

Charmi Kansara

Twitter Sentiment Analysis

Abstract - Nowadays, a great volume of information is produced or collected via social networking sites. Since microblogging sites only allow for brief and straightforward expressions, 90% of individuals in the globe share their thoughts there every day. We can now more easily obtain any type of data at any time and from any location thanks to the numerous devices, mobiles, laptops, tablets, and other IoT data gadgets that generate enormous volumes of data and are powered by Microservices-based web applications. We can voice our opinions on goods and services on social media as well. It is possible to compile the reviews and ratings of millions of social media users, extract their attitudes and sentiments toward any goods or services, and utilize that data to influence the future market and business analysis or domain research. It is challenging to extract user opinions from social media; there are many different techniques to do it. Using an open-source methodology, we have collected tweets from the Twitter API and used R to pre-process, analyze, and visualize these tweets. R programming is a statistical technique that we are using to assess the sentiments of tweets. This sentiment analysis is based on text data retrieval from the streamed web and then classifying people's perspectives into three unique sentiments (positive, negative, and neutral).

INTRODUCTION

The popularity of social networking sites like Twitter is growing along with social awareness. Twitter is primarily used for exchanging information and viewpoints on popular events [1]. With 328 million monthly active members and millions of tweets published each day, Twitter is the ninth-largest social network in the world. Celebrities, politicians, people involved in the entertainment business, and athletes are the main users of Twitter. Users can tweet any content, including text, audio, and video, and other users can respond by posting their thoughts[2]. Today, there is a greater need for feedback on new goods, government programs and initiatives, executions, and international discussions so that businesses, the government, and other organizations may use this feedback to learn what the public thinks of their products and initiatives. The finest tweets and retweets are typically loud and represent knowledge on a variety of subjects and shifting mental states in an unedited and unstructured manner. Unstructured data analysis is a challenging endeavor in and of itself, and getting valuable information out of it is a major difficulty. Powerful tools and systems that can manage millions of tweets and extract emotion from them are required for this. There are numerous options for how to approach this. In this study, we do sentiment analysis using the R programming language [3]. R is an open-source program used to perform sentiment analysis and text mining on web reviews [4].

Finding and classifying the emotions expressed in a text source is the technique of sentiment analysis. Tweet analysis can yield a substantial amount of sentiment data. These numbers enable us to comprehend how people feel about a variety of situations. Sentiment analysis is a method for automatically locating and classifying subjective information in text data. This could be a stance, a judgment, or an emotion regarding a particular topic or product feature.

In other words, Sentiment analysis is the process of locating and classifying the emotions conveyed in the text source. Examining tweets often yields a significant amount of sentiment data. This information aid in understanding public opinion on a variety of topics. The most common type of sentiment analysis, polarity identification, entails classifying comments as Positive, Negative, or Neutral. The opinion is significant because we consider others' perspectives whenever we must make a decision [5]. Subjectivity analysis, sentiment mining, review mining, effect analysis, and other subtasks can all be used for sentiment analysis[4].

The proposed methodology, titled "Sentiment Analysis," aims to analyze the vast amount of data that is readily accessible through social media. The algorithm calculates a general sentiment score for the supplied topic based on whether it is favorable, negative, or neutral. In addition, it attempts to determine how frequently certain words are used. Additionally produced is a word cloud, which is a visual depiction of terms based on how frequently they appear in the text. Due to its component-rich, thorough, and expressive capabilities for quantifiable information, R was used to actually perform the calculation.[6]

PURPOSE

This project, "Twitter Sentiment Analysis Using R," is a big data analytics-based sentiment analysis project. We will be able to analyze sentiment using text data supplied by Twitter thanks to this initiative. It can convey our sentimental mood regarding a certain person, a particular subject, or anything else we like.

R Programming Language: A Simple Approach to Twitter Sentiment Analysis

R is a coding language and software system utilized for the analysis of statistical data, the representation of charts, graphs, and reporting. R language was developed by Ross Ihaka and Robert Gentleman at Auckland University New Zealand. R is available freely under a public license. The name of the programming language R was derived from the first letter of the first name of the two R developers (Robert Gentleman and Ross Ihaka).

