COVID19 TWITTER DATA ANALYSIS

A TWITTER SENTIMENT ANALYSIS REPORT

Submitted by- Shreya Garg

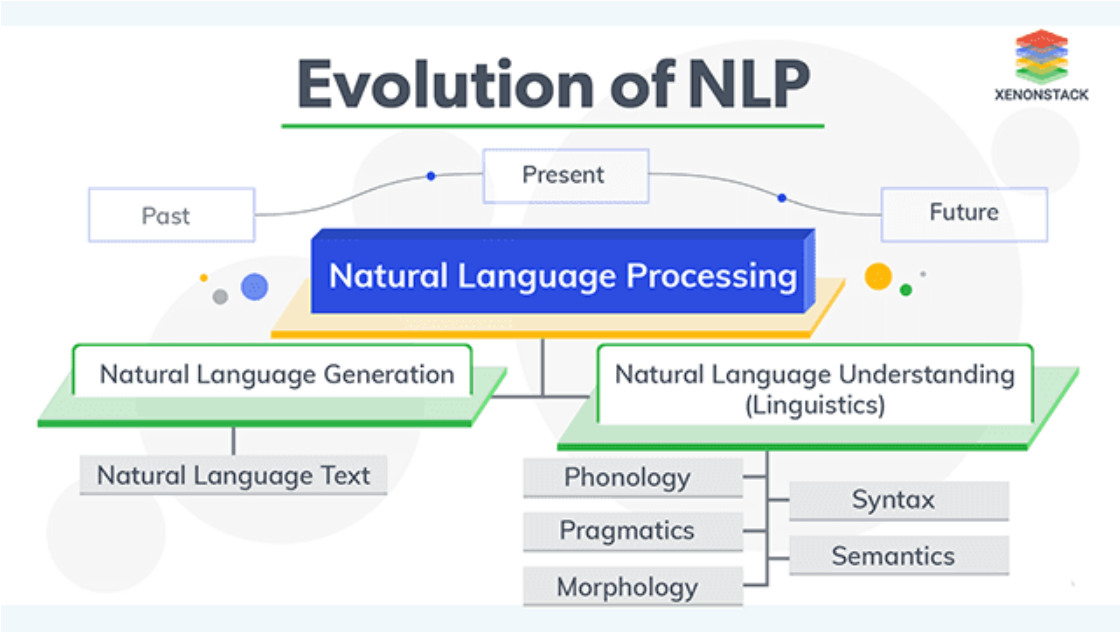
LIST OF TABLES

1. SOFTWARE TRAINING WORK UNDERTAKEN

1.1 THEORETICAL EXPLANATION

- Natural Language Processing

  Natural language processing is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data.

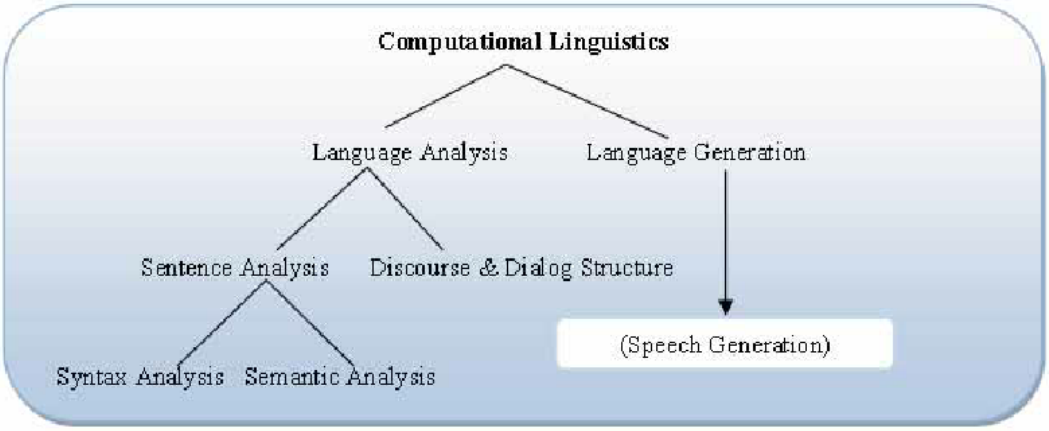


- Text Analysis

  Text analysis, also known as text mining, is the process of automatically classifying and extracting meaningful information from unstructured text. It involves detecting and interpreting trends and patterns to obtain relevant insights from data in just seconds.

