

**NAME: NIDHI VIG**

**UNIVERSITY ROLL NO: 2016874**

**SEMESTER: 3**

**PROJECT REPO: //GITLINK**

**PROBLEM STATEMENT:**

Analyzing twitter sentiments on latest twitter data.

**MOTIVATION:**

Sentimental analysis is an extremely common machine learning problem and is used in a lot of activities like product predictions, movie recommendations, and several others.

Currently for every machine learner new to this field, like myself, exploring this domain has become an integrate facet, and so I explored this topic and tried to build a project of my own. I thought it would be a good idea to obtain insights into how Twitter users felt about different domains.

I enjoyed the process and learned many things through my journey.

**ABSTRACT**

Nowadays, people from all around the world use social media to share their views or any information. Twitter is an online news and social networking site where people communicate in short messages called tweets. Users share their daily lives, post their opinions on everything such as brands and places.

Companies can benefit from this massive platform in various ways:

* They can use twitter to increase their brand awareness
* Inform customers about the updates
* To get instant feedback about their product and services
* Collect data related to opinions on them

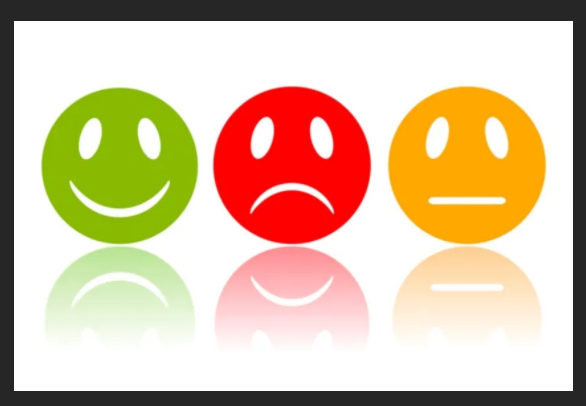
The most important thing for a company is to listen to market and meet the customers need. When companies grow it becomes difficult for them to pay attention to each customer’s review and know what people think about them, this is where SENTIMENTAL ANALYSIS comes to use.

**SENTIMENTAL ANALYSIS**

Sentimental analysis is the process of detecting positive and negative sentiment in text. It is also known as opinion mining.

Sentimental analysis has been acquiring a crucial role in both commercial and research applications because of their possible applicability to several different fields. Therefore, a huge number of companies have included the analysis of opinions and sentiments of customers as part of their mission.

One of the most interesting applications of these approaches involves the automatic analysis of social network messages, on the basis of the feelings and emotions conveyed.



**WHY SENTIMENTAL ANALYSIS?**

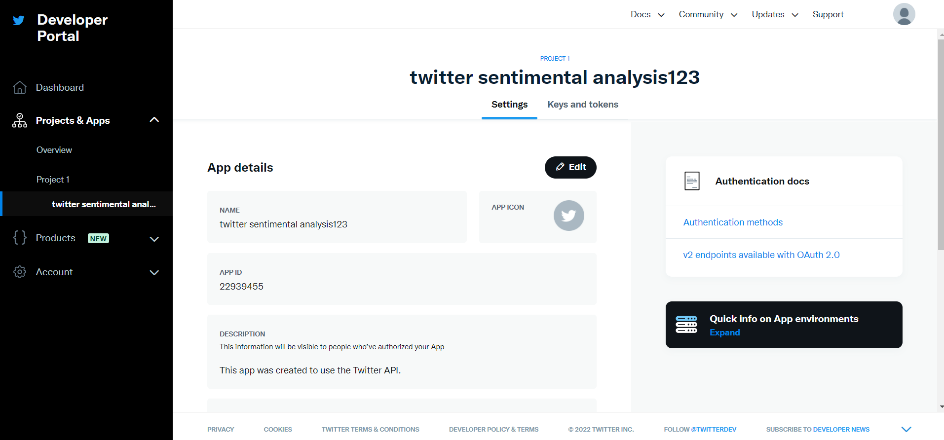
* **Business:** In marketing field companies use it to develop their strategies, to understand customer’s feelings towards products or brand, how people respond to their campaigns or product launches and why some of their products fail.
* **Politics:** In political field, it is used to keep track of political view, to detect consistency and inconsistency between statements and actions at the government level. It can be used to predict election results as well!
* **Public Actions:** It is also used to monitor and analyze social phenomena, and determing the general mood of the public.

