

**Stream Data Processing.**

SID:1819070

**MOD006125 Cloud Computing**

**BEng Computer Science**

**Submitted: May 2021**

Table of Contents

[Introduction 3](#_Toc67590974)

[Background 3](#_Toc67590975)

[Architecture 5](#_Toc67590976)

[Solution 6](#_Toc67590977)

[Conclusion 13](#_Toc67590978)

[Reference list 14](#_Toc67590979)

# Introduction

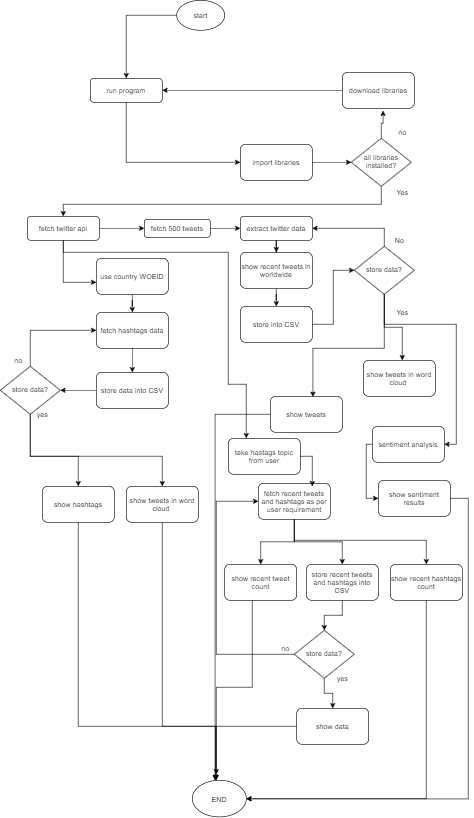
This paper will explain the data storing process, analyze twitter sentiment, retweets, likes, comments, and hash tags by using python language. This language has several visualization tools to make several kinds of plots and word cloud models from these datasets. Also, this program can handle large amounts of real-time data about global events, tweets, and hash tags. This program can fetch data through Twitter API. Also, this program can analyze twitter sentiments. This program provides an on-spot solution for a better user experience. This program fetches data from social networks and visualizes essential events, hash tags, tweets, retweets, and their services or products.

# Background

Twitter is the most popular social network in present days because all-powerful people or political parties start new trends or revolutions by posting a tweet. Also, on Twitter, every user can share their thoughts, information and ideas, and opinion on sports, music, movies, and other several topics. If powerful people post tweets to their followers who are motivated by this post then they can start trending by using hashtags (Bhavsar and Manglani, 2019). Several events are demonstrated in this world via social networks and trending effectiveness is more attractive to people doing social movements and covid disaster movements. This kind of program takes the analysis of these events (Daniel and Millimaggi 2020). This program can track every movement of Twitter users analyzing these activities. This Twitter is used to trends several kinds of topics worldwide. These trending topics proposed a sequential summary for serial chronology creation which presents several pieces of information.

This Twitter hashtag is a very popular thing in the present day because it can be used as business feedback. For these hashtags, an organization can serve actual products as per hashtags requirements. These hashtags can help to gain popularity and helps to increase followers. An organization can check Twitter sentiment by analyzing these sentiments because an organization can gain popularity through positive sentiment analysis. Several users proceed with their tweets by using a hash sign (Manguri *et al.* 2020). This Hash sign is for users to spread a specific page faster and track these events in real-time. Twitter provides the top 10 trending tweets every time all over the world to gain attention and coverage of news. An organization can fetch this Twitter by developing a Twitter analysis program. Through this program, an organization can stream public data, search old posts, and access trending posts in the world. Statistics Analysis these datasets and visualize these datasets which are helpful for market growth. To access this Twitter data by using python language which has data fetch feature from a Twitter API by using Twitter and tweepy module and store these data into a CSV file for visualization. Visualize this dataset this python language using several libraries such as matplotlib, seaborn, wordcloud, and many more. This matplotlib package is used for visualization data. This package is the cross-platform-based library for building 2D graph plots, plotting areas, and many more things. Seaborn is another visualization library that is used for high-level drawing interface and statistical graphics which is used for Twitter analysis. The word cloud is another library that is used to represent powerful visual objects by using text processing.

