**COURSE – PRINCIPLES OF BIG DATA MANAGEMENT**

**PROJECT - TWITTER STREAMING API’S USING TWITTER DATA COLLECTION**

**(Phase 2 of Project)**

**Team Members**

**Aarthi Nagireddy ( 16292448 )**

**Manaswini Vedula ( 16292800 )**

**Yamini Bommineni ( 16292447 )**

**Instructor**

**Dr. PRAVEEN RAO, Ph.D.**

**Abstract:**

Project Phase 2 deals with the below mentioned tasks

* To write analytical Queries on the twitter Data collected.
* To Visualize the Query results.

**Tools and Technologies used:**

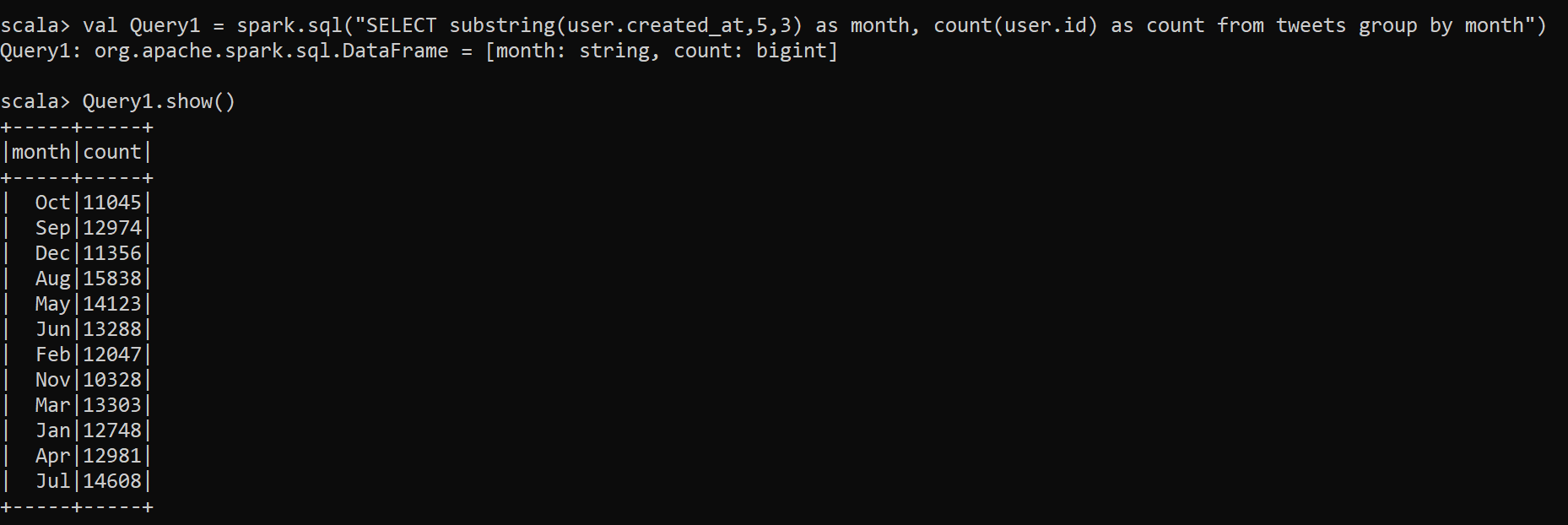
* Apache Spark
* Tableau

**Data:** Brands of Sportswear (Adidas, Puma, Nike, Reebok, UnderArmour)

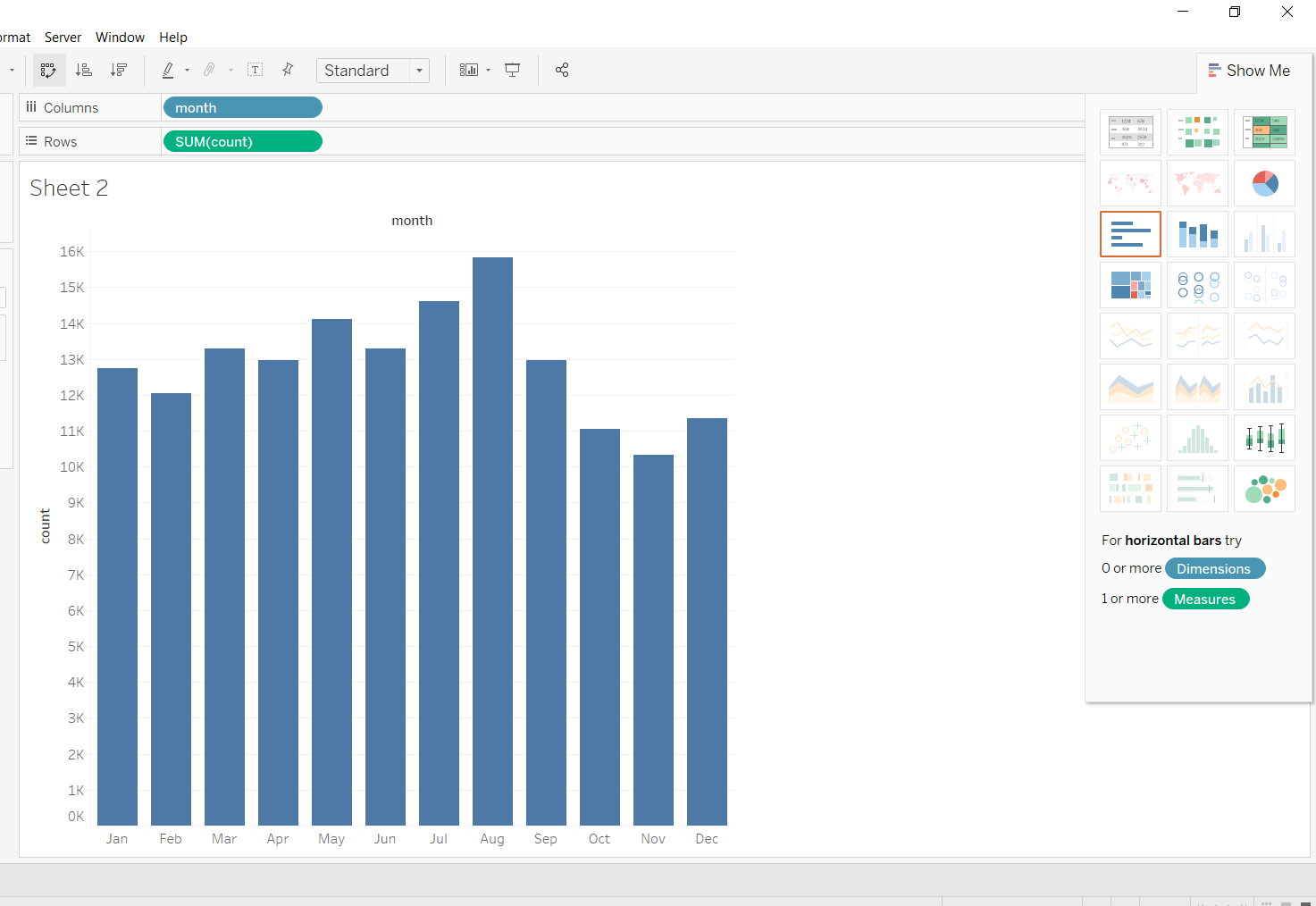
**Query Analysis and Visualization**

**Query 1 :** This Query returns the number of users posted tweets related to the Sportswear Brands in each month.

**Query :**

****

**Visualization :**

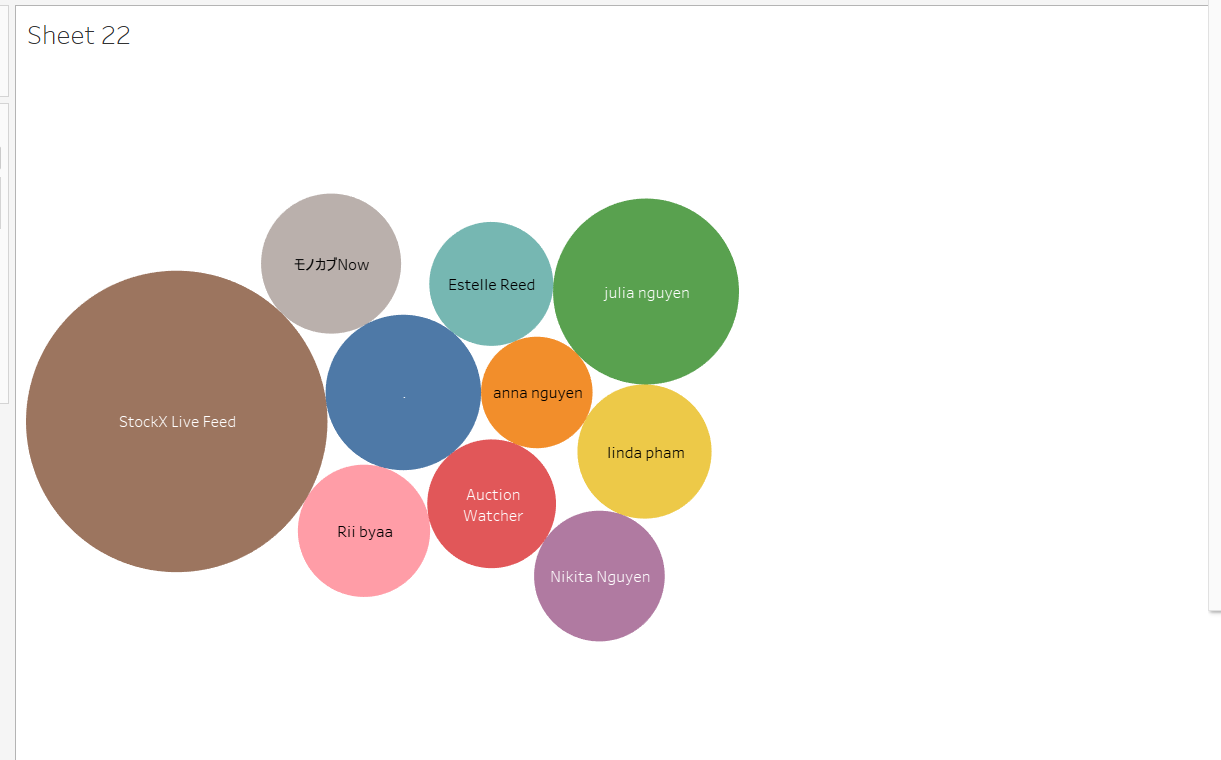


**Query 2 :** This Query returns the top 10 names of the users based on the number of tweets they have tweeted.

**Query:**

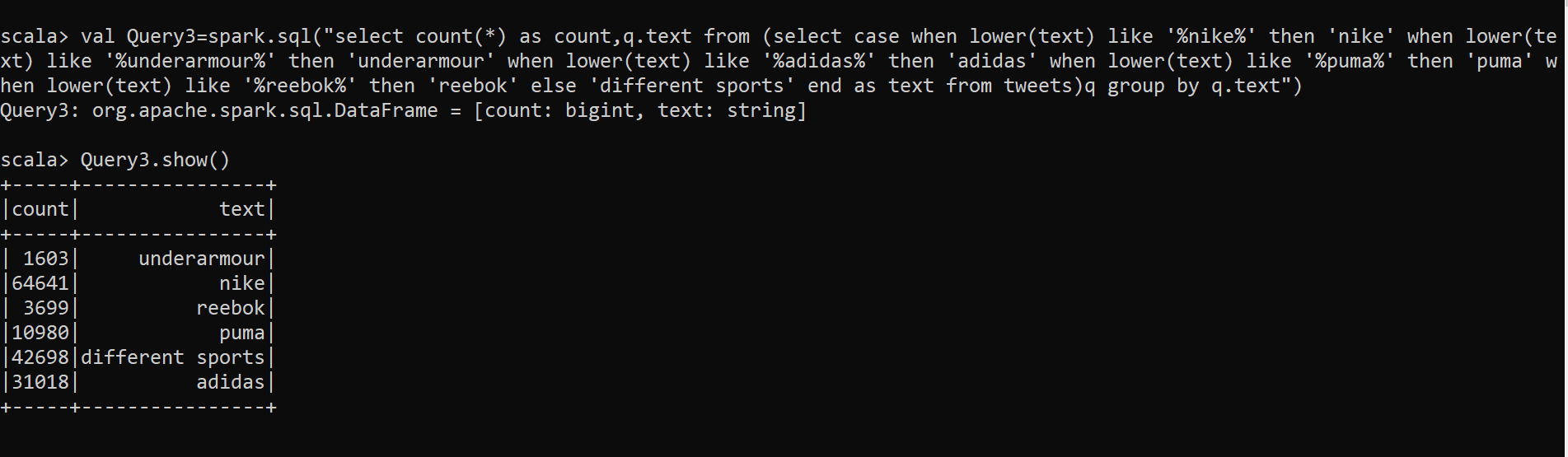


**Visualization :**

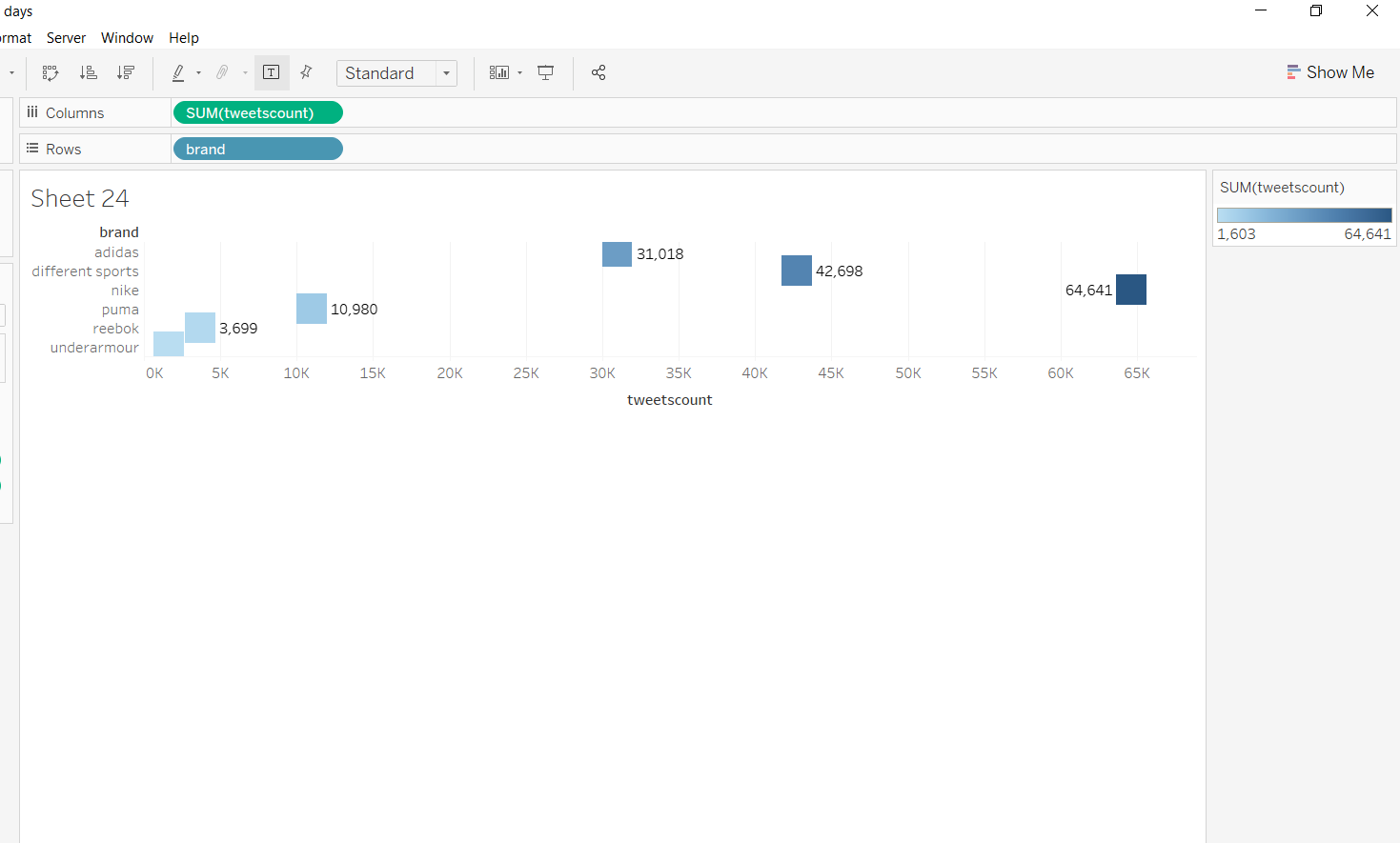


**Query 3 :** This Query returns the brand names and their corresponding tweets count.

**Query:**

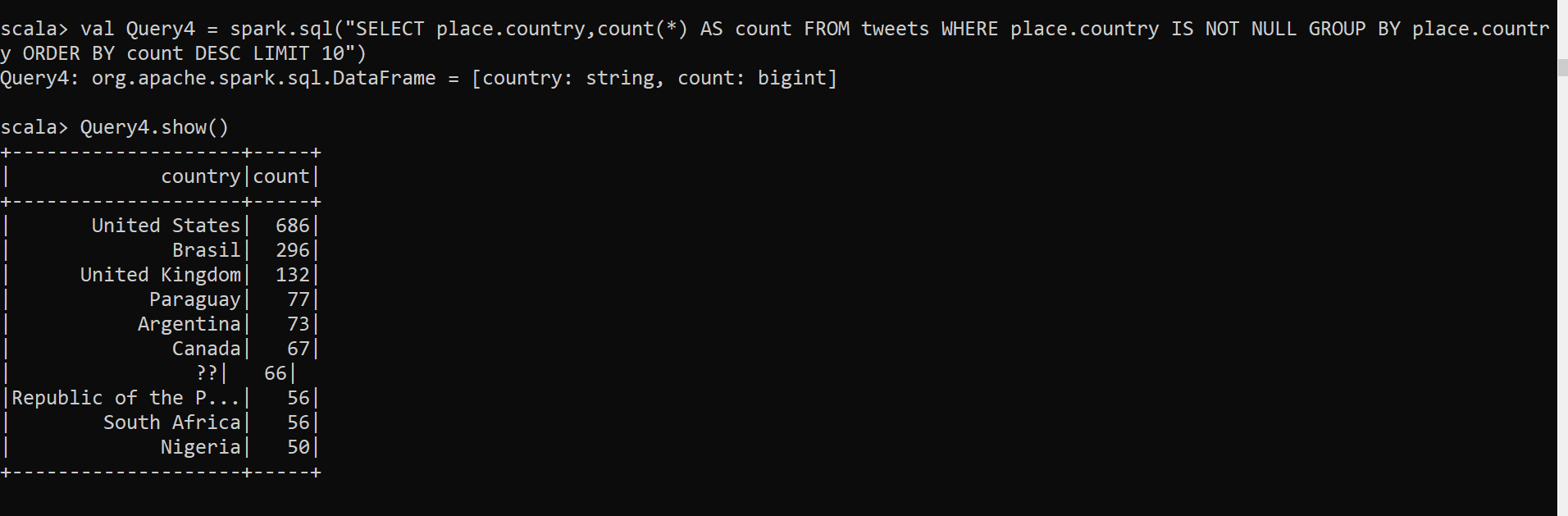


