## A

## INTERNSHIP REPORT

**on**

**TWITTER SENTIMENT ANALYSIS**

**Submitted for INTERNSHIP-I of**

**BACHELOR OF ENGINEERING**

in

**ARTIFICIAL INTELLIGENCE AND DATA SCIENCE**

by

**G. Sai Akshitha (1601-21-771-008) &**

**M. Tharshith (1601-21-771-048)**

**Under the guidance of**

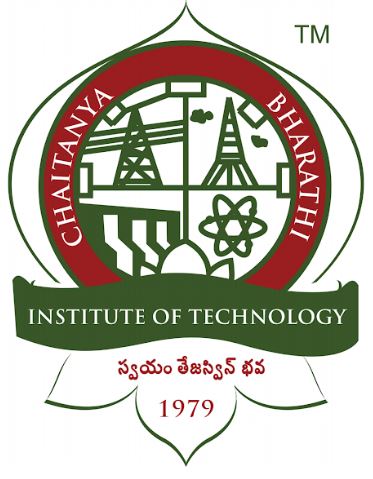
**Dr. Ramu Kuchipudi**

Associate Professor

&

### Mrs. K.H. Vijaya Kumari

Assistant Professor

****

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (A)**

**(Affiliated to Osmania University; Accredited by NBA(AICTE) and NAAC(UGC), ISO**

**Certified 9001:2015), Kokapet (V), GANDIPET(M), HYDERABAD – 500 075**

**Website:** [**www.cbit.ac.in**](http://www.cbit.ac.in)

**2022-2023**



**CERTIFICATE**

This is to certify that the project seminar report entitled **“Twitter Sentiment analysis”** submitted to **Chaitanya Bharathi Institute of Technology**, in partial fulfilment of the requirements for the award of degree of **B.E** (Artificial Intelligence and Data Science) during the academic year 2022-23 is a record of original work carried out by **G. Sai Akshitha (1601-21-771-008) and M. Tharshith (1601-21-771-048)** during the period of study of 2021-25 in the Dept. of IT, CBIT, Hyderabad.

**Technical guidance**

Saketh kallepu

CMO & Data science Mentor

Codegnan IT Solutions (OPC) Pvt Ltd.

**Mentor**  **Head of the Department**

Dr. Ramu Kuchipudi  Dr. K. Radhika

Associate Professor, Professor, IT Dept CBIT

IT Dept CBIT

**ACKNOWLEDGEMENT**

It is our privilege to acknowledge with deep sense of gratitude and devotion for keen personal interest and invaluable guidance rendered by our Project Guide **Sri Saketh Kallepu** .

Our respects and regards to **Dr. K. Radhika**, Professor, Department of Information Technology, Chaitanya Bharathi Institute of Technology, for his invaluable suggestions that helped us in successful completion of the project.

We are grateful to our Principal **Dr. P. Ravinder Reddy**. Chaitanya Bharathi Institute of Technology, for his cooperation and encouragement.

Finally, we also thank all the faculty of Dept. of IT, CBIT, our friends, and all our family members who with their valuable suggestions and support, directly or indirectly helped us in completing this project work.

**ABSTRACT**

**TWITTER SENTIMENT ANALYSIS**

Social media has become the core platform for people to share their opinions, ideologies and thoughts around the globe. Hence it becomes a crucial role for each platform to safeguard the public data and maintain proper tone of communication. With the rise of usage of internet they are various aspects like political, financial, emotional, etc being discussed on such platforms and here comes the need for the authorities to keep a check on the data. We have tried to scrape live twitter data using Tweepy for accessing the API. In this project we first extract the twitter data (web scrapping), identify the negative tweets by matching it with the data set and keep a count on them. This count will help us identify the types of opinions. This data is used to visually depict the various types of people on a platform. Twitter sentiment analysis allows you to keep track of what's being said about your product or service on social media, and can help you detect angry customers or negative mentions before they escalate. At the same time, Twitter sentiment analysis can provide valuable insights that drive decisions regarding further product upgrades, marketing and presentation methods. Sentiment analysis would help to secure the integrity of opinion of social media users by detecting and working on the unwanted negative opinions on platforms like Twitter. This could make the platforms become more wanted and attractive to the users where genuine opinions are respected and supported.

.

**TABLE OF CONTENTS**

***S. No Topics page no***

*Acknowledgements 3*

*Abstract 4*

*Table of Contents 5*

*List of figures**6*

1. **INTRODUCTION 7**
   1. *Overview 7*
   2. *Motivation 7*
   3. *Applications 8*
   4. *Problem Statement 9*
   5. *Aim of the project 9*
2. **SYSTEM REQUIREMENT SPECIFICATION 10**

*2.1 Functional Requirements 10*

*2.2 Non-Functional Requirements 10*

*2.3* Software Requirements *11*

*2.4 Hardware Requirements 11*

1. **PROPOSED SYSTEM 11**
   1. *Proposed work*  *11*
   2. *Algorithm 12*
   3. *Flowchart 13*
2. **IMPLEMENTATION 14**
3. **RESULTS 21**
4. **CONCLUSION 25**
5. **CONTRIBUTIONS 26**
6. **BIBLIOGRAPHY 27**

**LIST OF FIGURES**

3.3 Flowchart 10

4.1-4.11 Implementation snippets 14

5.1-5.9 Results 21

1. **INTRODUCTION**

**1.1 OVERVIEW**

**What is Sentiment analysis?**

**Sentiment analysis** is a task of natural language processing. Sentiment analysis, also referred to as opinion mining, is an approach to natural language processing **(**[**NLP**](https://www.techtarget.com/searchbusinessanalytics/definition/natural-language-processing-NLP)**)** that identifies the emotional tone behind a body of text. This is a popular way for organizations to determine and categorize opinions about a product, service, or idea. Sentiment analysis systems help organizations gather insights from unorganized and [unstructured text](https://www.techtarget.com/searchbusinessanalytics/definition/unstructured-text) that comes from online sources such as emails, blog posts, support tickets, web chats, social media channels, forums and comments.