# Architecture



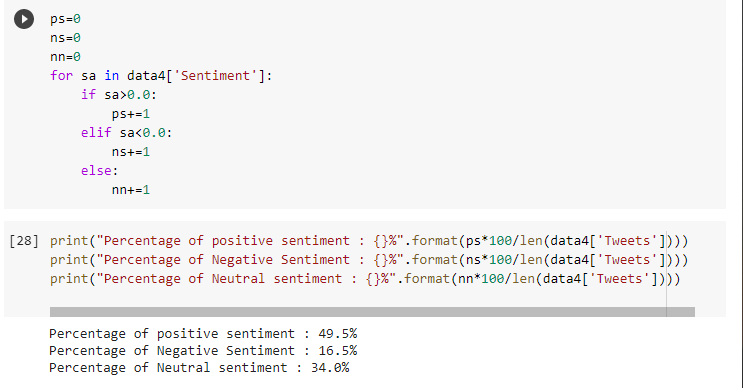
**Figure: The architecture of this twitter program**

(Source: created in draw.io)

This architecture shows how this program works step by step. At first, upload this program into the jupyter notebook or google colab. Then execute this program./ the IDE will check all imported libraries into the Python language, if not then download this "!pip install library-name" and download all libraries. When all libraries are installed then this IDE goes to the next option. Then this program fetches API. After that, this program stores 500 tweets into a CSV file and shows 500 trending tweets. This csv is used to show word cloud plots. Also, this data is used for sentiment analysis (Shelar and Huang, 2018). Then this program fetches API again with WOEID which helps to show country-wise hashtags. Then fetch these hashtags and store them into a CSV file. This csv file is used to show stored hashtags and make word cloud images. Again this program fetches Twitter API, then this program takes user hashtags input. After that, this program shows related topics, hashtags which post recently and counts them and shows this data. After that, this program shows a word cloud model.

# Solution

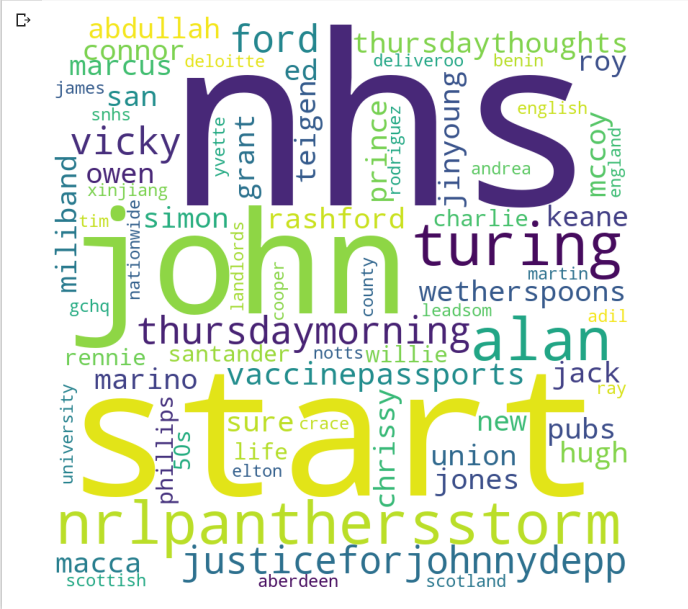
To build this program, it requires several libraries such as operator, requests, pickle, os, Twitter, NumPy, pandas, sys, JSON, matplotlib, seaborn, tweepy wordcloud, and ipython because matplotlib, seaborn, and wordcloud requires to visualize these data into several plots and graphs. JSON is used for work with a JSON file (Shobana *et al.* 2018). The Twitter library is a command-line tool that is used for viewing tweets, tweet lists, and checking replies. To use Twitter, this program needs 4 kinds of keys such as customer key, customer secret key, access token key, and access secret key. Tweepy module uses Twitter bot implementation. After fetch data, extract this data to show what kind of data fetch by this program (Siddharth *et al.* 2018). Also, read csv is used for reading data from CSV. The to\_CSV is used for saving extract into a specific CSV file, "plt" is used for creating a plot by providing figure size and face color. In this project, this program implemented several kinds of plots such as pie plot bar plot by using "tweetData[''].plot.pie()” and “tweets.plot.bar(subplots=True, figsize=(10,6),y” these methods