**METHODS AND TOOLS REQUIRED**

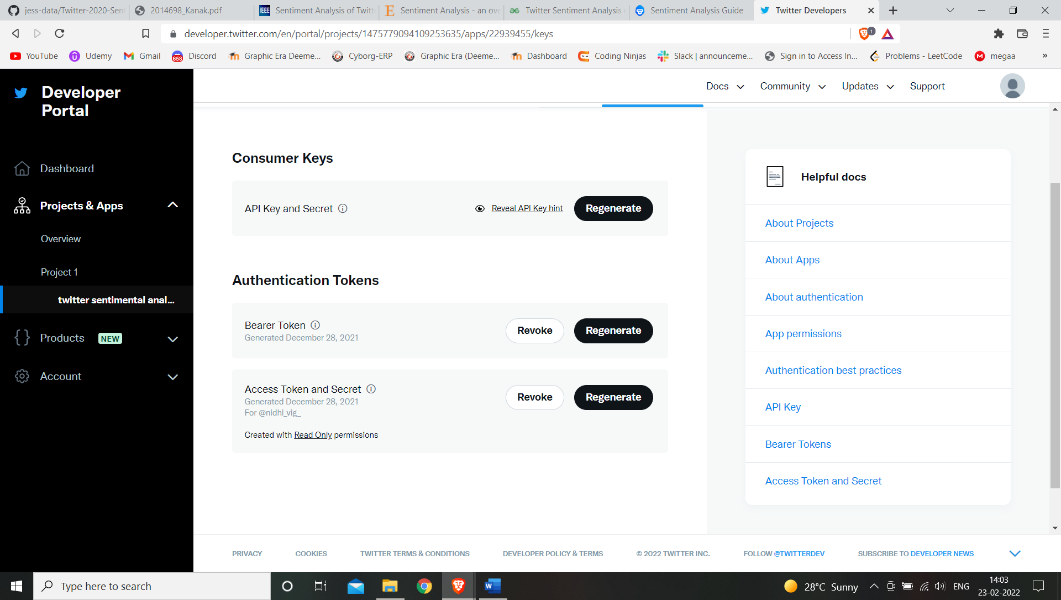
This project was done using NLP (Natural Language Processing) techniques. Twitter receives over 500 million tweets per day, across the globe. Hence, our work is to retrieve the data and analyze it.

**DEVELOPER ACCOUNT**

In order to fetch tweets through Twitter API, one needs to create a **“TWITTER DEVELOPER ACCOUNT”** from twitter developer portal and register an app through their twitter account.



Once the app is created, open the ‘Keys and Tokens’ tab, and copy ‘Consumer Key’, ‘Consumer Secret’, ‘Access token’ and ‘Access Token Secret’.



I carried out the following steps for the project:

* Import libraries
* Tweets mining
* Data cleaning
* Tweets processing
* Data exploration
* Sentimental Analysis

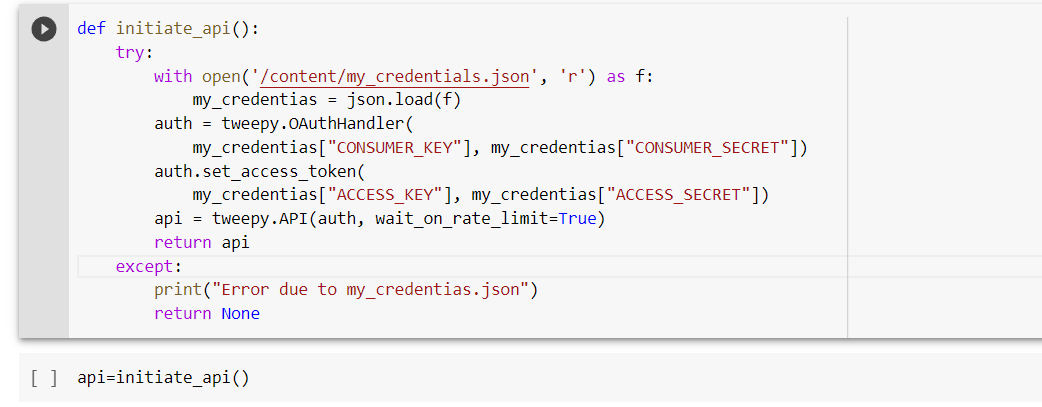
**IMPORTING LIBRARIES**

Python libraries like :-

* Tweepy :- for tweets mining
* Pandas :- for data cleaning/manipulation
* NLTK :- Natural Language Toolkit
* TextBlob :- for sentimental analysis
* MatPlotlib :- Data exploration
* word cloud :- Data exploration
* Plotly :- for data visulalization
* Re :- Regular expression, it lets you check is a particular string matches a given expression

**TWEETS MINING**

Authorize twitter API client.



my\_credentials.json file consists of all the keys and this step will authorize the API client.