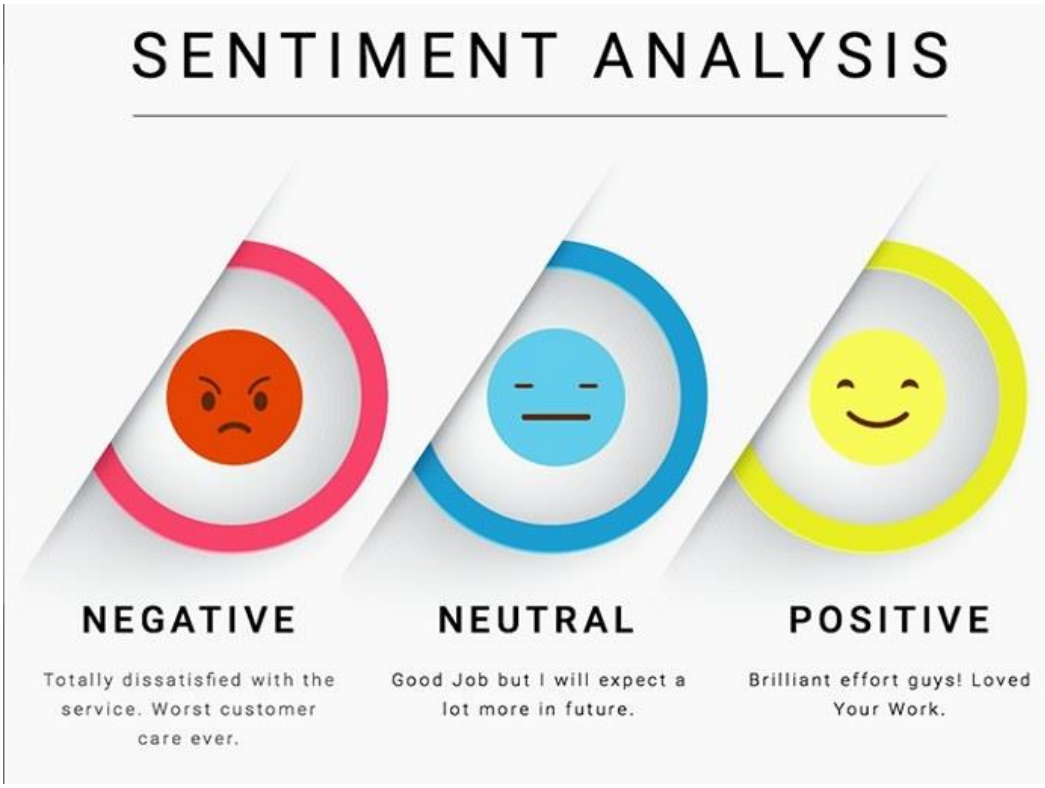
- Computational linguistics

  Computational linguistics is an interdisciplinary field concerned with the computational modelling of natural language, as well as the study of appropriate computational approaches to linguistic questions.



- Sentiment Analysis

  Sentiment analysis is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.



- Twitter Sentiment Analysis:

  Sentiment analysis refers to identifying as well as classifying the sentiments that are expressed in the text source. Tweets are often useful in generating a vast amount of sentiment data upon analysis. These data are useful in understanding the opinion of the people about a variety of topics.

1.2 SOFTWARE TOOLS LEARNED

- Anaconda:

  Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS.

- Jupyter:

  Project Jupyter is a nonprofit organization created to "develop open-source software, open-standards, and services for interactive computing across dozens of programming languages". Spun off from IPython in 2014 by Fernando Pérez, Project Jupyter supports execution environments in several dozen languages.

