Project report on Tweet analysis

By

Sri PraneethIyyapu (siqnf@mail.umkc.edu)

Pratap Rao Kadari (pkkv6@mail.umkc.edu)

Vihari Gorripati (vgxb2@mail.umkc.edu)

Submitted on 7th April, 2016

University of Missouri- Kansas City

(UMKC)

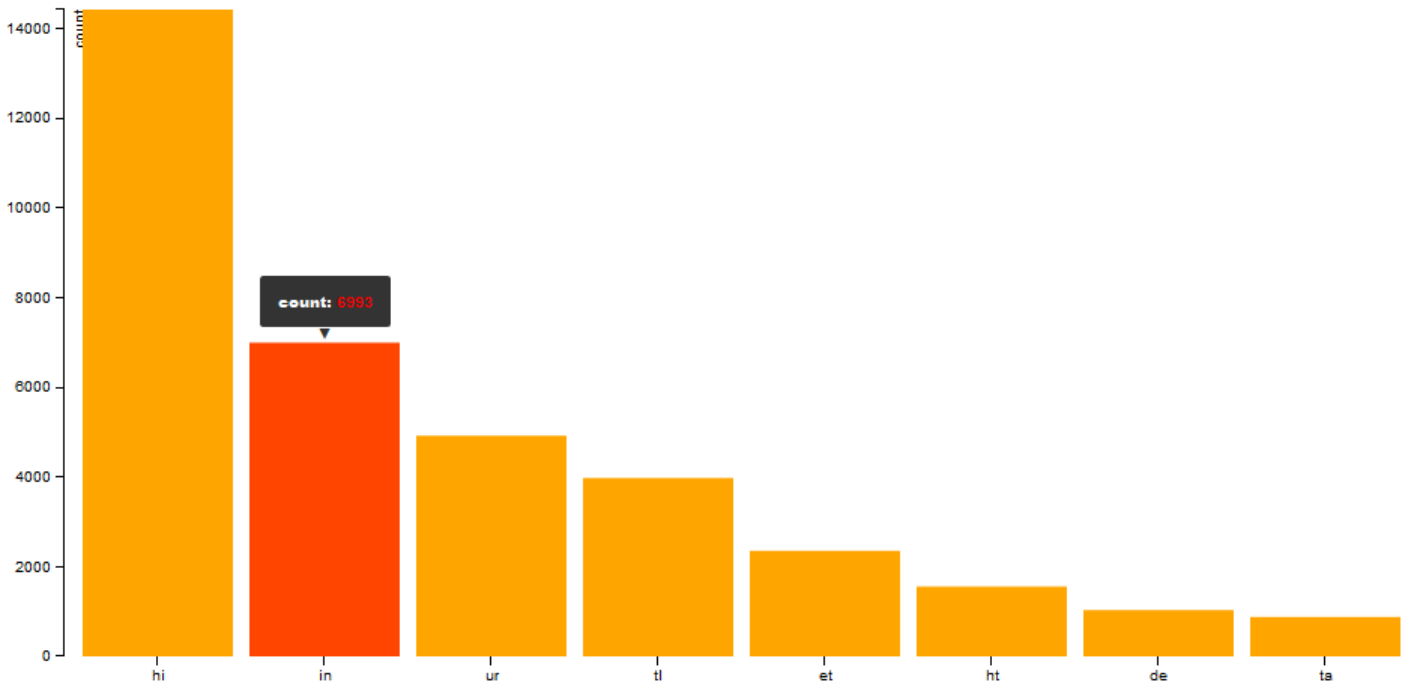
Queries developed:

Query1: Select user.time_zone AS timezone, count (user.time_zone) AS timezoneCount from tweets GROUP BY user.time_zone ORDER BY count (user.time_zone) desc limit 8

The above query analyzes all tweets and categorizes the number of users according to their time zone and displays the top 8 time zones.

