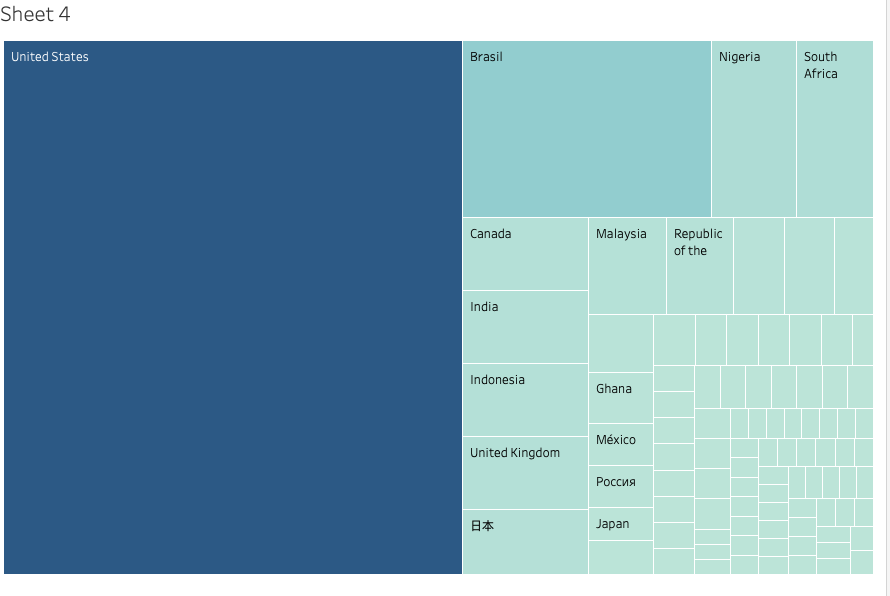
**Graph for Query 3:-**



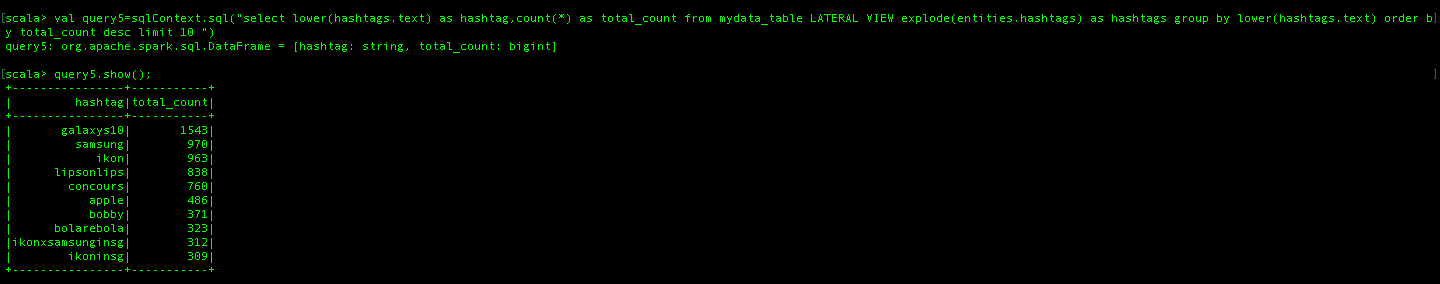
4.Query to fetch tweets from various locations based on country



