REPORT

On

→ Twitter SentiMental Analysis ←

Project Work On Course :-

PYTHON PROGRAMMING (INT 213)

Name	**CH RAHUL REDDY**
Registration Number	**12011984**
Program	**B.Tech (CSE)**
Semester	**3rd (Third)**
School	**Computer Science and Engineering**
University	**Lovely Professional University**
Submission Date	**30th NOVEMBER 2021**

PROJECT OUTLINE:

Twitter SentiMental Analysis Using Python.

Tools / Platform:

1	Operating System	WINDOWS 11
2	Language Used	PYTHON
3	Software Used	Anaconda
4	IDE Used	JUPYTER Notebook

GIT-HUB :

LINK:

Twitter_Sentimental_Analysis

Table of Contents

ABSTRACT

The Objective of my project is to analyze the sentiment of twitter. Social media Websites have emerged as one of the platforms to raise users opinions

The primary aim of the project is to extract Tweets and perform Sentimental Analysis on them and scoring them accordingly and dividing them into +ve ,-ve and neutral.

Sentiment analysis deals with identifying and classifying opinions or sentiments expressed in source text. Social media is generating a vast amount of sentiment rich data in the form of tweets, status updates, blog posts etc. Sentiment analysis of this user generated data is very useful in knowing the opinion of the crowd.

Twitter sentiment analysis is difficult compared to general sentiment analysis due to the presence of slang words and misspellings.

The maximum limit of characters that are allowed in Twitter is 140.

Twitter is one of the social media that is gaining popularity. Twitter offers organizations a fast and effective way to analyze customers' perspectives toward the critical to success in the marketplace.

Developing a program for sentiment analysis is an approach to be used to computationally measure customers' perceptions. This paper reports on the design of a sentiment analysis, extracting a vast amount of tweets. Prototyping is used in this development. Results classify customers' perspective via tweets into positive and negative,

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my teacher as well as **mentor- Prof. UPINDER KAUR Mam for her advice and inputs on this project.who spent countless hours listening and providing feedback.** *who gave me the golden opportunity to do this wonderful project on the topic* ***Twitter Sentiment Analysis*** *which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to them.*

Secondly I would also like to thank my parents and friends who helped me a lot in finalizing this project within the limited time frame.

Thank you for emboldening me.

INTRODUCTION

3.1 | CONTEXT

This project has been done as part of my course INT213 in my 3rd (Third) Semester at Lovely Professional University .

Supervised by Prof . UPINDER KAUR mam, I have 2 months to fulfill the requirements in order to succeed in this module.

The online medium has become a significant way for people to express their opinions and with social media, there is an abundance of opinion information available. Using sentiment analysis, the polarity of opinions can be found, such as positive, negative, or neutral by analyzing the text of the opinion.

3.2 | MOTIVATION

Recently, certain events, which affected the Government, have been triggered using the Internet. The social networks are being used to bring together people so as to organize mass gatherings and oppose oppression. On the darker side, the social networks are being used to insinuate people against an ethnic group or class of people, which has resulted in a serious loss of life. Thus, there is a need for Sentiment Analysis systems that can identify such phenomena and curtail them if needed.

Since the consumers have started using the power of the Internet to expand their horizons, there has been a surge of review sites and blogs, where users can perceive a product's or service's advantages and faults. These opinions thus shape the future of the product or the service.

3.3 | IDEA

To understand the public opinion on any organization, you'll have to analyze the news and emotion about it as well.
Nowadays, the age of the Internet has changed the way people express their views, opinions. It is now mainly done through blog posts, online forums, product review websites, social media ,etc. Nowadays, millions of people are using social network sites like Facebook, Twitter, Google Plus, etc. to express their emotions, opinion and share views about their daily lives. Through the online communities, we get an interactive media where consumers inform and influence others through forums.