2.1 Features of R Language

- R language is a well-developed, straightforward, and efficient programming language. It includes loops, conditionals, recursive functions, and input and output facilities.
- R has excellent storage and data handling facility.
- R provides a set of operators for vectors, arrays, and matrices.
- R provides a list of a wide collection of tools for data analysis.
- R is accessed through an interpreter based on the command line; it supports arithmetic operations that are matrix based. The data structure of R involves vectors, arrays, matrices, lists, and data frames. The extendable object scheme of R contains objects for regression models, time series, and geospatial coordinates. The scalar data type is not a data structure of R. As an alternative, a scalar is expressed as a vector that is of length one.
- Procedural programming is supported by R language with functions, and object-oriented programming with generic functions for particular functions. It is mainly utilized by

statisticians and mathematicians and needs an atmosphere for the analysis of statistical data and development of software, R language is also utilized as a toolbox for common matrix operations with performance standards similar to MATLAB or GNU octave.

2.2 Package

Performance of the R language can be enhanced through a package that is created by users and generally developed in C, C++, and java. For specific statistical methods, graphical plots (ggplots), Import/ Export abilities, reporting tools (knitr, sweave), etc. R has a core group of packages; it is provided through the installation, with more than 7,801 extra packages, these include Comprehensive R Archive Network (CRAN), Bio conductor, Omega hat, GitHub, etc.

Shiny WebApp was employed to create this project. Through the use of the R package Shiny, interactive web applications may be quickly created. Alternatively, you may embed standalone apps in R Markdown pages or create dashboards. All of these options are available. With CSS themes, HTML widgets, and JavaScript actions, you may further customize your Shiny apps.

Shiny: Graphic User Interface for the proposed work 'Sentiment Analysis' was developed utilizing the shiny package of RStudio. It is one of the strongest software supports provided by RStudio. Shiny is equipped with a lot of prominent interface-enhancing features. It's an interactive and user-friendly app development package.

In Shiny App, there are 2 types of creating GUI Based Web Applications:

1. Single-file (app.R)
2. Multi-file (ui.R, server.R)

This project is a Multi-file type, which contains 2 separate files for Front-End & Back-End Development. ui.R is for Front-End & server.R is for Back-End.

METHODOLOGY

Our Twitter sentiment analysis is based on the word we enter. Here, we've taken two files for positive and negative words. In order to analyze the sentiment associated with the word, the user needs to enter the data with "#". There is also an option for the number of recent tweets to use for analysis. It ranges between 5 to 1000 tweets. The final outcome would be generated from the data as well as the selected number of recent tweets.

The application has 7 major sections:

1. Top trending topics today
2. Word cloud
3. Histogram
4. Pie Chart
5. Table
6. Top Charts
7. Top Hashtags of User

1. Top trending topics today

- Here, we can get to see the top trending topics according to the location. Users can choose a location to extract the trending topics. It could be worldwide or based on countries.

2. Word cloud

- The most used words associated with the trending Hashtag is shown in this section as a visual representation. This style is helpful for finding a term alphabetically to establish its relative popularity as well as for rapidly identifying the most significant terms.

3. Histogram

- Histograms graphically represent the positive or negative sentiment people have toward a hashtag.
- Here, we're gonna have 3 histograms based on positive, negative, and overall scores.
 - 1) Histogram of positive sentiment
 - 2) Histogram of negative sentiment
 - 3) Histogram of score sentiment

4. Pie Chart

- An illustration of the sentiment of the hashtag is a circular statistical graphic called a pie chart, which is divided into slices. Each slice in a pie chart has an arc that is proportional to the quantity it represents, as well as an angle and area at its center. Here, the proportion of positive and negative sentiment is shown by the pie chart.

5. Table

- The table stores the tweets associated with the hashtag in tabular format.
- The scores for each sort of sentiment are displayed in this table to show the Sentiment of the Tweets (Positive, Negative, or Neutral) connected with the hashtag that was searched, and the positive as well as negative percentages.

6. Top charts

- This has been used to depict the top 20 users associated with the searched hashtag ie. who used that particular hashtag in the tweets. The table containing usernames and the number of tweets would also be depicted.