In our project we have extracted live data from Twitter platform and have classified the tweet data as negative or positive or neutral.

**1.2 MOTIVATION**

Twitter is a platform where people around the globe express their opinion on various topics. Our project analyses the tweet texts and determines the type of opinion people have on it. For product-based companies this analysis will allow to keep track of what's being said about your product or service on social media which can provide valuable insights that drive decisions regarding further product upgrades, marketing and presentation methods. This is an attempt to examine the types of people based on their opinions.

**1.3 APPLICATIONS**

Twitter data sentiment analysis is the process of using natural language processing (NLP) techniques to automatically identify and extract opinions, emotions, and attitudes expressed in tweets. Here are some common applications of Twitter data sentiment analysis:

Brand monitoring: Companies can use sentiment analysis to monitor the sentiment of their brand or product on Twitter. This can help them identify areas where they need to improve, and respond quickly to any negative feedback.

Crisis management: Twitter data sentiment analysis can help companies and organizations detect early warning signs of a crisis, such as a product recall, and respond quickly to address the situation.

Public opinion analysis: Twitter data sentiment analysis can help politicians and policymakers understand public opinion on various issues, and adjust their policies accordingly.

Market research: Twitter data sentiment analysis can provide valuable insights into consumer preferences and opinions, helping businesses make data-driven decisions about product development and marketing strategies.

Customer service: Companies can use sentiment analysis to monitor Twitter for complaints and feedback about their products or services, and respond quickly to address any issues.

Trend analysis: Twitter data sentiment analysis can help identify emerging trends in a particular industry or topic, providing businesses with insights into new opportunities and potential threats.

Overall, Twitter data sentiment analysis can be a valuable tool for businesses, organizations, and individuals to gain insights into public opinion and sentiment, and make data-driven decisions.

**1.4 PROBLEM STATEMENT**

Twitter is one of those social media platforms where people are free to share their opinions on any topic. Sometimes we see a strong discussion on Twitter about someone’s opinion that sometimes results in a collection of negative tweets. Sentiment analysis is a task of natural language processing. All social media platforms should monitor the sentiments of those engaged in a discussion. We mostly see negative opinions on Twitter when the discussion is political. So, each platform should continue to analyse the sentiments to find the type of people who are spreading hate and negativity on their platform.

**1.5 AIM OF PROJECT**

Twitter sentiment analysis project is based on the study of the various tweets by people on twitter platform. Overall, the aim of a Twitter sentiment analysis project is to gain insights into public opinion and sentiment on a particular topic, brand, or product, and to use these insights to inform business decisions or to generate new knowledge. Some common objectives for a Twitter sentiment analysis project could be: Identifying overall sentiment, Comparing sentiment across different groups, Monitoring sentiment over time, Identifying key themes and topics, Predicting sentiment. With the extracted data we would like to perform data analysis using python and visualize the data through plots.

**2. SYSTEM REQUIREMENT SPECIFICATION**

**2.1 Functional Requirements**

Here are some functional requirements for a Twitter data sentiment analysis project:

* Dataset: New live dataset is loaded every time we run the code.
* Tweepy package: With the help of twitter access keys and tokens we are able to access twitter live data.
* Re & emoji module: For cleaning the tweet data after extraction.
* Nltk and Sentiment Intensity Analyzer : To get the polarity scores of the tweet data.
* Plotly : For visualizing the data through graphs.

**2.2 Non-Functional Requirements**

Create developer account to access keys and token through which we get permission to access twitter live data.

**2.3 Software Requirements**

This project code is written in python. Python is installed in the system use Jupiter notebook to run the files. Linux- Ubuntu 16.04 to 17.10, or Windows 7 to 10, with 2GB RAM (4GB preferable).

**2.4 Hardware Requirements**

* x86 64-bit CPU (Intel/AMD Architecture)
* 4GB RAM
* 5GB Free Space

**3. PROPOSED SYSTEM**

**3.1 Proposed work**

A Twitter sentiment analysis is **the process of determining the emotional tone behind a series of words, specifically on Twitter**. A sentiment analysis tool is an automated technique that extracts meaningful customer information related to their attitudes, emotions, and opinions. In our project we have used tweepy module to extract the information and trained our code such that it is able to firstly organize the tweet data extracted and detects whether the tweet exclaims a positive or negative or neutral opinion. We have used nltk and Sentiment Intensity Analyzer to do the sentiment detection process.

VADER's SentimentIntensityAnalyzer() takes in a string and returns a dictionary of scores in each of four categories: negative, neutral, positive, compound score.

We have recorded these results and have presented a visual representation about the discussions being held on the platform, the opinions of people and their supporters.

For the visual representation we have used plotly and wordcloud.

