**Phase 2 Submission**

**Project Team Members:**

Hari Manasa Ganganagunta

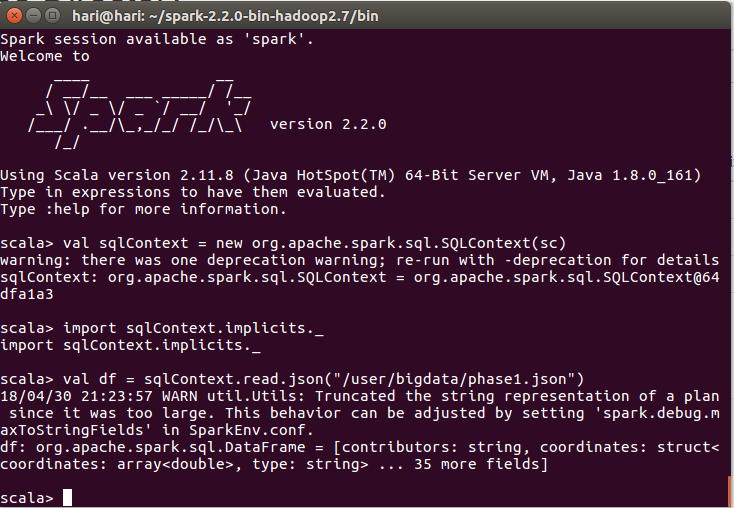
Sree Naga Sayana

Venkata Akhila Krishna Atluri

**Achievements:**

Storing the tweets in Spark SQL

We have uploaded the json file into the hdfs directory (/user/bigdata/phase1.json)





The above screen shot states that json file(extracted tweets) are converted into a table.

Analytic queries to explore and understand the data

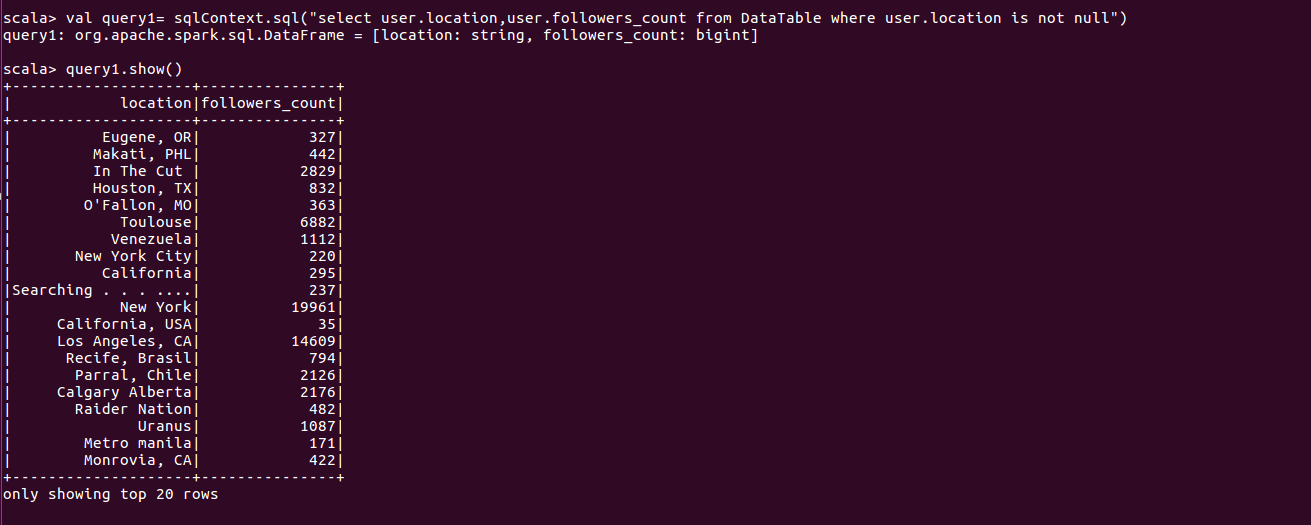
Through 11 different queries we have retrieved the data from the extracted json files of tweets.

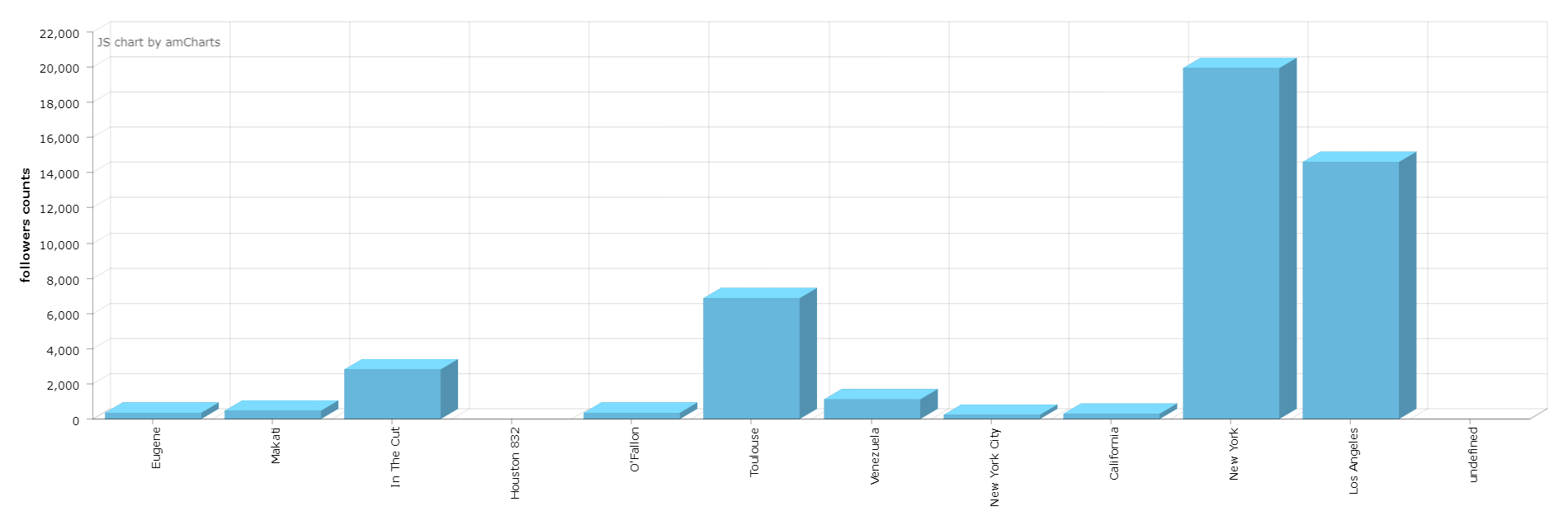
Detailed description of the queries and its visualization has been explained below.