7. Top hashtags of user

- If the app user wants to check the frequencies in the tweets of the Twitter user, that can be checked in the top hashtags of the user section. To do so, a user must need to enter the Twitter handle to be analyzed.

Prerequisites

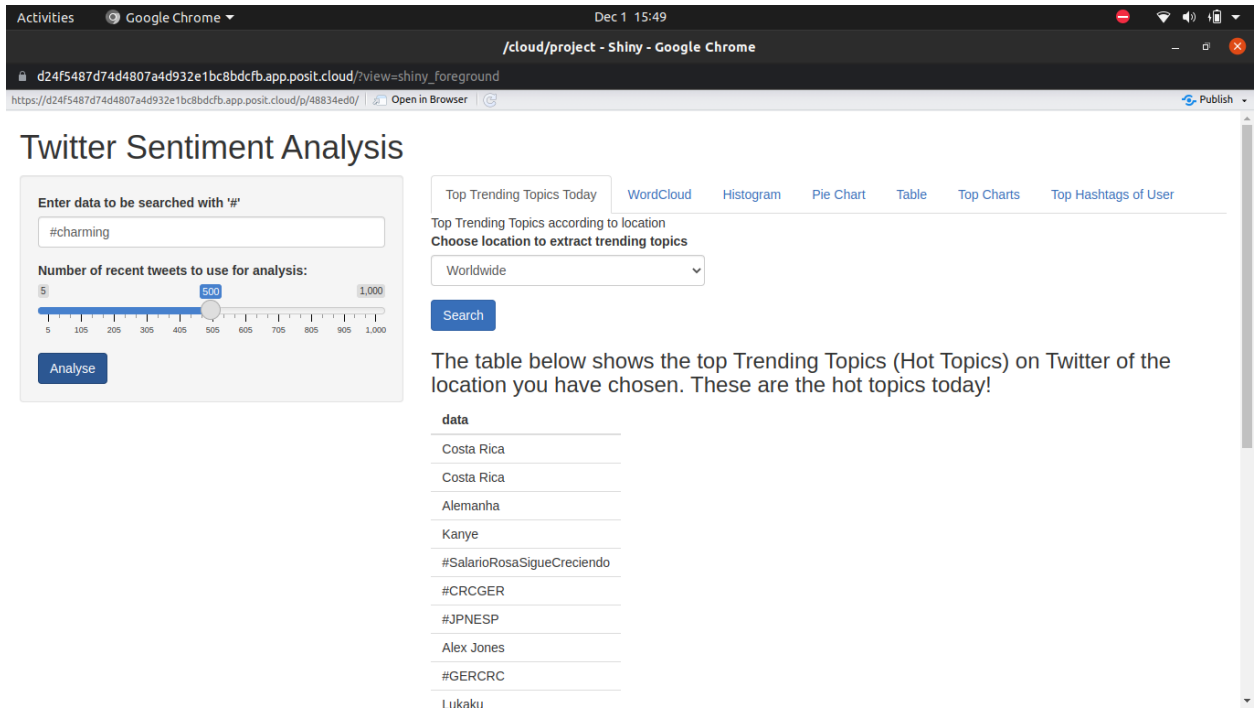
- Twitter API (Twitter development account)
- An IDE for R Programming (RStudio)
- Positive & Negative Words List (To compare Tweets for Sentiment)
- R Binary
- Shiny & Plotrix (R Libraries)

Steps to execute the application are:

- Firstly, create a Twitter development account. Once a development account request gets completed, create a Twitter app.
- Once the app creation is done, you will get Twitter API keys. (You need to save these credentials as they would be gone once you close that!)
 - API Key
 - API Key secret
 - Access Token
 - Access Token Secret
- Now, open the server.R file in Rstudio and paste the values of the API key, API key secret, access token, and access token secret in place of consumer_api_key, consumer_api_secret, access_token, and access_token_secret respectively.
- Finally, execute this script, using the "Run App" button provided in the above right-hand side section of RStudio.
- After running the script, you will be presented with an interactive dashboard for studying Twitter data sentiment analysis.
- The dashboard below, which is interactive, will appear once you run the ui.R/server.R script.