Social media is generating a large volume of sentiment rich data in the form of tweets, status updates, blog posts, comments, reviews, etc. Moreover, social media provides an opportunity for businesses by giving a platform to connect with their customers for advertising. People mostly depend upon user generated content over online to a great extent for decision making. For e.g. if someone wants to buy a product or wants to use any service, then they firstly look up its reviews online, discuss it on social media before taking a decision. The amount of content generated by users is too vast for a normal user to analyze. So there is a need to automate this, various sentiment analysis techniques are widely used

Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics.Being able to quickly see the sentiment behind everything from forum posts to news articles means being better able to strategise and plan for the future

CONTRIBUTIONS

(solo CONTRIBUTIONS)

↳ RAHUL REDDY CHIDIPUDI ↰

(1). Planning Project :

- Initiation of project.
- Gathering requirement information.
- Installing Software.

(2). Sentimental Analysis:

- Perform Analysis on Data. (Tweets)
- Sorting of Data.
  (cleaning without any hyperlink and special characters)
- Calculating % of +ve and -ve tweets.

(3). Extract Data from Twitter:

- Accessing Twitter Developer Account.
- Using API KEYS.
- Generating required APIS and accessing them for the use of the project.

(4). Coding:

- **Installing required LIBRARIES.**
- **Writing Code.**
- **Writing Comments and Steps along with Code.**
- **Display of Correct Outputs and word cloud visualization**
- **Plotting of Tweets.**

(5). GIT-HUB Maintenance:

- **Creating Repositories.**
- **Uploading files from jupyter Notebook.**

(6). Report Writing:

- **Generating Report after Execution of code.**
- **Used Google Docs.**

(7). PPT Making:

- **Used Microsoft Powerpoint.**
- **Used Animations and much more for attractive representation.**

LIBRARIES

[1].Tweepy :

Tweepy is better suited for collecting a richer set of metadata, allows for flexibility and potentially scalability as well for those using the official API.

Tweepy provides the convenient Cursor interface to iterate through different types of objects. Twitter allows a maximum of 3200 tweets for extraction. These all are the prerequisites that have to be used before getting tweets of a user.

[2].TextBlob :

When you give a new text for analysis, it uses NaiveBayes classifier to classify the new text's polarity in +ve and -ve probabilities.

TextBlob is the perfect library for you to get hands-on with. The best way to go through this article is to follow along with the code and perform the tasks.

[3].WordCloud :

Word Clouds are visual displays of text data – simple text analysis. Word Clouds display the most prominent or frequent words in a body of text

Word clouds have recently become a staple of data visualization. They are especially popular when analyzing text.

The larger a word's size in the cloud, the more frequently it is used. Creating a word cloud helps engage, educate, and quickly capture your audience's attention.

Word clouds allow viewers to quickly assess which text is the most popular or used in the selected content.

[4]. Pandas :

Pandas is the most popular python library that is used for data analysis. We will provide highly optimized performance with back-end source code with the use of Pandas.

We can use Pandas to perform various tasks like filtering your data according to certain conditions, or segmenting and segregating the data according to preference, etc.

[5]. Numpy :

NumPy is a general-purpose array-processing package.

It provides a high performance multidimensional array object, and tools for working with these arrays.It is the fundamental package for scientific computing with Python. As the whole project is based on whole complex stats ,we will use this fast calculations and provide results.

[6]. MatPlotLib.pyplot :

Matplotlib tries to make easy things easy and hard things possible. We will generate plots, histograms, scatterplots, etc.,to make our project more appealing and easier to understand

Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy.