We use this code to fetch tweets, and filter the retweets and links.

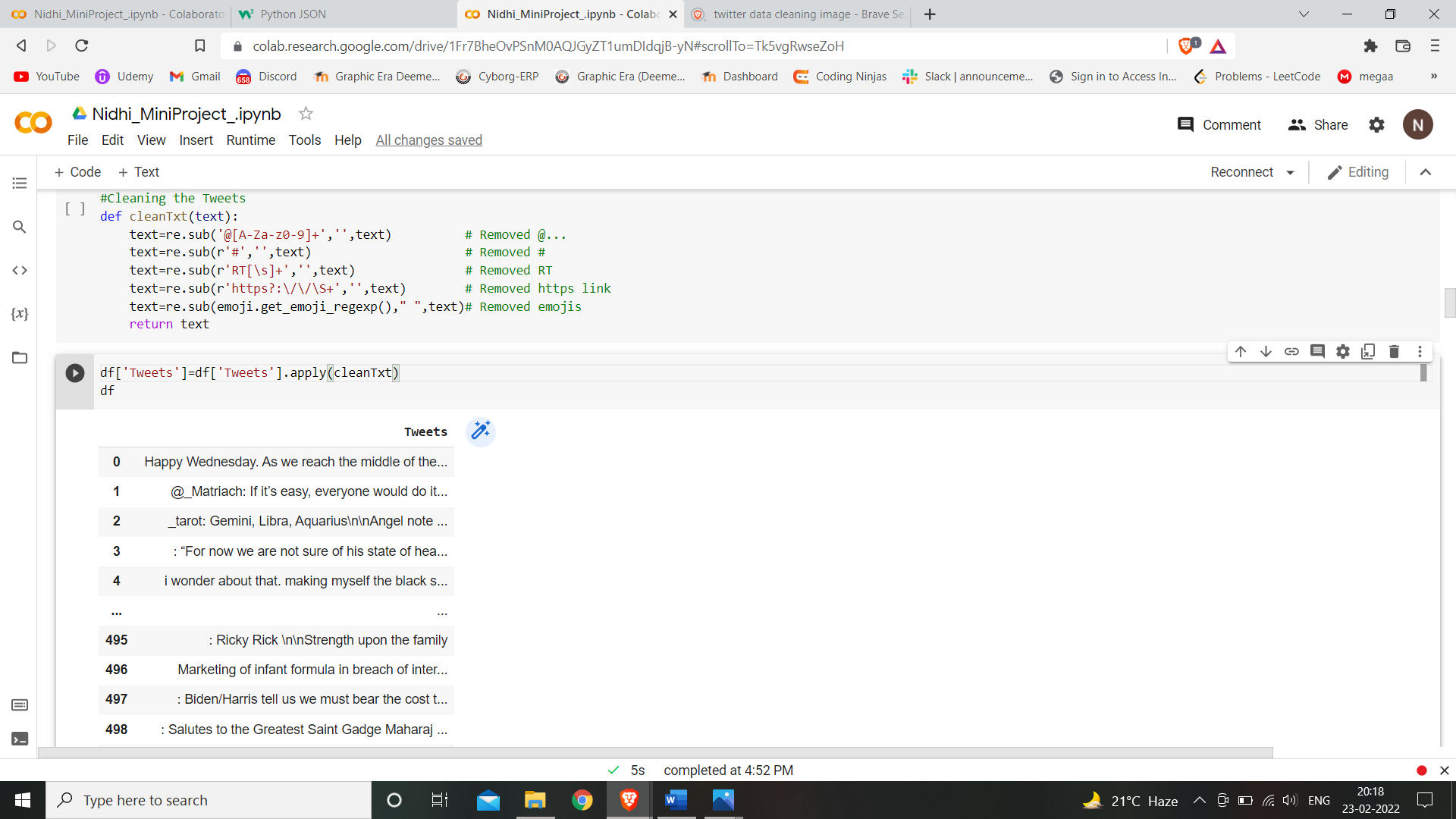
I created a dataframe using pandas library with a column tweets

**DATA CLEANING AND TWEETS PROCESSING**

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

The ultimate goal is to clean up the individual tweets.