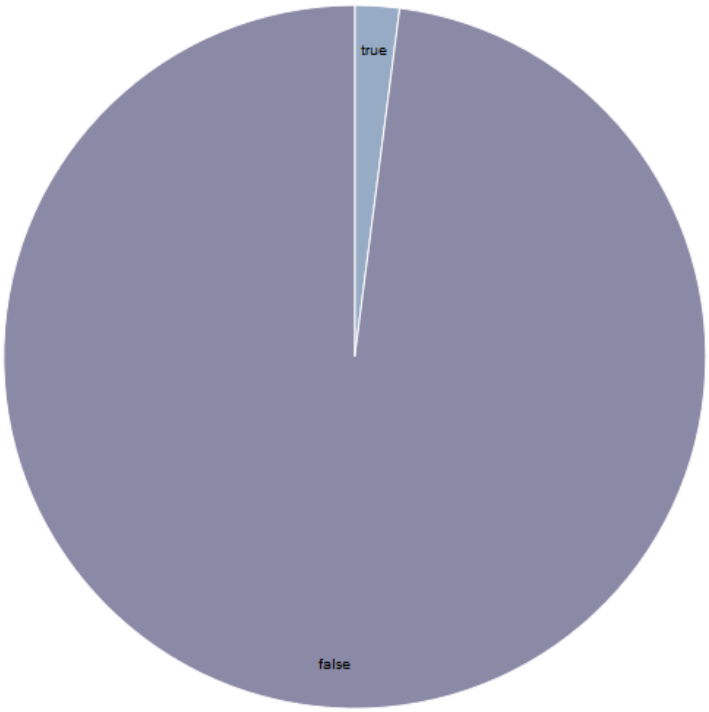
Query 2: Select lang as lang, count (*) as count from tweets group by lang having (lang <> 'en' and lang<> 'und') order by count desc limit 8

The above query displays the most tweeted languages other than English.



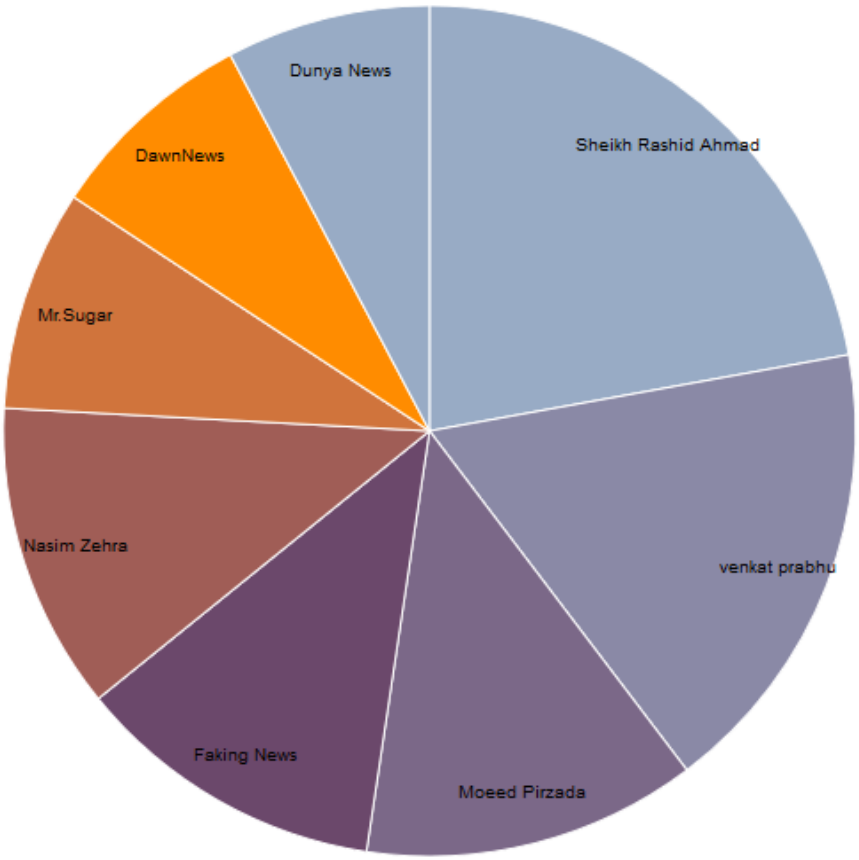
Query 3: Select possibly_sensitive as is_sensitive, count (*) as count from tweets where possibly_sensitive = false or possibly_sensitive = true group by possibly_sensitive

The above query lets us know how much percent of the tweets contain sensitive data.