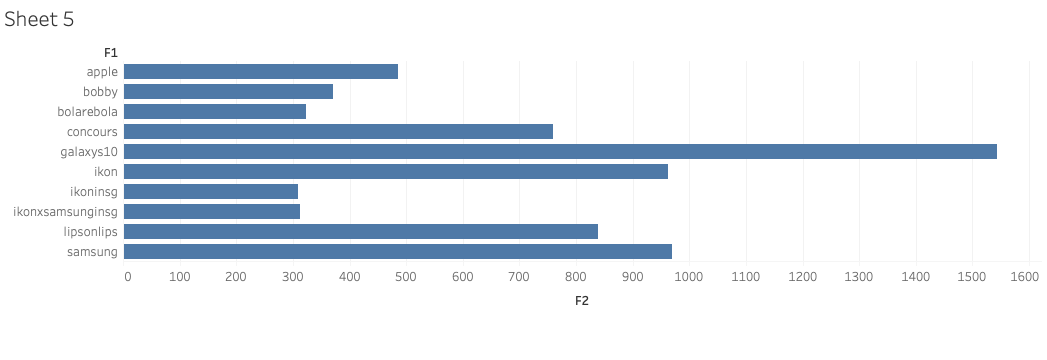
**Graph for Query-4:**

****

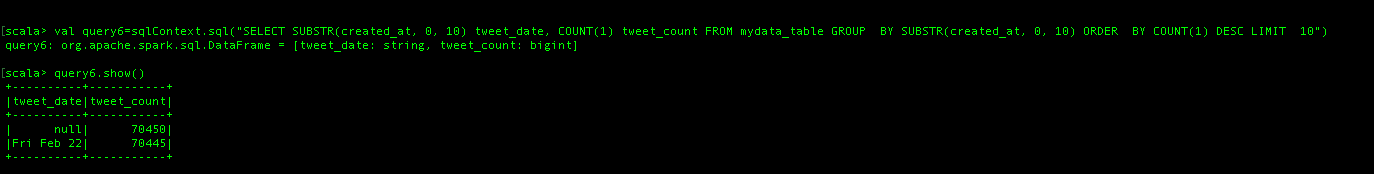
5.Query to fetch widely used hashtags along with their count



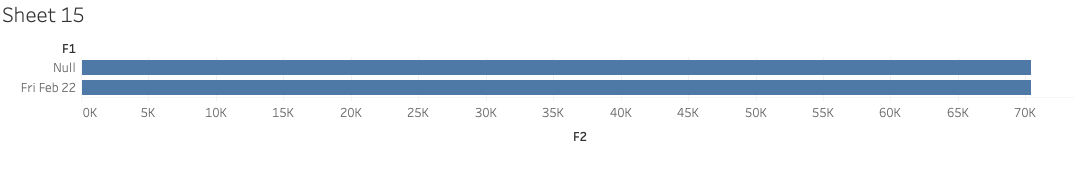
**Graph for Query 5**



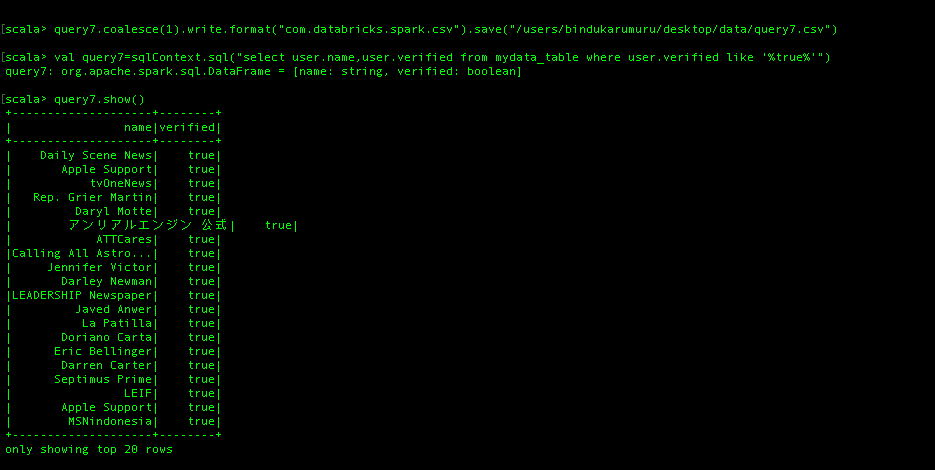
6.Query to fetch number of tweets made on specific date:



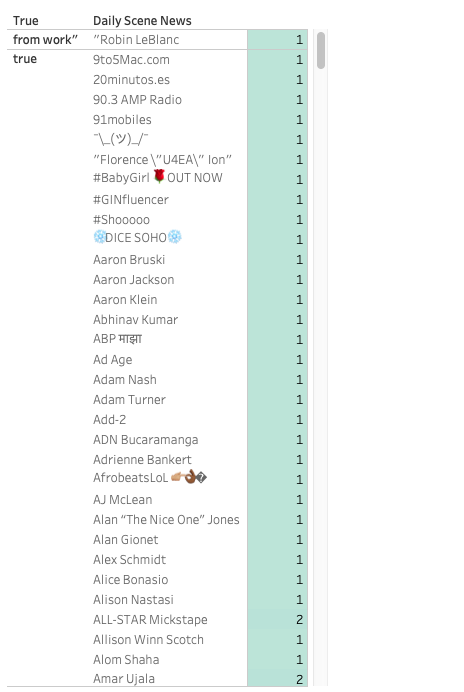
Graph for Query 6:



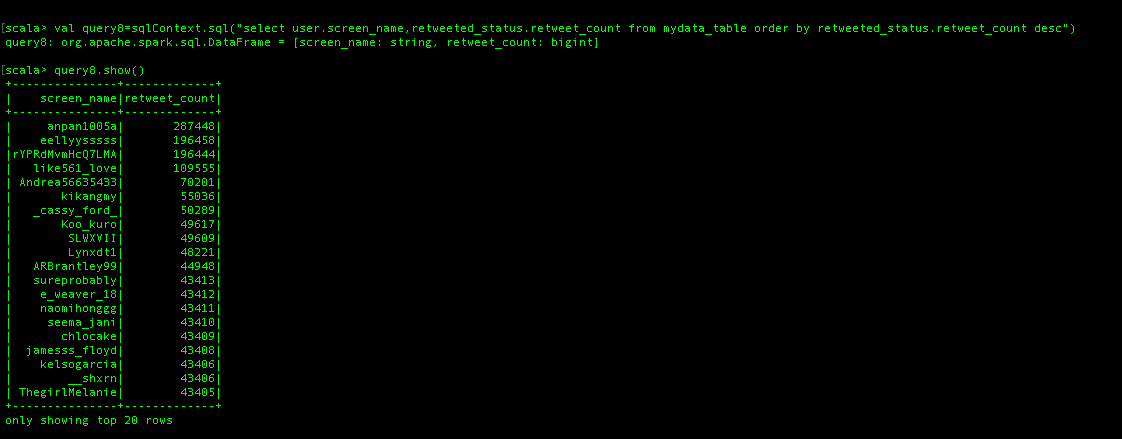
7.Query to fetch top 20 verified users.



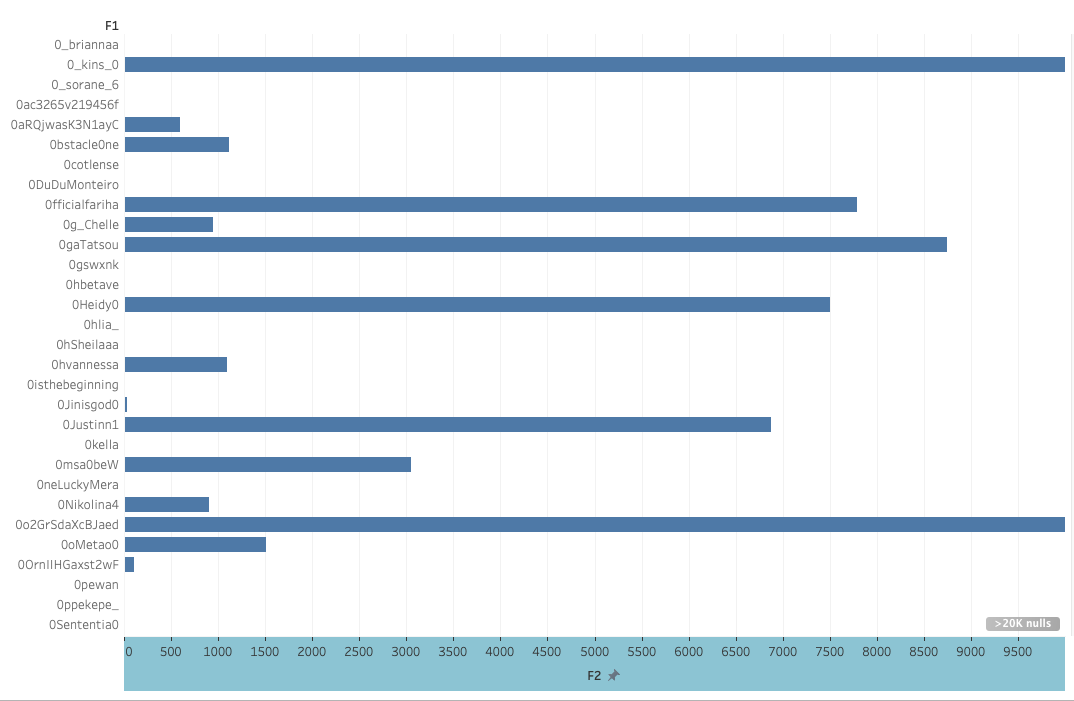
Graph for Query 7:-



8.Query to fetch retweeted count along with the user names.



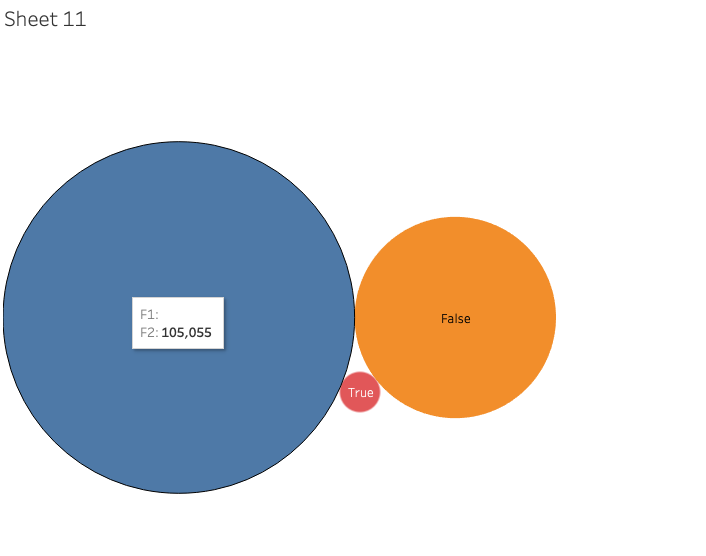
Graph for Query 8:-



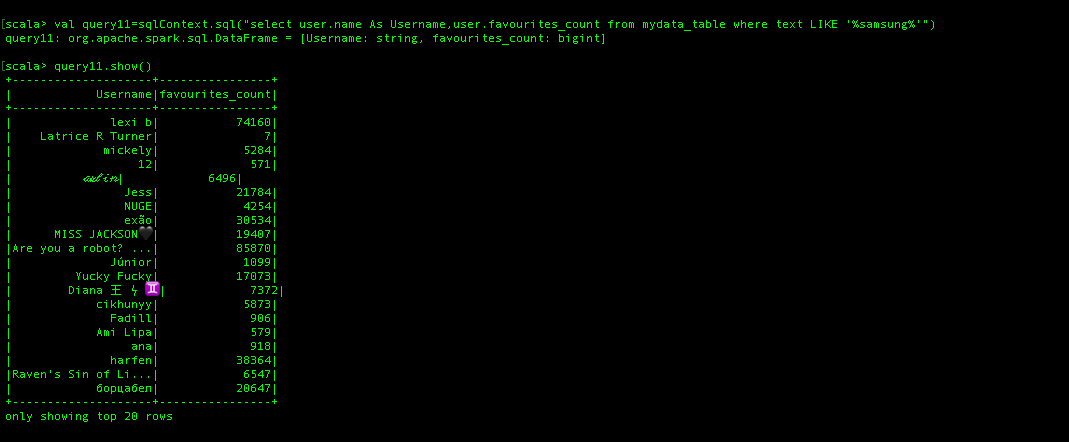
9. Query to extract the sensitivity of tweets:



Graph for Query 9



10.Query to fetch top tweeted user name based on count where text is Samsung:



**Graph for Query 10**