2. PROJECT WORK

2.1 DATA COLLECTION

- Data Type: Text
- Abstract: This data set consists of tweets from Twitter which are constantly updated.
- Dataset Taken: https://cdn.spotle.ai/zip/Tweeter_Data_IN.csv.zip

2.2 CONCEPT WORKING

There are 4 major steps implemented during the project:

- Importing the Data

  Read a comma-separated values (csv) file into DataFrame.

  pandas.read_csv(*filepath_or_buffer, sep=NoDefault.no_default, delimiter=None, header='infer', names=NoDefault.no_default, index_col=None, usecols=None, squeeze=False, prefix=NoDefault.no_default, mangle_dupe_cols=True, dtype=None, engine=None, converters=None, true_values=None, false_values=None, skipinitialspace=False, skiprows=None, skipfooter=0, nrows=None, na_values=None, keep_default_na=True, na_filter=True, verbose=False, skip_blank_lines=True, parse_dates=False, infer_datetime_format=False, keep_date_col=False, date_parser=None, dayfirst=False, cache_dates=True, iterator=False, chunksize=None, compression='infer', thousands=None, decimal='.', lineterminator=None, quotechar='"', quoting=0, doublequote=True, escapechar=None, comment=None, encoding=None, encoding_errors='strict', dialect=None, error_bad_lines=None, warn_bad_lines=None, on_bad_lines=None, delim_whitespace=False, low_memory=True, memory_map=False, float_precision=None, storage_options=None*)

- Generating word cloud based on the tweets

  Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. Significant textual data points can be highlighted using a word cloud. Word clouds are widely used for analyzing data from social network websites.
  For generating word cloud in Python, modules needed are – matplotlib, pandas and wordcloud. To install these packages, run the following commands :

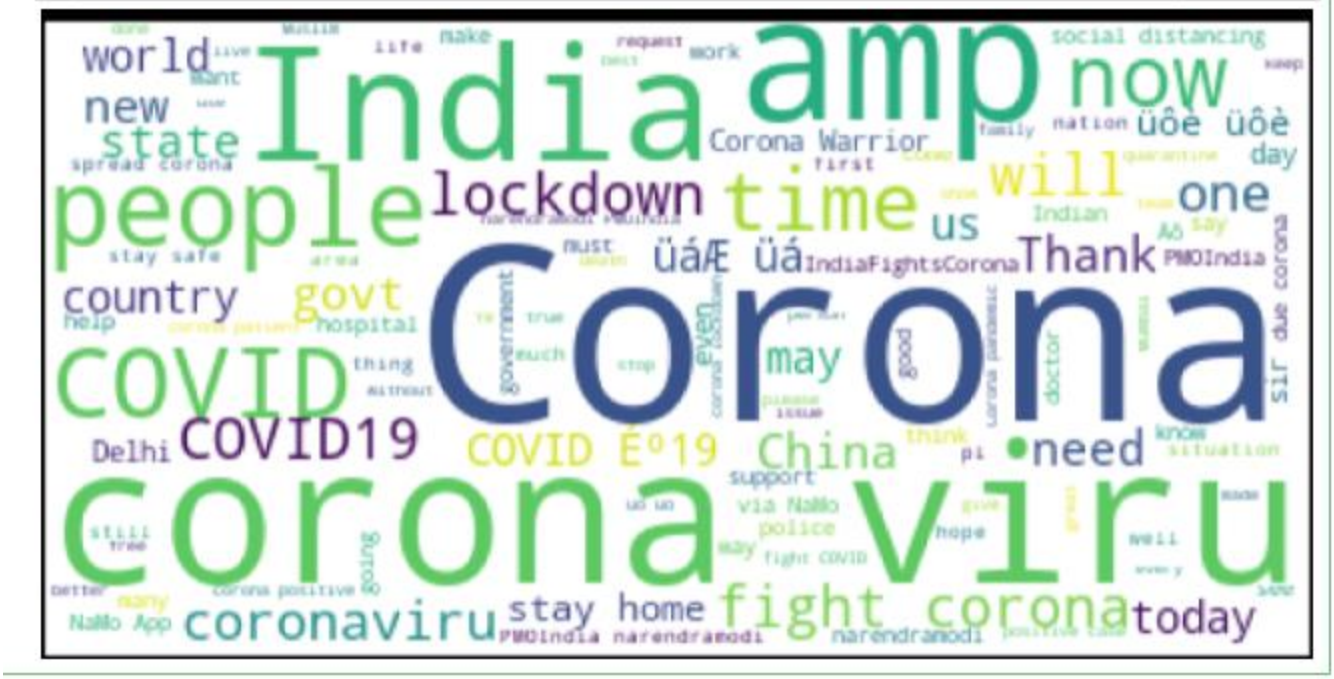
```
pip install matplotlib
pip install pandas
pip install wordcloud
```

  Advantages of Word Clouds :
  Analyzing customer and employee feedback.
  Identifying new SEO keywords to target.