To remove the mentions, #, and emojis, I created a function cleanTxt(text) which uses re library.



To make the cleaning more efficient, I converted all the tweets to lower case, removed the punctuation marks, or any irrelevant character and also removed stop words from the tokens, by using stop words library.

Stop words are the commonly used words which are irrelevant in text analysis like I, am, you, are, etc.

Additionally, I used a concept known as “Lemmatization”. This is a process of returning words to their “base” form.

|  |  |
| --- | --- |
| **BEFORE CLEANING** | **AFTER CLEANING** |
|  |  |

Note that, the more you clean your data, the more effective and accurate your result will be.

**DATA EXPLORATION**

**WORD CLOUD**

It is a visual representation of text data, which is often used to depict keyword metadata.

Using the WordCloud library, you can generate a Word Cloud based on the word frequency and superimpose these words on any image. In this case, I used a rectangular block and Matplotlib to display the image. The Word Cloud shows, words with higher frequency in bigger text size while “not-so” common words are in smaller text sizes.

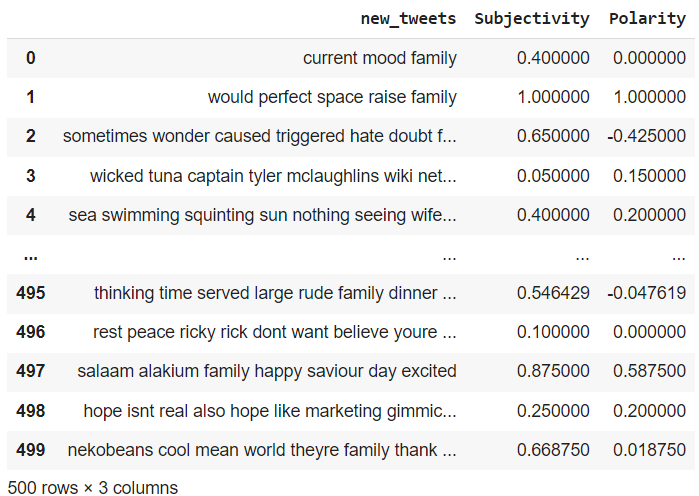
It can also be used to check whether our cleaning was successful or not, by taking a look at word cloud and seeing if the words make any sense or not.

**SENTIMENTAL ANALYSIS**

For this analysis, I went with TextBlob. Text Blob analyzes sentences by giving each tweet a Subjectivity and polarity score. Based on the Plolarity score, one can define which tweets were Positive, Negative, or Neutral.

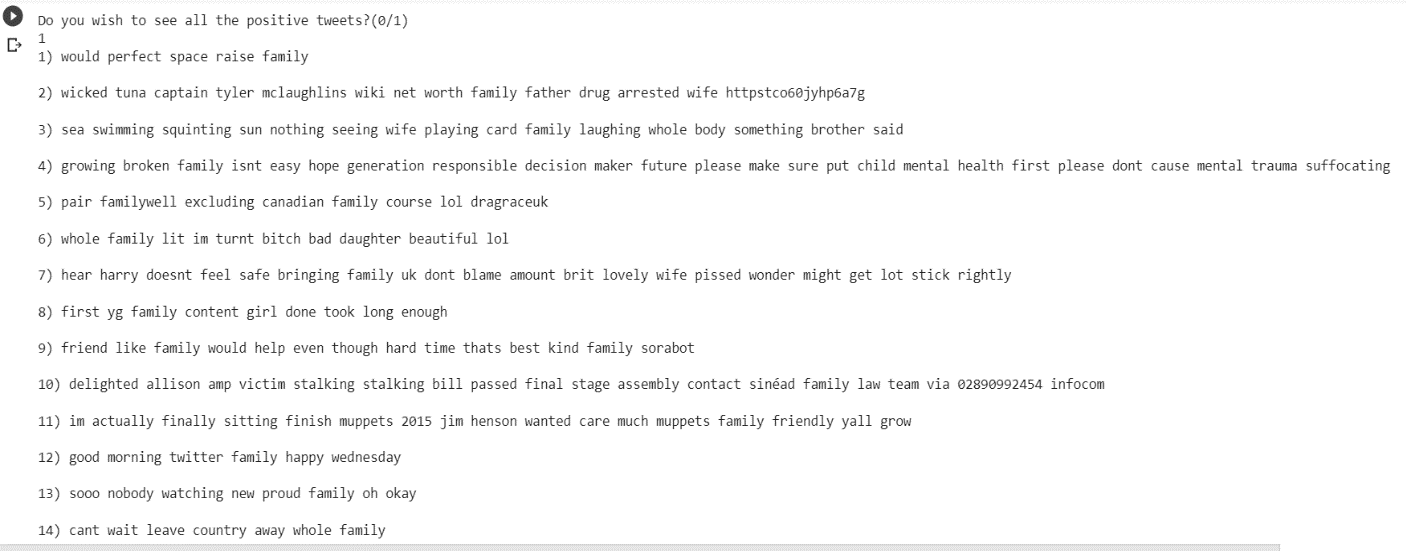
Polarity simply means emotions expressed in a sentence. Emotions are closely related to sentiments. The strength of a sentiment is typically linked to the intensity of certain emotions, e,g, joy and anger.

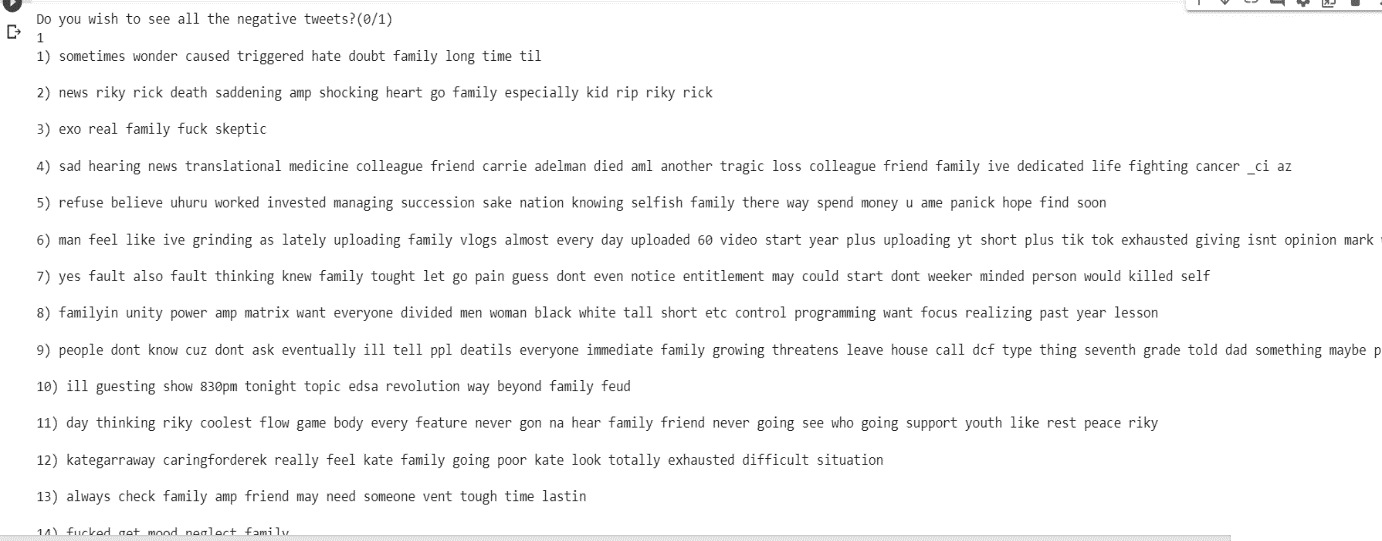
Subjectivity, subjective sentence expresses some personal feelings, views, or beliefs. A subjective sentence may not express any sentiment.

I created two columns of subjectivity and polarity in my dataframe.

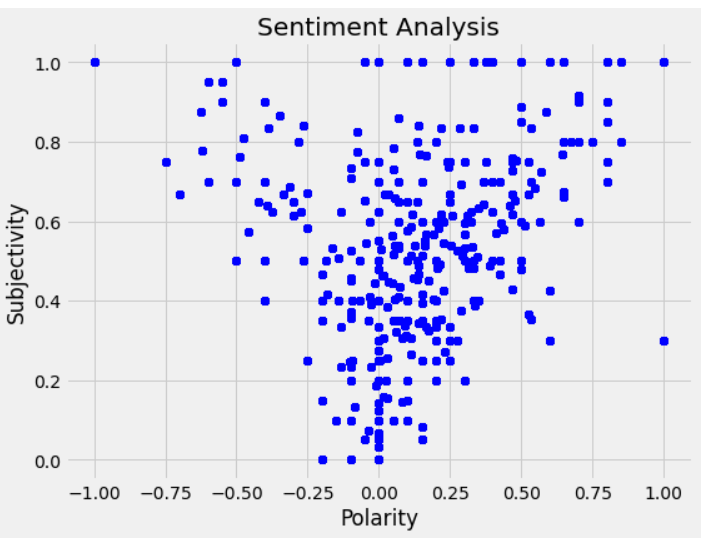
A polarity score of < 0 is Negative, 0 is Neutral while>0 is Positive. I used the “apply” method on the “Polarity” column in my dataframe to return the respective sentiment category. And create a column “Analysis”.

Now, to check whether the analysis has been done properly or not I have asked the user if he/she wish to see all the positive/negative tweets or not.

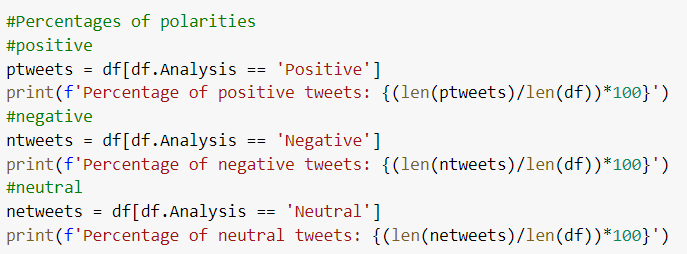
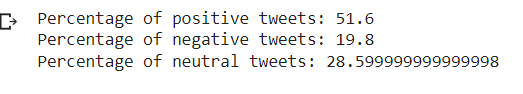
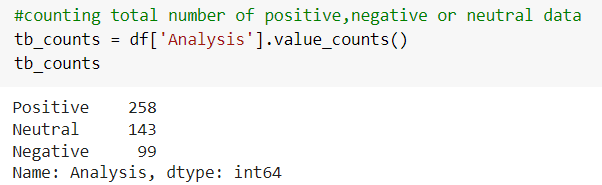


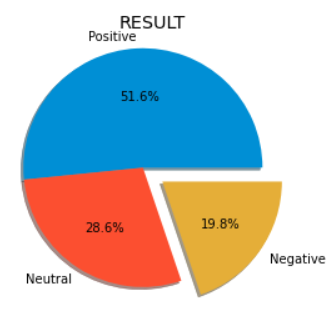
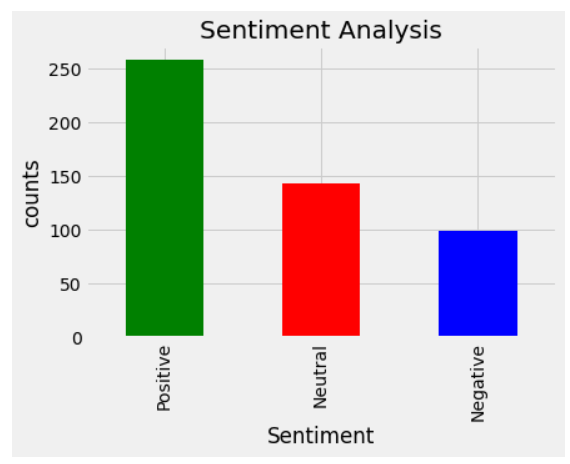


If you press 1 all the positive or negative tweets will be printed.

**POLARITY AND SUBJECTIVITY GRAPH**

**CALCULATING THE PERCENTAGE AND NUMBER OF POSITIVE, NEGATIVE, NEUTRAL TWEETS**

****

**DISTRIBUTION OF SENTIMENTS CATEGORY**

**Tweet sentiments:** The proportion of the sentiment categories shows that people love their families no matter what. 51.6% tweets are positive, which, is a real good proportion.