PROPOSED MODULES

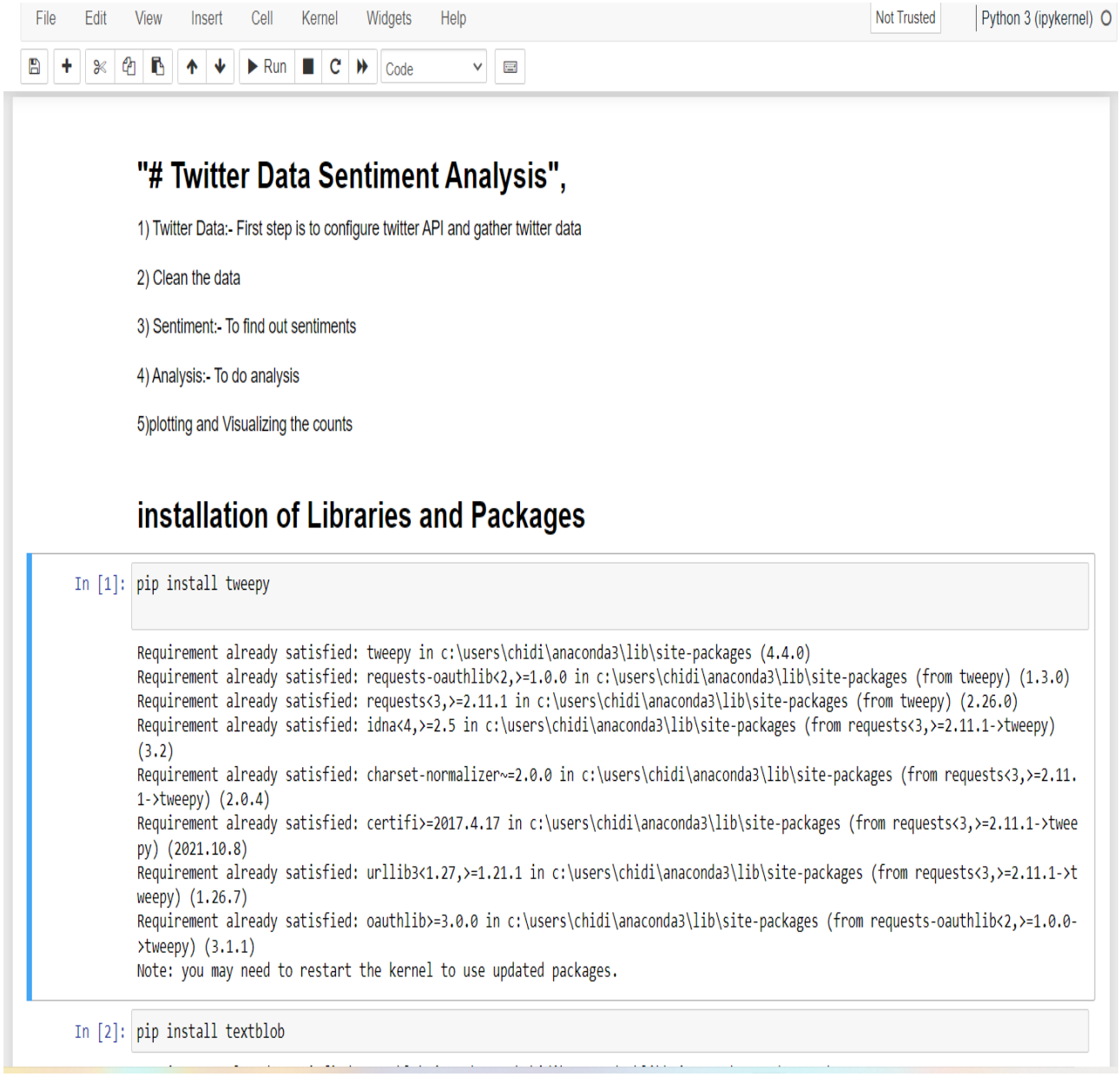
"Twitter Data Sentiment Analysis"

In this project, we try to implement a Twitter sentiment analysis model that helps to overcome the challenges of identifying the sentiments of the tweets. The necessary details regarding the dataset are:

- Import Necessary Dependencies
- Read and Load the Dataset
- Exploratory Data Analysis
- Data Visualization of Target Variables
- Data Preprocessing
- Splitting our data into Train and Test Subset
- Transforming Dataset using TF-IDF Vectorizer
- Function for Model Evaluation

CODE SCREENSHOTS

(WITH OUTPUT)



In [2]:
```
pip install textblob
```

```
Requirement already satisfied: textblob in c:\users\chidi\anaconda3\lib\site-packages (0.17.1)
Requirement already satisfied: nltk>=3.1 in c:\users\chidi\anaconda3\lib\site-packages (from textblob) (3.6.5)
Requirement already satisfied: click in c:\users\chidi\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (8.0.3)
Requirement already satisfied: joblib in c:\users\chidi\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (1.1.0)
Requirement already satisfied: regex>=2021.8.3 in c:\users\chidi\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (2021.
8.3)
Requirement already satisfied: tqdm in c:\users\chidi\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (4.62.3)
Requirement already satisfied: colorama in c:\users\chidi\anaconda3\lib\site-packages (from click->nltk>=3.1->textblob) (0.4.4)
Note: you may need to restart the kernel to use updated packages.
```

In [11]:
```
pip install wordcloud
```

```
Requirement already satisfied: wordcloud in c:\users\chidi\anaconda3\lib\site-packages (1.8.1)Note: you may need to restart the
kernel to use updated packages.
Requirement already satisfied: pillow in c:\users\chidi\anaconda3\lib\site-packages (from wordcloud) (8.4.0)
Requirement already satisfied: numpy>=1.6.1 in c:\users\chidi\anaconda3\lib\site-packages (from wordcloud) (1.20.3)
Requirement already satisfied: matplotlib in c:\users\chidi\anaconda3\lib\site-packages (from wordcloud) (3.4.3)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\chidi\anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.
3.1)
Requirement already satisfied: cycler>=0.10 in c:\users\chidi\anaconda3\lib\site-packages (from matplotlib->wordcloud) (0.10.0)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\chidi\anaconda3\lib\site-packages (from matplotlib->wordcloud)
(2.8.2)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\chidi\anaconda3\lib\site-packages (from matplotlib->wordcloud) (3.
0.4)
Requirement already satisfied: six in c:\users\chidi\anaconda3\lib\site-packages (from cycler>=0.10->matplotlib->wordcloud) (1.
16.0)
```

In [3]:
```
!pip install pandas
```

```
Requirement already satisfied: pandas in c:\users\chidi\anaconda3\lib\site-packages (1.3.4)
Requirement already satisfied: pytz>=2017.3 in c:\users\chidi\anaconda3\lib\site-packages (from pandas) (2021.3)
Requirement already satisfied: numpy>=1.17.3 in c:\users\chidi\anaconda3\lib\site-packages (from pandas) (1.20.3)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\chidi\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\chidi\anaconda3\lib\site-packages (from python-dateutil>=2.7.3->pandas) (1.
16.0)
```

In [4]:
```
pip install numpy
```

```
Requirement already satisfied: numpy in c:\users\chidi\anaconda3\lib\site-packages (1.20.3)
Note: you may need to restart the kernel to use updated packages.
```

Code begins

In [12]:
```
# Import the libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt


# Twitter Api Credentials
APIkey= "6e9NdMl4faDsuKZnzMeeUs7Of"
APISecreatKey= "SeJqKOeSp4tCZGYWqFhPxdlCoV5z3hWp5rGblwEsV5LdlGemxO"
accessToken= "1434032678970429441-XiwwPoHT4Gj6yZEn2hAJOD1vnfIjEE"
accessTokenSecreat= "VKfK3kTPSvWYqsWjlmoAfCnvPd7Wi1JV5nqkj4XN6mXul"

# Create the authentication object
authenticate = tweepy.OAuthHandler(APIkey,APISecreatKey)
# Set the access token and access token secret
authenticate.set_access_token(accessToken,accessTokenSecreat)
# Creating the API object while passing in auth information
api= tweepy.API(authenticate)
```

Extract and Analyze the Data

the 10 most recent tweets on screen_name = "Rahul"

In [31]:
```
# Extract 100 tweets from the twitter user
posts = api.user_timeline(screen_name="Rahul", count = 100,lang ="en", tweet_mode="extended")
i=1
for tweet in posts[:10]:
    print(str(i)+ ')'+ tweet.full_text +'\n')
    i=i+1
```

```
Unexpected parameter: lang

1)RT @roryreckons: People often ask me how I tell the difference between ADHD and Autism. It's hard - I know ADHD people ask this more, so he…

2)"we fixed it, now it's alphabetic and inclusive of non-binary people"

🙂 https://t.co/OlyN2Dj5C1

3)what kind of energy supermix is this, youtube https://t.co/VNqFaaHeMP

4)@bombsfall this thread in Reader mode is next level

5)https://t.co/3sQmAKyTfp

6)RT @DavidRickmann: This thread is ova rated.

7)@unknownrelic 😠

8)@imdanielholt @trulyaliem just as well, that house has no corners to sob in anyway
```

In [32]:
```
# Create a dataframe with a column called Tweets
df = pd.DataFrame([tweet.full_text for tweet in posts], columns=['Tweets'])
df
```

Out[32]:

	Tweets
0	RT @roryreckons: People often ask me how I tel...
1	"we fixed it, now it's alphabetic and inclusiv...
2	what kind of energy supermix is this, youtube ...
3	@bombsfall this thread in Reader mode is next ...
4	https://t.co/3sQmAKyTfp
...	...
95	RT @irondavy: Submitted for your approval: my ...
96	@aboodman I remember asking chrome devrel for ...
97	RT @PennyRed: 'Normal life will come back slow...
98	@aboodman @ErikArvidsson This is literally one...
99	@tesseralis https://t.co/FBNkyQlj8Z

100 rows × 1 columns

Create a function to clean the tweets

In [33]:
```
def cleanTxt(text):
 text = re.sub('@[A-Za-z0-9]+', '', text) #Removing @mentions
 text = re.sub('#', '', text) # Removing '#' hash tag
 text = re.sub('RT[\s]+', '', text) # Removing RT
 text = re.sub('https?:\/\/\S+', '', text) # Removing hyperlink

 return text


# Clean the tweets
df['Tweets'] = df['Tweets'].apply(cleanTxt)

# Show the cleaned tweets
df
```

Out[33]:

	Tweets
0	: People often ask me how I tell the differenc...
1	"we fixed it, now it's alphabetic and inclusiv...
2	what kind of energy supermix is this, youtube
3	this thread in Reader mode is next level
4	
...	...
95	: Submitted for your approval: my TOP FIFTY de...
96	I remember asking chrome devrel for the jQuer...
97	: 'Normal life will come back slow and strange...
98	This is literally one of my interview questi...
99	

100 rows × 1 columns

Sentimental Analysis

on a general statetement (test case)

In [36]:
```
analysis=TextBlob("Today was the wonderful  day")

analysis.sentiment
```

Out[36]:
```
Sentiment(polarity=1.0, subjectivity=1.0)
```

In [37]:
```
# Create a function to get the subjectivity
def getSubjectivity(text):
   return TextBlob(text).sentiment.subjectivity

# Create a function to get the polarity
def getPolarity(text):
   return  TextBlob(text).sentiment.polarity


# Create two new columns 'Subjectivity' & 'Polarity'
df['Subjectivity'] = df['Tweets'].apply(getSubjectivity)
df['Polarity'] = df['Tweets'].apply(getPolarity)

# Show the new dataframe with columns 'Subjectivity' & 'Polarity'
df
```

Out[37]:

	Tweets	Subjectivity	Polarity
0	: People often ask me how I tell the differenc...	0.520833	0.104167
1	"we fixed it, now it's alphabetic and inclusiv...	0.200000	0.100000
2	what kind of energy supermix is this, youtube	0.900000	0.600000
3	this thread in Reader mode is next level	0.000000	0.000000
4		0.000000	0.000000
...	...	...	...
95	: Submitted for your approval: my TOP FIFTY de...	0.600000	0.166667
96	I remember asking chrome devrel for the jQuer...	1.000000	-0.750000
97	: 'Normal life will come back slow and strange...	0.365000	0.060000
98	This is literally one of my interview questi...	0.700000	0.800000
99		0.000000	0.000000

100 rows × 3 columns

word cloud visualization

In [48]:

```
# word cloud visualization
allWords = ' '.join([i for i in df['Tweets']])
Cloud = WordCloud(width=2000, height=1000, random_state=33, max_font_size=333).generate(allWords)


plt.imshow(Cloud, interpolation="bilinear")
plt.show()
```

Analysis of Tweets

In [39]:
```
# Create a function to compute negative (-1), neutral (0) and positive (+1) analysis
def getAnalysis(score):
   if score <0:
      return 'Negative'
   elif score == 0:
       return 'Neutral'
   else:
        return 'Positive'

df['Analysis'] = df['Polarity'].apply(getAnalysis)
# Show the dataframe
df
```

Out[39]:

	Tweets	Subjectivity	Polarity	Analysis
0	: People often ask me how I tell the differenc...	0.520833	0.104167	Positive
1	"we fixed it, now it's alphabetic and inclusiv...	0.200000	0.100000	Positive
2	what kind of energy supermix is this, youtube	0.900000	0.600000	Positive
3	this thread in Reader mode is next level	0.000000	0.000000	Neutral
4		0.000000	0.000000	Neutral
...	...	...	...	...
95	: Submitted for your approval: my TOP FIFTY de...	0.600000	0.166667	Positive
96	I remember asking chrome devrel for the jQuer...	1.000000	-0.750000	Negative
97	: 'Normal life will come back slow and strange...	0.365000	0.060000	Positive
98	This is literally one of my interview questi...	0.700000	0.800000	Positive
99		0.000000	0.000000	Neutral

100 rows × 4 columns

% of +ve and -ve Tweets

In [44]:
```
# Print the percentage of positive tweets
ptweets = df[df.Analysis == 'Positive']
ptweets = ptweets['Tweets']
ptweets

round( (ptweets.shape[0] / df.shape[0]) * 100 , 1)
```

Out[44]: 39.0

In [45]:
```
# Print the percentage of negative tweets
ntweets = df[df.Analysis == 'Negative']
ntweets = ntweets['Tweets']
ntweets

round( (ntweets.shape[0] / df.shape[0]) * 100, 1)
```

Out[45]: 14.0

The most positive tweet is the 1 tweet.

In [40]:
```
# Printing positive tweets
print('Printing positive tweets:\n')
j=1
sortedDF = df.sort_values(by=['Polarity']) #Sort the tweets
for i in range(0, sortedDF.shape[0] ):
  if( sortedDF['Analysis'][i] == 'Positive'):
    print(str(j) + ') '+ sortedDF['Tweets'][i])
    print()
    j= j+1
```

```
Printing positive tweets:

1) : People often ask me how I tell the difference between ADHD and Autism. It's hard - I know ADHD people ask this more, so
he…

2) "we fixed it, now it's alphabetic and inclusive of non-binary people"

🙂

3) what kind of energy supermix is this, youtube

4) The consultant got out her notepad and checked a box. This facility was operating effectively and would meet its counter-p
roductivity targets for the quarter.

Yes, this would do. Another successful day as a DEI consultant on the books.

5) “What are they doing?”

The supervisor shrugged. “Not much. It keeps them occupied. That lever’s the key.”
```

The most negative tweet is the #1 tweet.

In [41]:
```
# Printing negative tweets
print('Printing negative tweets:\n')
j=1
sortedDF = df.sort_values(by=['Polarity'],ascending=False) #Sort the tweets
for i in range(0, sortedDF.shape[0] ):
  if( sortedDF['Analysis'][i] == 'Negative'):
    print(str(j) + ') '+sortedDF['Tweets'][i])
    print()
    j=j+1
```

```
Printing negative tweets:

1)  no but I might have to now!! this is just my average wednesday afternoon

2)   my favourite part about this was that it made me wonder what other field i could spend a month learning about only to have everything around me in the world that related to it ruined

3) She watched as one of the operators pulled back on their lever. A package popped out, dropped onto a conveyor belt, and was deposited a few rows down, where it disappeared into another worker's kiosk. She turned to the supervisor.

4) Looking out over the factory floor, the distraction, exhaustion and indifference consultant saw rows and rows of workers, pa
cked tightly together, staring at kiosks displaying shifting imagery. Each station had a button, a lever and a basket.

5) : Awaken the unholiest queen, ready to corrupt the core of your absolute being. 💋👑🩸
Samhain2021 HappyHalloween2021
(1/3…

6)  the concept of letting go of irrational annoyances :/

7) Stringer (shaking his head): "And if you don't?"

Prof: "Reduce price and increase market share."

Stringer: "That assumes low overhead."
```