  Drawbacks of Word Clouds :
  Word Clouds are not perfect for every situation.
  Data should be optimized for context.

Here is the code of wordcloud:

```
def word_cloud(tweets):
    stopwords = set(STOPWORDS)
    stopwords.update(["https", "co"])
    wordcloud = WordCloud(background_color="white", stopwords=stopwords, random_state = 5000).generate(" ".join([tw for tw i
    plt.figure( figsize=(40,20), facecolor='k')
    plt.imshow(wordcloud)
    plt.axis("off")
    plt.title("Twitter WordCloud")
word_cloud(tweets)
```



- <u>The relative popularity of the hashtags</u>

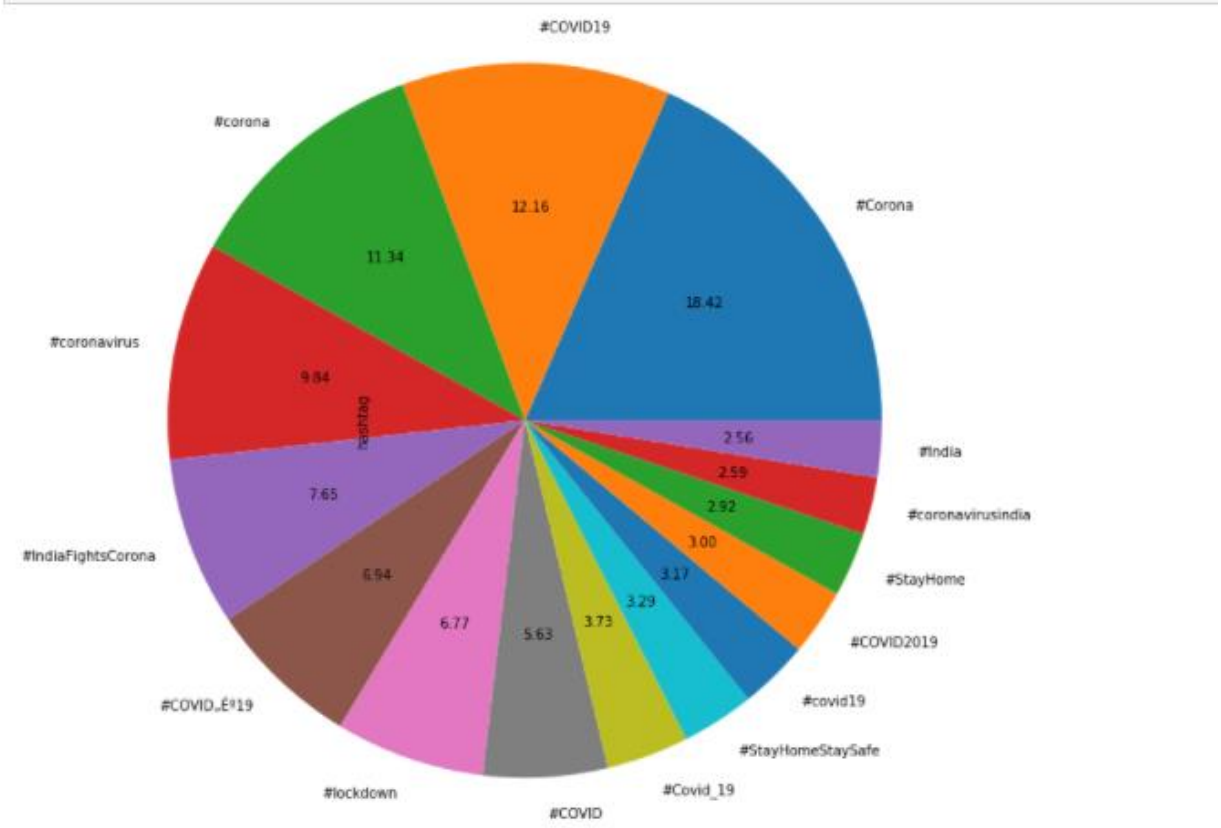
Hashtags are very common and they are used on most social media platforms such as Twitter, Instagram and Facebook. Therefore, hashtag monitoring is as important as any other social media strategy. A hashtag tracker is used to accurately track hashtags over time and on multiple social networks.

