

# Social Media Sentiment Analysis(Twitter)

## Overview:

Social media sentiment analysis is the process of analyzing user-generated content from platforms like Twitter to understand public sentiment towards specific topics, products, or events. This project leverages Natural Language Processing (NLP) techniques to preprocess text data, extract sentiment scores, and visualize sentiment trends over time.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import re
import string
import nltk
nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer

# Load the dataset
df=pd.read_csv("Twitter Sentiments.csv")
df

[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!

{"summary":{"\n  \"name\": \"df\", \n  \"rows\": 31962, \n  \"fields\": [\n    {\n      \"column\": \"id\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 9226, \n        \"min\": 1, \n        \"max\": 31962, \n        \"num_unique_values\": 31962, \n        \"samples\": [\n          12228, \n          14710, \n          19320\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"label\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 0, \n        \"min\": 0, \n        \"max\": 1, \n        \"num_unique_values\": 2, \n        \"samples\": [\n          1, \n          0\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"tweet\", \n      \"properties\": {\n        \"dtype\": \"string\", \n        \"num_unique_values\": 29530, \n        \"samples\": [\n          \"even if you like #windows 10, you should be at #microsoft\" #computer #spam #spyware #unreliable #os #ms #win \", \n          \"the countdown begins! 10 hours till the musical! \", \n          \"itsthefinalcountdown #youngcarers #rctcbc #rctcouncil #rct #carersweek \", \n          \"\n        ]\n      }\n    }\n  ]\n}}
```

```

{"semantic_type": "\n",\n      "description": "\n\n      }\n      }\n      ]\n      }", "type": "dataframe", "variable_name": "df"}

df.head() # read the first 5 rows

{"summary": "{\n  \"name\": \"df\",\n  \"rows\": 31962,\n  \"fields\": [\n    {\n      \"column\": \"id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 9226,\n        \"min\": 1,\n        \"max\": 31962,\n        \"num_unique_values\": 31962,\n        \"samples\": [\n          12228,\n          14710,\n          19320\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\n\n      }\n      },\n      {\n        \"column\": \"label\",\n        \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 0,\n          \"min\": 0,\n          \"max\": 1,\n          \"num_unique_values\": 2,\n          \"samples\": [\n            1,\n            0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\n\n      }\n      },\n      {\n        \"column\": \"tweet\",\n        \"properties\": {\n          \"dtype\": \"string\",\n          \"num_unique_values\": 29530,\n          \"samples\": [\n            \"even if you like #windows 10, you should be at #microsoft\" #computer #spam #spyware #unreliable #os #ms #win \",\n            \"the countdown begins! 10 hours till the musical! \",\n            \"itsthefinalcountdown #youngcarers #rctcbc #rctcouncil #rct #carersweek \",\n            \"\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\n\n      }\n      }\n      ]\n      }", "type": "dataframe", "variable_name": "df"}

df.tail() # read the last 5 rows

{"summary": "{\n  \"name\": \"df\",\n  \"rows\": 5,\n  \"fields\": [\n    {\n      \"column\": \"id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 1,\n        \"min\": 31958,\n        \"max\": 31962,\n        \"num_unique_values\": 5,\n        \"samples\": [\n          31959,\n          31962,\n          31960\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\n\n      }\n      },\n      {\n        \"column\": \"label\",\n        \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 0,\n          \"min\": 0,\n          \"max\": 1,\n          \"num_unique_values\": 2,\n          \"samples\": [\n            1,\n            0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\n\n      }\n      },\n      {\n        \"column\": \"tweet\",\n        \"properties\": {\n          \"dtype\": \"string\",\n          \"num_unique_values\": 5,\n          \"samples\": [\n            \" to see nina turner on the airwaves trying to wrap herself in the mantle of a genuine hero like shirley chisolm. #shame #imwithher\",\n            \"thank you @user for you follow \",\n            \"\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\n\n      }\n      }\n      ]\n      }", "type": "dataframe"}

df.info() # Display Datatype info

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 31962 entries, 0 to 31961
Data columns (total 3 columns):
#   Column   Non-Null Count  Dtype
---  -
0   id        31962 non-null  int64
1   label     31962 non-null  int64
2   tweet     31962 non-null  object
dtypes: int64(2), object(1)
memory usage: 749.2+ KB

df.isnull().sum()

id        0
label     0
tweet     0
dtype: int64

df.shape

(31962, 3)

# Text cleaning function
def clean_text(text):
    text = text.lower()
    text = re.sub(r'http\S+', '', text) # Remove URLs
    text = re.sub(r'@\w+', '', text)   # Remove mentions
    text = re.sub(r'[^a-zA-Z\s]', '', text) # Remove special
characters
    return text.strip()
data['cleaned_text'] = data['text'].apply(clean_text)

```

## 1. Remove pattern in the input text

```

def remove_pattern(input_txt, pattern):
    r = re.findall(pattern, input_txt)
    for word in r:
        input_txt = re.sub(word, "", input_txt)
    return input_txt

df

{"summary": "{\n  \"name\": \"df\",\n  \"rows\": 31962,\n  \"fields\": [\n    {\n      \"column\": \"id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 9226,\n        \"min\": 1,\n        \"max\": 31962,\n        \"num_unique_values\": 31962,\n        \"samples\": [\n          12228,\n          14710,\n          19320\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"label\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 0,\n        \"min\": 0,\n        \"max\": 0,\n        \"num_unique_values\": 1,\n        \"samples\": [\n          0\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    }\n  ]\n}"

```

```

{"min\\": 0,\\n          \\\"max\\": 1,\\n          \\\"num_unique_values\\": 2,\\n
\\\"samples\\": [\\n          1,\\n          0\\n          ],\\n
\\\"semantic_type\\": \\\"\\\",\\n          \\\"description\\": \\\"\\\"\\n          }\\n
n      },\\n      {\\n          \\\"column\\": \\\"tweet\\\",\\n          \\\"properties\\": {\\n
n          \\\"dtype\\": \\\"string\\\",\\n          \\\"num_unique_values\\":
29530,\\n          \\\"samples\\": [\\n          \\\"\\\"\\\"even if you like
#windows 10, you should be at #microsoft\\\" #computer #spam
#spyware #unreliable #os #ms #win \\\",\\n          \\\"the countdown
begins! 10 hours till the musical! \\u00f0\\u009f\\u0098\\u008a\\
u00f0\\u009f\\u0098\\u0084\\u00f0\\u009f\\u0098\\u0085
#itsthefinalcountdown #youngcarers #rctcbc #rctcouncil #rct
#carersweek \\u00f0\\u009f\\u0098\\u0081\\\"\\n          ],\\n
\\\"semantic_type\\": \\\"\\\",\\n          \\\"description\\": \\\"\\\"\\n          }\\n
n      }\\n      ]\\n}\\", \"type\": \"dataframe\", \"variable_name\": \"df\"}

```

## 2. Remove the twitter handles (@user)

```

df['clean_tweet'] = np.vectorize(remove_pattern)(df['tweet'], "@[\\w]*")

df.head()

{"summary": "{\\n  \\\"name\\": \\\"df\\\",\\n  \\\"rows\\": 31962,\\n  \\\"fields\\":
[\\n    {\\n      \\\"column\\": \\\"id\\\",\\n      \\\"properties\\": {\\n
\\\"dtype\\": \\\"number\\\",\\n      \\\"std\\": 9226,\\n      \\\"min\\": 1,\\n
\\\"max\\": 31962,\\n      \\\"num_unique_values\\": 31962,\\n
\\\"samples\\": [\\n        12228,\\n        14710,\\n        19320\\n
],\\n      \\\"semantic_type\\": \\\"\\\",\\n      \\\"description\\": \\\"\\\"\\n
}\\n    },\\n    {\\n      \\\"column\\": \\\"label\\\",\\n      \\\"properties\\":
{\\n      \\\"dtype\\": \\\"number\\\",\\n      \\\"std\\": 0,\\n
\\\"min\\": 0,\\n      \\\"max\\": 1,\\n      \\\"num_unique_values\\": 2,\\n
\\\"samples\\": [\\n        1,\\n        0\\n        ],\\n
\\\"semantic_type\\": \\\"\\\",\\n      \\\"description\\": \\\"\\\"\\n      }\\n
n    },\\n    {\\n      \\\"column\\": \\\"tweet\\\",\\n      \\\"properties\\": {\\n
n      \\\"dtype\\": \\\"string\\\",\\n      \\\"num_unique_values\\":
29530,\\n      \\\"samples\\": [\\n        \\\"\\\"\\\"even if you like
#windows 10, you should be at #microsoft\\\" #computer #spam
#spyware #unreliable #os #ms #win \\\",\\n      \\\"the countdown
begins! 10 hours till the musical! \\u00f0\\u009f\\u0098\\u008a\\
u00f0\\u009f\\u0098\\u0084\\u00f0\\u009f\\u0098\\u0085
#itsthefinalcountdown #youngcarers #rctcbc #rctcouncil #rct
#carersweek \\u00f0\\u009f\\u0098\\u0081\\\"\\n        ],\\n
\\\"semantic_type\\": \\\"\\\",\\n      \\\"description\\": \\\"\\\"\\n      }\\n
n    },\\n    {\\n      \\\"column\\": \\\"clean_tweet\\\",\\n
\\\"properties\\": {\\n      \\\"dtype\\": \\\"string\\\",\\n
\\\"num_unique_values\\": 29528,\\n      \\\"samples\\": [\\n        \\\"
says who?! your iq is so low u have 2stand on a highstool 2raise your
iq. we both know #blm is a\\u00e2\\u0080\\u00a6 \\\",\\n        \\\"i'll

```