Few insights on the dashboard:

1. Search input & Location Selection (By default location is set to Worldwide)



Twitter Sentiment Analysis

Enter data to be searched with '#'

#charming

Number of recent tweets to use for analysis:

5 500 1,000

Analyse

Top Trending Topics Today WordCloud Histogram Pie Chart Table Top Charts Top Hashtags of User

Top Trending Topics according to location

Choose location to extract trending topics

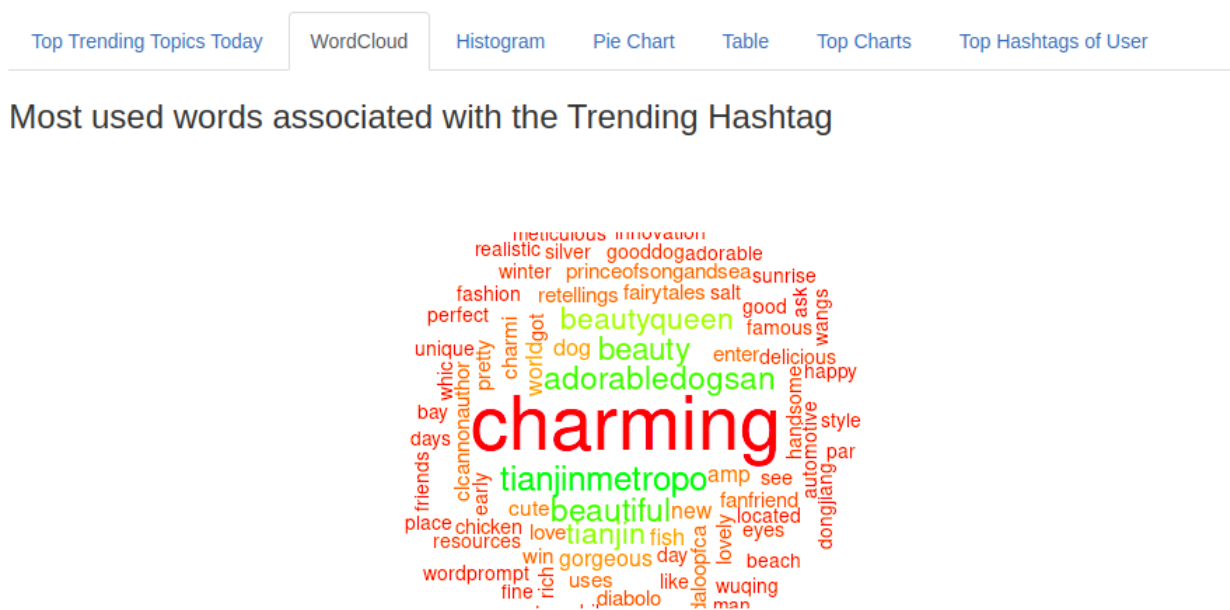
Worldwide

Search

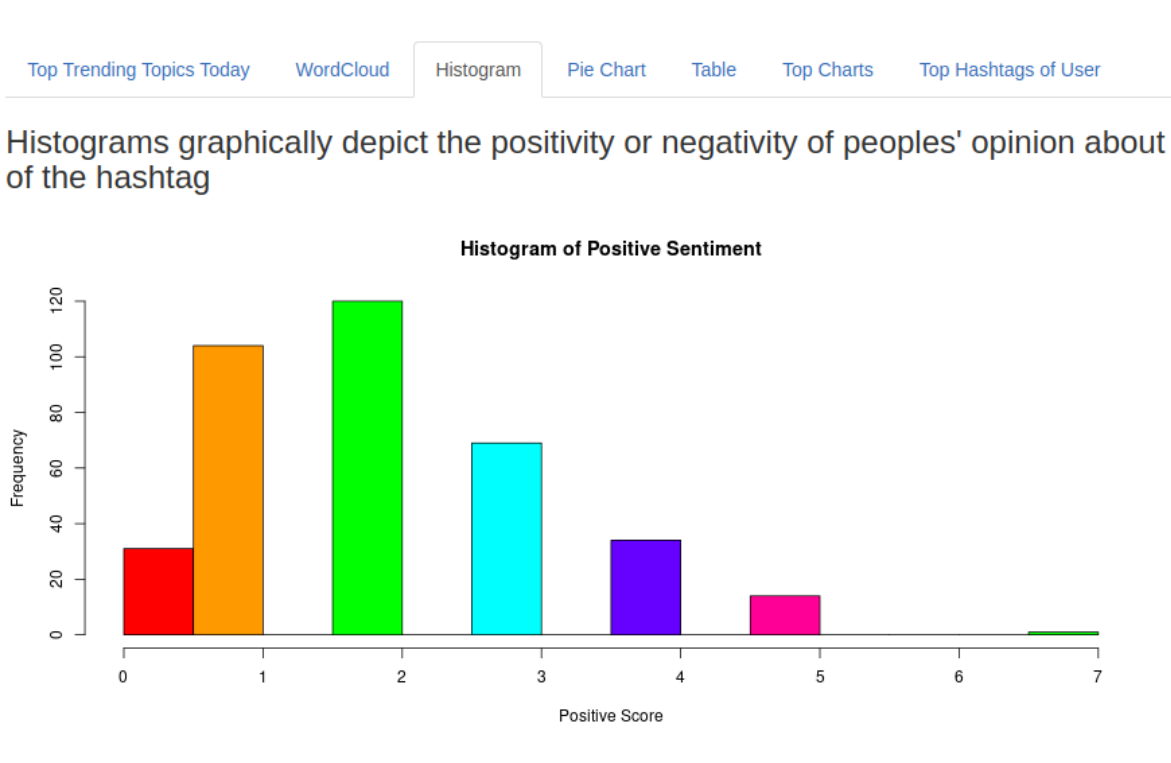
The table below shows the top Trending Topics (Hot Topics) on Twitter of the location you have chosen. These are the hot topics today!

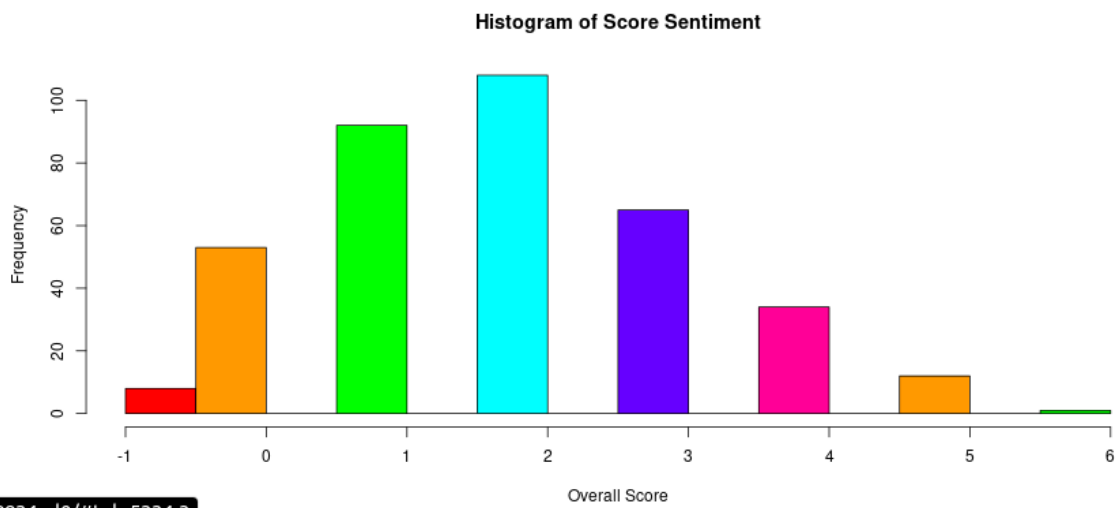
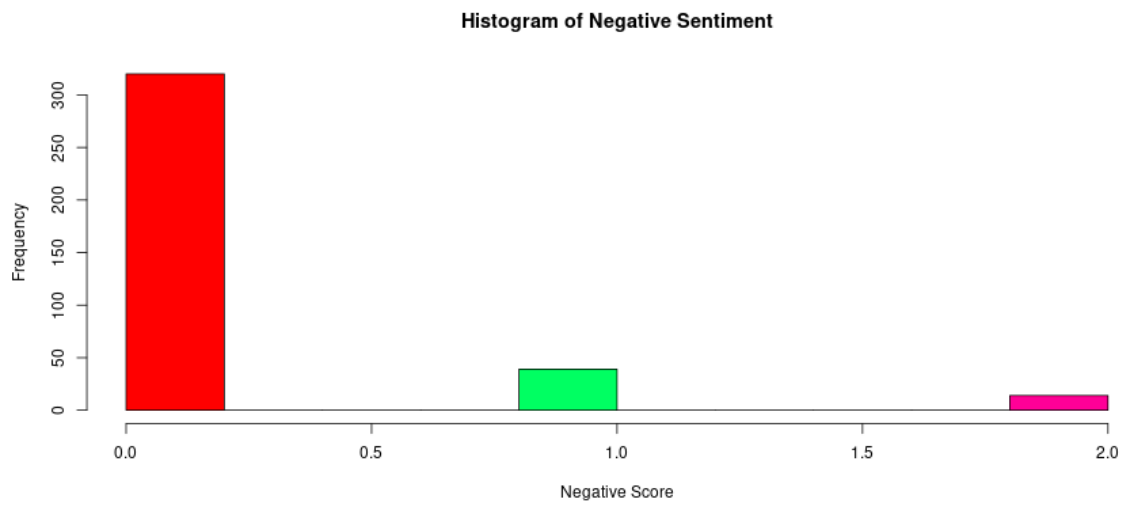
data
Costa Rica
Costa Rica
Alemanha
Kanye
#SalarioRosaSigueCreciendo
#CRCGER
#JPNESP
Alex Jones
#GERCRC
I ukaku