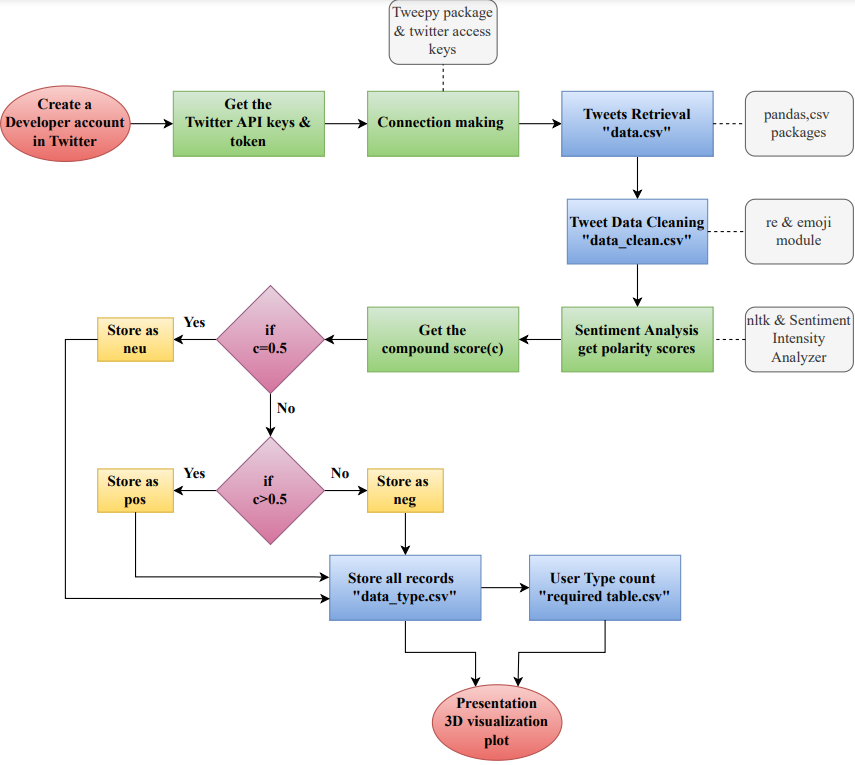
We have chosen Python as our programming language for analysing and visualizing our observations in this project.

In our project we have searched words related to “INDIA” and have performed sentiment analysis on these tweets.

**3.2 Algorithm:**

* Create a developer account and collect the access keys and tokens.
* Import Tweepy package to access and extract Twitter live data and csv package and pandas for storing the extracted data inti csv files..
* Create a new file “data.csv” and create the required columns in the csv file to store the data under each category like location, source, follower count, etc for each tweet we have extracted.
* Using the access keys and tokens and with the counter limit set to 100 start the extraction process with the search word provided. Extract this decoded information and write it into csv file.
* Import re and emoji module for the cleaning of data by removing the emojis, urls, \n, @, etc. and write the cleaned data into another file “data\_clean.csv”. Fill the empty values with nan.
* Using nltk and Sentiment Intensity Analyzer perform sentiment analysis on tweet data column to get the polarity scores and based on the compound score categorize it into positive or negative or neutral under another separate column and save it into another file “data\_type.csv”.
* For analyzing the based on the count of type of user create another file “required table.csv” to store the counts of each type.
* Plot various graphs using plotly from the data in the files “data\_type.csv” and “required table.csv”.

**3.3 Flowchart:**

Representing the project flow and tools used at every step.****

*Fig:3.3 Flowchart depecting the workflow*

**4. IMPLEMENTION**

1. **Accessing Twitter API keys:**

To access the Twitter API keys we have created a developer account and have generated the access keys, tokens and API keys, tokens.



*Fig:4.1*

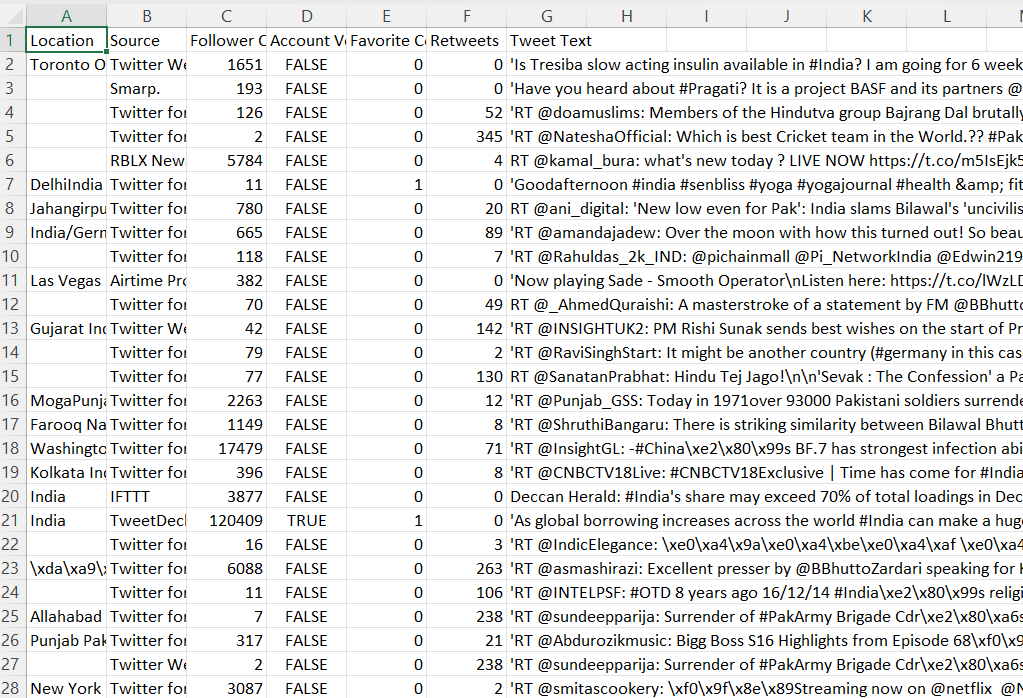
1. **Extraction of data:**

* Install required packages like os, tweepy, pandas, csv.
* Make a connection by using tweepy package and twitter acess keys.
* We have extracted 100 tweets per search and create a dataframe from this extracted twitter data.

We have stored this extracted data in **‘data.csv’** file.



*Fig:4.2*



*Fig:4.3* TABLE: data.csv

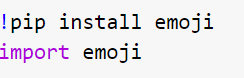
1. **Cleaning of data:**

The various problems encountered in this process is that the tweet data consists of some **emojis, urls, \n, @** which have to be cleaned. We have also experienced that some of the data after extraction also remains encoded. We have used **re module** and **emoji module** for such cleaning process by using **re.sub** method.

