**(a)Introduction:**

The world is now passing through the unprecedented time because of Corona Virus pandemic. Many people lost their lives, but many of us successfully defeated this new strain i.e. Covid-19. World Health Organization has declared the virus as a pandemic.

In this project I am going to use a Dataset consisting of data related to the tweets from 24th of july,2020 to 30th of august 2020 with covid19 hashtags. The data set has contained total 179108 rows and 13 columns. This project will analyze various types of “Tweets” gathered during pandemic times.

**The features include the following:**

user\_name, user\_location, user\_description, user\_created, user\_followers, user\_friends user\_favourites, user\_verified, date, text, hashtags, source, is\_retweet**.**

**For a better look:**

**user\_name user\_location ... …. source is\_retweet**

0 ᏉᎥլꂅϮ astroworld ... … Twitter for iPhone False

1 Tom Basile 🇺🇸 New York, NY ... … Twitter for Android False

2 Time4fisticuffs Pewee Valley, KY ... Twitter for Android False

3 ethel mertz Stuck in the Middle ... Twitter for iPhone False

4 DIPR-J&K Jammu and Kashmir ... Twitter for Android False

**(b) Methodology:**

Here I am going to use python to apply sentimental analysis on the tweets. Aim is to observe the people reactions to the pandemic throughout the mentioned time.

**Tools:** Python, Spyder, re, string, Sentiment Analyzer, nltk.corpus (stopwords), Counter, plotly, seaborn, vaderSentiment, TextBlob, SentimentIntensityAnalyzer.

**Steps:**

**Part 1:** (Introduction and Loading our Dataset)

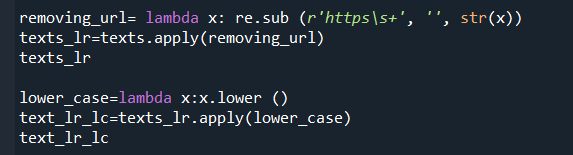
Firstly, the data set has been loaded by importing the csv file.

**Part 2:** Text preprocessing

The preprocessing of the text data is an important step as it makes the raw text ready for mining. The purpose of this step is to clean noise those are less relevant to find the sentiment of tweets such as punctuation(.,?,” etc.), special characters(@,%,&,$, etc.), numbers(1,2,3, etc.), tweeter handle, links(HTTPS: / HTTP:)and terms which don’t carry much weightage in context to the text.

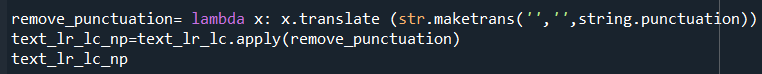
**Removing URL and make lower cases:**

Firstly, the URL has been removed from the text. Additionally, we are converting all the uppercase letters in the string with the lowercase letters.



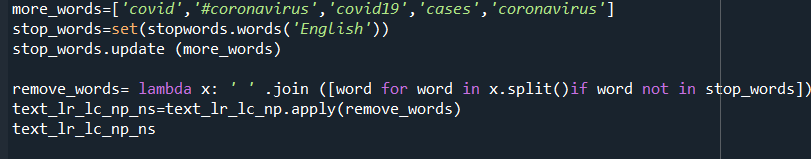
**Removing Punctuation**

This can be clubbed with step of removing special characters. Removing punctuation is fairly easy. It can be achieved by using string. punctuation and keeping everything which is not in this list.