2. Word Cloud Formation



3. Plotting Histogram of Sentiments





40024-10/#5-b-5224-2

4. Pie Chart for Sentiment Percentage

[Top Trending Topics Today](#)

[WordCloud](#)

[Histogram](#)

[Pie Chart](#)

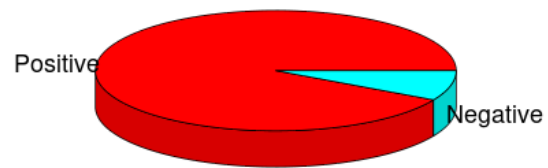
[Table](#)

[Top Charts](#)

[Top Hashtags of User](#)

Pie Chart

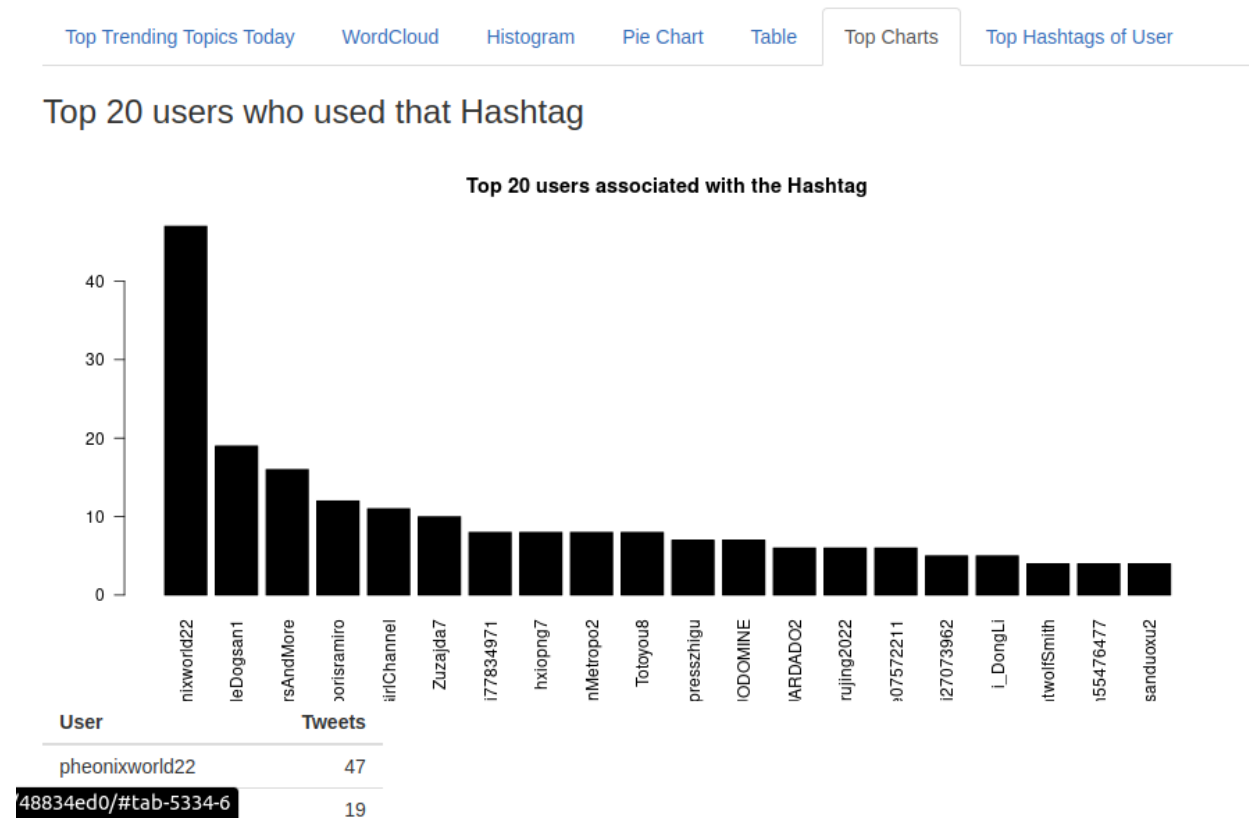
Sentiment Analysis



5. Table of Tweets with Analysis Related to the Searched Hashtag

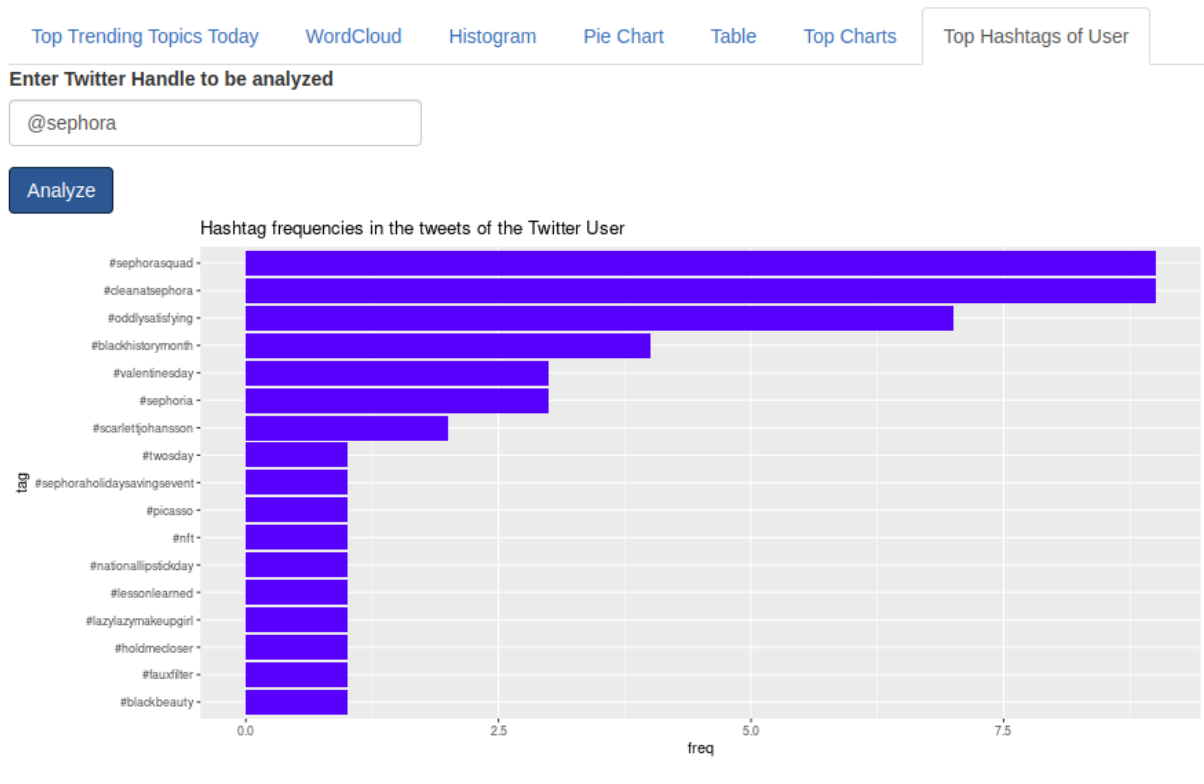
Top Trending Topics Today WordCloud Histogram Pie Chart Table Top Charts Top Hashtags of User					
Storing the Tweets associated with the Hashtag in Tabular Format					
Text	Positive	Negative	Score	PosPercent	NegPercent
RT @AdorableDogsan1: Henlo, we are here to steal the internet [by wendyviola] #love #charming /NFCLUqI3Fd	2	1	1	66.67	33.33
RT @MonitorsAndMore: Cashier at my local Ace Hardware [by haddiejustice7979] #love #charming /Pnz0mzhNna	2	0	2	100.00	0.00