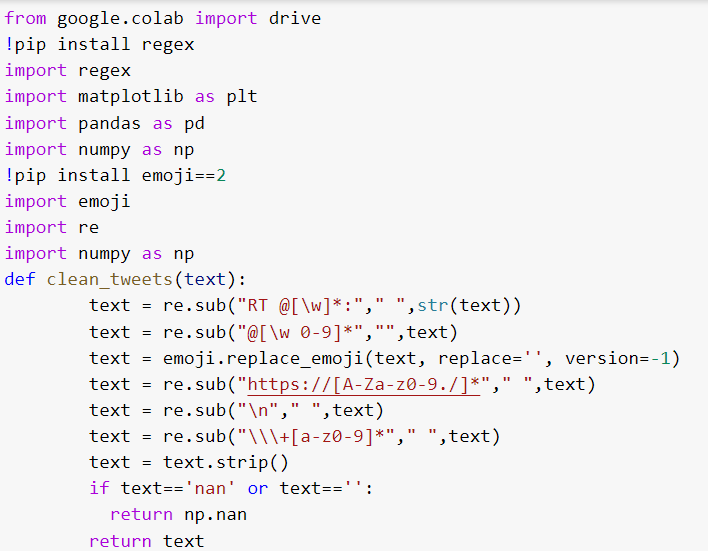
We have made a function clean\_tweets to remove:

* Old style retweet ‘RT’.
* Hyperlinks.
* Words started with @ and replace with space.
* Emojis.
* Some encoded data.
* \n and replace with space.

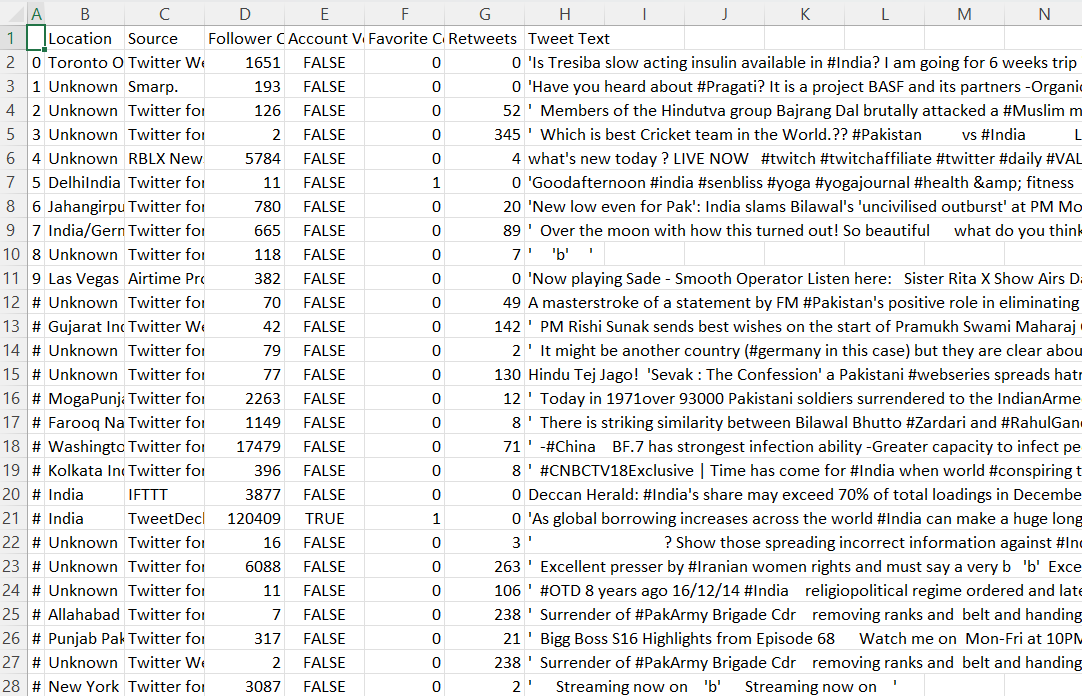
We have stored this cleaned data in **‘data\_clean.csv’** file.