**Query1**: User’s location and followers count were extracted from the data table

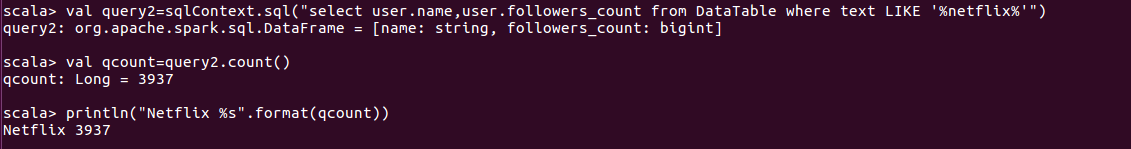
And given the constrain as location can not be null.

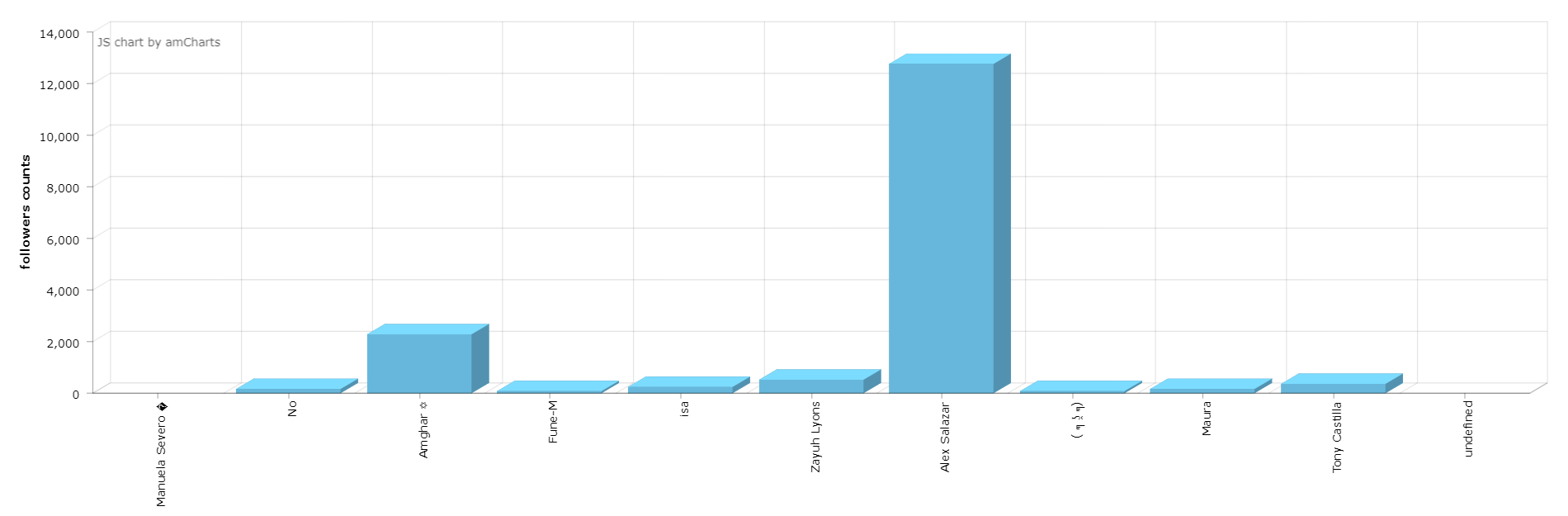
A screenshot is provided below which shows the followers count from different locations like Eugene, Houston, New York city etc., Each location has got their respective follower’s count.



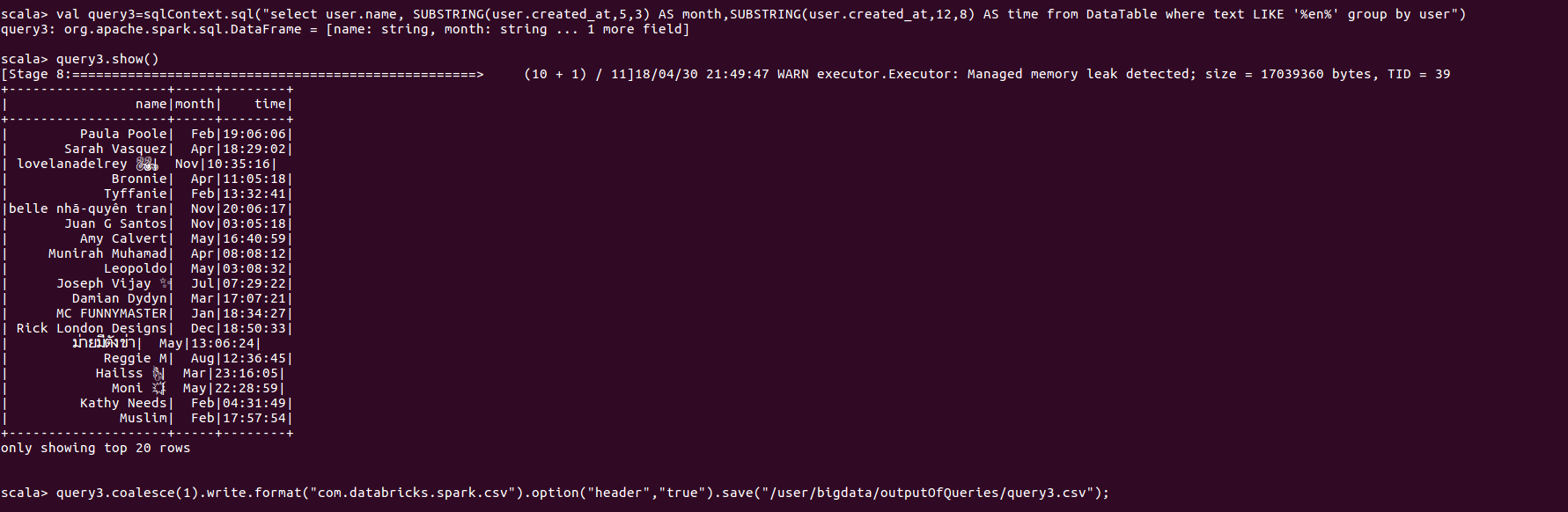


**Query2**: In this query we have retrieved the follower’s count who have tweeted as Netflix which can found from the below provided screenshot.

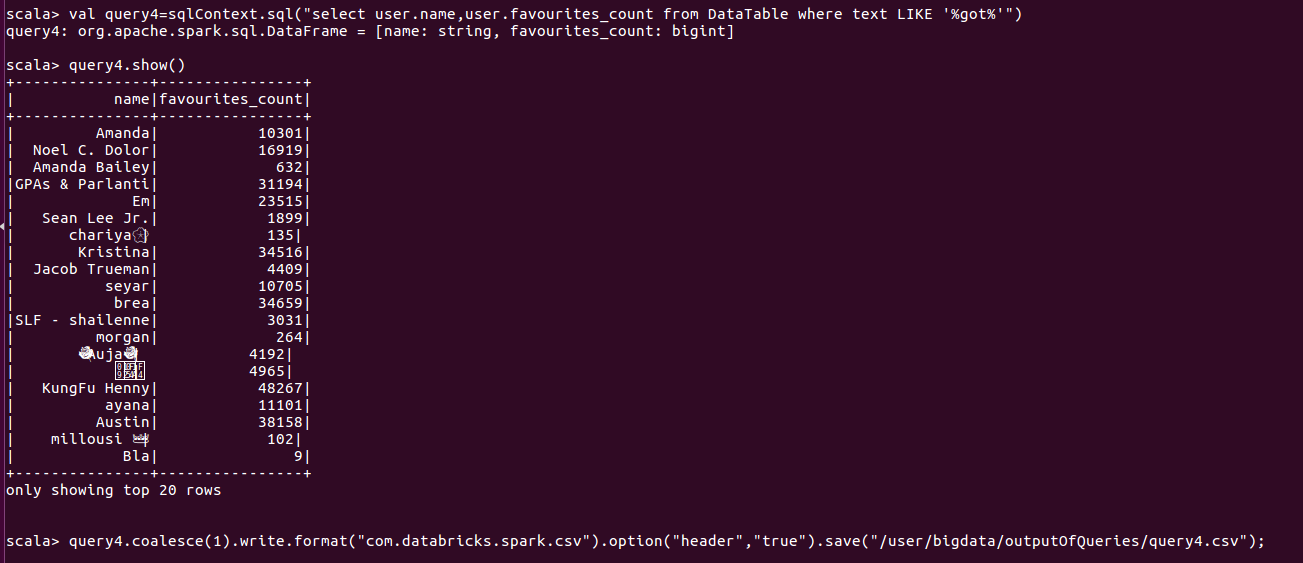


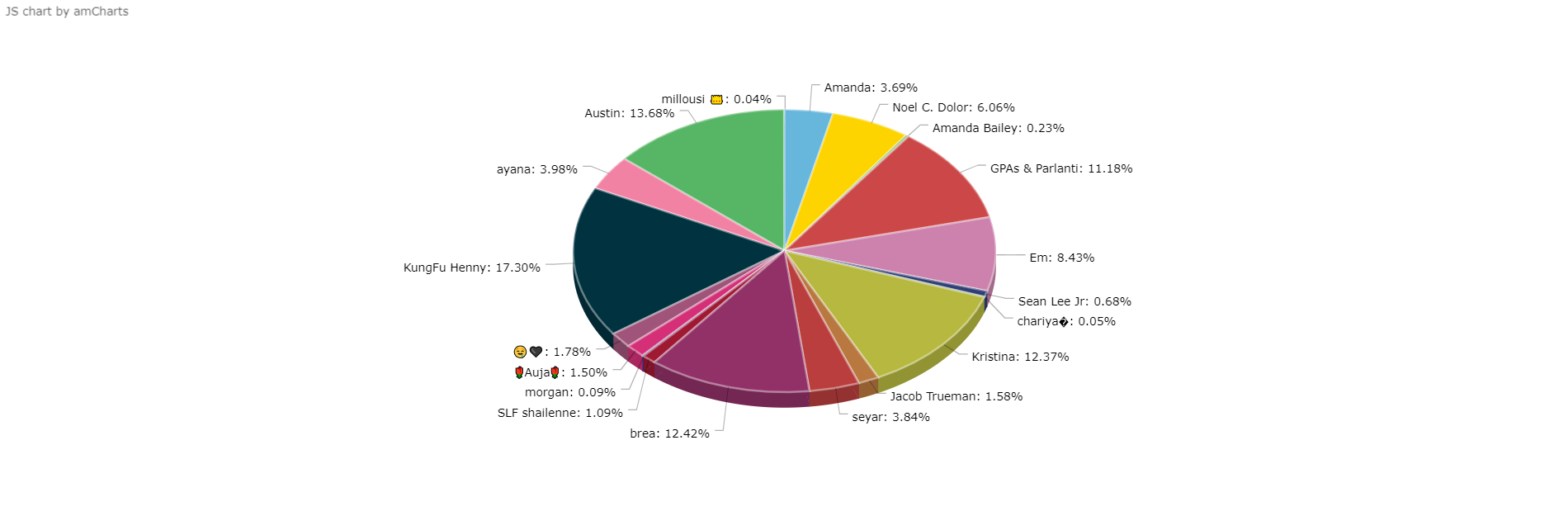


**Query3**: Query states about the name, month and time from the table DataTable who has tweeted language as English.

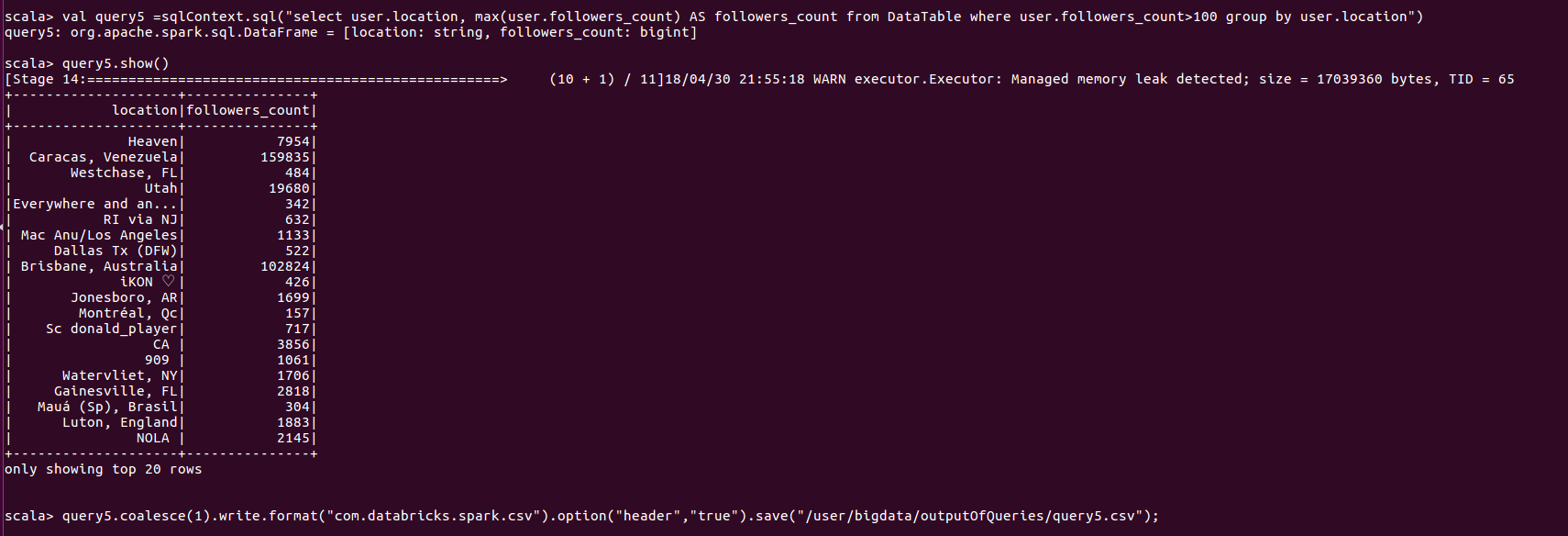


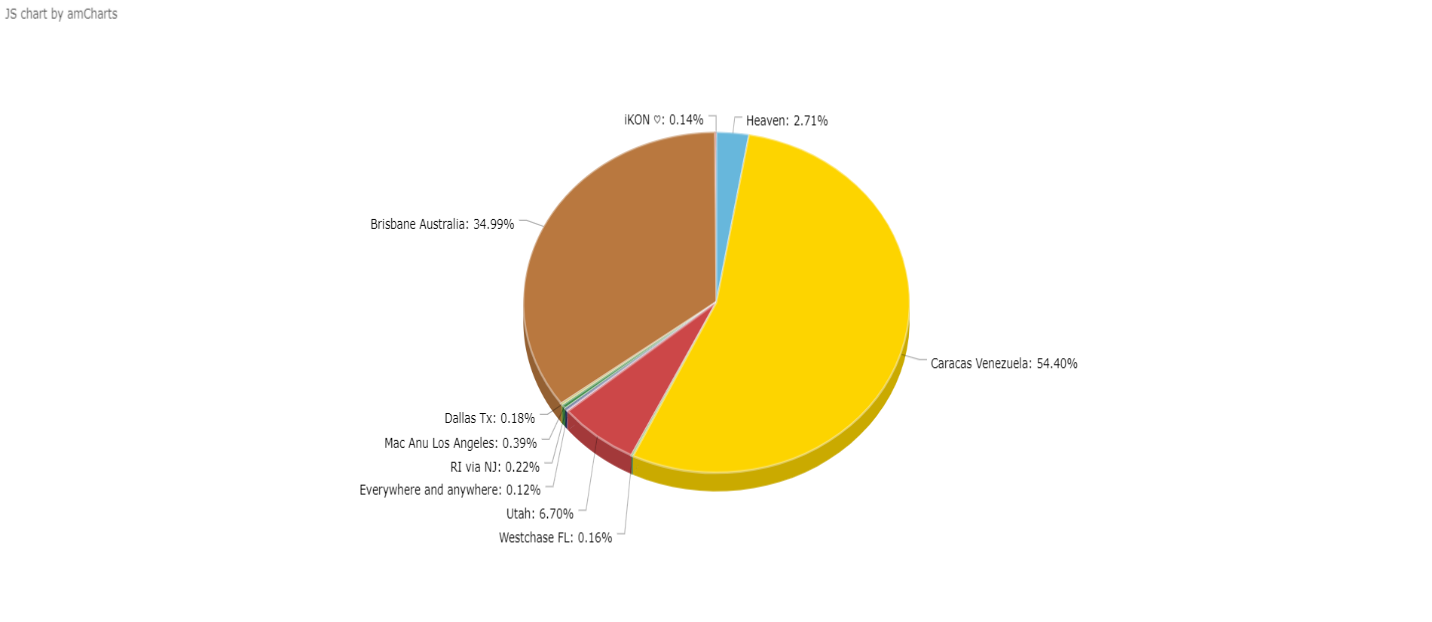
**Query4**: The below query states the users name and favourites who has tweeted text as got



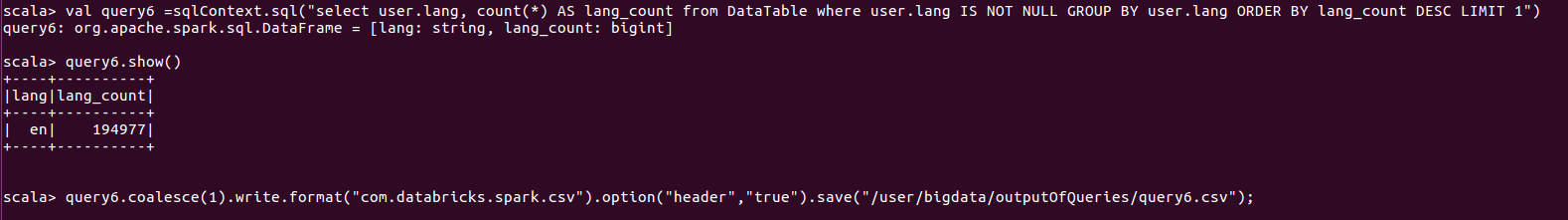


**Query5**: Query states the location and followers count in which followers count is greater that 100 and grouped with location.

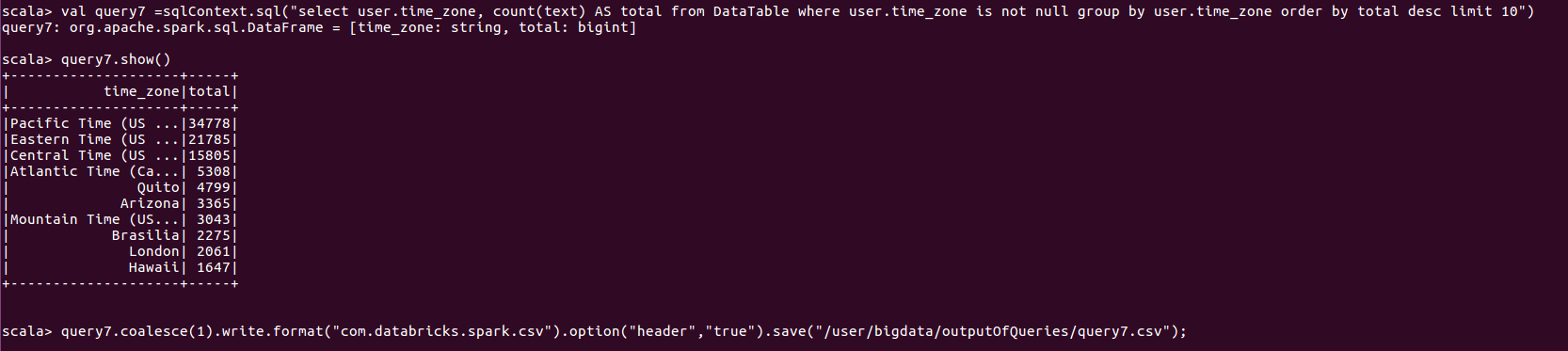


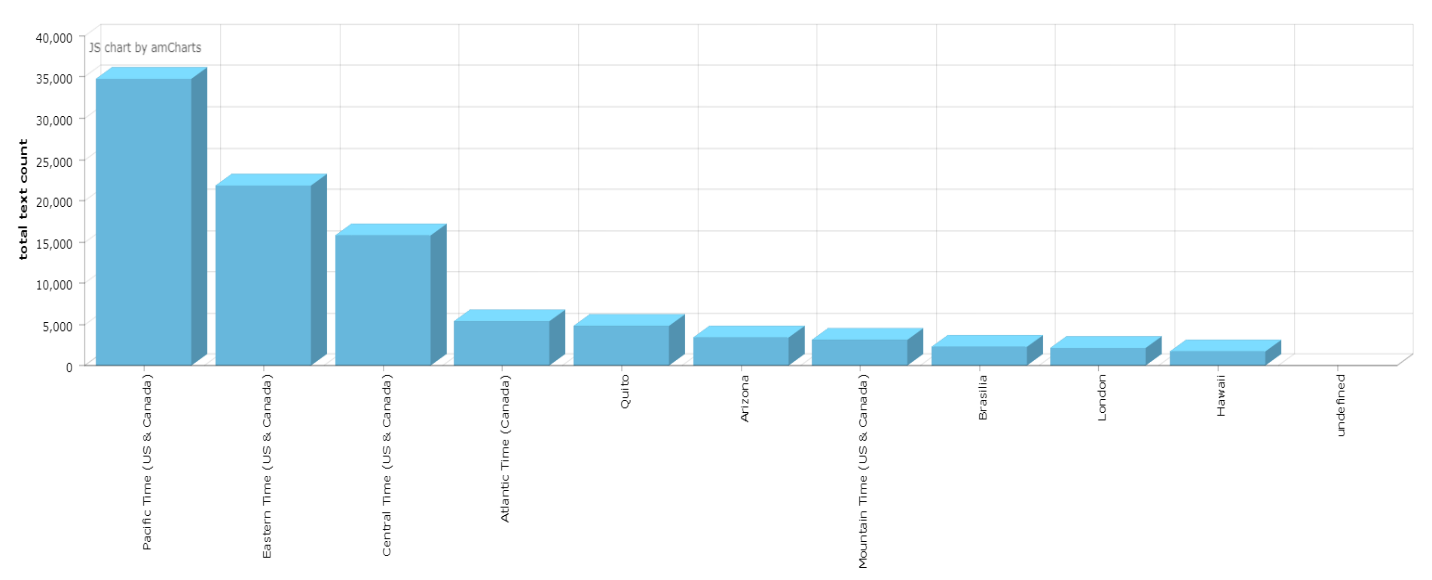


Query6: Query is about the any one language total count

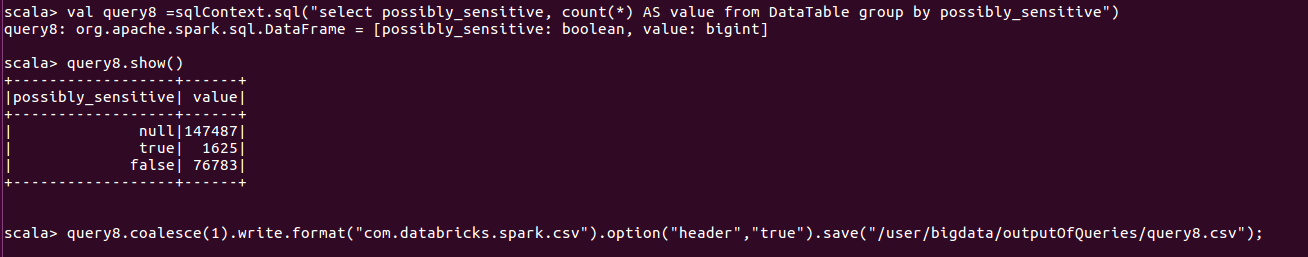


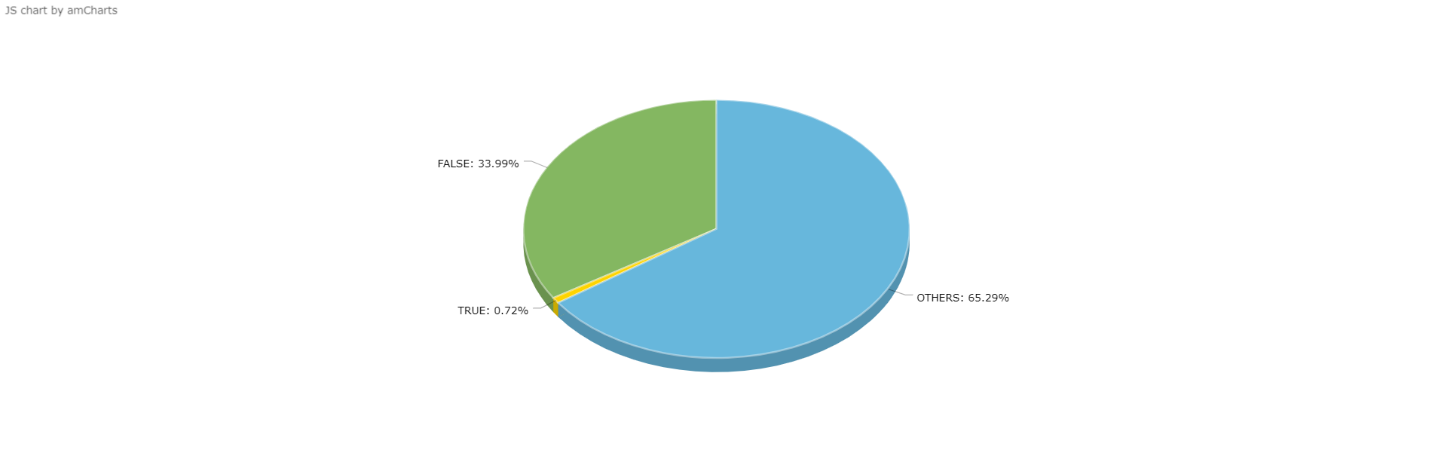
Query7: Query states time\_zone and the count of the text with a limit of 10





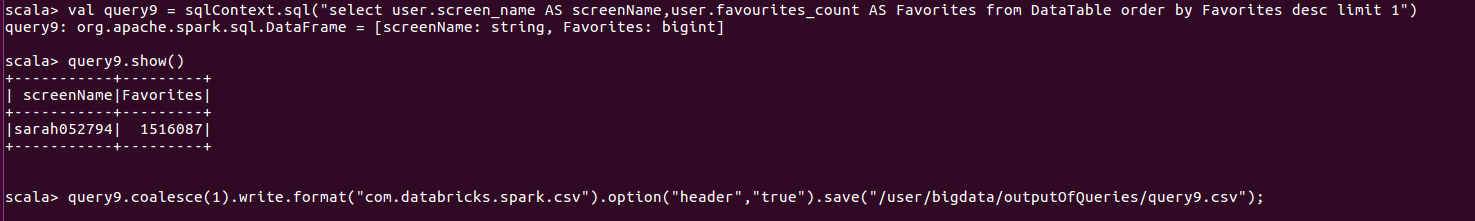
Query8: Counted the possibly\_sesntitive emotions from the DataTable





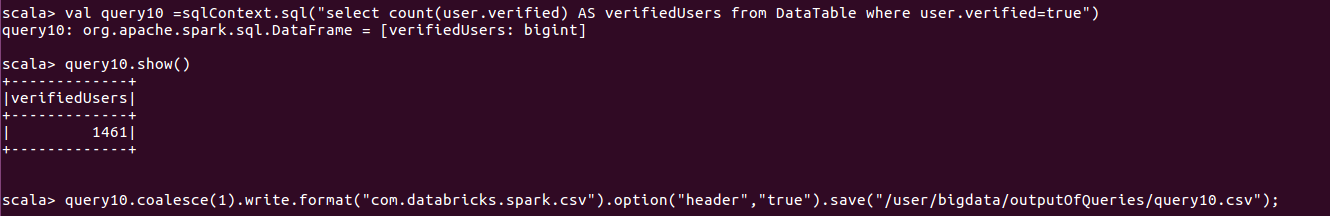
**Query9**:

Top tweeted user name from the DataTable which is calculated on the bases of favourites\_count



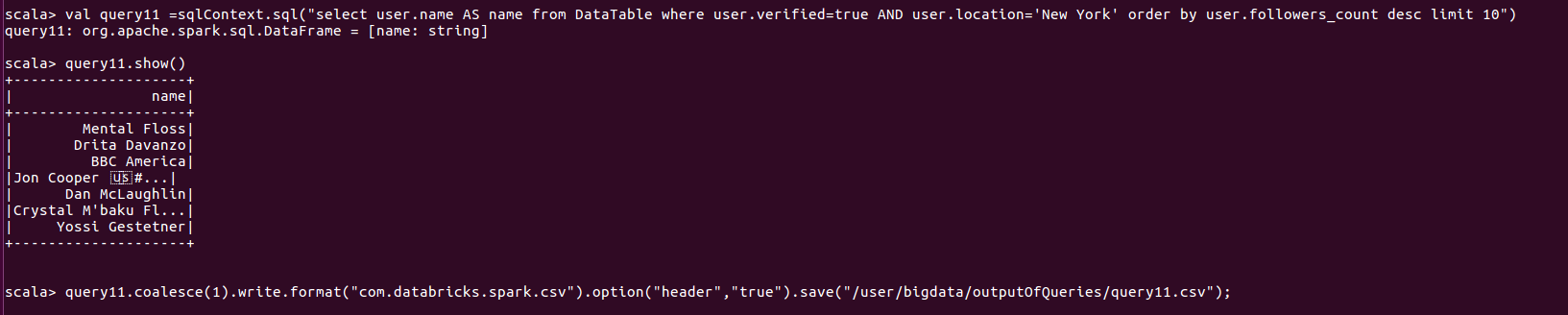
Query10:

Number of verified users from the collected data

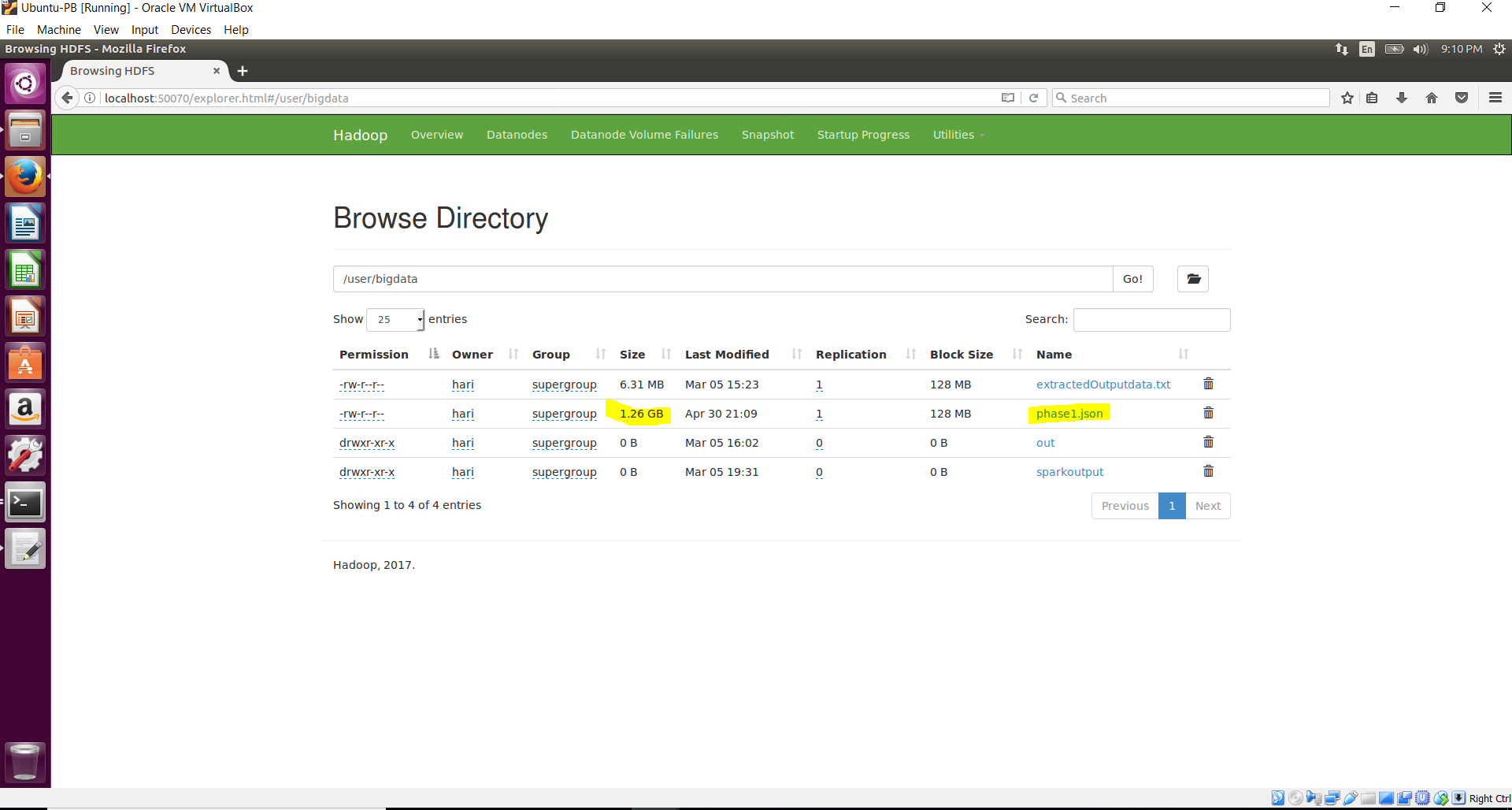


Query11:

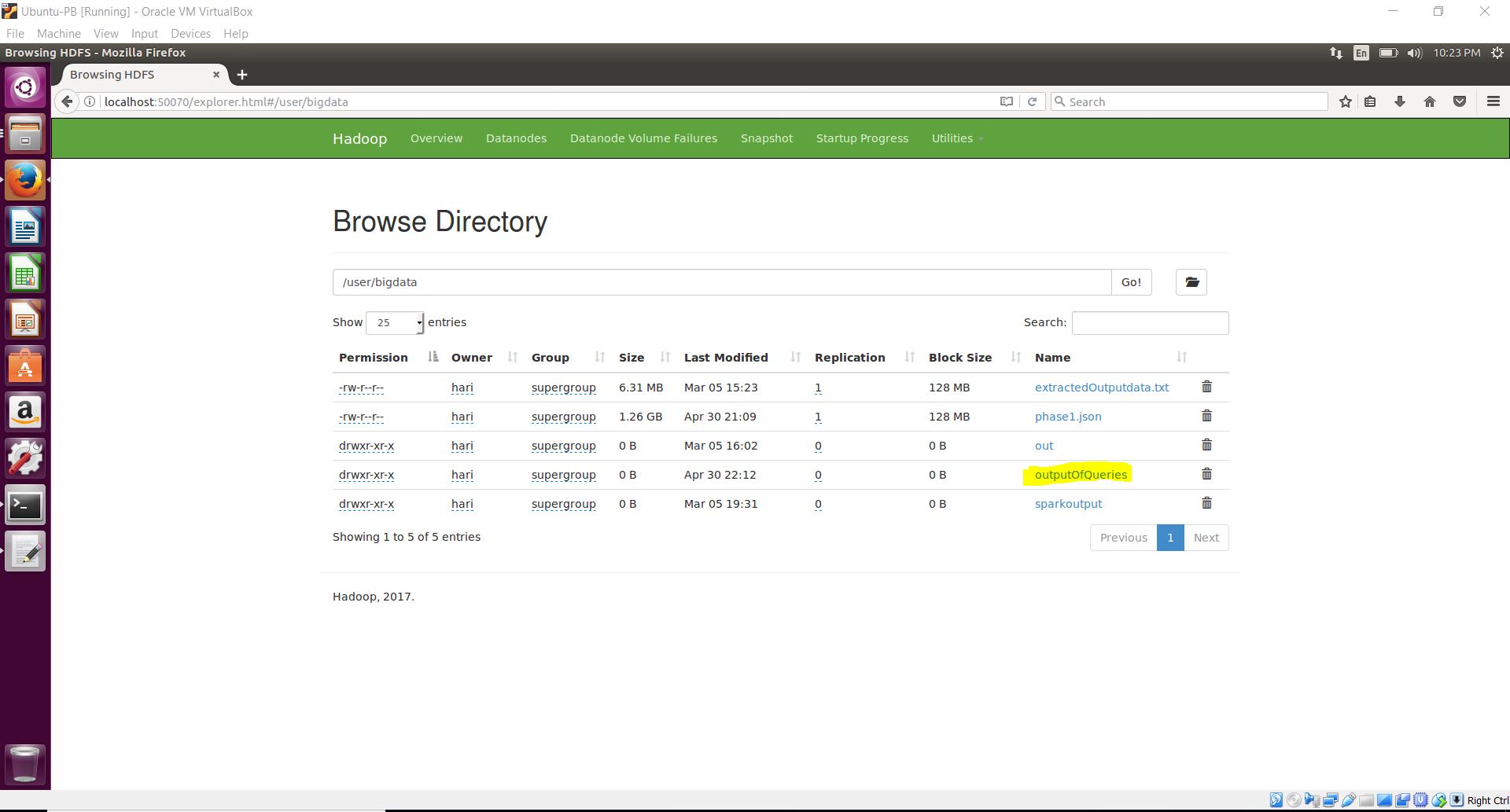
Top 10 verified user names.



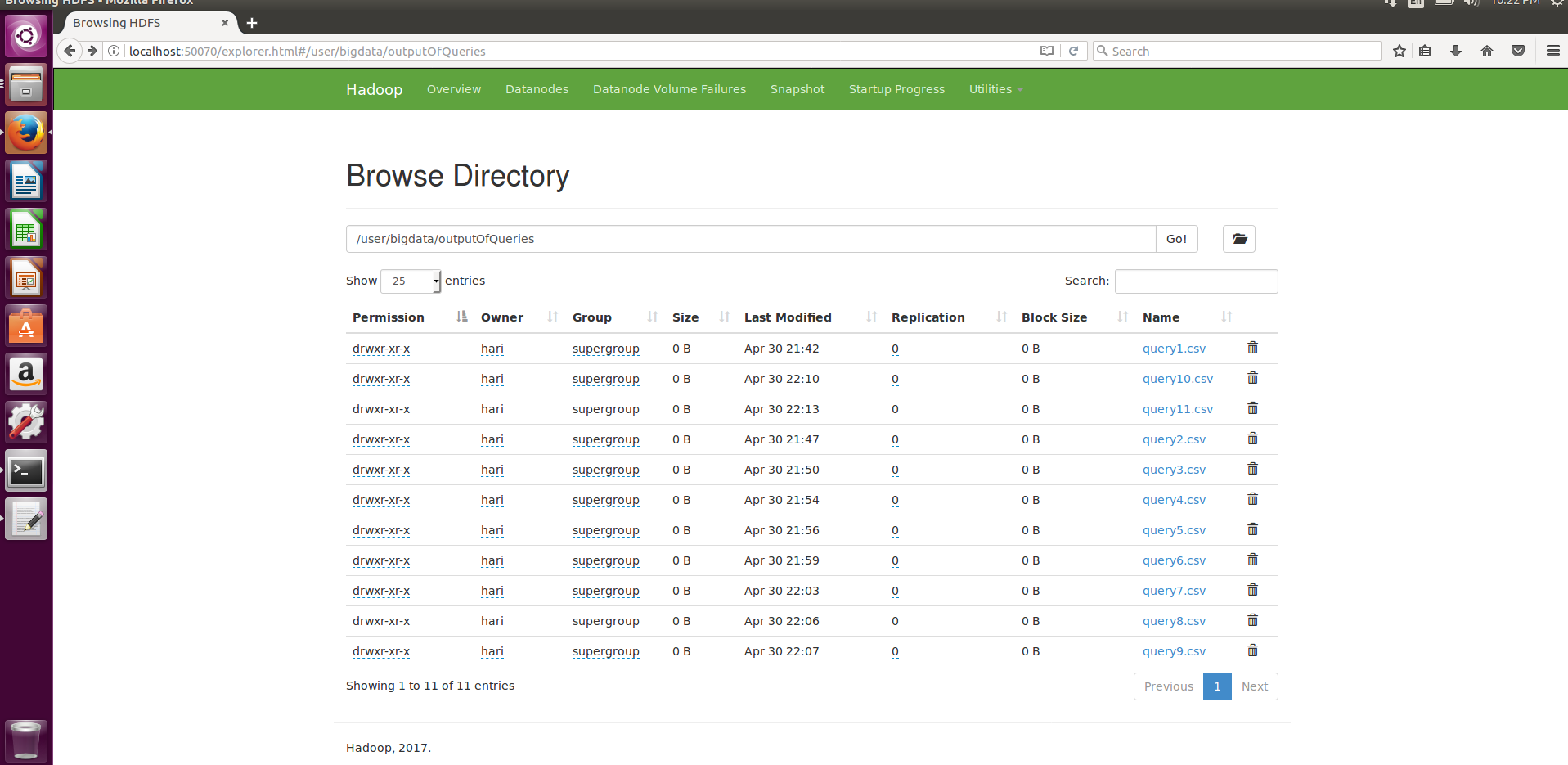
This is the uploaded json file (copyFromLocal) from which table is created



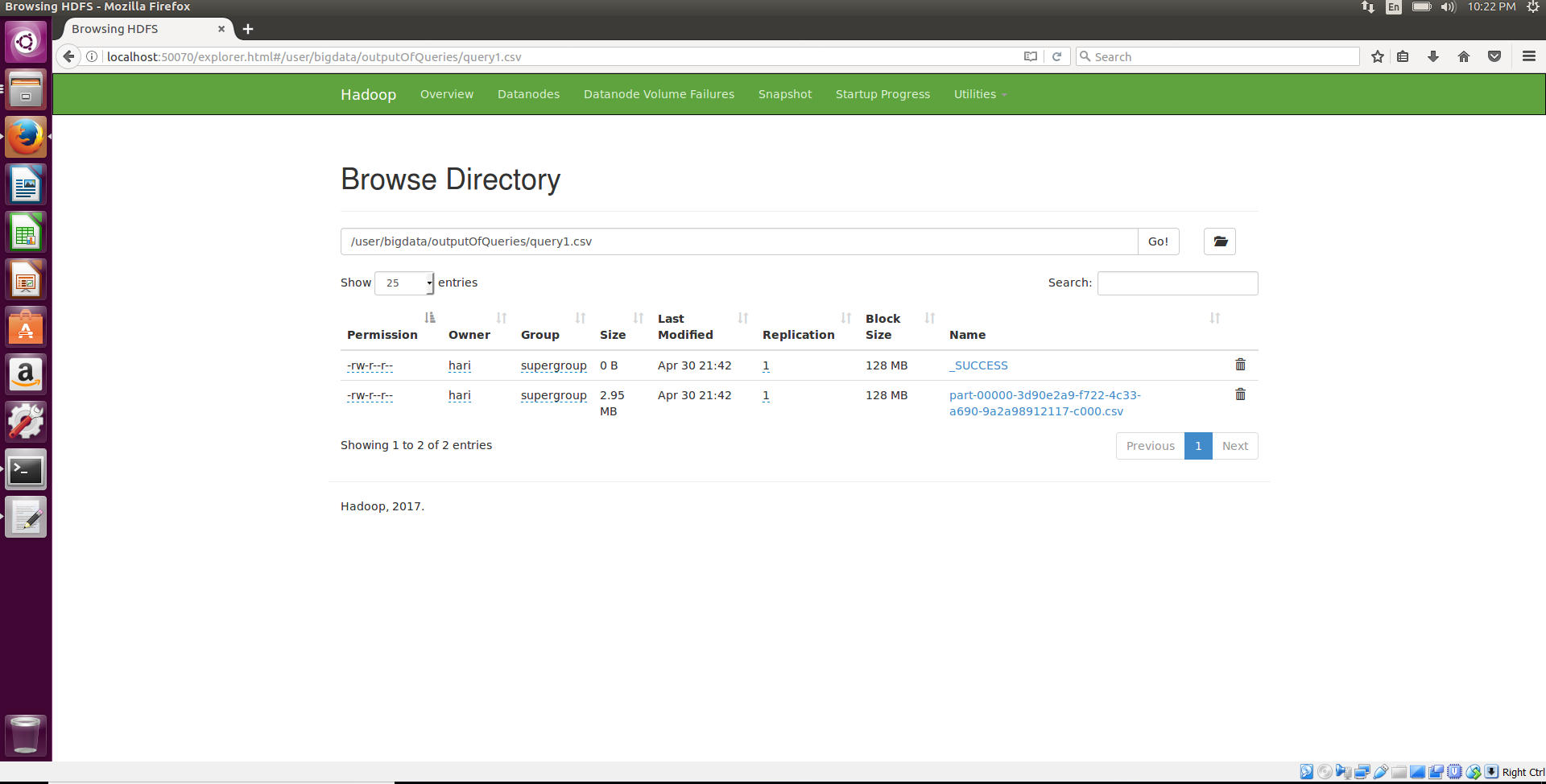
Created a directory naming outputOfQueries.



Screenshot of extracted csv files:



Sample screenshot of extracted csv file



Work has been distributed equally by all the team members

**Source code URL:**

https://drive.google.com/drive/folders/1IXl0piaaTtfonXgAeatsyeSx3V0TIBYH