```
be there!! \\u00f0\\u009f\\u0091\\u008d\\u00f0\\u009f\\u008f\\u00bd
#ical6 \\n      ],\\n      \\\"semantic_type\\\": \\\"\\\",\\n
\\\"description\\\": \\\"\\\"\\n      }\\n      ]\\n
n}\\\", \"type\": \"dataframe\", \"variable_name\": \"df\"}
```

### 3. Remove a special characters, numbers and punctuations

```
df['clean_tweet'] = df['clean_tweet'].str.replace("[^a-z A-Z#]", " ")
df.head()

{"summary": "{\\n  \\\"name\\\": \\\"df\\\",\\n  \\\"rows\\\": 31962,\\n  \\\"fields\\\":
[\\n    {\\n      \\\"column\\\": \\\"id\\\",\\n      \\\"properties\\\": {\\n
\\\"dtype\\\": \\\"number\\\",\\n      \\\"std\\\": 9226,\\n      \\\"min\\\": 1,\\n
\\\"max\\\": 31962,\\n      \\\"num_unique_values\\\": 31962,\\n
\\\"samples\\\": [\\n        12228,\\n        14710,\\n        19320\\n
],\\n      \\\"semantic_type\\\": \\\"\\\",\\n      \\\"description\\\": \\\"\\\"\\n
}\\n    },\\n    {\\n      \\\"column\\\": \\\"label\\\",\\n      \\\"properties\\\":
{\\n        \\\"dtype\\\": \\\"number\\\",\\n        \\\"std\\\": 0,\\n
\\\"min\\\": 0,\\n        \\\"max\\\": 1,\\n        \\\"num_unique_values\\\": 2,\\n
\\\"samples\\\": [\\n        1,\\n        0\\n      ],\\n
\\\"semantic_type\\\": \\\"\\\",\\n      \\\"description\\\": \\\"\\\"\\n    }\\n
n    },\\n    {\\n      \\\"column\\\": \\\"tweet\\\",\\n      \\\"properties\\\": {\\n
n      \\\"dtype\\\": \\\"string\\\",\\n      \\\"num_unique_values\\\":
29530,\\n      \\\"samples\\\": [\\n        \\\"\\\"\\\"even if you like
#windows 10, you should be at #microsoft\\\" #computer #spam
#spyware #unreliable #os #ms #win \\\",\\n        \\\"the countdown
begins! 10 hours till the musical! \\u00f0\\u009f\\u0098\\u008a\\
u00f0\\u009f\\u0098\\u0084\\u00f0\\u009f\\u0098\\u0085
#itsthefinalcountdown #youngcarers #rctcbc #rctcouncil #rct
#carersweek \\u00f0\\u009f\\u0098\\u0081\\\"\\n      ],\\n
\\\"semantic_type\\\": \\\"\\\",\\n      \\\"description\\\": \\\"\\\"\\n    }\\n
n    },\\n    {\\n      \\\"column\\\": \\\"clean_tweet\\\",\\n
\\\"properties\\\": {\\n        \\\"dtype\\\": \\\"string\\\",\\n
\\\"num_unique_values\\\": 29528,\\n        \\\"samples\\\": [\\n
\\n        says who?! your iq is so low u have 2stand on a highstool 2raise your
iq. we both know #blm is a\\u00e2\\u0080\\u00a6 \\\",\\n        \\\"i'll
be there!! \\u00f0\\u009f\\u0091\\u008d\\u00f0\\u009f\\u008f\\u00bd
#ical6 \\\"\\n      ],\\n      \\\"semantic_type\\\": \\\"\\\",\\n
\\\"description\\\": \\\"\\\"\\n    }\\n    }\\n  ]\\n
n}\\\", \"type\": \"dataframe\", \"variable_name\": \"df\"}
```

## 4. Remove a twitter handles (@user)

```
df['clean_tweet'] = df['clean_tweet'].apply(lambda x : " ".join([w for w in x.split() if (len(w)>2)]))
df.head()
```

```
{"summary":{"\n  \"name\": \"df\",\n  \"rows\": 31962,\n  \"fields\": [\n    {\n      \"column\": \"id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 9226,\n        \"min\": 1,\n        \"max\": 31962,\n        \"num_unique_values\": 31962,\n        \"samples\": [\n          12228,\n          14710,\n          19320\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"label\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 0,\n        \"min\": 0,\n        \"max\": 1,\n        \"num_unique_values\": 2,\n        \"samples\": [\n          1,\n          0\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"tweet\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 29530,\n        \"samples\": [\n          \"even if you like #windows 10, you should be at #microsoft\" #computer #spam #spyware #unreliable #os #ms #win \",\n          \"the countdown begins! 10 hours till the musical! \",\n          \"itsthefinalcountdown #youngcarers #rctcbc #rctcouncil #rct #carersweek \",\n          \"personalised new... gbp 6.99 get here: #shop #cool #home #fun\",\n          \"one our happy troopers with his fun sock day school #fun #troopers #sil...\",\n          \"factsguide:, society, now, #motivation\"\n        ]\n      }\n    }\n  ],\n  \"column\": \"clean_tweet\",\n  \"properties\": {\n    \"dtype\": \"string\",\n    \"num_unique_values\": 28568,\n    \"samples\": [\n      \"personalised new... gbp 6.99 get here: #shop #cool #home #fun\",\n      \"one our happy troopers with his fun sock day school #fun #troopers #sil...\",\n      \"factsguide:, society, now, #motivation\"\n    ]\n  }\n}, \"type\": \"dataframe\", \"variable_name\": \"df\"}
```

## Individual words are considered tokens

```
tokenized_tweet = df['clean_tweet'].apply(lambda x : x.split())
tokenized_tweet.head()
```

```
0    [when, father, dysfunctional, and, selfish, dr...
1    [thanks, for, #lyft, credit, can't, use, cause...
2                                [bihday, your, majesty]
3    [#model, love, take, with, all, the, time, urð...
4                                [factsguide:, society, now, #motivation]
Name: clean_tweet, dtype: object
```

## Stem the word in the sentence

```
from nltk.stem.porter import PorterStemmer # Import the Porter
Stemmer for stemming words
stemmer = PorterStemmer() # Create an instance of the Porter Stemmer

# Apply stemming to each word in each sentence of the tokenized_tweet
column
tokenized_tweet = tokenized_tweet.apply(lambda sentence:
[stemmer.stem(word) for word in sentence])

# Display the first few rows after stemming
tokenized_tweet.head()

0    [when, father, dysfunct, and, selfish, drag, h...
1    [thank, for, #lyft, credit, can't, use, caus, ...
2                                [bihday, your, majesti]
3    [#model, love, take, with, all, the, time, urđ...
4                                [factsguide:, societi, now, #motiv]
Name: clean_tweet, dtype: object
```

## Combine words into single sentence

```
for i in range(len(tokenized_tweet)):
    tokenized_tweet[i] = " ".join(tokenized_tweet[i])

df['clean_tweet'] = tokenized_tweet
df.head()

{"summary":{"\n  \"name\": \"df\",\n  \"rows\": 31962,\n  \"fields\":
[\n    {\n      \"column\": \"id\",\n      \"properties\": {\n
\"dtype\": \"number\",\n      \"std\": 9226,\n      \"min\": 1,\n
\"max\": 31962,\n      \"num_unique_values\": 31962,\n
\"samples\": [\n        12228,\n        14710,\n        19320\n
],\n      \"semantic_type\": \"\",\n      \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"label\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 0,\n
\"min\": 0,\n        \"max\": 1,\n        \"num_unique_values\": 2,\n
\"samples\": [\n        1,\n        0\n      ],\n      \"semantic_type\": \"\",\n
      \"description\": \"\"\n    }\n  },\n  {\n      \"column\": \"tweet\",\n      \"properties\": {\n
n        \"dtype\": \"string\",\n        \"num_unique_values\":
29530,\n        \"samples\": [\n          \"even if you like
#windows 10, you should be at #microsoft\" #computer #spam
#spyware #unreliable #os #ms #win \",\n          \"the countdown
begins! 10 hours till the musical! \\\u00f0\\\u009f\\\u0098\\\u008a\\\
u00f0\\\u009f\\\u0098\\\u0084\\\u00f0\\\u009f\\\u0098\\\u0085
#itsthefinalcountdown #youngcarers #rctcbc #rctcouncil #rct
```

```
#carersweek  \\u00f0\\u009f\\u0098\\u0081\\n      ],\\n
\\\"semantic_type\\\": \\\"\\\",\\n      \\\"description\\\": \\\"\\\"\\n      }\\n
n    },\\n    {\\n      \\\"column\\\": \\\"clean_tweet\\\",\\n
\\\"properties\\\": {\\n      \\\"dtype\\\": \\\"string\\\",\\n
\\\"num_unique_values\\\": 28560,\\n      \\\"samples\\\": [\\n      \\\"all
day today. danceswithfilm #dwf19 #danceswithfilm #treadingyesterday
tcl\\u00e2\\u0080\\u00a6\\\",\\n      \\\"do more the thing that make
\\u00f0\\u009f\\u0098\\u0081\\u00f0\\u009f\\u0098\\u0081\\u00f0\\u
009f\\u0098\\u0081 getting\\u00e2\\u0080\\u00a6\\\"\\n      ],\\n
\\\"semantic_type\\\": \\\"\\\",\\n      \\\"description\\\": \\\"\\\"\\n      }\\n
n    }\\n  ]\\n}\\\", \"type\": \"dataframe\", \"variable_name\": \"df\"}
```

```
data = pd.read_csv('Twitter Sentiments.csv') # Ensure the correct
file path
print(data.head()) # Ensure 'data' is defined and contains text
```

	id	label	tweet
0	1	0	@user when a father is dysfunctional and is s...
1	2	0	@user @user thanks for #lyft credit i can't us...
2	3	0	bihday your majesty
3	4	0	#model i love u take with u all the time in ...
4	5	0	factsguide: society now #motivation

## Exploratory Data Analysis

```
!pip install pandas numpy matplotlib seaborn wordcloud vaderSentiment
```

```
Requirement already satisfied: pandas in
/usr/local/lib/python3.11/dist-packages (2.2.2)
Requirement already satisfied: numpy in
/usr/local/lib/python3.11/dist-packages (1.26.4)
Requirement already satisfied: matplotlib in
/usr/local/lib/python3.11/dist-packages (3.10.0)
Requirement already satisfied: seaborn in
/usr/local/lib/python3.11/dist-packages (0.13.2)
Requirement already satisfied: wordcloud in
/usr/local/lib/python3.11/dist-packages (1.9.4)
Collecting vaderSentiment
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl.metadata (572
bytes)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.11/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.11/dist-packages (from pandas) (2025.1)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.11/dist-packages (from pandas) (2025.1)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib) (1.3.1)
```



```

Requirement already satisfied: cycloper>=0.10 in
/usr/local/lib/python3.11/dist-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.11/dist-packages (from matplotlib) (4.56.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib) (1.4.8)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.11/dist-packages (from matplotlib) (24.2)
Requirement already satisfied: pillow>=8 in
/usr/local/lib/python3.11/dist-packages (from matplotlib) (11.1.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib) (3.2.1)
Requirement already satisfied: requests in
/usr/local/lib/python3.11/dist-packages (from vaderSentiment) (2.32.3)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2-
>pandas) (1.17.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests-
>vaderSentiment) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.11/dist-packages (from requests-
>vaderSentiment) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests-
>vaderSentiment) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests-
>vaderSentiment) (2025.1.31)
Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl (125 kB)
126.0/126.0 kB 1.8 MB/s eta
0:00:00
ent
Successfully installed vaderSentiment-3.3.2

```

## Visualize the frequent words

```

all_words = " ".join([sentence for sentence in df['clean_tweet']])

from wordcloud import WordCloud
wordcloud = WordCloud(width = 800, height = 500 , random_state = 42,
max_font_size = 100).generate(all_words)

# Plot the graph
plt.figure(figsize= (15,8))
plt.imshow(wordcloud, interpolation = 'bilinear')
plt.axis('off')
plt.show()

```



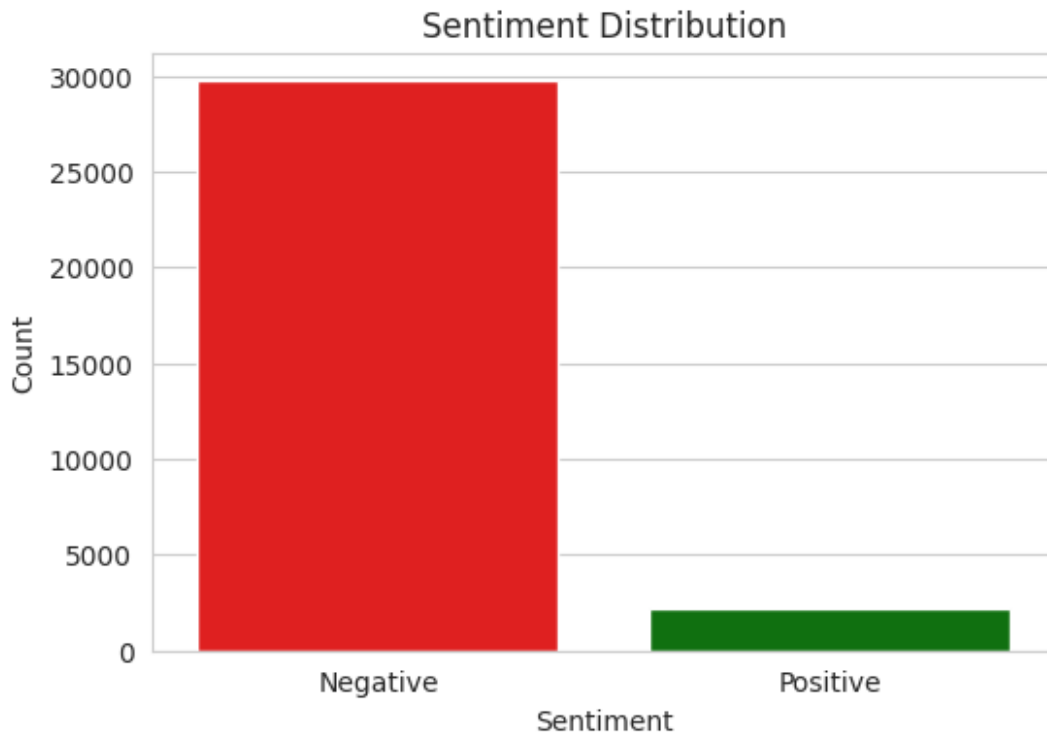






```
# Set plot style
sns.set_style("whitegrid")

# Sentiment distribution
plt.figure(figsize=(6,4))
sns.countplot(data=df, x='label', hue='label', palette=['red',
'green'], legend=False)
plt.xticks(ticks=[0,1], labels=['Negative', 'Positive'])
plt.title("Sentiment Distribution")
plt.xlabel("Sentiment")
plt.ylabel("Count")
plt.show()
```



```
#Check the column exists
```

```
df.columns
```

```
Index(['id', 'label', 'tweet', 'clean_tweet'], dtype='object')
```

## Extract the hashtag

```
def hashtag_extract(tweet):  
    hashtags = []  
    # Loop words in the tweet  
    for tweets in tweet:  
        ht = re.findall(r"#(\w+)", tweets)  
        hashtags.append(ht)  
    return hashtags
```

## 1. Extract hashtags from non racist/sexist tweets

```
ht_positive = hashtag_extract(df['clean_tweet'][df['label'] == 0])  
ht_positive
```

```
[['run'],
 ['lyft', 'disapoint', 'getthank'],
 [],
 ['model'],
 ['motiv'],
 ['allshowandnogo'],
 [],
 ['school', 'exam', 'hate', 'imagin', 'actorslif', 'revolutionschool',
 'girl'],
 ['allin', 'cav', 'champion', 'cleveland', 'clevelandcavali'],
 ['gr8'],
 ['ireland', 'blog', 'silver', 'gold', 'forex'],
 ['orlando',
 'standwithorlando',
 'pulseshoot',
 'orlandoshoot',
 'biggerproblem',
 'selfish',
 'heabreak',
 'valu',
 'love'],
 ['80day', 'gettingf'],
 ['got7', 'junior', 'yugyoem', 'omg'],
 ['thank', 'posit'],
 ['friday', 'cooki'],
 [],
 ['euro2016'],
 ['badday', 'coneofsham', 'cat', 'piss', 'funni', 'laugh'],
 ['wine', 'weekend'],
 ['tgif', 'fff', 'gamedev', 'indiedev', 'indiegamedev', 'squad'],
 ['upsideofflorida', 'shopalyssa', 'love'],
 ['smile', 'media', 'pressconfer', 'antalya', 'turkey', 'throwback'],
 ['ical6'],
 [],
 ['rip', 'orlando'],
 [],
 ['alohafriday', 'time', 'not', 'exist', 'positivevib', 'hawaiian'],
 [],
 ['goodnight', 'badmonday'],
 ['taylorswift1989'],
 ['travelingram', 'dalat', 'ripinkylif'],
 ['photoshop',
 'enoughisenough',
 'dontphotoshopeveryth',
 'wheresallthenaturalphoto'],
 ['cedarpoint'],
 ['thank', 'posit'],
 ['bookworm', 'ontothextnovel'],
 [],
 [],
```

```
[],  
['flow', 'grow', 'garden', 'iphonesia', 'bliss', 'bloom',  
'basilicabotanica'],  
['i_am', 'posit', 'affirm'],  
['model'],  
['whenev', 'someth', 'wrong'],  
['illustr'],  
[],  
['abc2020', 'pulseclub', 'prayfororlando'],  
['bihday', 'nose', 'job', 'bihday', 'petunia'],  
['albanpilgrimag'],  
[],  
['wedding'],  
[],  
['video', 'father', 'day', 'rayo', 'world', 'hotvideo', 'video'],  
['ascot', 'fashion', 'monochrom', 'style', 'instahappyday'],  
['selfi', 'yolo', 'xoxo', 'like4lik'],  
['work', 'mindset'],  
['christinarip', 'voic', 'christinagrimmi'],  
['roar', 'preschool', 'student', 'proudâ'],  
[],  
[],  
['chatiado'],  
[],  
[],  
['travel', 'yeah', 'thejourneybegin', 'helloâ'],  
['luv', 'hottweet', 'venusexchangeâ'],  
['awork', 'solut'],  
['orlando', 'disneygatorattack'],  
[],  
['proud'],  
['udtapunjab', 'amarinder', 'aap'],  
['senseaboutmath'],  
[],  
['golfstrengthandcondit', 'strong', 'felixfoisgolf'],  
['greathonour', 'careerconvo'],  
['innov', 'wateringhol', 'cave', 'mountaintop', 'campfir'],  
[],  
['heal', 'altwaystoh', 'healthi', 'peace'],  
[],  
[],  
['midweek', 'newmus', 'watchthisspac', 'guitar'],  
['up'],  
[],  
['snapchat', 'redhead', 'vermillionr'],  
['peopl'],  
['goodtim', 'histori'],  
['summeim', 'memori'],  
['memori'],  
[],
```

```
['model'],
[],
['hump', 'hump', 'humpersð'],
['shop', 'cool', 'home', 'fun'],
[],
[],
['nashvilletour2016', 'nashvilleforev'],
['eur', 'blog', 'silver', 'gold', 'forex'],
[],
['thank', 'posit'],
['christinagrimmi', 'prayfororlando'],
['eyelid'],
['vigilfororlando', 'harp', 'clonakilti', 'rip'],
['ihavenofriend', 'someonecomewithm'],
['weekend', 'relax', 'icon', 'woman', 'sundaymorn', 'sunday',
'marilynmonro'],
[],
['marvel', 'song', 'musica', 'weed', 'ripchristina', 'adel', 'danc'],
['under', 'blog', 'silver', 'gold', 'forex'],
['healthi', 'altwaystoheal'],
['model'],
['ofw', 'pinoy', 'followm', 'iger', 'instagood', 'smile', 'love'],
['toradora', 'anim', 'animeedit', 'breakup', 'alon', 'lone'],
[],
['cotd'],
[],
['footbal', 'snapchat'],
['fringes', 'qcbags', 'summ'],
['nochebuena', 'lasvega', 'usa'],
['instaboy',
'instaman',
'sefie',
'septum',
'friend',
'smiles',
'blackandwhit'],
['haiku', '3line', 'micropoetri'],
['jewelrymak', 'button'],
['weston'],
['eo', 'fun', 'cawaii', 'love', 'strawberri', 'tomato'],
['me', 'boy', 'guy'],
['social', 'analyt', 'photooftheday', 'iger', 'like'],
['doj', 'fbi', 'emailg', 'hillary2016', 'shame', 'disgrac',
'ridicul'],
['stupid', 'negligence', 'sachintendulkar'],
['â', '21thbihday', 'bestfriend', 'loveð'],
['enjoy', 'pooh4u'],
['chrisevan', 'bihdayâ'],
['florida'],
['newalbum', 'indi', 'guitar', 'echobelli'],
```



```
['tgif', 'webmaret', 'seo', 'commun', 'manag', 'weekend'],
['nzd', 'blog', 'silver', 'gold', 'forex'],
['bihdayweeksuck', 'bithday27', 'tâ'],
[],
['disneysmagickingdom', 'disney', 'magickingdom', 'disneyland',
'orlandoâ'],
[],
[],
[],
['healthi'],
['love', 'commit', 'trust', 'faith'],
['newarkfestiv'],
['mochithepug', 'puglif'],
[],
['justalil', 'butstillontim'],
['goodmorn', 'friday', 'girl', 'love', 'florida'],
['aap', 'punjabi', 'drugaddicts'],
['tracerequest', 'touch'],
['guy', 'mato', 'chapon'],
['friend'],
['model'],
['bloke'],
[],
['vfest', 'summer'],
['bihday', 'shane', 'watson', 'australia', 'cricker'],
['whoknows'],
['radiol', 'coldplay'],
[],
[],
['gloucesterrugbi'],
[],
['sad',
'togeth',
'alon',
'up',
'down',
'fine',
'dandi',
'lone',
'exuber',
'depress',
'halfful',
'halfempti'],
['mikeashley', 'sposdirectsham', 'workersright', 'euref', 'remain'],
['bird', 'stella', 'bad', 'princess'],
[],
['realli'],
[],
['orlando', 'heabreak'],
['runbeck', 'happi', 'quot', 'inspir'],
```

```
['acne', 'altwaystoh', 'healthi', 'healing'],
['beat', 'makingbeat', 'thursday', 'work', 'istandard', 'akaiproâ'],
['wed', 'weddingpay', 'love', 'forev', 'weddingdressâ'],
['cpd', 'onemoreday'],
['twitch', 'stream'],
['verbal', 'abus', 'hu', 'love', 'adult', 'teen'],
[],
['thank', 'posit'],
['dejeun', 'lunch', 'happi', 'love', 'amoureuxâ'],
['lawofattract', 'quiz', 'love', 'happi', 'motiv', 'success', 'joy',
'lifeâ'],
['pole', 'bgp', 'rozthediva', 'blackgirlspol', 'bodyandpol', 'live',
'love'],
['argentina2015'],
[],
['nohcarolina'],
['build'],
['ksleg'],
['orlando'],
['aly_alshahawi',
'ad',
'cute',
'summer',
'photographerdubai',
'sun',
'hair',
'mydubaiâ'],
[],
[],
['i_am', 'posit', 'affirm'],
['thank', 'posit'],
[],
['wimbledon2016'],
['success',
'model',
'entrepreneur',
'speaker',
'ff',
'fresh',
'confid',
'fashionâ'],
['gala', 'day', 'kid', 'bipolar', 'frikiegirl2016'],
[],
['amaz', 'dad', 'enjoy'],
['magnettherapi', 'altwaystoh', 'heal', 'healthi'],
[],
['smile', 'egsmil', 'torio'],
[],
[],
['bu', 'alamindah', 'white', 'music', 'carrieunderwood'],
```

```
['model'],
['5'],
['propey', 'ocean', 'summer', 'life', 'famili'],
['quot',
'quot',
'love',
'pakistan',
'allin216',
'girl',
'boy',
'fashion',
'feel'],
['thank', 'posit'],
['countdown', 'nolefam', '22092016'],
[],
[],
['wakow', 'buzz', 'moment', 'ride', 'love', 'crush', 'date',
'hookup', 'app'],
[],
['risay18', 'risaypl'],
['thursday'],
['i_am', 'posit', 'affirm'],
[],
['thank', 'bless', 'jeffreycwilliam'],
['buttru'],
['cute', 'smile', 'babygirl', 'beauti', 'niec', 'bless', 'xo'],
['t', 'oneteamlincoln'],
['rockthepark', 'timetogetmyjesuson'],
[],
['happi'],
['sanjapan', 'animeconvent'],
['tech', 'announc', 'moto_', 'question'],
['mondaymotiv', 'livelifetothe fullest', 'healthi', 'everydaycount'],
['classmat', 'friend'],
['brentwood', 'homesta'],
['kitten', 'cat', 'blackkittenâ'],
[],
[],
[],
['pharrellwilliam'],
['workâ'],
['gut', 'rosi'],
['fever'],
['mad', 'teeth', 'bite', 'piss', 'dog', 'bite', 'dogbit',
'dogtreat'],
['dreamcatch'],
['englishteach', 'teacher', 'ecuador'],
['strong', 'greatquotesâ'],
[],
['itako', 'alone', 'it'],
```

```
['letthembelittl', 'childhoodunplug'],
['calm', 'sunni', 'chubbygirl', 'glass', 'finnishgirl', 'finn'],
['yuhuuuuu', 'tmaâ'],
['nationalbestfriendsday', 'true'],
[],
['ugh'],
['model'],
['game', 'game', 'girlgam', 'voez'],
['weekend'],
[],
['beyou', 'ofbham'],
['fathersday', 'love'],
['æ'],
[],
['positivevib', 'fresh'],
['conjuring2', 'readyornot', 'truestori'],
[],
['iwearbanana'],
['scare', 'redditchbizhour'],
[],
['smile', 'corbinrayp'],
['scare', 'missyou', 'whatwasthat'],
['mondaymorn', 'mondaymotiv', 'monday', 'indraloka'],
['sanasafinaz', 'noexchangepolici'],
['ending'],
['littlebear',
'joy',
'toddlerfriend',
'puresweet',
'latergram',
'toddlerfun'],
[],
['orlando'],
[],
['pharrel'],
[],
[],
['real', 'tru'],
['slut',
'snapshot',
'hot',
'nasti',
'naughti',
'sexi',
'horni',
'shi',
'porn',
'nude',
'kinki',
'xxx'],
```

```
'y'],
['thankfulthursday'],
['i_am', 'posit', 'affirm'],
['canadianrav'],
['michaelkor', 'luckygirl', 'bihday', 'liverpool'],
[],
['freakshak', 'freak', 'yum', 'dalston'],
[],
['thank', 'posit'],
[],
['vegan',
'organ',
'fresh',
'fruit',
'organ',
'healthi',
'healthyliv',
'life'],
['sad'],
[],
['blackpool'],
['fathersday'],
[],
['newblood16'],
['funer'],
[],
[],
['usa'],
['stephcurry', 'nbafin'],
['grate', 'affirm'],
['missyou', 'mybrotherskeep'],
[],
['friday', 'weird'],
[],
['direction',
'niall',
'nialhoranfact',
'niallerwin',
'hot',
'justindrewbieb',
'justindb',
'believetou'],
['father', 'sanya', 'whererefreshingbegin'],
['lifeaftersurgeri',
'dog',
'dogslif',
'labrador',
'labradorretriev',
'lifeofsam'],
['glastofest'],
```

```
[],  
[],  
['mep', 'webseri'],  
['model'],  
['juneteenth',  
  'independenceday',  
  'food4thought',  
  'rich',  
  'ancestr',  
  'heritag'],  
['river'],  
['pathet', 'ripgop'],  
[],  
[],  
['smile', 'instalik', 'instamood', 'instap'],  
['graffiti',  
  'dublin',  
  'ireland',  
  'music',  
  'vimeo',  
  'movi',  
  'video',  
  'europ',  
  'peopl',  
  'flow',  
  'fun',  
  'rafalwojcicki'],  
['oitnbseason4'],  
['iamposs', 'extraordinaryladyspeak', 'staraward'],  
['breakfast', 'sun', 'day', 'sunnyday', 'sun', 'blueâ'],  
['yellowjacket', 'neshl', 'bhl', 'hockey'],  
['tonyaward'],  
['fathers', 'day', '2016'],  
['countri', 'rock', 'bluegrass', 'beatl', 'love', 'pop', 'romanc'],  
['project', 'blog', 'silver', 'gold', 'forex'],  
['300dayshappi',  
  'dayseventyseven',  
  'day77',  
  'happi',  
  'love',  
  'lifequot',  
  'lifeâ'],  
[],  
['bihday'],  
['thank', 'posit'],  
[],  
['quiet', 'july4th', 'firework'],  
[],  
['paratyphoid'],  
['thevoic', 'stoptheh', 'ripchristina'],
```

```
[],
[],
['up', 'enjoy', 'music', 'today', 'free', 'app', 'free', 'music'],
['lawofattract', 'resourc', 'healing', 'altwaystoh', 'healthi',
'idwp'],
[],
['music', 'therapi', 'healing', 'altwaystoh', 'healthi'],
[],
['salud', 'ch', 'smile'],
['nofilt'],
['overwhelm', 'happi'],
[],
['buffalo'],
[],
['thebachelorett', 'empti', 'nowin'],
[],
['good', 'faith', 'vaitacacommafiadv'],
['love', 'success', 'work', 'businessâ'],
['torrington', 'lunch', 'afterschool', 'instadaili', 'yummi',
'withmom'],
[],
['cycl', 'sunset', 'ride', 'bike', 'fit', 'cardio', 'deseâ'],
[],
['tagsforlik',
'me',
'food',
'friends',
'summer',
'swag',
'quot',
'love',
'god',
'girl',
'weekend',
'polaroid',
'followm',
'teen',
'win',
'job'],
['payintheusa'],
[],
['stellar', 'girl', 'dreami', 'goodmorn'],
[],
[],
['yay', 'wednesdaywisdom'],
['forev'],
['bee'],
['beauti', 'smooth', 'lifestyl', 'skin', 'instas', 'f4f', 'fashion'],
['lawofattract', 'heal', 'loa', 'healthi', 'well', 'idwp'],
['oil', 'blog', 'silver', 'gold', 'forex'],
```

```
['newblood'],
['first',
 'bihday',
 'puppi',
 'elois',
 'sweetbabin',
 'dog',
 'growingup',
 'cute',
 'bihdaygirl'],
[],
['prayfororlando', 'guncontrol'],
['smile', 'lovethemâ'],
[],
['ombr', 'blue'],
['warcraft'],
['family', 'fun'],
[],
[],
['lover', 'friend', 'astrolog', 'love'],
[],
['sdl326am'],
['summev', 'zoo'],
['heal', 'altwaystoh', 'healthi'],
['roar', 'tiger', 'annaswelshzoo', 'bihdaycelebr'],
['laugh', 'xtra'],
['positivevib', 'sunshin', 'sunday'],
[],
['masterkeyexperi'],
['face', 'pizza'],
['sisteradventur', 'sister', 'ilovemysisterâ'],
[],
['usn', 'navi'],
['euro'],
['happi', 'minion'],
['bethesda', 'e3', 'e32016'],
['tiggi', 'jax', 'soa', 'throwbackthursday'],
['eurogam', 'pump', 'eurogamer2016'],
['staytun', 'tapfoag'],
['sun',
 'love',
 'outdoor',
 'picnic',
 'summer',
 'famili',
 'daughter',
 'father',
 'dess',
 'icecream'],
['retailtherapi', 'riverisland', 'shoppingqueen'],
```



```
'alwayscheersyouup'],
['clean', 'laundryday'],
['ebay', 'woohoo', 'sell', 'sell', 'buy', 'shop', 'shop'],
['thank', 'posit'],
['chill'],
['instagood', 'i', 'a', 'idola', 'idol', 'a', 'followm', 'selfi',
'a'],
['christinagrimmi', 'orlando', 'america'],
['brexit', 'blog', 'silver', 'gold', 'a'],
['eachbpm', 'busi'],
['alohabeachbu', 'hawaii', 'explor'],
[],
['great'],
[],
['bird', 'download', 'film', 'il', 'ita'],
['sunday',
'photo',
'model',
'urbex',
'topless',
'tit',
'mono',
'retweet',
'sexi',
'heel',
'ace'],
[],
['shegotmarri'],
['model'],
[],
['soccer', 'spring', 'season'],
[],
['sun', 'fun', 'beachpark'],
['euros2016'],
['anim', 'love', 'stori', 'not', 'end', 'like4likea'],
['poet'],
['i_am', 'posit', 'affirm'],
['2pm', 'kpop'],
['essentialoil', 'anxieti', 'healthi', 'peac', 'altwaystoh'],
[],
['gettingold', 'becomingbor', 'notborderland', 'goblizzard'],
['fatkidinacandystor'],
['snapshot',
'young',
'shi',
'kinki',
'wet',
'porn',
'hot'],
```