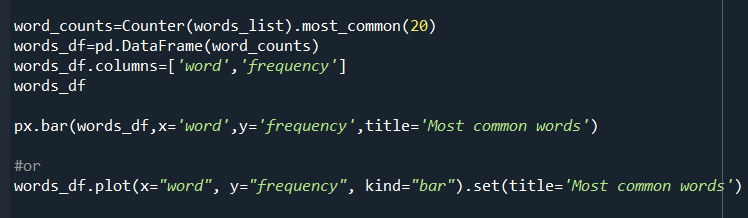


**Remove Stop Words:**

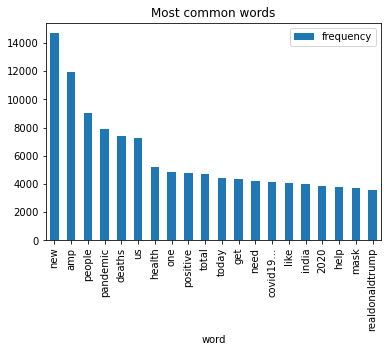
 Many words are frequently used but are only meaningful in a sentence. These are called stop words. We need to remove these stop words from tweets. Stop words are those words in natural language that have very little meaning, such as “is”, “an”, “the”, etc. To remove stop words from a sentence, I divide the text into words and then remove the word if it exists in the list of stop words provided by NLTK.



**(C)Result:**

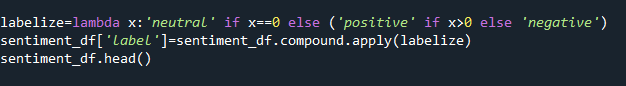


There are some words like ‘new’, ‘people’, having the maximum frequency in our dataset. We can see it from the following most frequent word graph. There are various #hashtags in the tweets column. But they are almost the same in all sentiments hence they are not giving us meaningful full information.



**Figure:** Most common 20 words plot

In this last part, I am doing the Sentimental Analysis. In this task NLTK module has been used to produce Polarity scores for each tweet.



**Visualizing the Result:**

After completing the sentimental analysis, it is seen that most of the peoples (40 percent) are having positive sentiments about various issues shows us their optimism during pandemic times. Then nearly 35 percent people are neutral. And one forth (25 percent) of the people have negative idea, they have more fear and anger and negatives thoughts about Covid-19.

The result has been visualized for analysis using line chart (exact number)

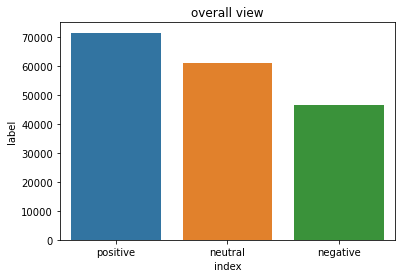
index label

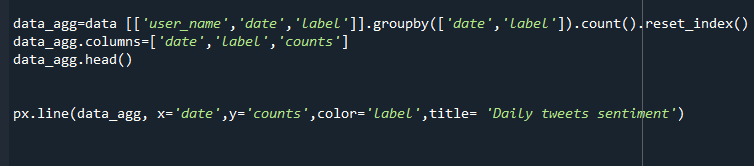
positive 71397

neutral 61094

negative 46617

**In graph:**





**date label counts**

2020-07-24 negative 84

2020-07-24 neutral 95

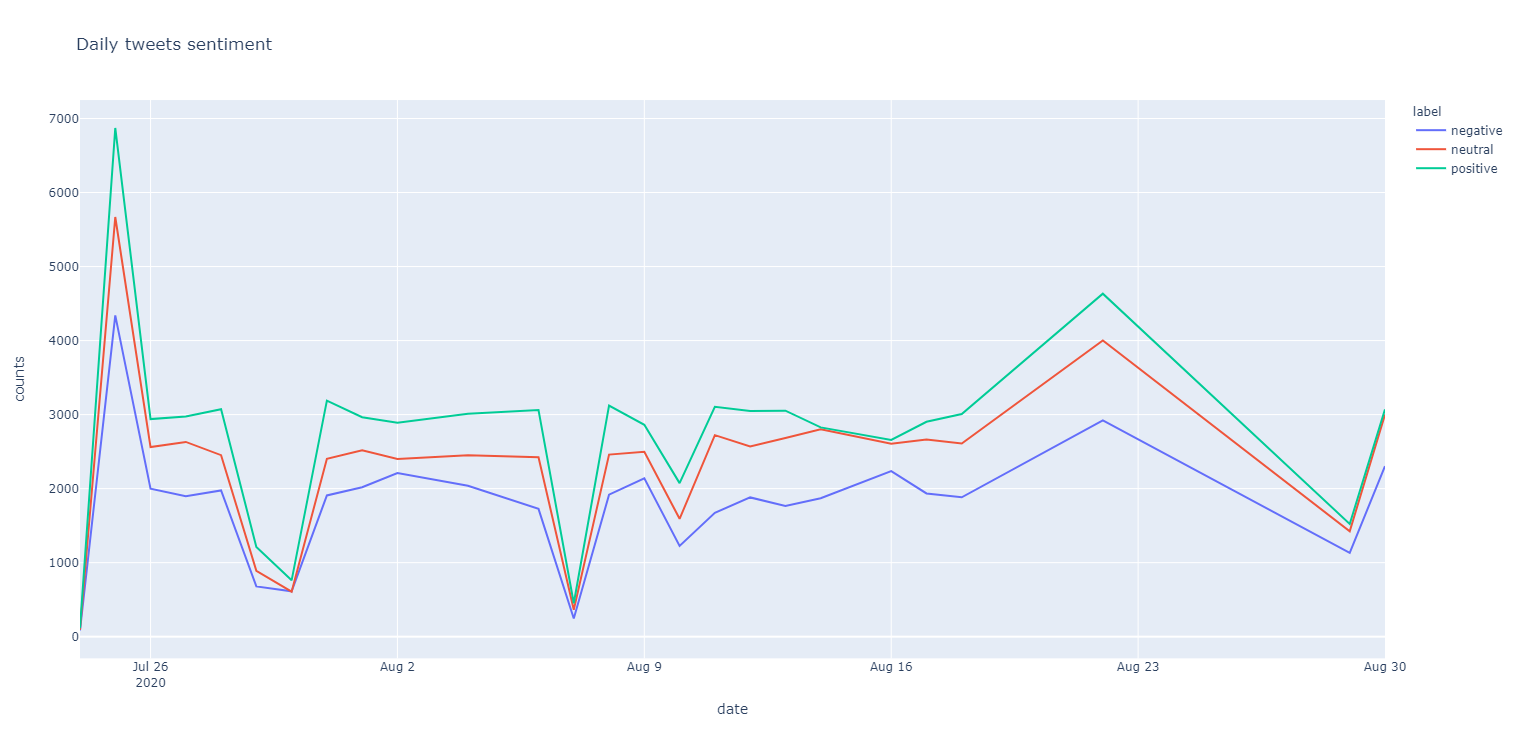
2020-07-24 positive 116

2020-07-25 negative 4340

2020-07-25 neutral 5669

… … … …

**Below the graph:**



**Figure:** For Daily tweets sentiment (Counts vs Date)

From the above graph we can see that the people’s sentiment of the specific date. Here the green, red, blue color represents respectively positive, neutral, and negative sentiment in a specific date. From the graph, we can notice that people’s sentiment can vary from time to time. Sometimes people has expressed themselves more and sometimes people expressed less. Similarly, we noticed that some days people tested more positive (covid tested positive) and some days there are less positive number (means more covid negative). These might be a fact of expressing the people’s feelings.

**(d) Conclusion:**

Our findings present the high prevalence of keywords and associated terms among tweets during COVID-19. I have labeled the tweets as Positive, Negative, and neutral. The emotional level has been considered as positive, negative, and neutral. The mental and physical condition of mass people is found to be directly proportional to this pandemic disease. Then the result has been visualized to see the people reactions on twitter. The study can be beneficial for various reasons. For example, Government can make use of this information in policymaking as they can able to know how people are reacting to this new strain, what all challenges, tasks they are facing such as food scarcity, public mental condition, panic attacks, etc.

**References:**

1.Data **Link:** (<https://www.kaggle.com/gpreda/covid19-tweets>)

2.https://towardsdatascience.com/nlp-for-beginners-cleaning-preprocessing-text-data-ae8e306bef0f