Project Report On

TWITTER SENTIMENT ANALYSIS

Submitted in partial fulfillment of requirements for the course

CSE18R468-Big Data Analytics

Bachelor's of Technology In Computer Science and Engineering

DONE BY

V.Nishitha-9920004347

Under the Guidance

DR.S.WILSON PRAKASH

Kalasalingam Academy of Research and Education
Anand Nagar, Krishnankoil-626126.
Department of Computer Science and Engineering

ABSTRACT

Twitter Sentiment analysis is the process of understanding how people are using Twitter and how it affects people's lives and it will tell that tweet is positive(or)negative(or)neutral. Python and Power BI are two powerful tools that can help in this process.

DECLARATION

I hereby declare that the work presented in this report entitled "Project name", in partial fulfilment of the requirements for the course CSE18R468-Big Data Analytics and submitted in Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education (Deemed to be University) is an authentic record of our own work carried out during the period from DEC 2023 under the guidance of DR.S.WILSON PRAKASH.

The work reported in this has not been submitted by me for the award of any other degree of this or any other institute.

V.Nishitha-9920004347

ACKNOWLEDGEMENT

I express my deep gratitude to DR.S.WILSON PRAKASH for their valuable guidance throughout my training.

V.Nishitha-9920004347

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

Twitter Sentiment analysis is the process of understanding how people are using Twitter and how it affects people's lives and it will tell that tweet is positive(or)negative(or)neutral. Python and Power BI are two powerful tools that can help in this process.

The first step in performing Twitter analysis is to collect the data. This is done by using the Twitter API or by scraping the website. The data collected can include tweets, Polarity, Subjectivity and Score.

Once the data is collected, it must be cleaned and organized. This involves removing any irrelevant or duplicate data, and formatting the data for analysis.

Chapter 2

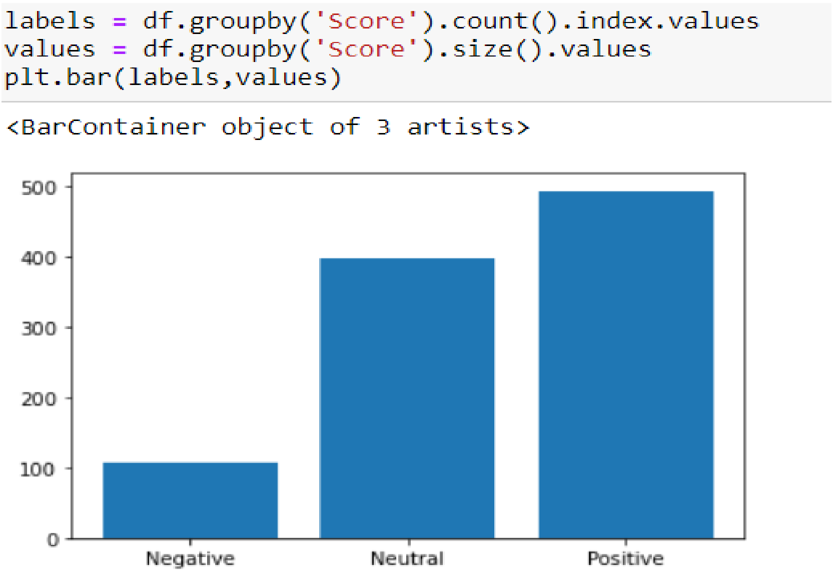
ABOUT PROJECT

Large Twitter datasets will be analysed using Python and Power BI for this project. The data must be pre-processed to remove pertinent information like user IDs, hashtags, and timestamps. A list of popular subjects, trends, and the tone of the tweets will be the result in the end