```
'xxx',
'horni',
's'],
['love', 'girl', 'snapchat', 'flower', 'instasmil', 'instaloveâ'],
['i_am', 'posit', 'affirm'],
[],
['acne', 'altwaystoh', 'healthi', 'healing'],
['msgapparelstoday'],
['lovecamp', 'brownwood'],
['grate', 'affirm'],
['forev',
'karaok',
'friend',
'friendship',
'love',
'nofilt',
'instacool',
'instagood'],
['318'],
['thealter'],
['u', 'blog', 'silver', 'gold', 'forex'],
['fathersday', 'love'],
['motiv'],
['violent', 'world', 'uk'],
['funni', 'truestori'],
['monday', 'menmodelsâ'],
[],
[],
['oitnbseason4'],
[],
['noooo', 'whyyyy', 'loveyoudesi'],
['jordan', 'paid', 'wear', 'lowtopsâ'],
['i_am', 'posit', 'affirm'],
['salut'],
['nyc', 'newyorkc', 'myview', 'wine', 'fridaynightâ'],
[],
['love', 'bless'],
[],
['danian', 'love', 'colour', 'rainbow'],
['aiel', 'aielmaharashtra', 'irrit'],
['rocksteddi', 'band', 'enjoy', 'idol', 'omg'],
['cheapshot', 'lowblow', 'shame'],
['parent', 'tantrum', 'toddler', 'calm', 'mind', 'strong'],
['i_am', 'posit', 'affirm'],
['siri', 'maco', 'wwdc2016'],
['diy', 'dresser', 'nicknack', 'newelectricsscrewdriv'],
[],
[],
['choic', 'alway'],
```

```
['ë', 'ê', 'ì'],
['app', 'june', 'issu', 'bazaarmagâ'],
[],
[],
[],
['enjoy',
'greec',
'crete',
'sabinedrift',
'chillinn',
'blondegirl',
'bluedressâ'],
[],
['gamereadi'],
['summer', 'life', 'no_school', 'tahiti', 'like'],
['stupidthought', 'quot', 'lmao'],
['downasquadmemb', 'treeout'],
['japan', 'blog', 'silver'],
['bbuk'],
['good', 'morn'],
['fridayfeel', 'relax', 'bristol'],
[],
['rap', 'hiphop', 'rock', 'funni', 'comedi', 'pop', 'kanyewest'],
['merri', 'christma', 'new', 'year', 'freepik', 'ad', 'decor'],
[],
['test', 'tire', 'annoy', 'whi'],
['anxieti', 'altwaystoh', 'heal', 'peac'],
[],
[],
['periscope', 'collegebound'],
['nervou', 'curiou', 'bringiton'],
['lifegroup'],
[],
['lawofattract', 'resourc', 'healing', 'altwaystoh', 'healthi',
'idwp'],
['surfday', 'fathersday', 'dad', 'novyspos', 'novytowel',
'surfgear'],
[],
[],
[],
[],
['showyourjeepday', 'jeepmafia', 'jeeplif', 'jeep'],
[],
['ä½'],
['gratitud', 'laurasworld'],
[],
['nbafin'],
['uot', 'savecharity', 'tragic', 'rip'],
['space', 'blue', 'fabricsourc', 'finish', 'pillow', 'cushion'],
```

```
[],
[],
['fathersday', 'dad', 'financi', 'future', 'save', 'insur',
'invest'],
['saturday', 'weekend'],
['fathersday'],
['footbal'],
['ee', 'buxton', 'grumpi'],
[],
['troopingthecolour', 'thequeenat90'],
['work', 'fusiona', 'develop', 'chase'],
['hamster', 'rip', 'imissy'],
['love'],
['love', 'boyfriend', 'smile'],
['icrs2016'],
['fun'],
['lichfield'],
['surpris', '2like', 'unexpect'],
[],
[],
[],
['embarrass'],
['internationaldayofyoga2016',
'yogalov',
'yoga',
'relax',
'healthi',
'positiveâ'],
[],
[],
['coldplay', 'conc', 'hampden', 'fave', 'coupl', 'cute'],
['heal', 'altwaystoh', 'healthi', 'peace'],
['flagday', 'texa', 'householdnam', 'takingoverâ'],
[],
['cat', 'furry', 'persiancat', '4798'],
['tgif', 'kiss', 'kiss', 'friday', 'fridayfun'],
['bless', 'thank'],
['euro2016'],
[],
['theriveourberlin', 'theboss'],
['nowplay', 'sick', 'littl', 'world', 'live', 'set'],
['2ndamend'],
['dog'],
['boohiss'],
['whenrealtorscompeteyouwin', 'callm', 'gotheextramileâ'],
['poaventura', 'whatajok'],
['dcvote'],
['eur', 'blog', 'silver', 'gold', 'forex'],
['elections2016', 'america', 'smh', 'usa'],
```

```
['anim', 'puppi', 'vine', 'magic', 'subscrib', 'youtub'],  
[],  
[],  
['dad', 'fathersday', '2016', 'intellicr'],  
[],  
['pwcproud'],  
['love', 'instagood', 'tbt', 'cute', 'me', 'beauti', 'followm'],  
['weekend', 'famili', 'home'],  
['tragic', 'puls', 'orlandostaystrong', 'thoughtsandprayersâ'],  
['yrkkh', 'uglyspat'],  
[],  
['thekil'],  
[],  
['travel',  
  'tim',  
  'instamood',  
  'moment',  
  'love',  
  'mylif',  
  'russian',  
  'girl',  
  'model',  
  '4likeâ'],  
['tuesday',  
  'jacki',  
  'quot',  
  'taketim',  
  'soul',  
  'enjoylif',  
  'smile',  
  'believ',  
  'dream',  
  'havefun'],  
['edinburgh', 'scotland', 'homesweethom', 'famili'],  
[],  
[],  
['thank', 'posit'],  
['studiotim'],  
[],  
['cfisdrrr'],  
['prettypussi', 'sexybbw', 'realmvp'],  
['excitâ'],  
[],  
[],  
[],  
['stone', 'gift', 'amaz', 'beauti'],  
['lfc', 'ynwa', 'englishpremierleagu'],  
['rogergoodell', 'fathersday', 'goodellsucksâ'],  
['orb', 'blog', 'silver', 'gold', 'forex'],
```

```
[],  
[],  
['love', 'fathersday', 'sunday', 'father', 'daddi'],  
[],  
['actor', 'seeklearn', 'stafresh'],  
[],  
['bastard'],  
['socialmedia'],  
['smaphon'],  
['lifeisbeauti', 'loveit', 'workhard', 'trainhard', 'bless',  
'fitlifeâ'],  
['spo', 'dure', 'lui', 'oscar'],  
['njoyhit40', 'lhitmusicst', 'nice', 'nyc'],  
['personaltrain'],  
[],  
['thank', 'posit'],  
['melancholi', 'melancholymus'],  
['tvk3fathersday'],  
['thank', 'posit'],  
['thank', 'posit'],  
[],  
['love', 'sinkthepink', 'music', 'instagay', 'celebspot'],  
['ccmf2016', 'mychurch', 'frontrow', 'upshow'],  
['stupid', 'africa'],  
['gold', 'blog', 'silver', 'gold', 'forex'],  
['horrif', 'unforgiv'],  
['i_am', 'posit', 'affirm'],  
['instatravel',  
  'instamo',  
  'igdaili',  
  'instagramhub',  
  'instagood',  
  'instalik'],  
['lonely'],  
['girliguessimwithh'],  
['mobil', 'whitepap'],  
[],  
['work', 'mindset'],  
['rochest'],  
[],  
['tattoosleev'],  
[],  
['tell', 'river', 'complain', 'philosophi', 'proverb'],  
['thank', 'posit'],  
[],  
['happi', 'lifecoach', 'reinventimposs'],  
[],  
['thank', 'posit'],  
['outsid',
```

```
'thepeel',
'work',
'homophob',
'taxi',
'driver',
'guy',
'kiss',
'gay',
'melbourn',
'13cab'],
[],
['smile', 'kind'],
['windows10', 'browser', 'customcasetab', 'noedit'],
[],
['whi'],
['totalpolitician', 'outoftouch', 'disgrac'],
['fathersday'],
['truth', 'suicid', 'blackboy'],
['homesweethom',
'amaz',
'love',
'istanbul',
'girl',
'beauti',
'instalik',
'meâ'],
[],
[],
['swing', 'mybabygirl', 'babi', 'littlegirl', 'park', 'love',
'cute'],
['winter', 'bathtim', 'cozi'],
['i_am', 'posit', 'affirm'],
[],
['sunflow', 'natur', 'garden', 'photograpi'],
[],
['growth', 'hack', 'confer', 'growthwithhubspot'],
['dog', 'sunshin', 'dogsoftwitt'],
['sleepnumb', 'bonustolongrest', 'dnd'],
['satisfi', 'hair', 'workit', 'smile', 'faceâ'],
['sunday', 'weekend', 'selfi', 'longhair', 'girl', 'selfporait',
'loveâ'],
['nature_perfect'],
['model'],
['039'],
['golinuntern'],
['grimm', 'nadalind'],
[],
['disrespectful', 'frustrat'],
['friday', 'pub', 'ilovehim', 'instagay', 'beer'],
```

```
[],
[],
[],
['fruit', 'dess', 'instamo', 'photooftheday', 'yvr', 'sweet',
'vancitybuzzâ'],
[],
['iqbal', 'galib', 'wasi', 'faraz', 'mohsin', 'no'],
['armaanamaalontkss'],
['royal', 'cleveland', 'disgrac'],
[],
['latina'],
['orlando'],
['triathlet', 'triathlon', 'coach', 'vacat', 'fearless', 'rulâ'],
['reflect'],
['gratitud'],
[],
['takalot', 'bengreenfield'],
['saysometh'],
[],
[],
['slugg'],
['awesom',
'beat',
'cool',
'dj',
'danc',
'edm',
'hous',
'goodmus',
'madnessâ'],
['jpy', 'blog', 'silver', 'gold', 'forex'],
[],
['steakandshak', 'not'],
['grayneedstogo'],
['leadership', 'inspir', 'quot', 'inspir'],
['usd', 'blog', 'silver', 'gold', 'forex'],
['rlrockallday', 'nois'],
['loveherbreakup', 'hotgoss'],
['sunshine', 'sunshine'],
['game', 'gamer', 'xboxon', 'playstat', 'e32016'],
['bringbackgladiators'],
['humpday',
'life',
'lifestyl',
'wednesday',
'midweek',
'snapchat',
'walru',
'funnyâ'],
```



```
[],
['archiveday', 'weekendblogshar', 'happi', 'quot', 'tshi'],
['goodtogreat', 'inspirem', 'love'],
['london', 'puppi', 'thursdaymotiv'],
['draw', 'blackandwhit'],
['romances', 'horror', 'moviegenr', 'ambianc'],
['wildlifefestiv'],
['ewabeach', 'magicalhappydrink'],
['pleasedontchang'],
['holiday', 'croatia', '13daystogo'],
[],
['newyork', 'newyorkc', 'travel', 'bigappl'],
['motiv'],
['fashionblog'],
['becauseofhim', 'inlov'],
['smile', 'love', 'iminlov', 'iam', 'im', 'in', 'nice', 'day'],
['monaco',
 'resto',
 'chezedgard',
 'grandprixfl',
 'essais',
 'sun',
 'lewissuppo',
 'ö',
 'amour',
 'love'],
['toread', 'letsmak'],
['sukhbir', 'badal', 'eastcoast', 'canada'],
[],
['shop', 'cool', 'home', 'fun'],
[],
['scottieandrussel', 'work', 'love', 'picofthedayâ'],
['rain', 'water', 'tbt', 'me', 'weather', 'bribieisland'],
[],
['wed', 'love'],
['icymi', 'debelop', 'gun'],
['schoolsuppli'],
['babiescat', 'cat', 'newborn', 'congratul'],
['pretending',
 'to',
 'b',
 'when',
 'you',
 'in',
 'pain',
 'is',
 'just',
 'an',
 'exampl',
```

```
'of',
'how',
'strong',
'you',
'are',
'as',
'a',
'personö',
'nightyö'],
[],
['mentalhealth'],
['weekend'],
['thank', 'posit'],
['cssday'],
['bihday', 'mom', 'fondant', 'pumpkin', 'chocol', 'delici', 'beauti',
'amaz'],
['bigbang', 'babi'],
['tgifriday', 'goodvibesonli', 'becauseimhappi', 'smilemor'],
['letsgo', 'agent', 'tooeasi', 'dlc', 'gamingforeveryon'],
['tissu'],
[],
['fowoh', 'goldenretriev', 'lcck9comfodog', 'workingdog'],
[],
['edc', 'music', 'musiclif', 'edm', 'danc', 'girlsâ'],
['funni', 'humankind', 'news', 'bad', 'good', 'future'],
[],
[],
['wish', 'dad', 'father', 'day', 'rock'],
['abrahamhick', 'lawofattract', 'healthi', 'heal', 'joytrain'],
['model'],
[],
['smile'],
['success', 'quot'],
[],
['friends', 'sweden', 'reunion', 'grateful', 'lov'],
['pizzaslice',
'pizzaslice2',
'aoyama',
'pizza',
'yummy',
'delicia',
'goodtimeswithgoâ'],
['beer', 'craftbee', 'dog', 'malinoislov', 'malinoi', 'passion',
'friend'],
['i_am', 'posit', 'affirm'],
['fatherä', 'day', 'dad', 'skinni', 'and', 'singl'],
[],
['cleveland', 'home', 'calm'],
['gun', 'mentalil', 'gang', 'orlando'],
[],
```

```
['bless', 'lipstick', 'makeup', 'fun', 'grate', 'share'],
[],
[],
[],
['saturday', 'morning', 'thetrust', 'comus', 'indi'],
['father', 'day'],
['good'],
[],
['orlando', 'vengeanc'],
['live_long'],
['graduat', 'mechtechplumb', 'futureplumb', 'proudmama'],
['model'],
['orlando', 'christinagrimmi', 'prayer'],
['innovative_n'],
['social', 'analyt', 'instalik', 'tbt', 'yoga', 'summer', 'fun'],
[],
['newflat'],
['love', 'chees', 'shotermlongdist'],
['friendship'],
['poor', 'sin', 'jesu', 'faith', 'believ', 'wisdom'],
['lovequot'],
['posit'],
['orlandoshoot', 'gunviol'],
['zaz', 'umbrela'],
[],
['tcashowroom'],
[],
['relax', 'positivevib', 'california', 'cakesbytheocean'],
['eur', 'blog', 'silver', 'gold', 'forex'],
['brokedonald', 'lyindon'],
['friday', 'fridayfeel', 'friyay', 'goodweek', 'client', 'hairâ'],
[],
[],
['me', 'selfi', 'myfriend', 'loveh', 'giannena', 'smile',
'thecomeback'],
[],
[],
['burnsurvivor', 'dolbf', 'motiv', 'motiv'],
['greatfultogod', 'beendoinggreat', 'selfi'],
['dog'],
['nosie', 'year', 'wrc', 'rallypoland'],
['music', 'edinburgh', 'scotland', 'band', 'musicvideo'],
['pulseshoot'],
[],
['33'],
['drakeandjosh'],
[],
['openrp', 'bore'],
['manchest', 'gin'],
```

```
['trailn2v3'],
['kikid', 'miabbook'],
['inspir', 'mi', 'lyf', 'baba'],
['beauti', 'homeown'],
[],
[],
['brickingit'],
[],
[],
[],
[],
[],
['shop', 'cool', 'home', 'fun'],
['doubledown', 'veganbeerfest', 'vegan'],
[],
['bjmckenzi', 'born2win', 'bjmuzic'],
[],
['jaibalayya'],
['kingsland', 'highlandreserv', 'hous', 'famili', 'build',
'ourhomeâ'],
[],
['decor'],
['soldâ'],
[],
['marb', 'euro2016', 'depress', 'bringbackhappi'],
[],
['gdragon',
'gd',
'kwonjiyong',
'jiyong',
'vip',
'bigbang',
'handsom',
'cute',
'loveâ'],
[],
[],
[],
['develop', 'python'],
['womenareamaz', 'induoev', 'superwomen', 'follow', 'dowhatyoulov'],
[],
['oitnb', 'myfavorit', 'season4', 'weekendplan'],
['christinagrimmi', 'loss', 'rip', 'miss', 'fan', 'foreverinourhea'],
['sundayrewind', 'hongkong', 'travel', 'ladi', 'lawyerâ'],
['inger',
'followm',
'follow',
'follow4follow',
'beach',
'pool',
```