Sydney Sweeney #SydneySweeney #SydneySweeneyHot #SydneySweeneyFans #SydneySweeneyLovers #beauty #beautiful /AEyNSQw5zj	2	0	2	100.00	0.00
Henlo, we are here to steal the internet [by wendyviola] #love #charming /NFCLUqI3Fd	2	1	1	66.67	33.33
Cashier at my local Ace Hardware [by haddiejustice7979] #love #charming /Pnz0mzhNna	2	0	2	100.00	0.00
Do you have a character good at charming people in your group? #charm #charming #potion #potions /JWaMgKKnKj	4	0	4	100.00	0.00
Margot Robbie #MargotRobbie #MargotRobbieHot #MargotFans #MargotLovers #beauty #beautiful #beautyqueen /6TRzk8nOsL	2	0	2	100.00	0.00
See more: /rAh9RulsI8	0	0	0	0.00	0.00
RT @TianjinMetropo2: #Charming #Traditionalculture #Tianjin	3	0	3	100.00	0.00

6. Top 20 users of that hashtag in a bar graph



User	Tweets
pheonixworld22	47
AdorableDogsan1	19
MonitorsAndMore	16
borisramiro	12
NiceGirlChannel	11
Zuzajda7	10
fengwei77834971	8
hxiopng7	8
TianjinMetropo2	8
Totoyou8	8
impresszhigu	7
QUODOMINE	7
MARCOAGUARDADO2	6
ruchirujing2022	6
zhangle07572211	6
hexi27073962	5
i_DongLi	5
NightwolfSmith	4
Sachin55476477	4

7. Examining Twitter handles (Hashtag Frequencies by a particular User)



REFERENCES

1. K. Arun, A. Srinagesh, and M. Ramesh, "Twitter Sentiment Analysis on Demonetization an Initiative Government of India," *Int. J. Recent Trends Eng. Res.*, vol. 4, no. 6, pp. 252–258, 2017.
2. D. Suganthi and A. Geetha, "Twitter Sentiment Analysis on GST tweets using R tool," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 2, no. 5, pp. 793–796, 2017.
3. Syed Tanzeel, and Qamar Rayees "Quantifying Suicidal Ideation on Social media" [Unpublished]
4. R. Wagh and P. Punde, "Survey on Sentiment Analysis using Twitter Dataset," *Proc. 2nd Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2018*, no. Iceca, pp. 208–211, 2018.
5. M. Edison and A. Aloysius, "Concepts and Methods of Sentiment Analysis," pp. 16288–16296, 2016.
6. Shruti Kohli, HimaniSingal, "Data Analysis with R", 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing