Name- Harsh Gupta

Section- CCIS

University Roll.no- 2014670

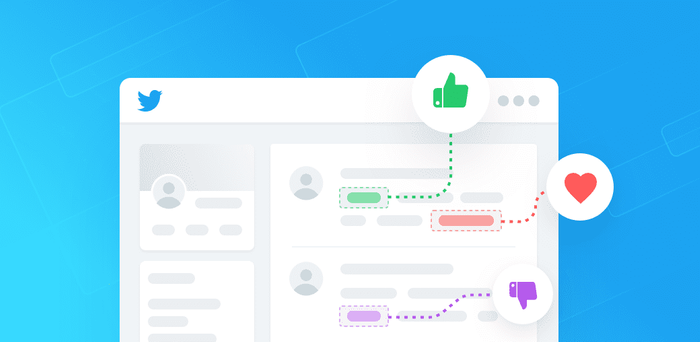
Contact- 9140943594

Class Roll no- 25

Gmail- harshc2303@gmail.com

What is Sentiment Analysis?

Sentiment Analysis examines the problem of studying texts, like posts and reviews, uploaded by users on microblogging platforms, forums, and electronic businesses, regarding the opinions they have about a product, service, event, person, or idea. Mainly the sentiments are of three types:

* Positive Sentiment
* Neutral Sentiment
* Negative Sentiment

Libraries Used

* **tweepy**

Tweepy is an open-source Python package that gives you a very convenient way to access the Twitter API with Python. Tweepy includes a set of classes and methods that represent Twitter's models and API endpoints.

* **textblob**

TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

* **pandas**

Pandas is an open-source Python package that is most widely used for data science/data analysis and machine learning tasks. It is built on top of another package named Numpy, which provides support for multi-dimensional arrays.

* **matplotlib.pyplot**

matplotlib. pyplot is a collection of functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc.

* **re**

A regular expression (or RE) specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression (or if a given regular expression matches a particular string, which comes down to the same thing).

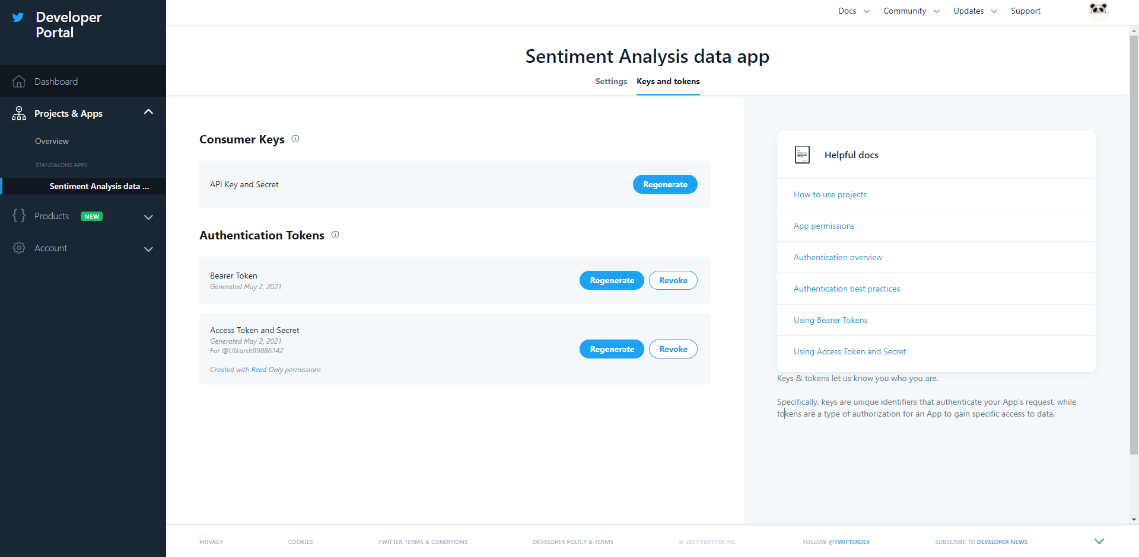
* **wordcloud**

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.

Importing the Twitter API

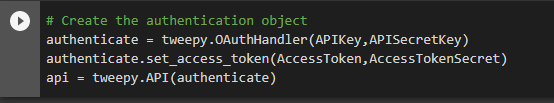
For the Twitter API, we need the Developer Account of Twitter and then we need to create an app.

I have created an app named the Sentiment Analysis data app.



By creating this app, we can access the API keys with the help of which we can access the tweets from Twitter.

Creating the authentication object and accessing the Tweets with the help of API



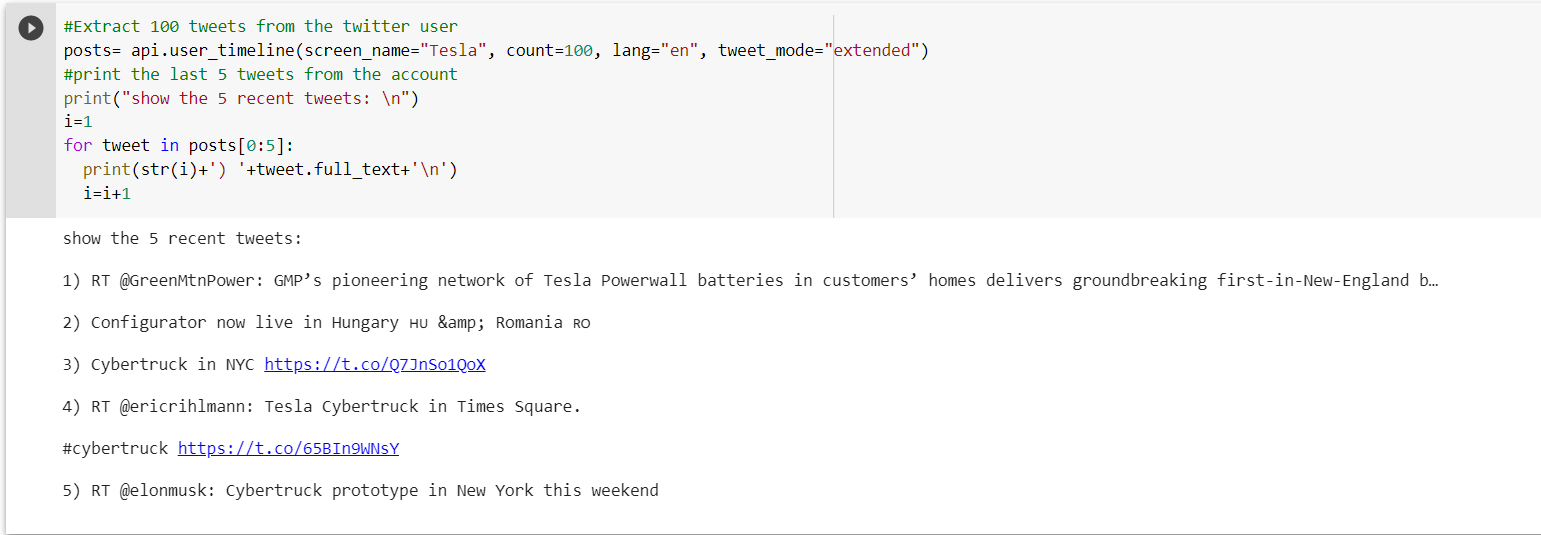
Here I am authenticating the keys and connecting to Twitter with the help of OAuthHandler() function present in the tweepy library.

Gathering the tweets

By using the user\_timeline() function present in the tweepy library I have got the tweets from Twitter on a specific topic.

For example:

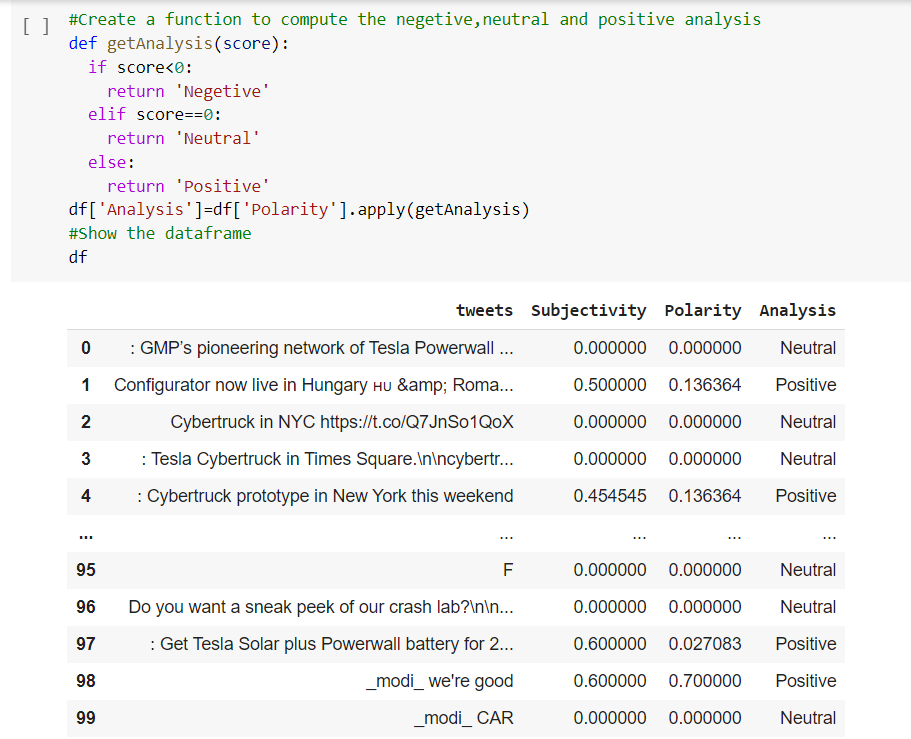
I have searched 100 tweets related to Technology in the English Language by using the user\_timeline() function and printed them using a for loop.



Create a DataFrame with a column called tweets

Now after the extraction of the tweets I have created a dataframe having a column named Tweets with the help of pandas library.

A DataFrame is a 2-dimensional labelled data structure.



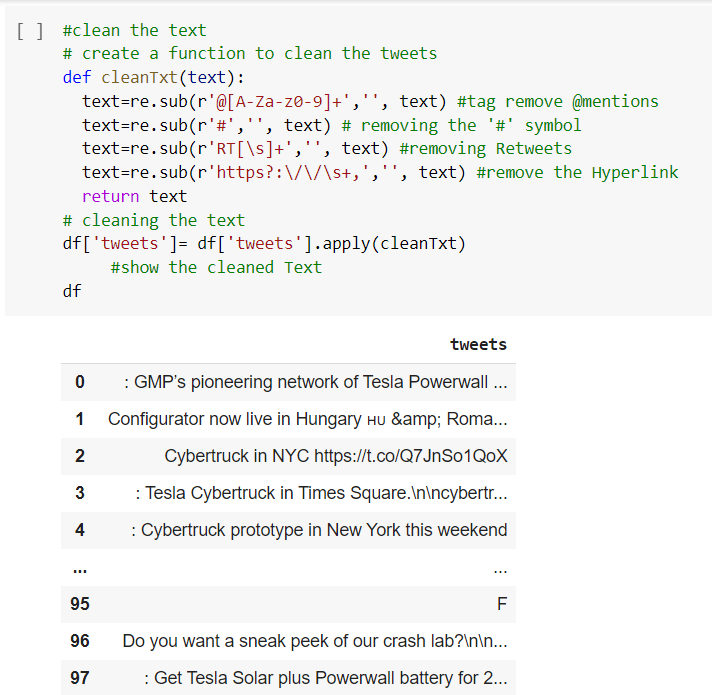
Cleaning the data

The tweets have some amount of data that we don’t need during the sentiment analysis. Such as “#”, “@” or mentions, hyperlinks, retweets, etc.

I have created a function **cleanText(text)** in which I have all such types of elements present in the tweets.

After creating this function I have applied it to the dataframe df:

After applying the cleanText() function the dataframe look like this:



Calculating the subjectivity and polarity of tweets

The main part of sentiment analysis is the Subjectivity and Polarity from which we can differentiate between the positive, negative, and neutral tweets.

**SUBJECTIVITY**

Subjectivity is a float value that lies in the range of [-1,1]. It describes whether the data is a fact or an opinion. Subjective sentences generally refer to opinion, emotion, or judgment.

**POLARITY**

Polarity is a float value which lies in the range of [-1,1] where 1 means a positive statement and -1 means a negative statement.

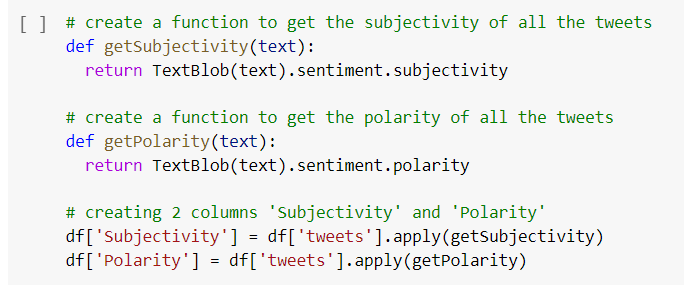
If the value of polarity is less than 0 then the Sentiment of the data is Negative.

If the value of polarity is greater than 0 then the Sentiment of the data is Positive.

If the value of polarity is equal to 0 then the Sentiment of the data is Neutral.

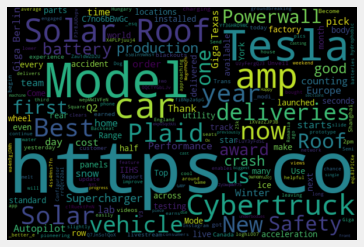
I have created two functions.

* getSubjectivity(): For calculating the Subjectivity.
* getPolarity(): For calculating the Polarity.



Word Cloud Visualization

A word cloud (or tag cloud) is a word visualization that displays the most used words in a text from small to large, according to how often each appears. The largest words are the most used, and stop words (*a, and, the*, etc.) are automatically removed.

****

Creating a function to compute Negative, Neutral and Positive

I have created a function gteAnalysis() to analyze the score of the data by differentiating it into three parts :

* Negative
* Neutral
* Positive

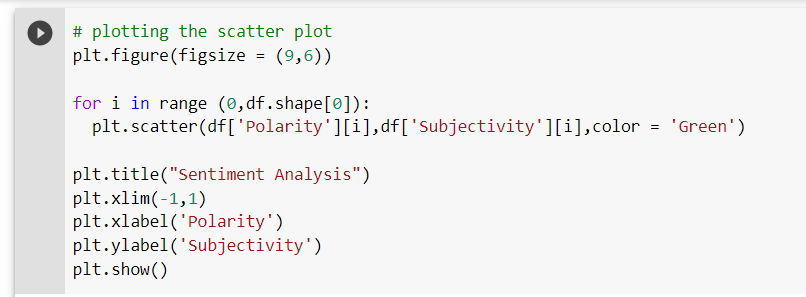
And displaying it into the column “Polarity” in the dataframe.

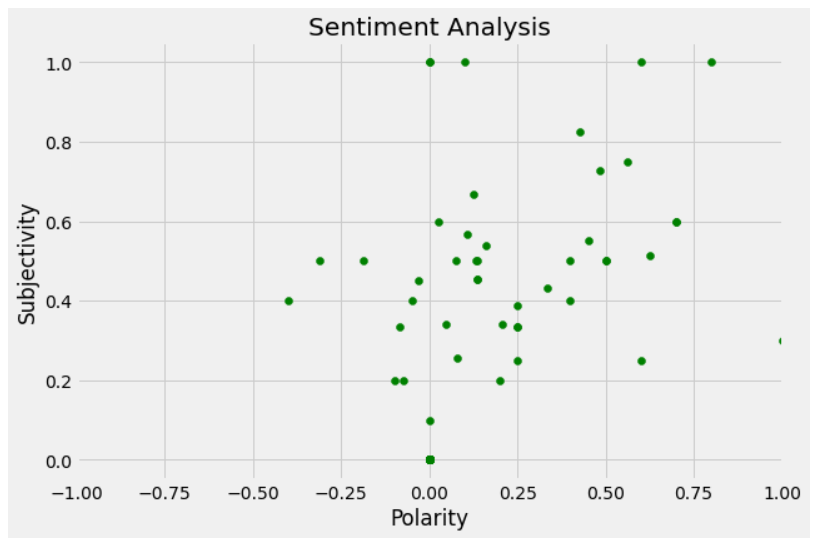
Plot the Scatter Plot

What is a Scatter Plot?

A scatter plot is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data.

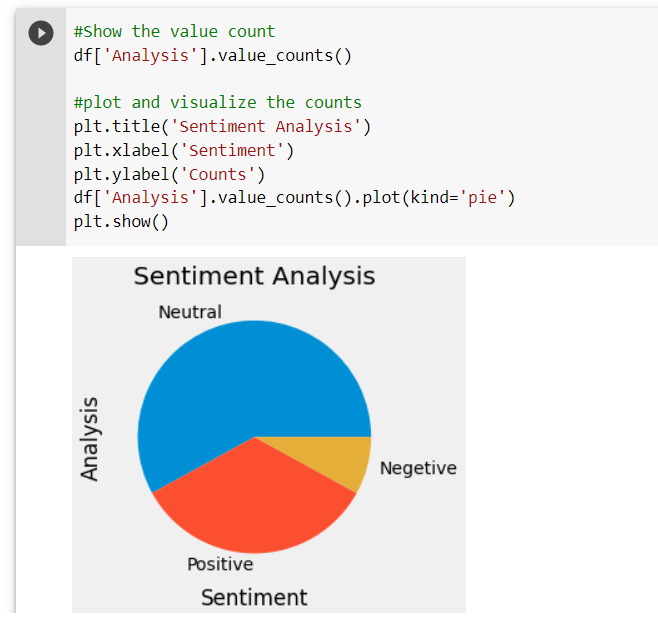
I have plotted a Scatter plot between Polarity ( x-axis ) and   
Subjectivity( y-axis ) and the title is Sentiment Analysis.





Plotting Bar Graph

A bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally.



Conclusion

The field of sentiment analysis is an exciting new research direction due to a large number of real-world applications where discovering people’s opinion is important in better decision-making. The development of techniques for document-level sentiment analysis is one of the significant components of this area. Recently, people have started expressing their opinions on the Web that increased the need of analyzing opinionated online content for various real-world applications. A lot of research is present in literature for detecting sentiment from the text. Existing sentiment analysis models can be improved further with more semantic and commonsense knowledge.

References

[www.google.com](http://www.google.com)

[www.wikipedia.org](http://www.wikipedia.org)

<https://stackoverflow.com>

[www.analyticsvidhya.com](http://www.analyticsvidhya.com)

[www.youtube.com](http://www.youtube.com)