*Fig:4.4*



*Fig:4.5*

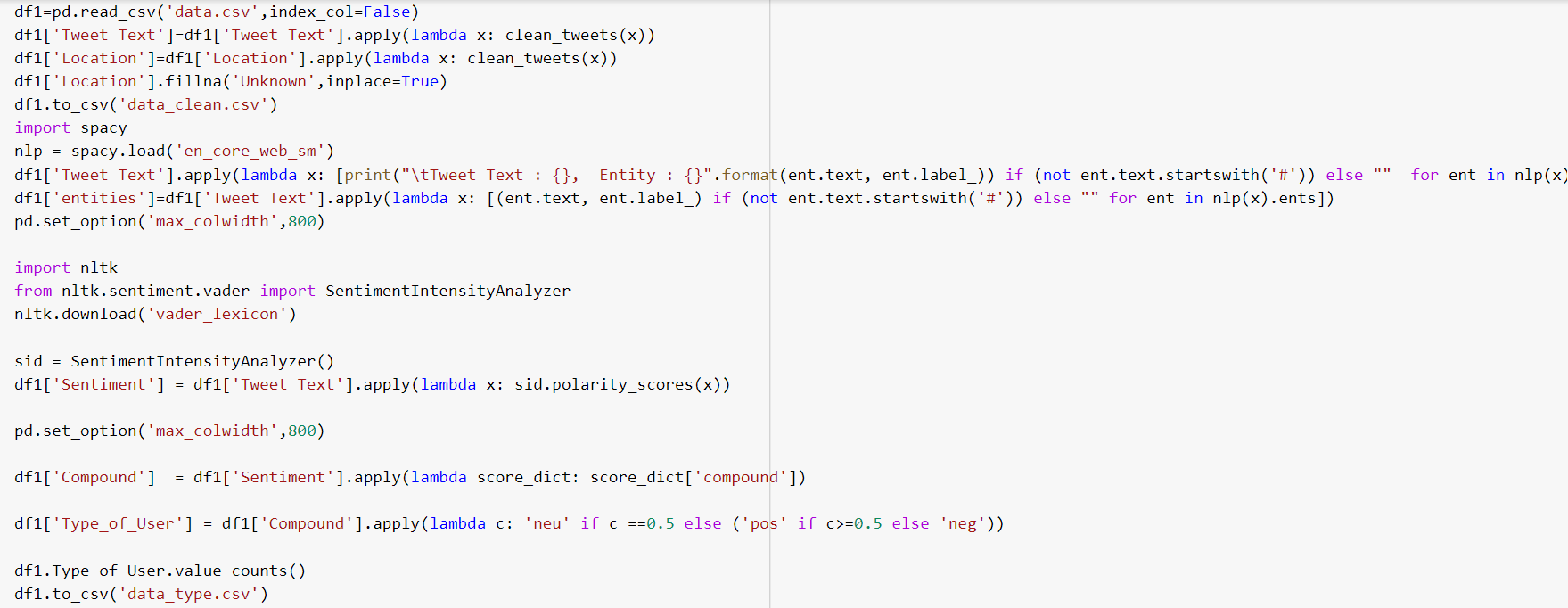


*Fig:4.6* TABLE: data\_clean

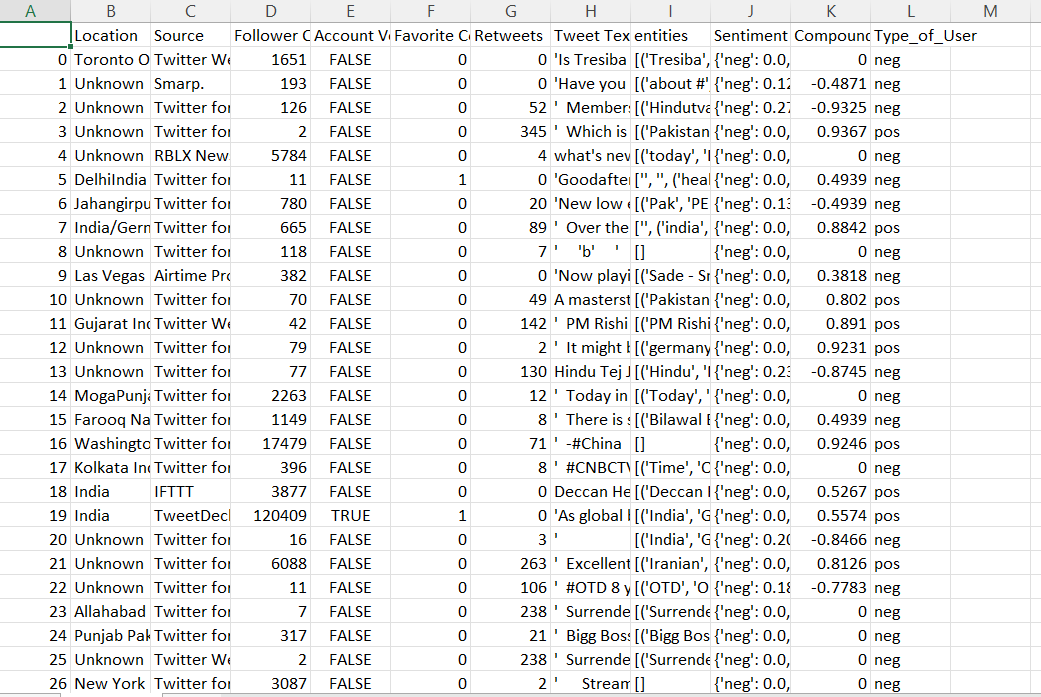
1. **Sentiment analysis using NLP package:**

* Import nltk and Sentiment Intensity Analyzer.
* Apply lambda function on cleaned text column to extract polarity scores in new column called sentiment.
* Create new column to get compound column from sentiment column
* After getting the compound score we have segregated the sentiment of the text by considering compound score equals to 0.5 as neutral, compound score greater than 0.5 as positive and compound score less than 0.5 as negative.
* Created an another new column ‘Type\_of\_User’ to store the neu, pos, neg values.

We have stored this data consisting of type of users description in **‘data\_type.csv’** file.



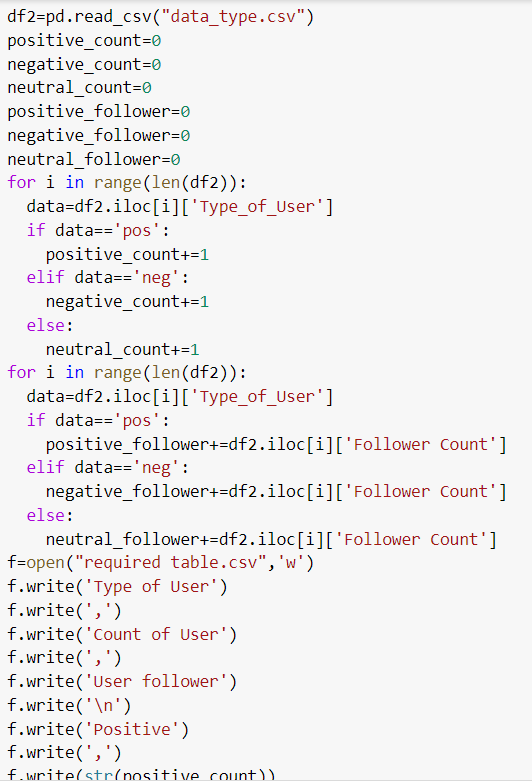
*Fig:4.7*



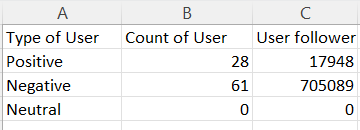
*Fig:4.8* TABLE: data\_type

1. **Counting the types of people:**

After creating the ‘**data\_type.csv’** file we have created another file which would consists of the count corresponding to each type of user called as **‘required table.csv’** file.