Chapter 3

OUTPUT SCREENSHOTS

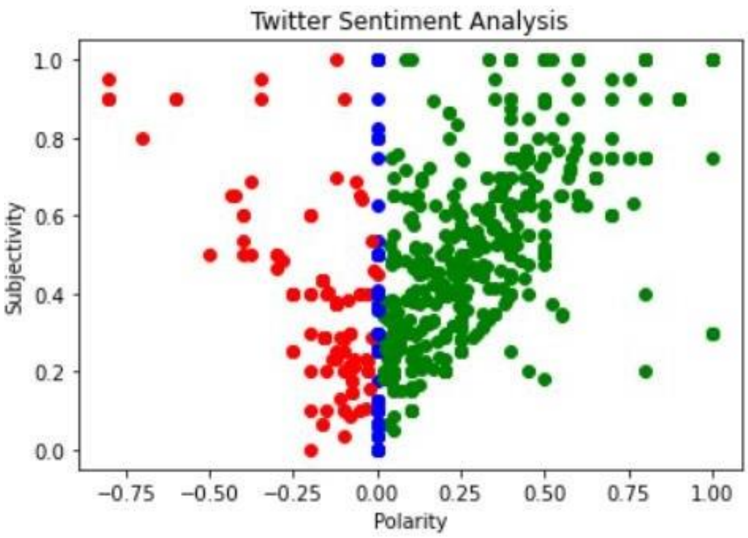
BAR GRAPH



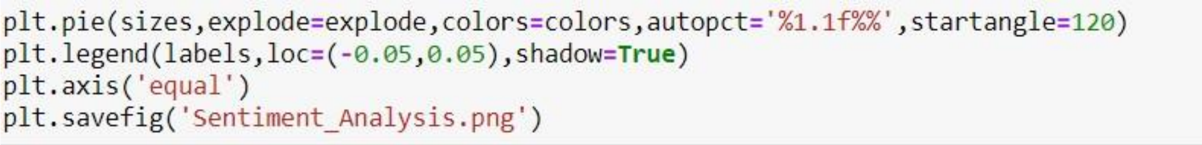
```
labels = df.groupby('Score').count().index.values
values = df.groupby('Score').size().values
plt.bar(labels,values)
```

```
<BarContainer object of 3 artists>
```

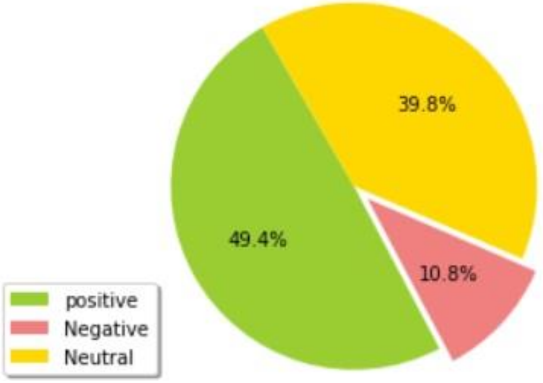
SCATTER PLOT



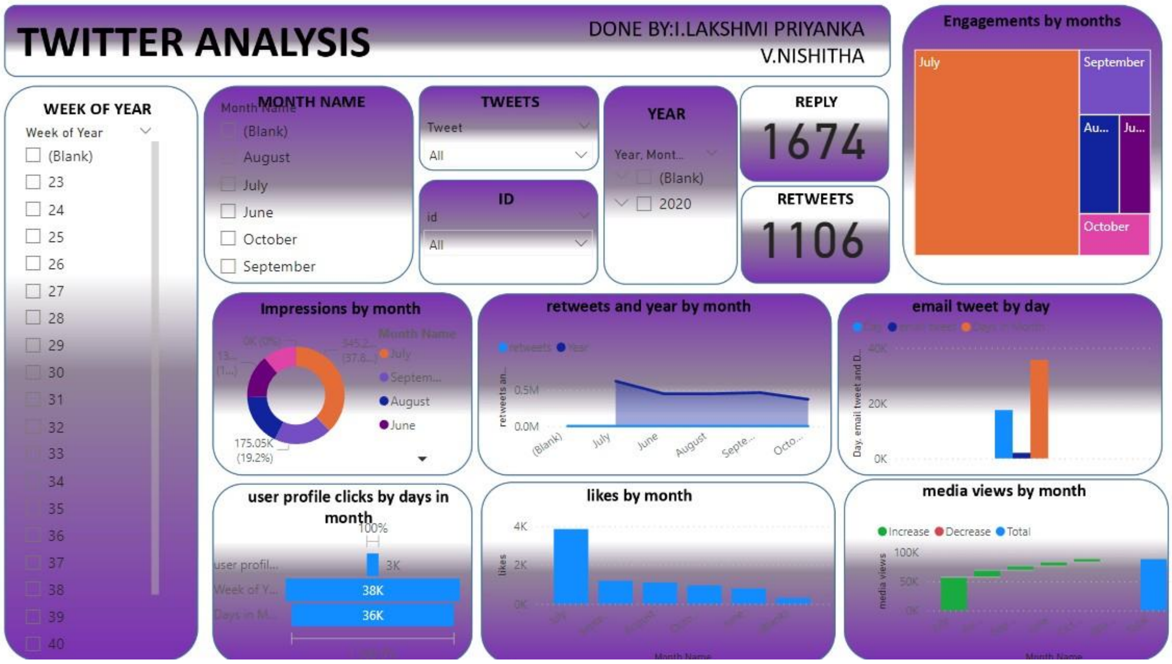
PIE PLOT



```
plt.pie(sizes,explode=explode,colors=colors,autopct='%1.1f%%',startangle=120)
plt.legend(labels,loc=(-0.05,0.05),shadow=True)
plt.axis('equal')
plt.savefig('Sentiment_Analysis.png')
```