COUNTS OF +VE, -VE AND NEUTRAL

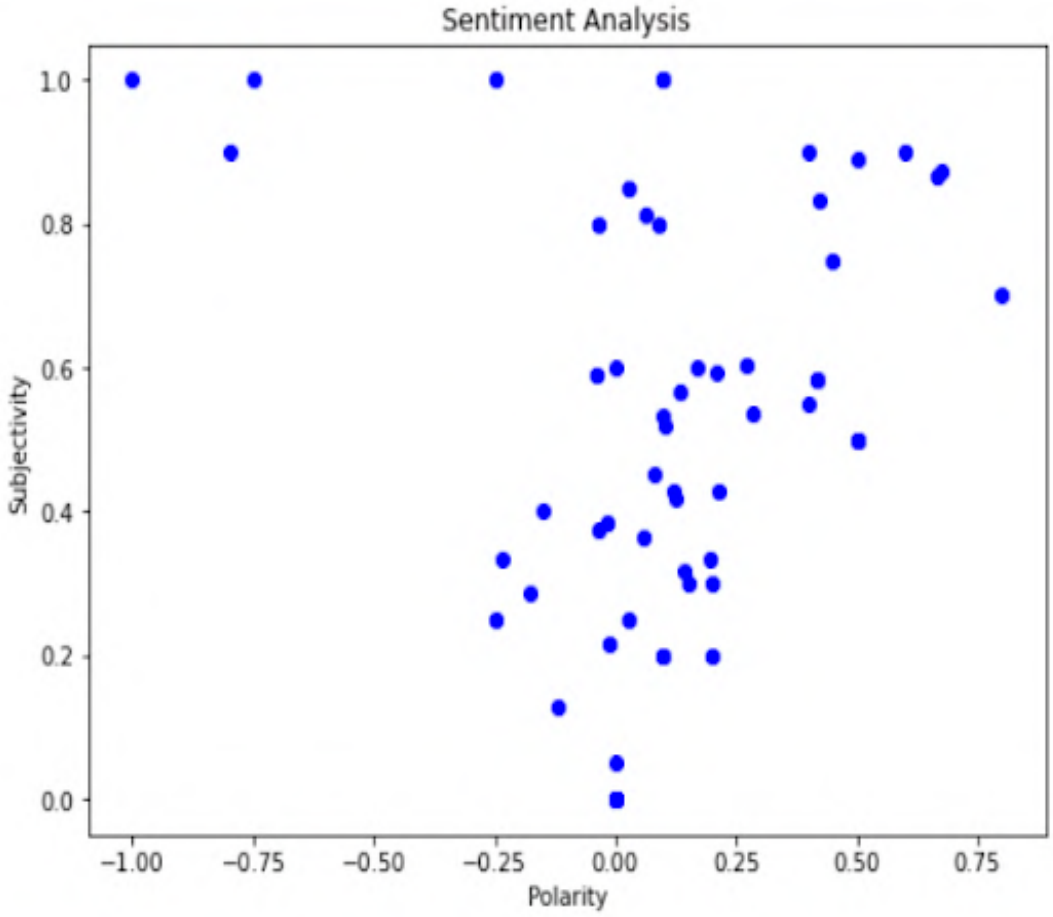
In [42]:
```
# Show the value counts
df['Analysis'].value_counts()
```

Out[42]:
```
Neutral     47
Positive    39
Negative    14
Name: Analysis, dtype: int64
```

Plotting of tweets

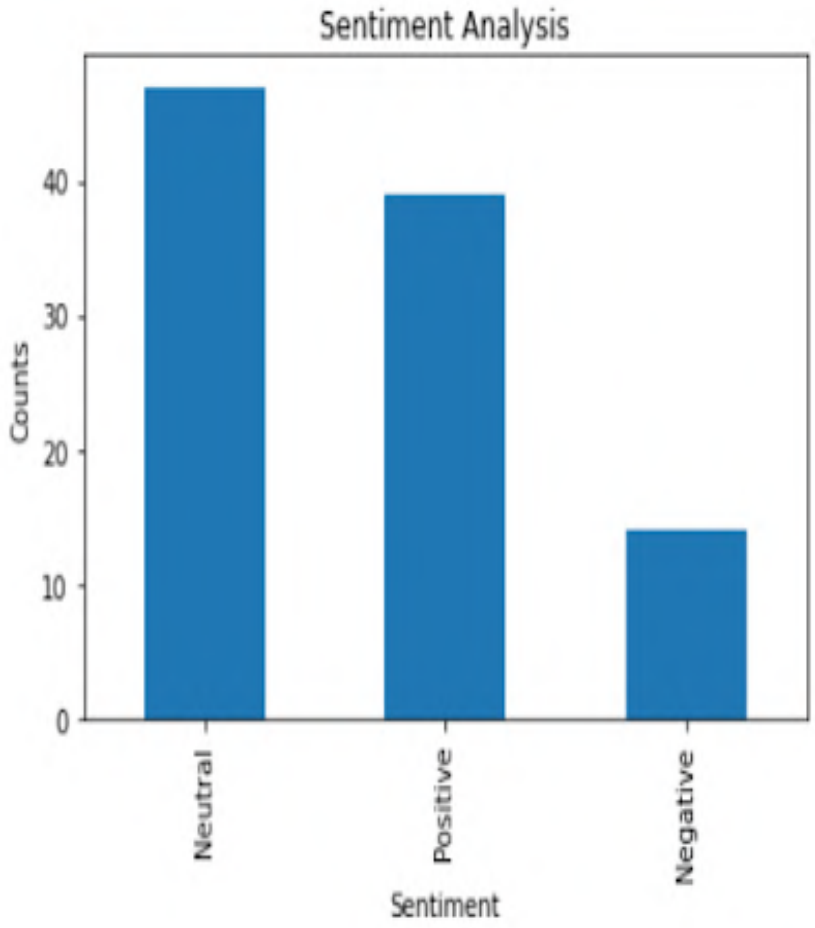
according to (subjectivity and polarity)

In [43]:
```
# Plotting
plt.figure(figsize=(8,6))
for i in range(0, df.shape[0]):
  plt.scatter(df["Polarity"][i], df["Subjectivity"][i], color='Blue')
# plt.scatter(x,y,color)
plt.title('Sentiment Analysis')
plt.xlabel('Polarity')
plt.ylabel('Subjectivity')
plt.show()
```



Polarity

In [47]:
```
#Plotting and visualizing the counts
plt.title('Sentiment Analysis')
plt.xlabel('Sentiment')
plt.ylabel('Counts')
df['Analysis'].value_counts().plot(kind = 'bar')
plt.show()
```



CONCLUSION

Therefore, sentiment analysis is a hot topic in machine learning. We are still far from detecting the sentiments of the corpus of texts very accurately because of the complexity in the English language and even more if we consider other languages such as Chinese, Hindi,korean etc..,

In this project I started with the basics of Twitter Data Analysis. OAuth & Tweepy were used for the authentication from Twitter App (API Keys)

I tried to show the basic way of classifying tweets into positive or negative categories using Naive Bayes as a baseline and how language models are related to the Naive Bayes and can produce better results.

REFRENCES


WIKIPEDIA
The Free Encyclopedia


YouTube


ResearchGate


ANACONDA


jupyter


word
clouds