**Figure: sentiment percentage**

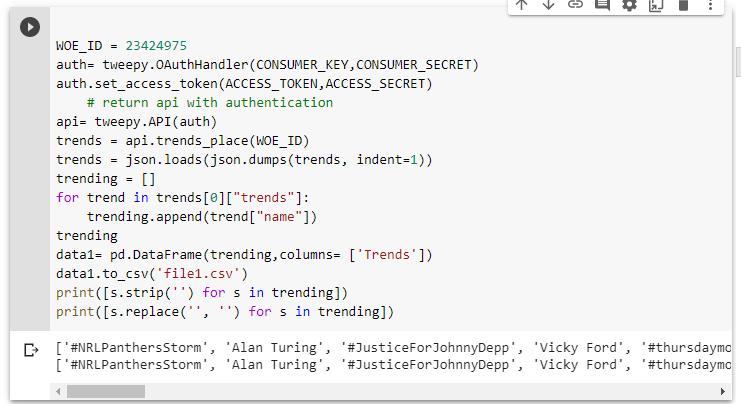
(Source: created in Google colab)





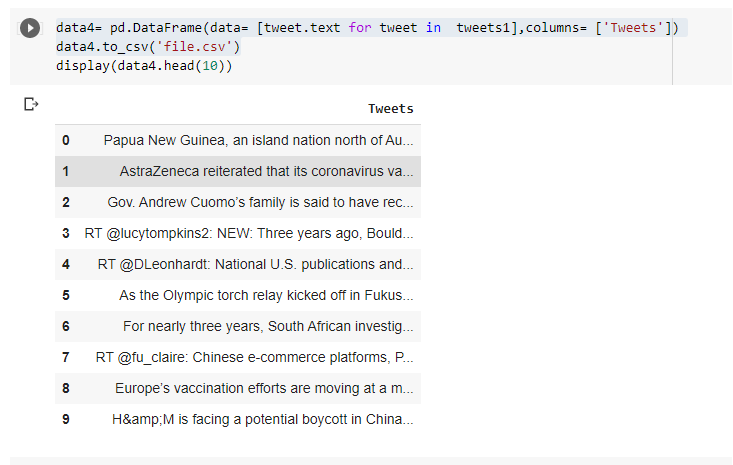
**Figure: Word cloud implementation**

(Source: created in Google colab)



**Figure: hash tags using WOEID**

(Source: created in Google colab)



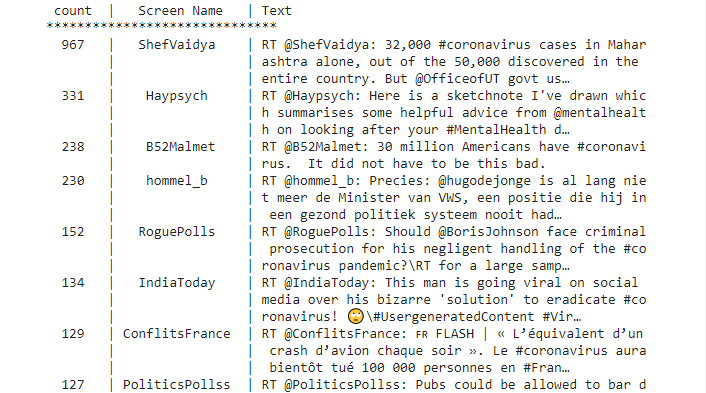
**Figure: Store data into CSV**

(Source: created in Google colab)