**Visualization :**

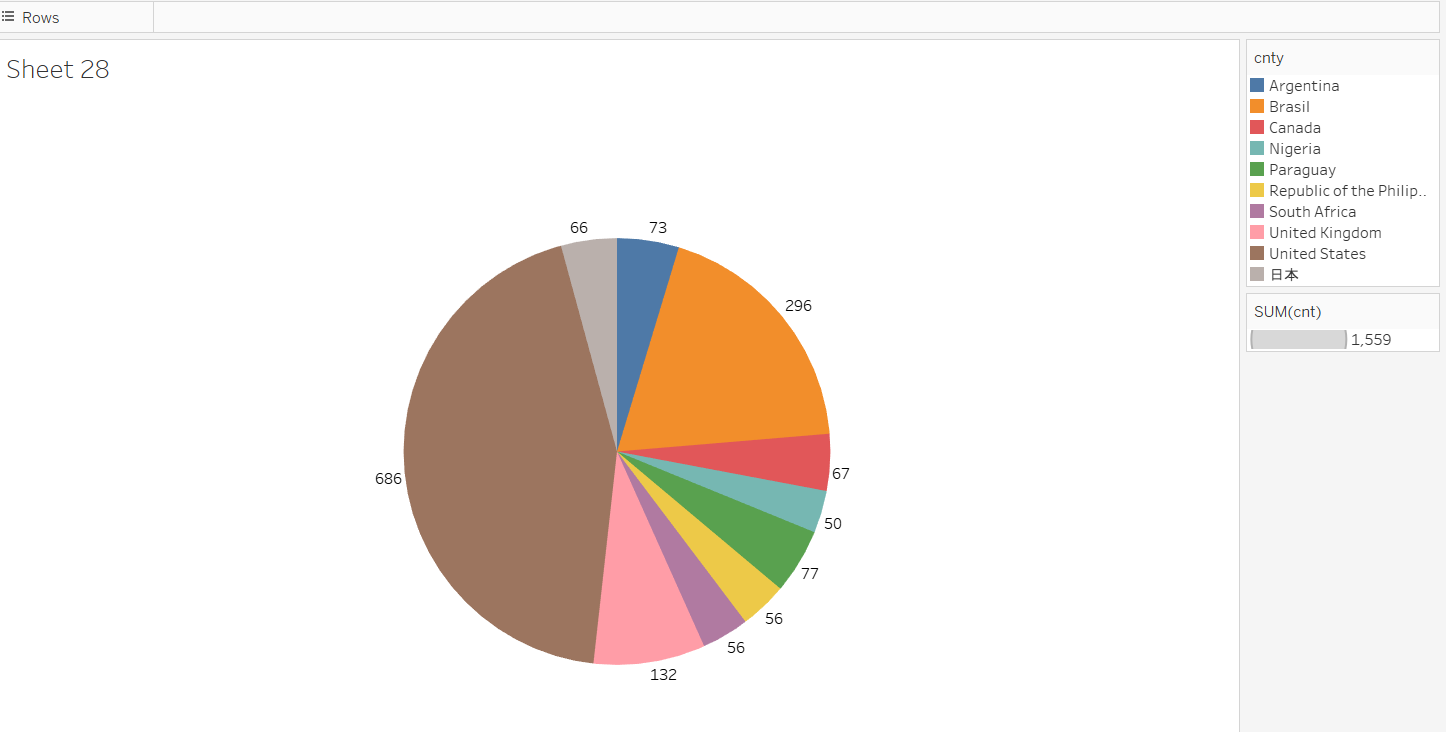


**Query 4 :** This Query returns the top 10 countries tweets posted by their people related to the Brands.

**Query:**

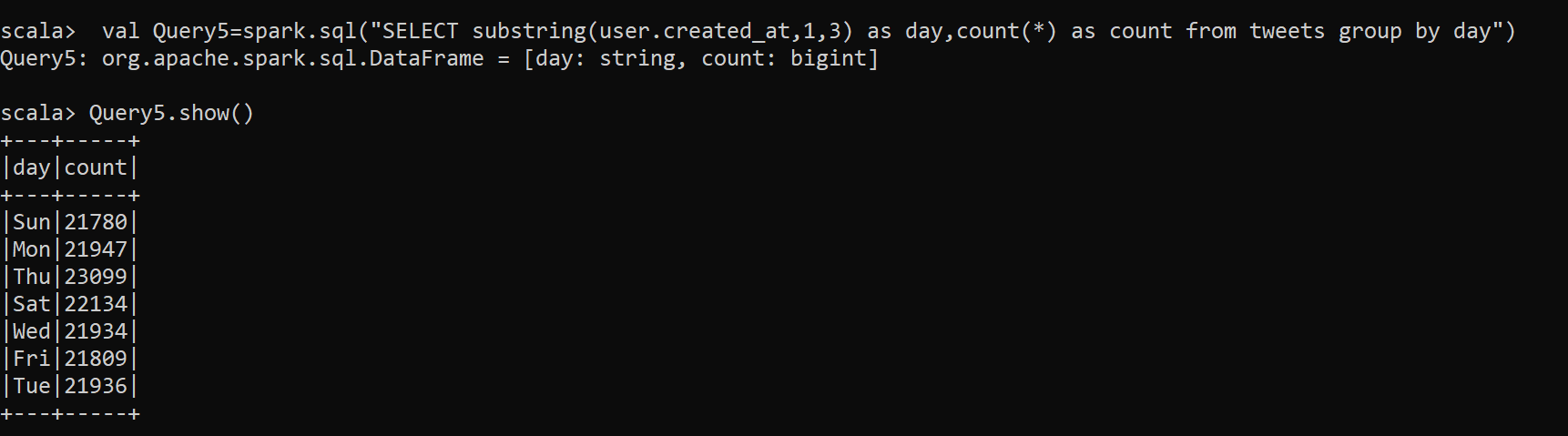


**Visualization :**

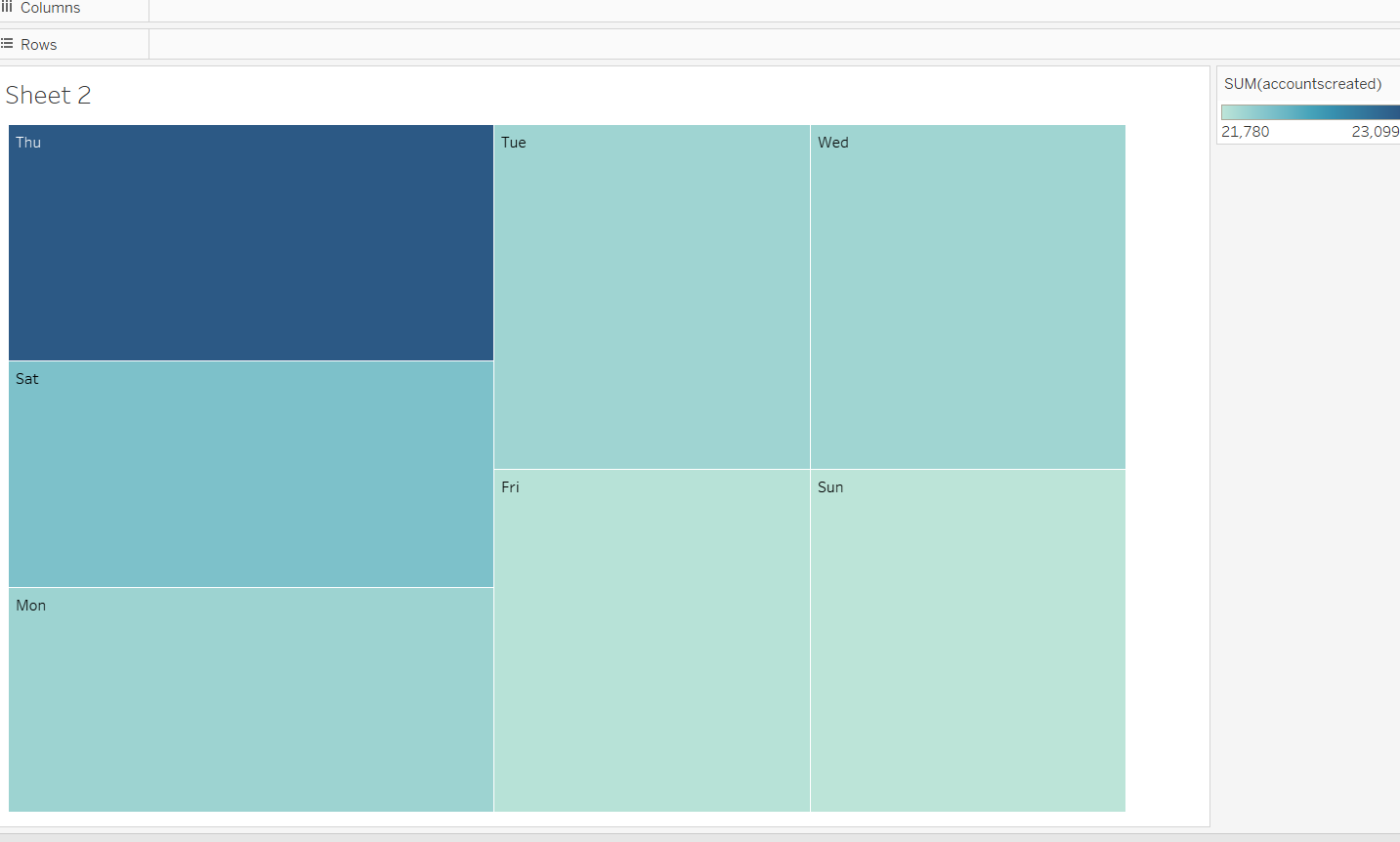


**Query 5 :** This Query returns the count of tweets in a days of a week.

**Query:**

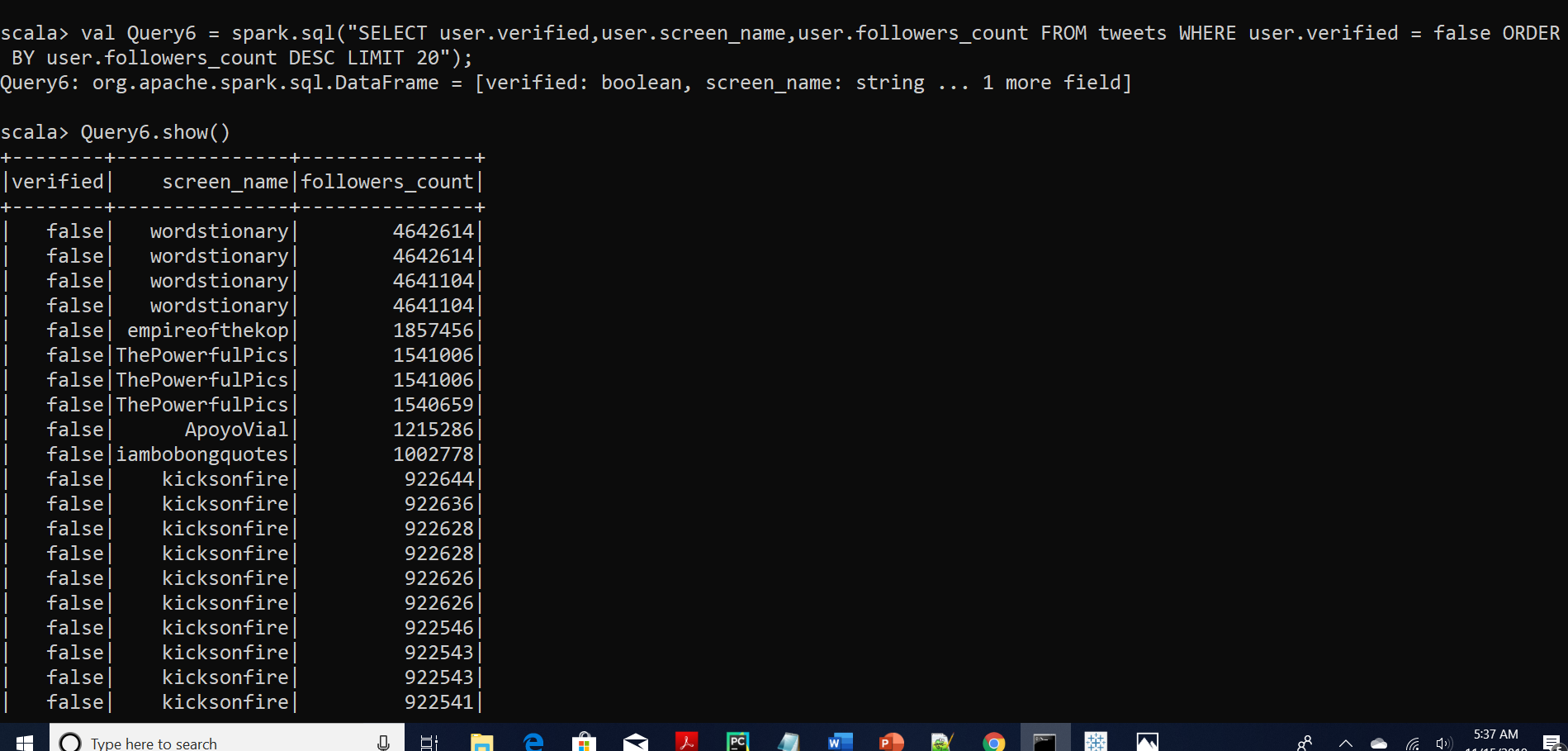


**Visualization :**

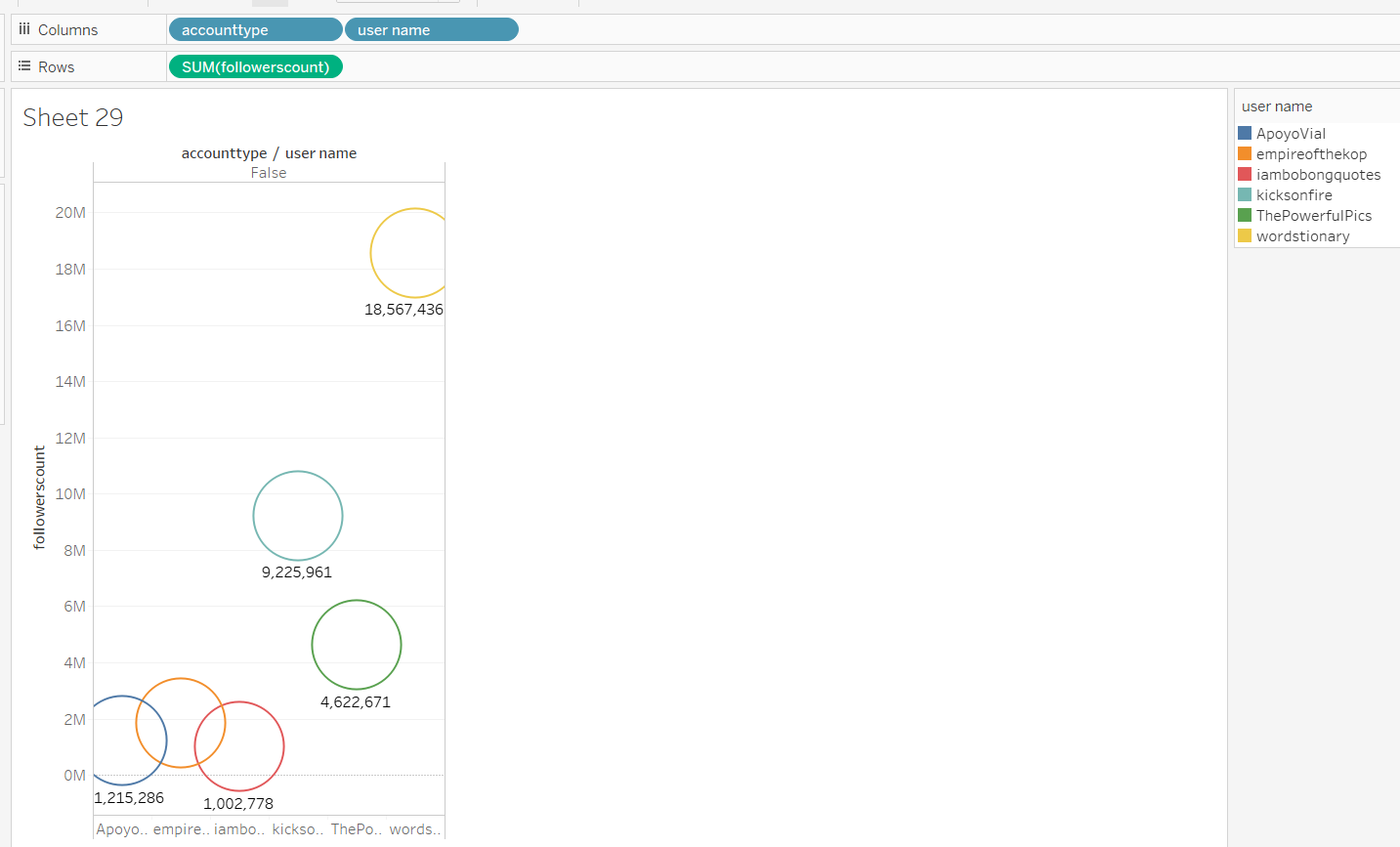


**Query 6 : This Query returns the top 20 unverified user names and their followers count**

**Query:**

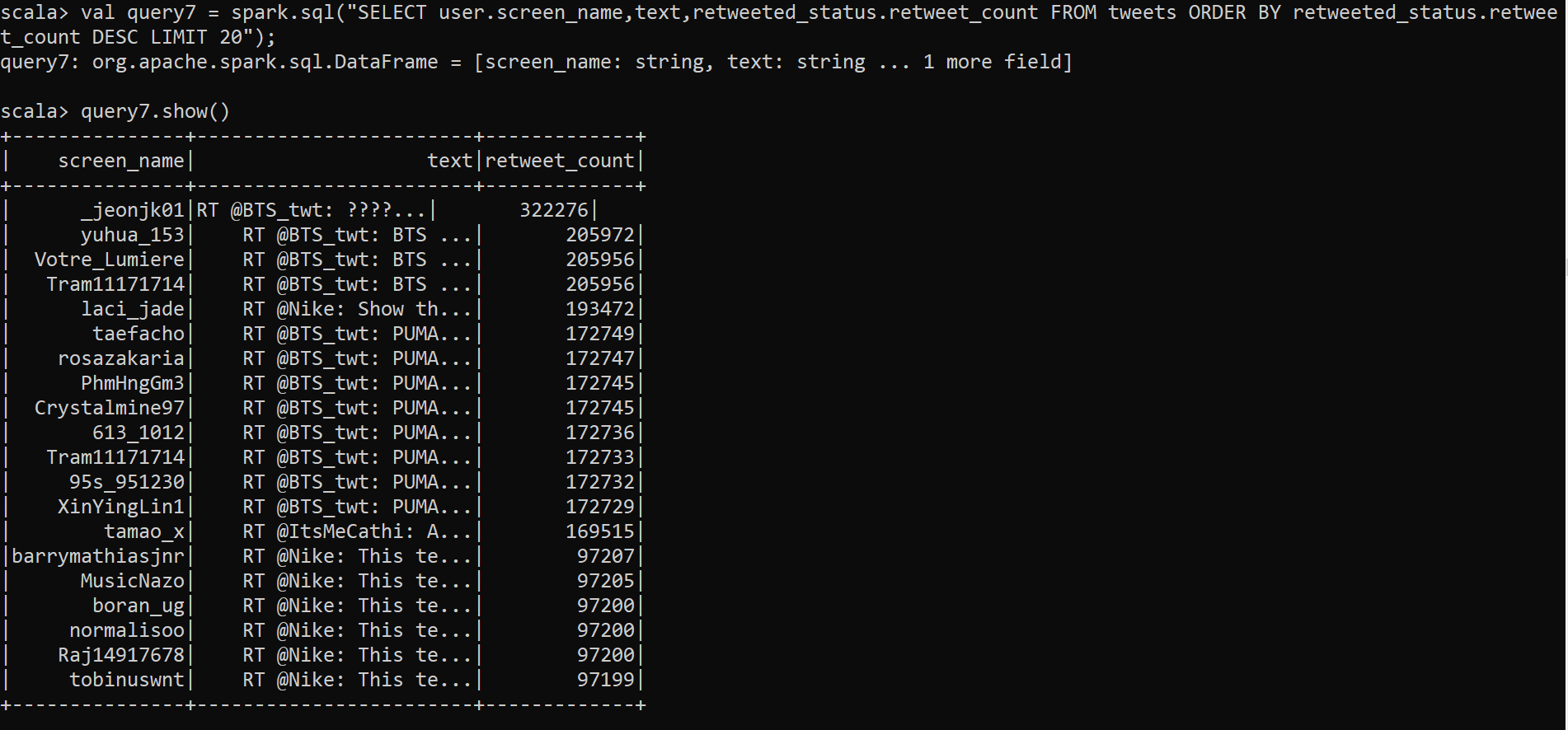


**Visualization:**

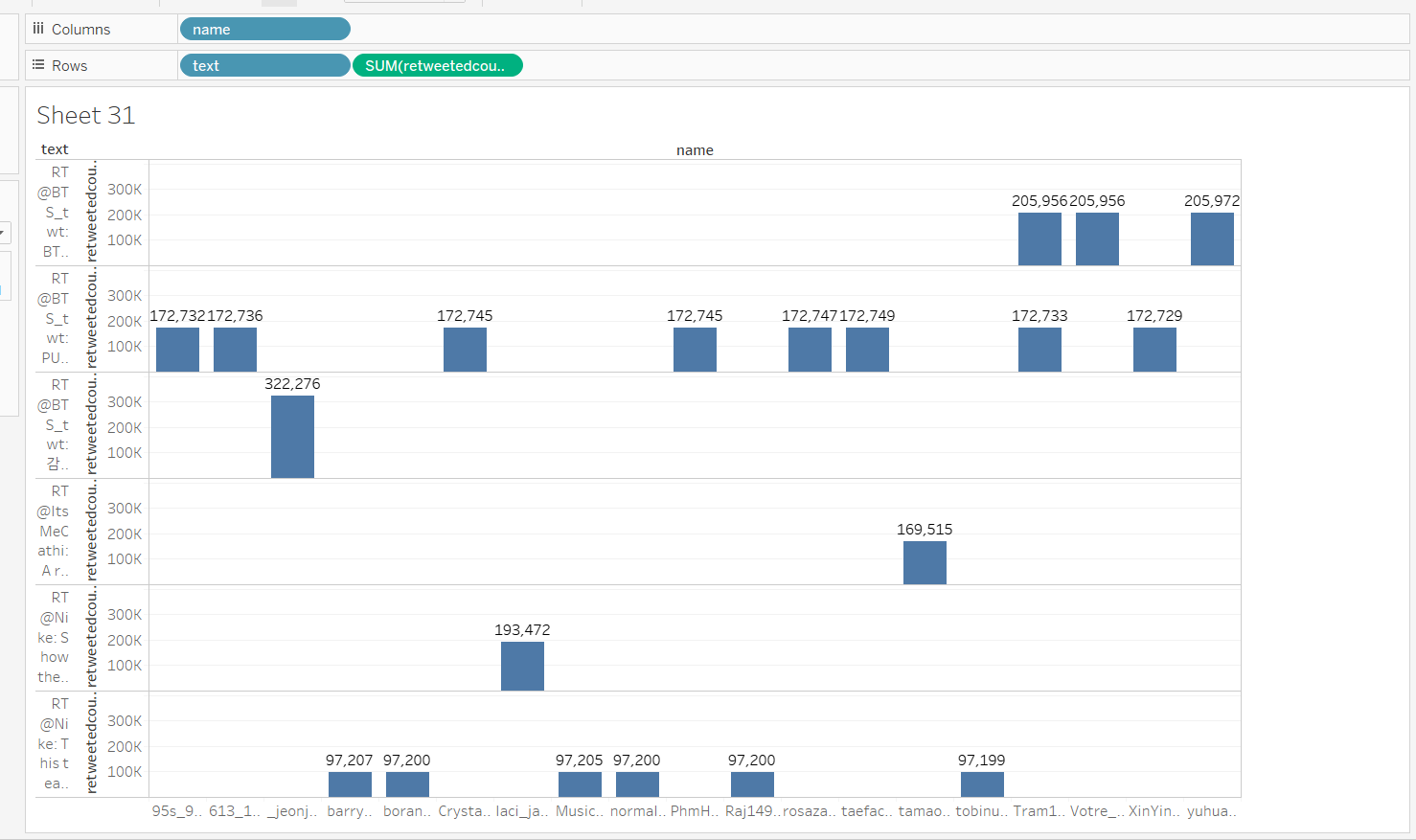


**Query 7 :This Query returns the top 20 users retweet count and their tweet text**

**Query:**

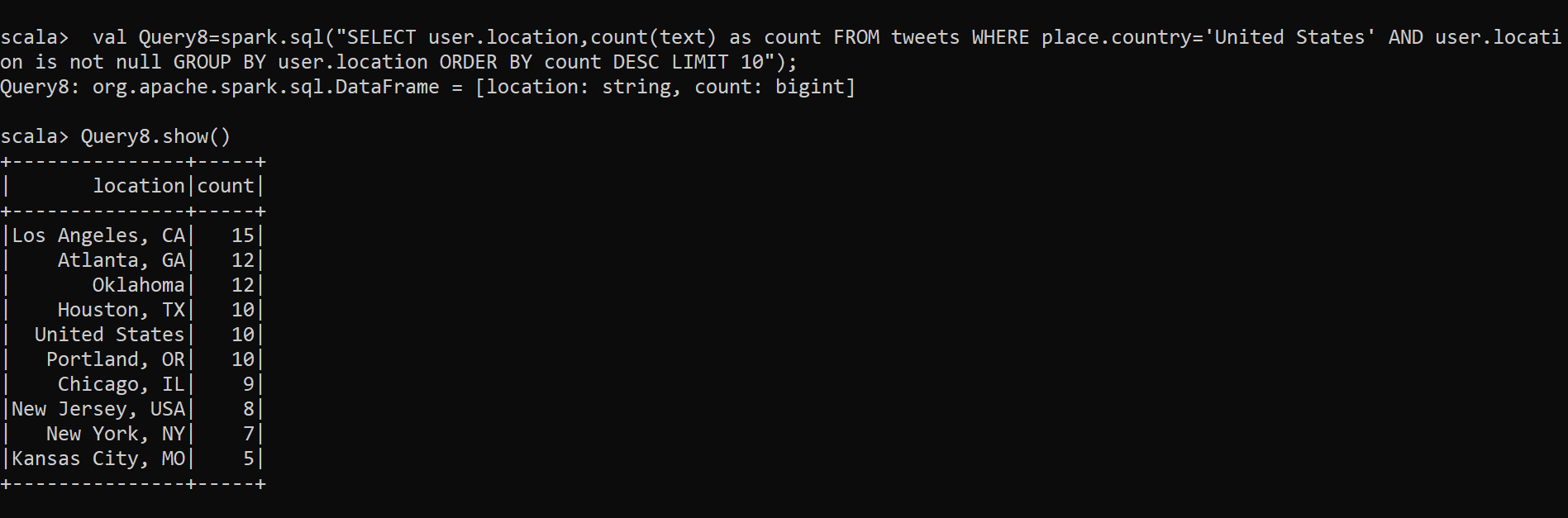


**Visualization:**

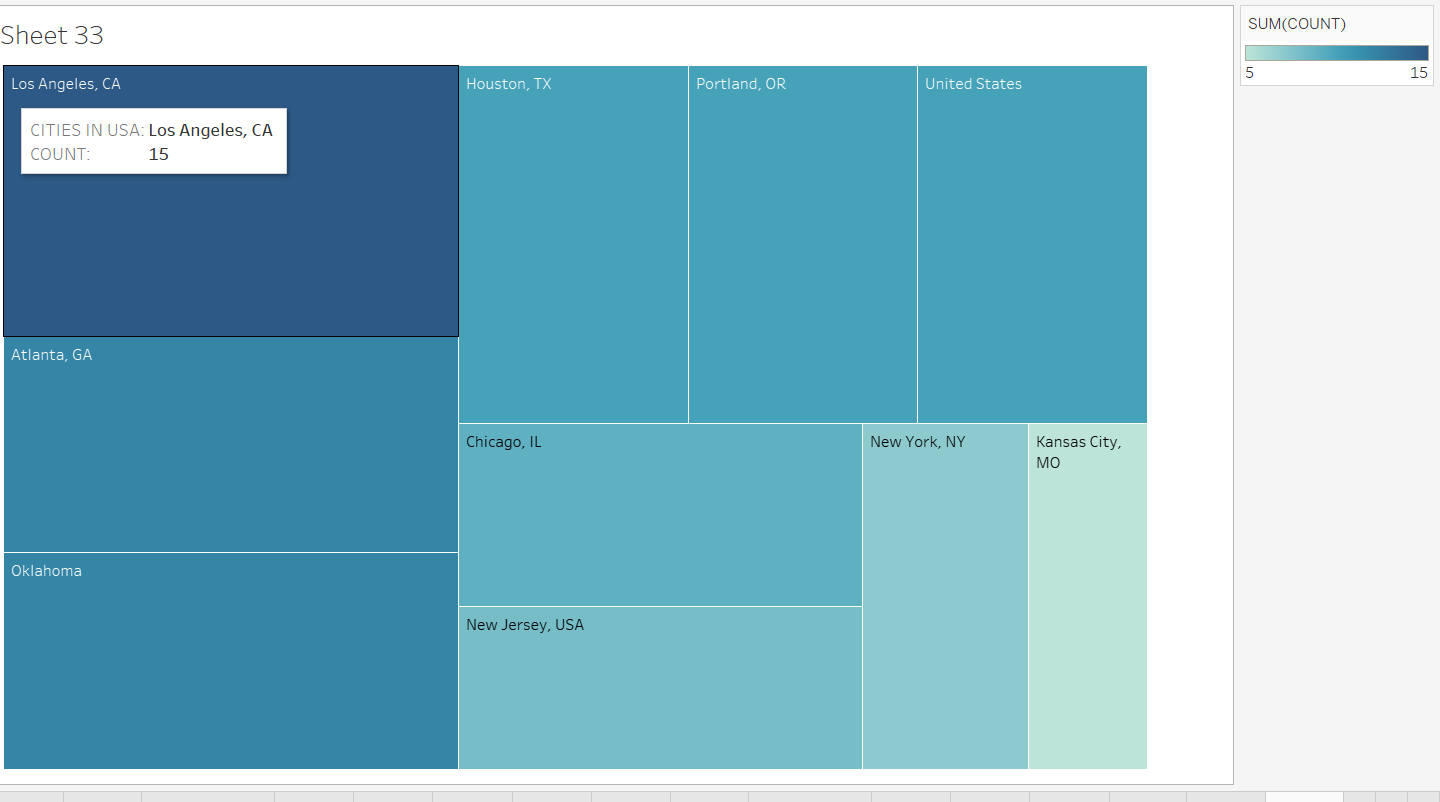


**Query 8 :** This Query returns the top 10 count of tweets posted by people from USA and their corresponding location.

**Query:**

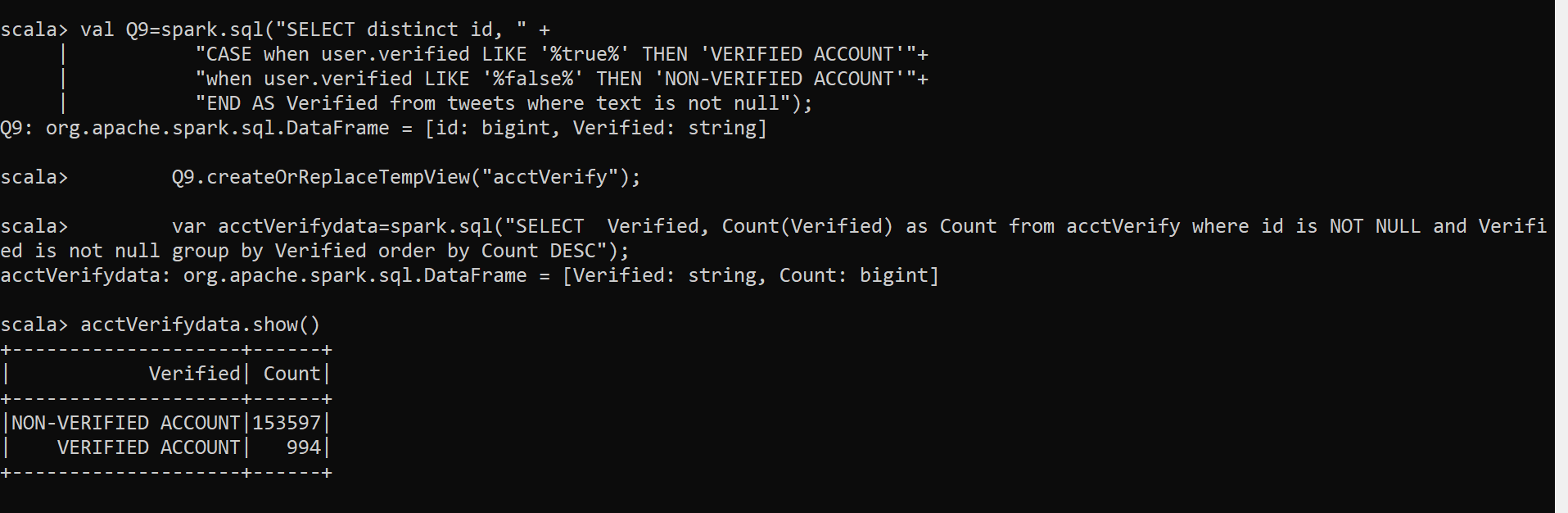


**Visualization :**

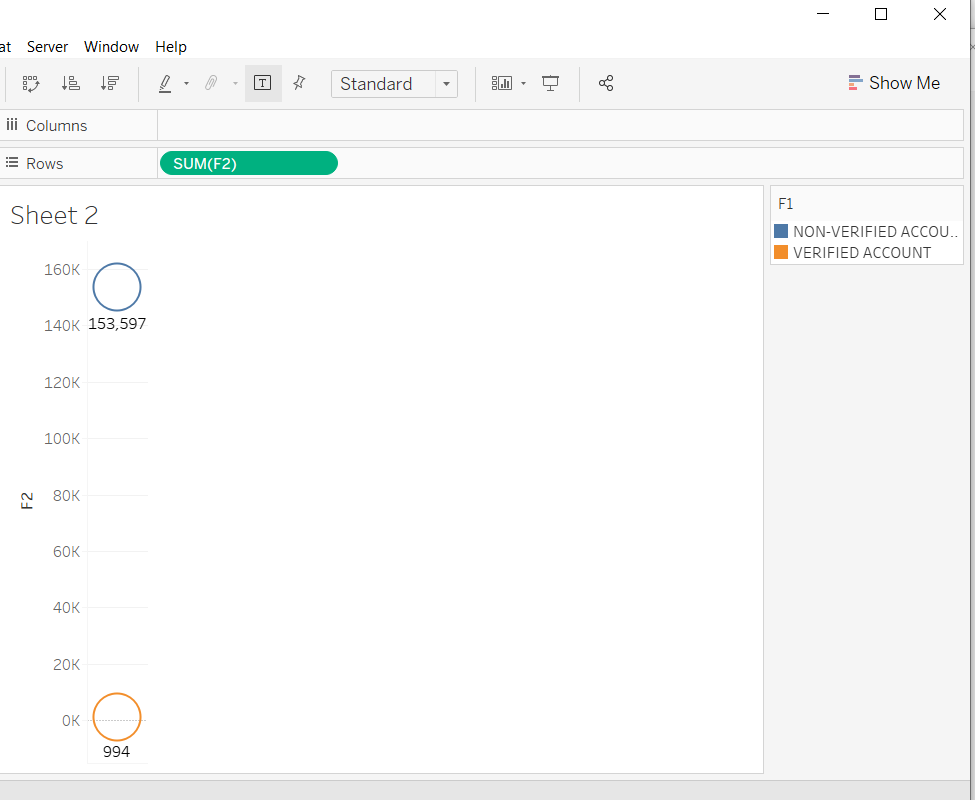


**Query 9:This query returns the count of verified and non-verified accounts from the collected twitter data**

**Query:**

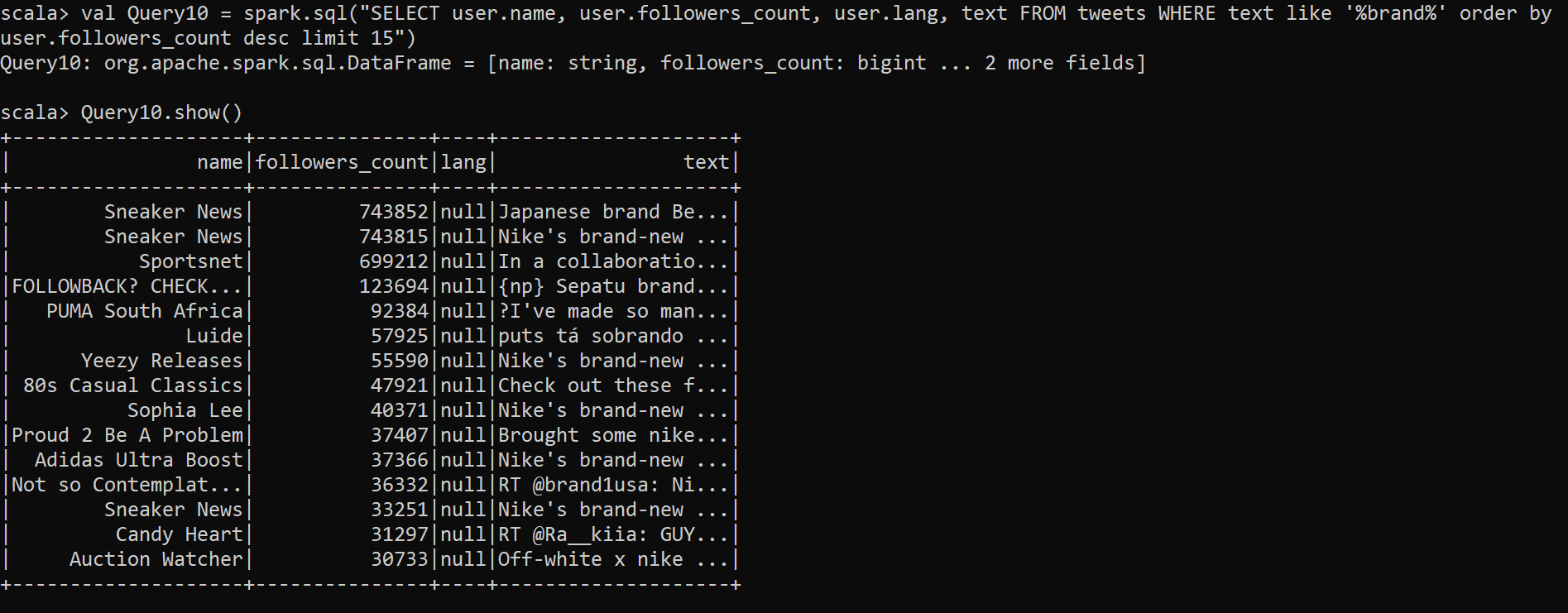


**Visualization:**

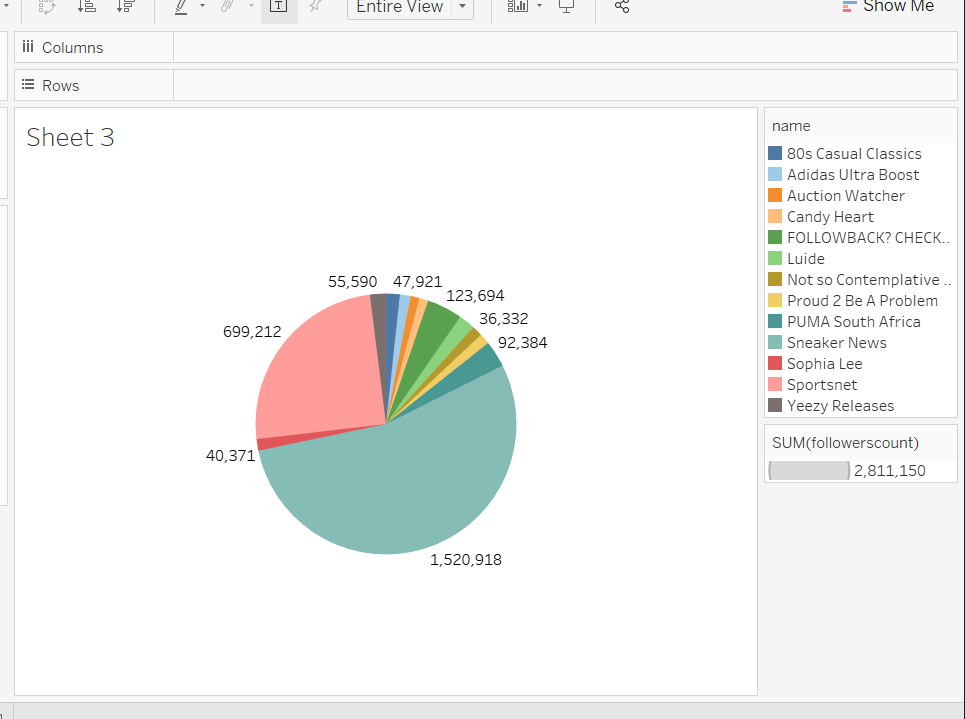


**Query 10 :**

**Code :This Query returns the top 10 users account name, followes count and text they have posted related to the brands**



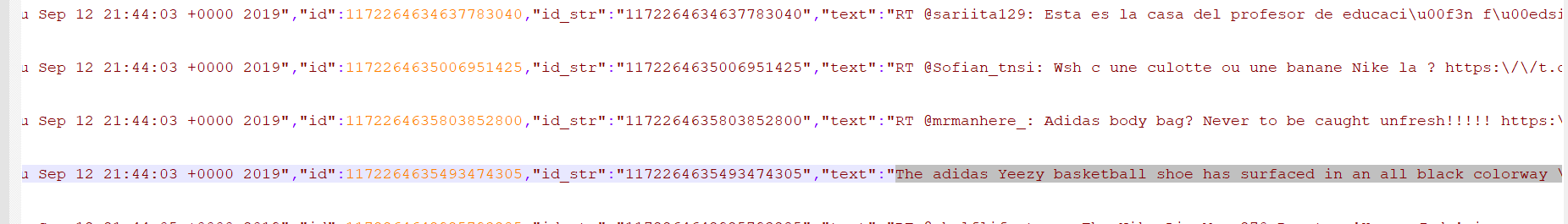
**Visualization :**



**Debugging :**

We tested the results manually by selecting some text from the data collected and then found the tweet in twitter.

Example : Let’s take “ The Adidas Yeezy basketball has surfaced in an all black colorway” from the json data collected.

****

We got the below tweet , related to the search.



**Github link:** <https://github.com/manaswinivedula/PB-phase2>

**Json data link: https://drive.google.com/open?id=18i0UiYj3nnVygQteli7jZiikHLojiL7h**

**References:**

• <https://www.tutorialspoint.com>

• <https://stackoverflow.com/>

• <http://text-processing.com/api/sentiment>