Here is the code used for finding most popular hashtags:

```
raw = ' '.join(tweets)
tags = [re.sub(r"(\W+)$", "", j) for j in [i for i in raw.split() if i.startswith("#")]]
df = pd.DataFrame({"hashtag": tags})
print(df['hashtag'].value_counts().head(20))
```

```
#Corona                 2676
#COVID19                1766
#corona                 1648
#coronavirus            1430
#IndiaFightsCorona      1111
#COVID„Éº19             1009
#lockdown                983
#COVID                   818
#Covid_19                542
#StayHomeStaySafe        478
#covid19                 460
#COVID2019               436
#StayHome                424
#coronavirusindia        376
#India                   372
#Covid19                 345
#stayhome                326
#staysafe                305
#CoronavirusOutbreak     297
#CoronaWarriors          272
Name: hashtag, dtype: int64
```

Here is the pie graph for the most popular hashtags:



- Top 10 twitter handlers which are most active

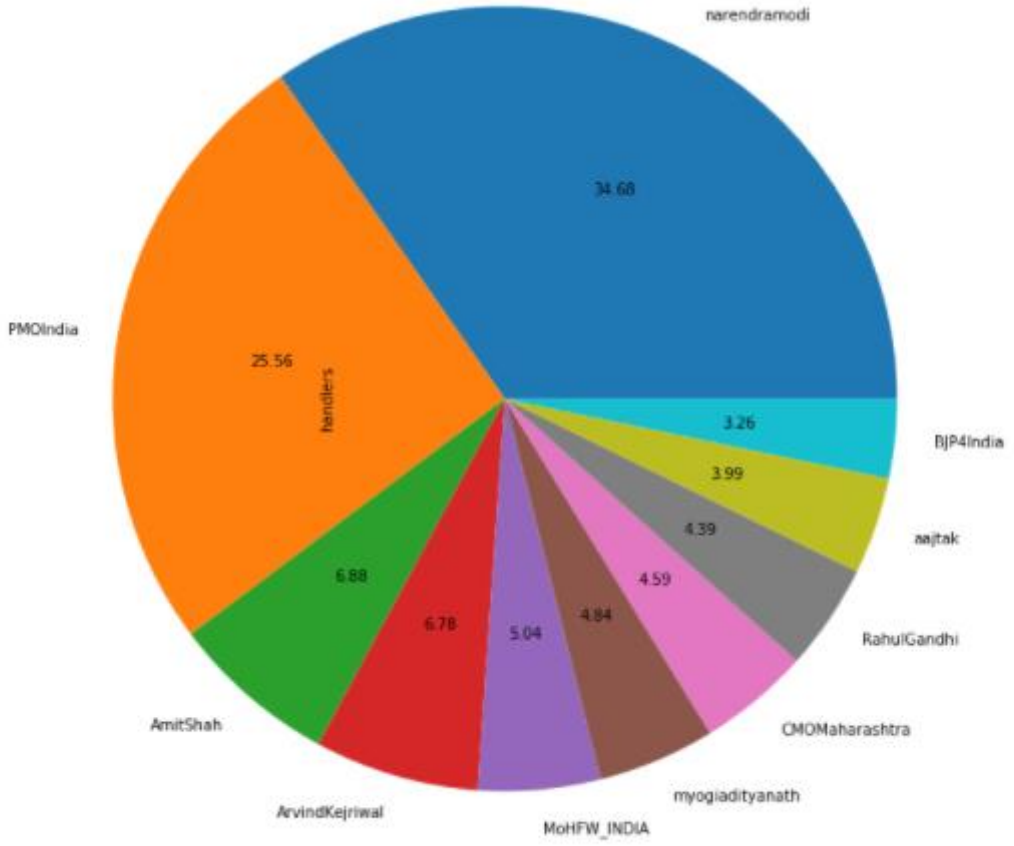
  Twitter's definition is based on user interactions with others, but focuses less on their actual activities on the site; an active user is someone who follows at least 30 accounts, and at least one third of those accounts follows them back.