*Fig:4.9*



*Fig:4.10* TABLE: required table

1. **Plotting the graphs:**

We have used the files ‘**data\_type.csv’** and **‘required table.csv’** to plot the graphs for better visualization and analyzation of data.

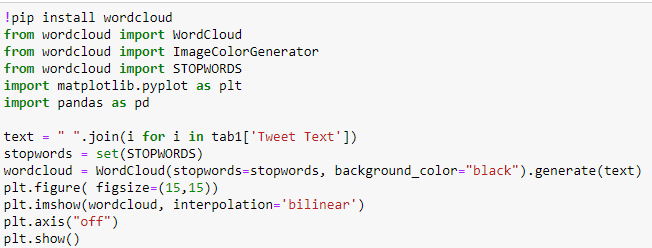


*Fig:4.11*

**5. RESULTS**

**Graphs:**

1. **Word Cloud:**

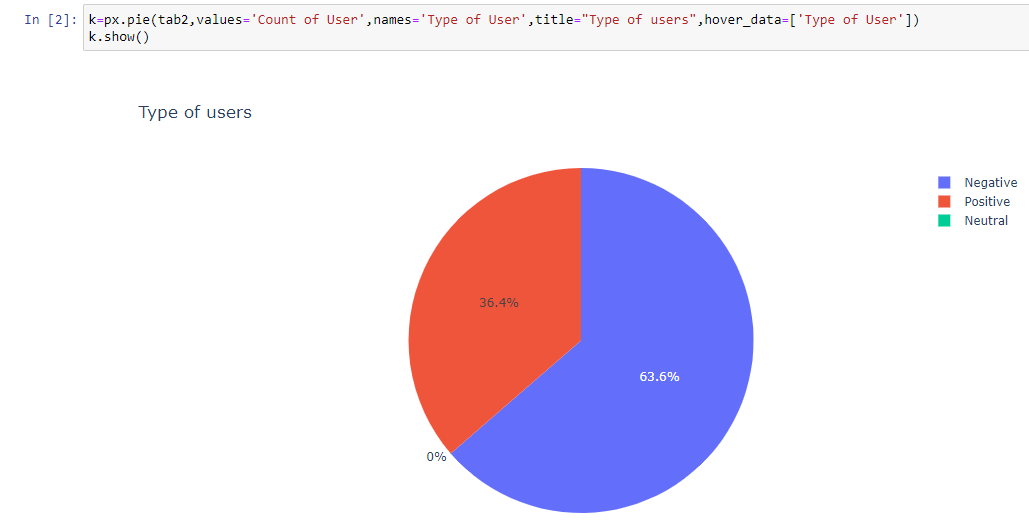


*Fig:5.1*

****

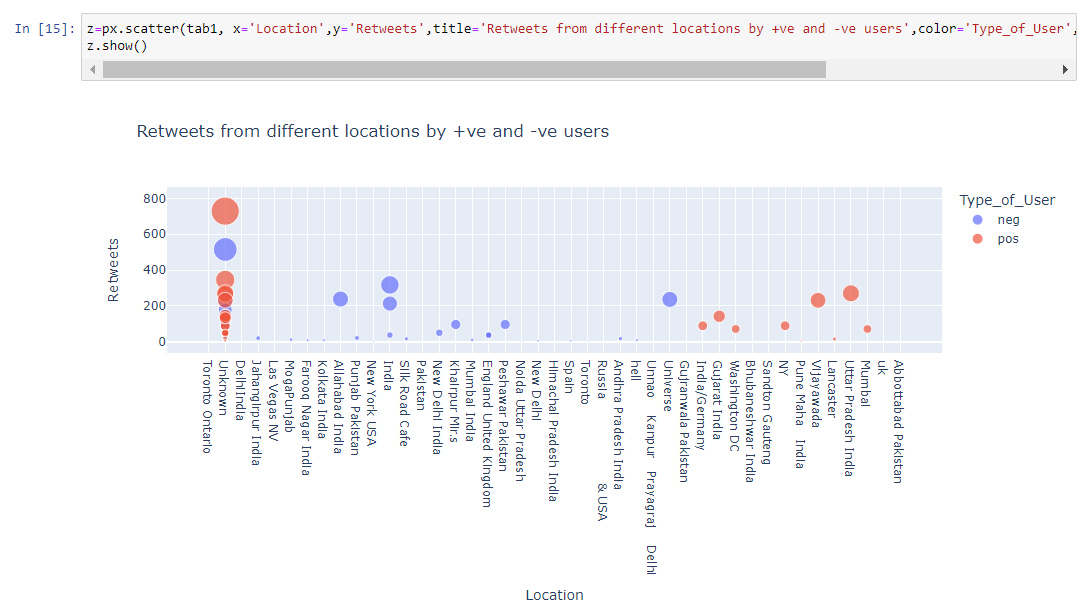
*Fig:5.2*

1. **Pie chart**:

****

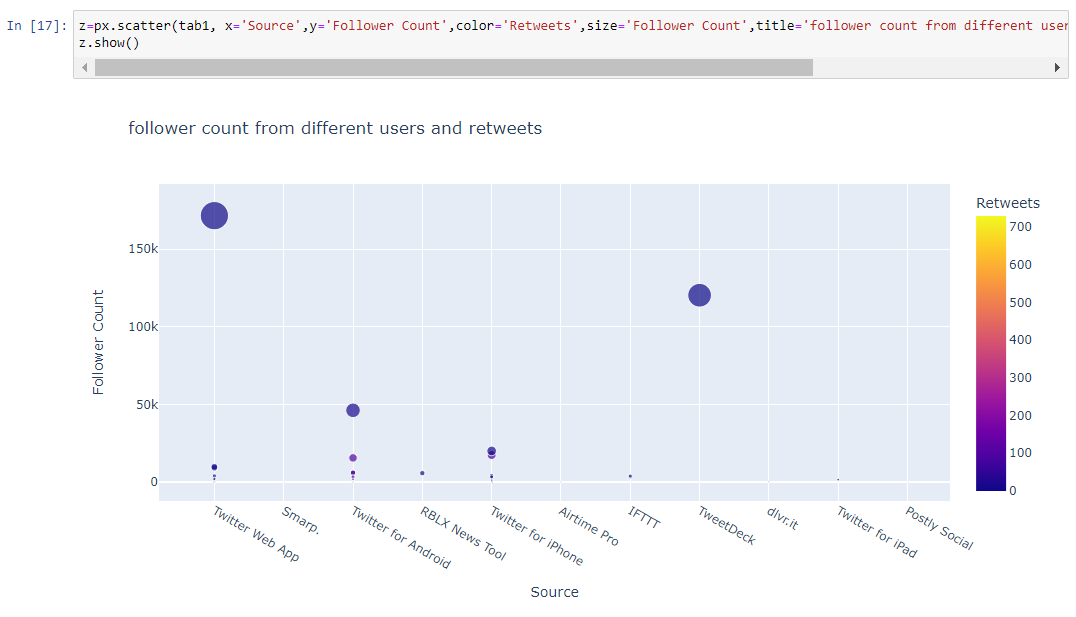
*Fig:5.3*

1. **Scatter plot**



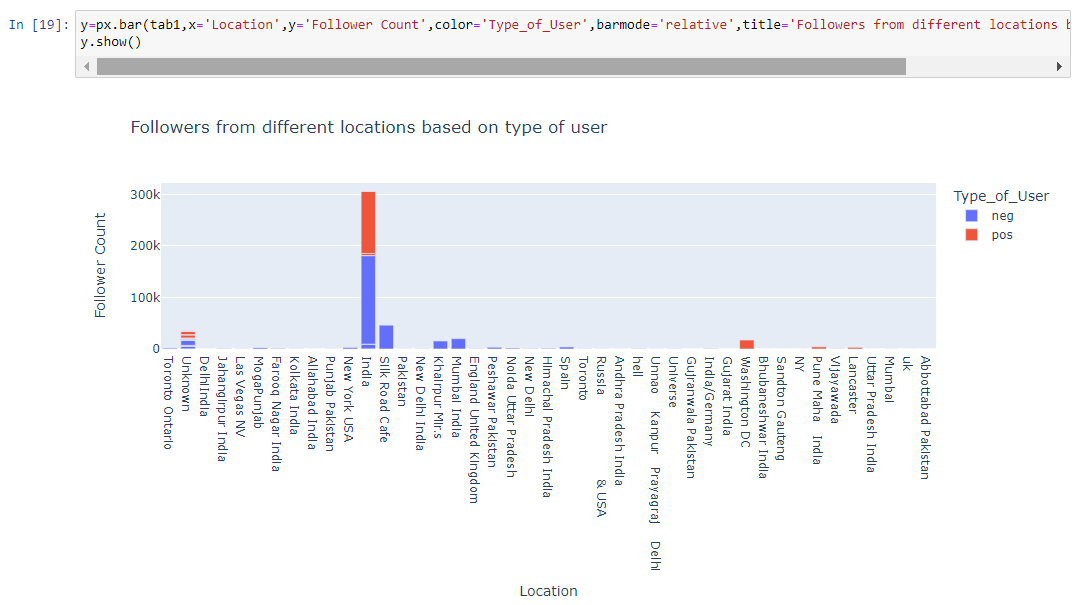
*Fig:5.4*

1. **Scatter plot**



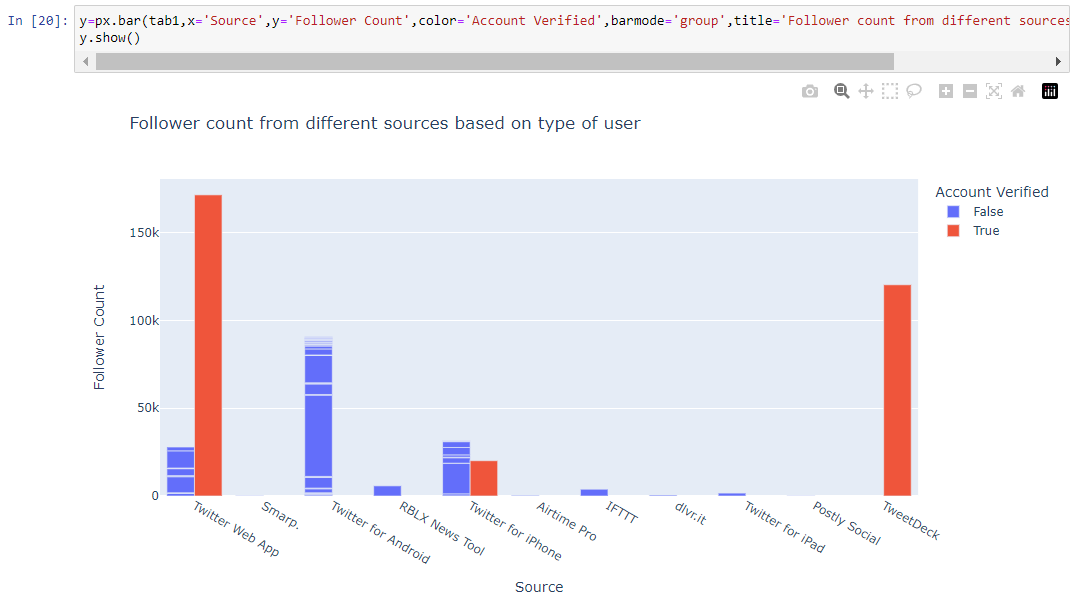
*Fig:5.5*

1. **Bar plot**



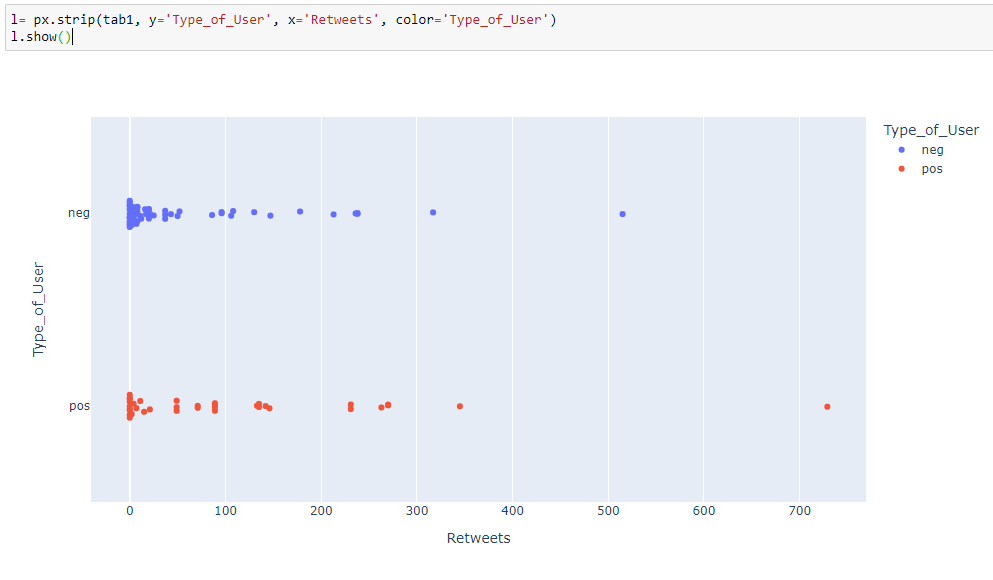
*Fig:5.6*

1. **Bar plot**



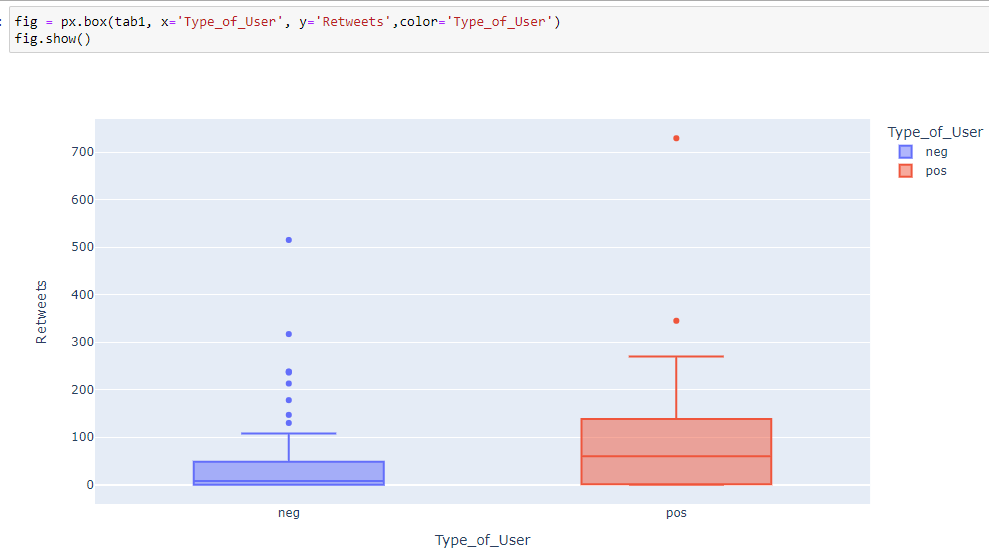
*Fig:5.7*

1. **Strip plot:**

****

*Fig:5.8*

1. **Box plot:**

****

*Fig:5.9*

**6. CONCLUSION**

Extracting live twitter data by using access keys, tokens and cleaning the tweet text was a challenging task as every time we run the program new information is extracted. Sentiment analysis plays a crucial task as it provides **Brand Experience Insights, Patient Insights, Improve Customer Service, Multilingual Insights, News Trend Analysis, Customer Feedback, Online Reputation Management, Data-Driven Marketing Insights, Competitor Analysis, Employee Experience Insights, etc. As we have chosen INDIA for searching the tweets we are able to draw various conclusions from our graphs.** Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. With wordcloud we have showed the entire set of word used on the Twitter platform to describe India. Using plotly we have depicted further details about the location from where we have received positive or negative tweets about India, the number of followers such people have, are their accounts verified or not, the number of retweets and the sources of such tweets. Sentiment analysis system is an active field of research and we can still further improve our system.

1. **CONTRIBUTIONS**

The entire code was a combined effort of both of us teammates.

**G. Sai Akshitha (1601-21-771-008):**

Have worked on collecting the access keys and tokens and have extracted the live data and stored into the csv files. Performed cleaning of data so that each tweet text is properly decoded.

**M. Tharshith (1601-21-771-048):**

Have worked on creating multiple files gathering data required for plotting and visualization from the already clean data file. Has plotted the graphs depicting various insights from the entire data extracted.

1. **BIBILOGRAPHY**
2. <https://pypi.org/project/requests/>
3. <https://pandas.pydata.org/>
4. <https://numpy.org/>
5. <https://pypi.org/project/selenium/>
6. [https://plotly.com/python /](https://plotly.com/python%20/)
7. <https://www.tweepy.org/>
8. <https://www.nltk.org/>