POWER BI DASHBOARD



TWITTER ANALYSIS

DONE BY:I.LAKSHMI PRIYANKA
V.NISHITHA

REPLY 1674

RETWEETS 1106

Chapter 4

CONCLUSION

Twitter analysis using Python and Power BI is a powerful way to gain insights into Tweet, subjectivity, polarity. By collecting, cleaning, analysing, processing, and visualizing the data, it is possible to gain valuable insights into if the tweet is positive or negative or neutral.

APPENDICES

Source Code

```
!pip install tweepy
!pip install textblob

import tweepy
from textblob import TextBlob
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt

config = pd.read_csv("C:/Users/86395/Desktop/Config.csv")

twitterApiKey= config['twitterApiKey'][0]
twitterApiSecret=config['twitterApiSecret'][0]
twitterApiAccessToken=config['twitterApiAccessToken'][0]
twitterApiAccessTokenSecret=config['twitterApiAccessTokenSecret']［0］

auth=tweepy.OAuthHandler(twitterApiKey,twitterApiSecret)
auth.set_access_token(twitterApiAccessToken,twitterApiAccessTokenSecret)
twitterApi=tweepy.API(auth,wait_on_rate_limit=True)

twitterAccount = 'KamalaHarris'
```

```
tweets = tweepy.Cursor(twitterApi.user_timeline,
screen_name=twitterAccount,
count=None,
since_id=None,
max_id=None,train_user=True,exclude_replies=True,contributor_details=False,
include_entities=False).items(1000);

df.head()


def cleanUpTweet(txt):
   txt = re.sub(r'@[A-Za-z0-9]+',",txt)
   txt = re.sub(r'#',",txt)
   txt = re.sub(r'RT : ',",txt)
   txt = re.sub(r'https?:\/\/[A-Za-z0-9\.\/]+',",txt)
   return txt
df['Tweet']=df['Tweet'].apply(cleanUpTweet)
def getTextSubjectivity(txt):
   return TextBlob(txt).sentiment.subjectivity
def getTextPolarity(txt):
   return TextBlob(txt).sentiment.polarity
df['Subjectivity']=df['Tweet'].apply(getTextSubjectivity)
df['Polarity']=df['Tweet'].apply(getTextPolarity)
```

```
df.head(1000)


def getTextAnalysis(a):
    if a<0:
        return "Negative"
    elif a==0:
        return "Neutral"
    else:
        return "Positive"
df['Score']=df['Polarity'].apply(getTextAnalysis)


positive=df[df['Score']=='Positive']
print(str(positive.shape[0]/(df.shape[0])*100)+'% of positive tweets')
pos=positive.shape[0]/df.shape[0]*100


negative=df[df['Score']=='Negative']
print(str(negative.shape[0]/(df.shape[0])*100)+'% of Negative tweets')
neg=negative.shape[0]/df.shape[0]*100


Neutral=df[df['Score']=='Neutral']
print(str(Neutral.shape[0]/(df.shape[0])*100)+'% of Neutral tweets')
```

```
neutral=Neutral.shape[0]/df.shape[0]*100

explode=(0,0.1,0)
labels='positive','Negative','Neutral'
sizes=[pos,neg,neutral]
colors=['yellowgreen','lightcoral','gold']
```

Box plot:

```
plt.pie(sizes,explode=explode,colors=colors,autopct='%1.1f%%',startangle=120)
plt.legend(labels,loc=(-0.05,0.05),shadow=True)
plt.axis('equal')
plt.savefig('Sentiment_Analysis.png')
```

Bar graph:

```
labels = df.groupby('Score').count().index.values
values = df.groupby('Score').size().values
plt.bar(labels,values)
```

Scatter Plot:

```
for index, row in df.iterrows():
    if row['Score']=='Positive':

plt.scatter(row['Polarity'],row['Subjectivity'],color='green')
    elif row['Score']=='Negative':
        plt.scatter(row['Polarity'],row['Subjectivity'],color='red')
    elif row['Score']=='Neutral':
        plt.scatter(row['Polarity'],row['Subjectivity'],color='blue')
plt.title('Twitter Sentiment Analysis')
plt.xlabel('Polarity')
plt.ylabel('Subjectivity')
plt.show()
```