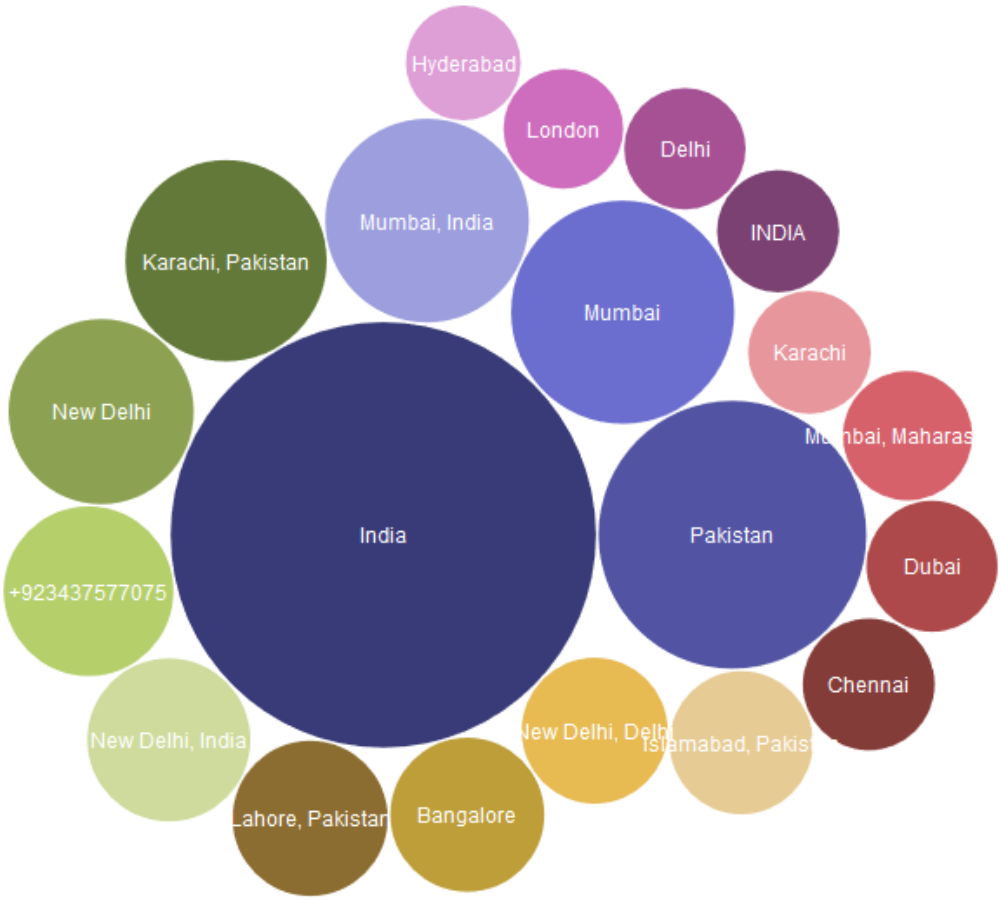
Query 4: select user.name as username, max (user.followers_count) as followers from tweets where user.verified = false group by user.name order by followers desc limit 8

The above query displays un verified users who tweeted maximum



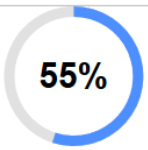
Query 5: Select user.location, count(user.id) as NumberOfUsers from tweets WHERE user.location is NOT NULL group by user.location order by count(user.id) desc limit 20

The above query displays the number of users according to their geographic locations and top 20 records are chosen.

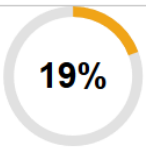


Query 6 :
"SELECT source, COUNT(*) AS total_count FROM tweets WHERE source IS NOT NULL GROUP BY source ORDER BY total_count DESC LIMIT 6"

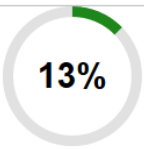
The above query displays the most used devices among the twitter users.