  Here is the code used for the top 10 twitter handlers:

```
raw = ' '.join(tweets)
tags = [re.sub(r"(\W+)$", "", j[1:]) for j in [i for i in raw.split() if i.startswith("@") and len(i) != 1 ]]
df = pd.DataFrame({"handlers": tags})
print(df['handlers'].value_counts().head(10))
```

```
narendramodi       3691
PMOIndia           2720
AmitShah            732
ArvindKejriwal      722
MoHFW_INDIA         536
myogiadityanath     515
CMOMaharashtra      488
RahulGandhi         467
aajtak              425
BJP4India           347
Name: handlers, dtype: int64
```

Here is the pie graph for the top 10 twitter handlers:



3. RESULTS AND DISCUSSIONS

- Analyzing thousands of texts in just a few seconds and automatically get information such as topic, sentiment, or language.

- 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. What do customers love about your brand? What aspects get the most negative mentions?

- If these tweets were comments on your product or service you might be happy to read the positive ones but you would be better spending your time looking at those with a negative sentiment to find out what the problem is. It would be simple enough to filter out the tweets that are particularly negative for special attention. Of course, it is not impossible that all of your feedback will be positive — but in the real world that is unlikely. Sentiment Analysis allows you to get an overview of how your customers feel and can allow you to spot problems before they get out of hand.

4. CONCLUSIONS AND FUTURE SCOPE

4.1 CONCLUSIONS

- Improve Customer Service
- Improve Media Perceptions
- Improve Crises Management
- Develop Quality Products
- Discovering New Marketing Strategies

4.2 FUTURE SCOPE

- As a result of deeper and better understanding of the feelings, emotions and sentiments of a brand or organization's key, high-value audiences, members of these audiences will increasingly receive experiences and messages that are personalized and directly related to their wants and needs. Rather than segment markets based on age, gender, income and other surface demographics, organizations can further segment based on how their audience members actually feel about the brand or how they use social media.
- Social media analytics helped predict and explain the emotions of concerned parties behind Brexit and the 2016 US election, which has spurred a number of non-brand organizations to investigate how sentiment analysis can be used to predict outcomes and map out the emotional landscape of people, voters and the like.
- With recent years bringing big leaps in machine learning and artificial intelligence, many analytics solutions are looking to these technologies to replace algorithms. Unfortunately for organizations looking to leverage sentiment analysis to measure audience emotions, machine learning isn't yet ready to tackle the complex nuances of text and how we talk, especially on social media channels that are rife with slang, sarcasm, double meanings and misspellings. These make it difficult for artificial intelligence systems to accurately sort and classify sentiments on social media. And, with any analysis project, accuracy is crucial.

5. REFERNCES

- https://en.wikipedia.org/wiki/Natural_language_processing
- https://www.xenonstack.com/hubfs/Imported_Blog_Media/evolution-of-nlp-xenonstack-2-1-1.png
- https://monkeylearn.com/blog/what-is-text-analysis/#:~:text=Text%20analysis%2C%20also%20known%20as,from%20data%20in%20just%20seconds.&text=Another%20term%20you%20may%20have%20heard%20is%20text%20analytics.
- https://datawider.com/wp-content/uploads/2019/10/what-is-text-analysis.png
- https://en.wikipedia.org/wiki/Computational_linguistics
- https://www.researchgate.net/profile/Md-Mostafa-Rashel/publication/228346098/figure/fig6/AS:301981391441941@1449009382749/Structural-Diagram-of-Computational-Linguistics.png
- https://en.wikipedia.org/wiki/Sentiment_analysis
- https://www.kdnuggets.com/images/sentiment-fig-1-689.jpg
- https://www.analyticsvidhya.com/blog/2021/06/twitter-sentiment-analysis-a-nlp-use-case-for-beginners/
- https://miro.medium.com/max/450/1*p3Ste5R_iJzi5IcSmFkmtg.png
- https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)
- https://en.wikipedia.org/wiki/Project_Jupyter
- https://pandas.pydata.org/pandas-docs/dev/reference/api/pandas.read_csv.html
- https://www.geeksforgeeks.org/generating-word-cloud-python/#:~:text=Word%20Cloud%20is%20a%20data,indicates%20its%20frequency%20or%20importance.
- https://brandmentions.com/hashtag-tracker/
- http://www.seekvisibility.com/2016/05/active-on-social-media/
- https://monkeylearn.com/blog/sentiment-analysis-of-twitter/
- https://towardsdatascience.com/sentiment-analysis-of-tweets-167d040f0583
- https://www.commsights.com/benefits-of-sentiment-analysis-for-businesses/
- https://www.linkedin.com/pulse/future-sentiment-analysis-shahbaz-anwar/