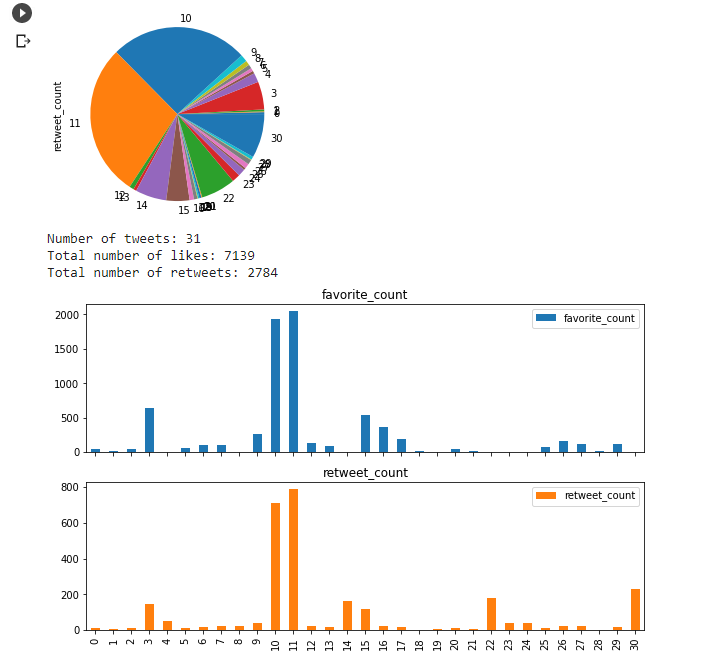
**Figure: Show 500 recent tweets**

(Source: created in Google colab)



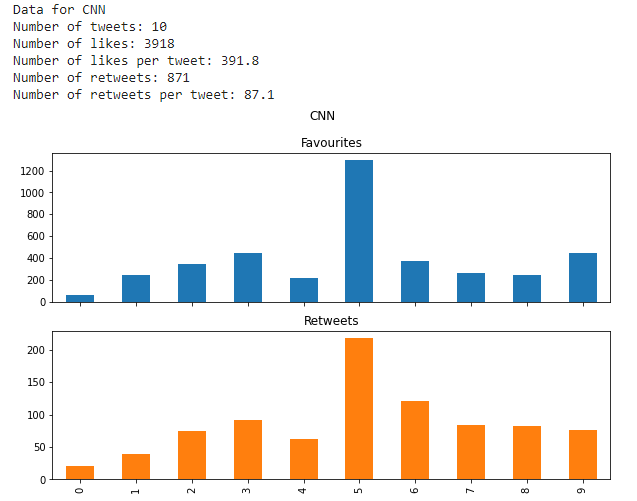
**Figure: Worldwide #coronavirus related tweets**

(Source: created in Google colab)



**Figure: graph plots**

(Source: created in Google colab)



**Figure: total Twitter details of CNN**

(Source: created in Google colab)

# Conclusion

In this report, there are discussions that this program can fetch real-time tweets through Twitter API and store these data into one CSV file. This csv file helps to store data and view this data. It also helps in graph plots and wordcloud plots by using CSV columns. This program helps to fetch hashtags county-wise by using WOEID. This WOEID provides several unique ids of every country. Also, these graph plots help to understand easily which kinds of hashtags are mostly used in the present time. Also, this program can create three kinds of sentiment analysis such as positive analysis, neutral analysis, and negative analysis. This analysis helps to understand how many users support a specific topic.

# Reference list

**Journals**

Bhavsar, H. and Manglani, R., 2019. Sentiment Analysis of Twitter Data using Python. *International Research Journal of Engineering and Technology (IRJET)*, *6*(3), pp.510-527.

Daniel, F. and Millimaggi, A., 2020. On Twitter bots behaving badly: A manual and automated analysis of Python code patterns on GitHub.

Manguri, K.H., Ramadhan, R.N. and Amin, P.R.M., 2020. Twitter sentiment analysis on worldwide COVID-19 outbreaks. *Kurdistan Journal of Applied Research*, pp.54-65.

Shelar, A. and Huang, C.Y., 2018, December. Sentiment analysis of twitter data. In *2018 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1301-1302). IEEE.

Shobana, G., Vigneshwara, B. and Maniraj Sai, A., 2018. Twitter sentimental analysis. *International Journal of Recent Technology and Engineering (IJRTE)*, *7*.

Siddharth, S., Darsini, R. and Sujithra, D.M., 2018. Sentiment Analysis on Twitter Data Using Machine Learning Algorithms in Python‖. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol*, *5*.