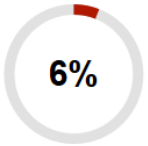
Twitter for Android



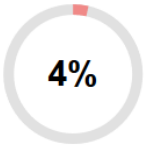
Twitter for iPhone



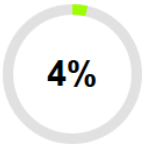
Twitter Web Client



IFTTT



Mobile Web (M2)

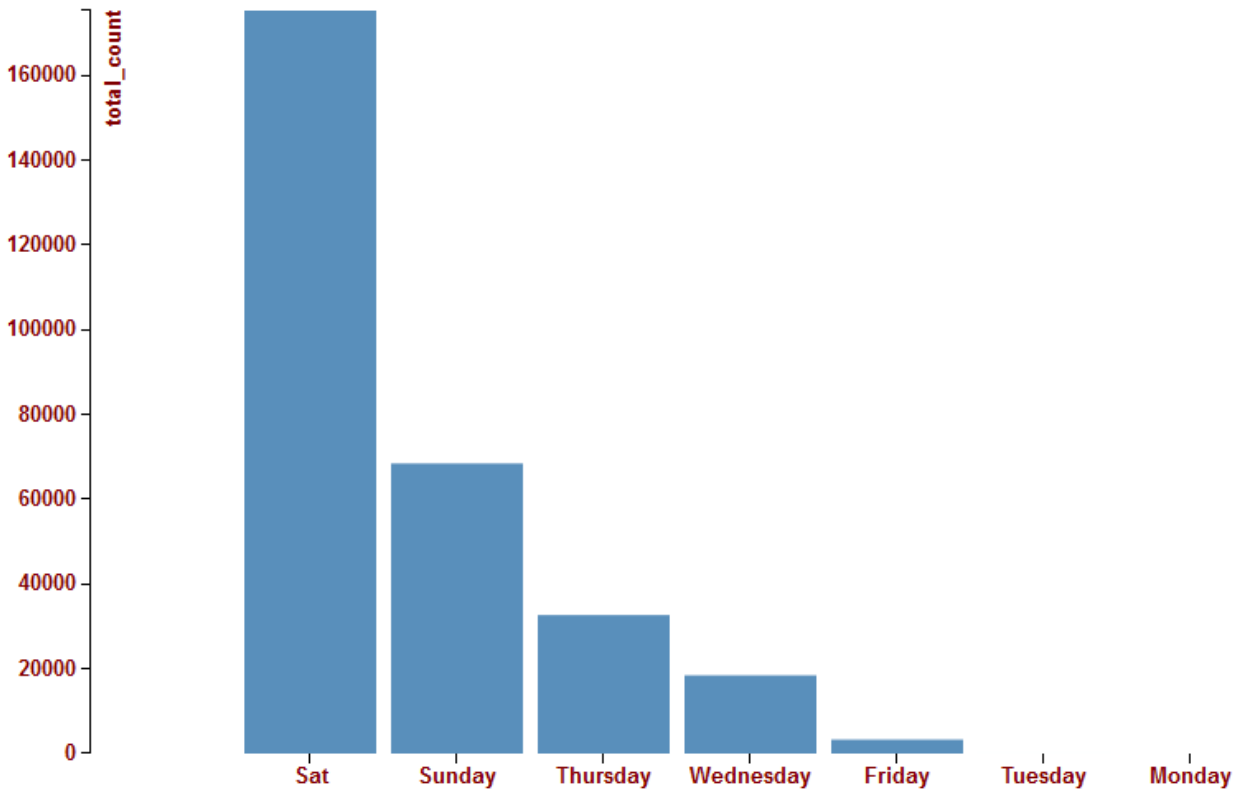


TweetDeck

Query 7:

"Select 'Saturday' AS DAY, count(*) as total_count from tweets where created_at like 'Sat%'"

The above query displays the tweets count on a particular day. We have implemented it for each day of week.


total_count
160000
140000
120000
100000
80000
60000
40000
20000
0
Sat
Sunday
Thursday
Wednesday
Friday
Tuesday
Monday

Query 8 :

SELECT entities.hashtags[0].text as popular_tags, count(*) as total_count FROM tweets group by entities.hashtags[0].text order by total_countdesc limit 8

This query displays most famous hashtags.