```
'relax',
'enjoy',
'coupleâ'],
['family', 'freeevent', 'festival', 'food', 'stal'],
['awesom', 'gopro', 'mount'],
['booktim', 'muchlov'],
['summereven', 'patioweath'],
[],
['imi'],
['model'],
['theshallow'],
['offici', 'poster', 'poalorigin', 'summer'],
['wwdc', 'wwdc2016'],
[],
['blessed', 'thank'],
['sundaymorn', 'sunday', 'peac', 'love'],
['quot', 'latenight'],
[],
['selflov',
'love',
'ilovem',
'quot',
'quot',
'quoteoftheday',
'happi',
'life',
'wisdom',
'wordsofwisdom'],
['boy', 'father', 'sweet', 'love', 'memori'],
[],
['life',
'impoant',
'human',
'spirit',
'truth',
'love',
'give',
'spiritu',
'thinkbigsundaywithmarshað'],
['father'],
[],
[],
[],
['wed', 'love'],
['friday'],
[],
['fun', 'eatâ'],
[],
['blogger', 'vloggerâ'],
```

```
['edinburgh'],
[],
[],
['passthecourvoisi'],
['rescuepup', 'rexi', 'pupsofsg', 'suicidegirl', 'beauti',
'puppylov'],
['tech', 'augment', 'futur', 'mobil'],
[],
[],
['hellosunday', 'beauti', 'sunday', 'sundayfunday'],
[],
['nzd', 'blog', 'silver', 'gold', 'forex'],
[],
['bihday', 'rip', 'tupacâ'],
[],
['fl', 'motionpost', 'mupparimanam', 'nervou'],
[],
['compleanno',
'toa',
'cake',
'famili',
'selfi',
'me',
'smile',
'girl',
'summerâ'],
['anim', 'anime', 'tags', 'anime', 'fana'],
['memori', '2year', 'lotstocom', 'love'],
['spoonieproblem'],
['bless'],
['eur', 'blog', 'silver', 'gold', 'forex'],
['ngitingtagumpay'],
['welcomeludovica'],
['cloudchas'],
['vlicob',
'xoxo',
'love',
'grace',
'amen',
'share',
'passion',
'church',
'lifeâ'],
['job', 'jobsearch', 'chandigarh'],
['restaur', 'eatallyoucan'],
['nervou', 'gameofthron'],
[],
['intruduceyourself'],
['ramadan', 'ramadankareem'],
```

```
[],
['success', 'growth', 'balanc', 'optim'],
[],
['thank', 'posit'],
[],
['beach', 'loveit'],
[],
['tgif', 'ff', 'gamedev', 'indiedev', 'indiegamedev', 'squad'],
[],
['goodmorn'],
[],
['yum', 'thespicesndteaexchang', 'dc', 'healthyâ'],
['followm'],
['love',
'like',
'live',
'life',
'enjoy',
'healthi',
'sun',
'tan',
'bikini',
'summerâ'],
['teenageson', 'fathersday'],
[],
[],
['getalif'],
[],
['nothappi'],
['closeencount', 'fatkind', 'fatkid', 'dese', 'sick', 'tire',
'maybemono'],
['gymmorgon', 'gym', 'morn', 'strength', 'energi', 'train'],
[],
['truthseek', 'empower', 'selfimprov', 'heal', 'ihearadio'],
['deletetweet'],
['view',
'oregon',
'poland',
'usa',
'unitedst',
'holiday',
'dj',
'produc',
'houseâ'],
['vote', 'share', 'you', 'smile', 'poll'],
[],
['travelwis', 'europe2016', 'london', 'pari'],
['thank', 'posit'],
[],
```

```
[],
['smile', 'pearli'],
['gentleman'],
['love'],
[],
[],
['2016neotour'],
[],
['movinghouâ'],
['music'],
['barbequ', 'hostel', 'summer', 'surf', 'kitesurf', 'wave', 'fun',
'holiday'],
['cnnlect', 'onpoint'],
['masterkeyexperi'],
['crazi',
'fashion',
'fun',
'color',
'gown',
'fashiondesign',
'design',
'illustrationâ'],
['fit', 'fresh'],
[],
[],
[],
['liber'],
[],
['newyear', 'newbless', 'thank', 'grate'],
['cymruambyth'],
['teamandr', 'idontcareifhesbritish'],
['hurryup'],
['happi', 'realtimechem', 'chemistri', 'educ'],
['sunday',
'weekend',
'drink',
'la',
'night',
'view',
'summer',
'fullmoon',
'dtla',
'fun'],
[],
['euro2016'],
['windows10', 'browser', 'lumia950xl', 'lg'],
['lighttherapi', 'depression', 'altwaystoh', 'healthi', 'happi'],
['bless', 'thank', 'optimist', 'posit', 'faith', 'strong', 'motiv'],
['littlewond'],
```



```
[],
[],
['atlanta'],
['balconyview',
 'brunch',
 'sundaybrunch',
 'sunday',
 'istanbul',
 'love',
 'girl',
 'instalikeâ'],
['columbu', 'mbb', 'mygrind', 'mrbaileybabi', 'tshiâ'],
['william', 'blog', 'silver'],
[],
[],
['kid', 'destruct'],
['helpm', 'imnotusetothi', 'perfecttim', 'weekend'],
['cantwait', 'latvia', 'hereicom', 'midsumm', 'familli', 'friendsâ'],
['love', 'women'],
['educationfest'],
[],
['reviv', 'sunris', 'rebelhea', 'morn', 'pielcanela'],
[],
['lighttherapi', 'depression', 'altwaystoh', 'healthi', 'happi'],
['peac', 'inspir'],
['motiv'],
['teammacb', 'fitmom', 'fitlifeâ'],
[],
['tranni', 'ladyboy', 'asianladyboy', 'thailand', 'usa',
'lasvegasâ'],
[],
[],
['johnxsafari', 'eastcapehunt', 'famili', 'proud', 'africa'],
['selfi', 'freetim', 'smile', 'instaviã', 'sonrisa', 'ust'],
['loop', 'shop', 'vine'],
['coverlov', 'bookporn'],
['euref'],
['destini'],
['i_am', 'posit', 'affirm'],
['heal', 'healthi', 'idwp', 'joytrain'],
['run', 'park', 'summer', 'exercis', 'fit', 'helthi'],
['hangingon'],
[],
[],
[],
['ea', 'hate', 'game', 'survey'],
['lighttherapi', 'depression', 'altwaystoh', 'healthi', 'happi'],
['smile', 'laugh', 'breathe', 'lov'],
...]
```

## 2. Extract hashtags from racist/sexist tweets

```
ht_negative= hashtag_extract(df['clean_tweet'][df['label'] == 1])
ht_negative
```

```
[['cnn', 'michigan', 'tcot'],
 ['australia',
  'opkillingbay',
  'seashepherd',
  'helpcovedolphin',
  'thecov',
  'helpcovedolphin'],
 [],
 [],
 ['neverump', 'xenophobia'],
 ['love', 'peac'],
 [],
 ['race', 'ident', 'medâ'],
 ['altright', 'whitesupremaci'],
 ['linguist', 'race', 'power', 'raciolinguist'],
 ['brexit'],
 ['peopl', 'trump', 'republican'],
 ['michelleobama'],
 ['knick', 'golfâ'],
 ['jewishsupremacist'],
 ['libtard', 'sjw', 'liber', 'polit'],
 ['trash', 'hate'],
 [],
 ['â'],
 ['stereotyp', 'prejudic', 'hope', 'hate', 'conflictâ'],
 ['pol', 'bluelivesmatt', 'draintheswamp', 'ferguson'],
 ['2016'],
 ['antisemit', 'hocoschool', 'columbiamd', 'hocomd'],
 ['nazi', 'hatred', 'bigotri'],
 ['libtard', 'sjw', 'liber', 'polit'],
 ['p2', 'p21', 'fyi', 'tcot'],
 ['topoli'],
 ['blacklivesmatt'],
 ['ushistori', 'eugen'],
 ['biher', 'bigotri', 'altright', 'potu', 'theresist'],
 ['crime', 'officerâ'],
 [],
 ['black', 'retweet', 'tampa', 'miamiâ'],
 ['allahsoil'],
 [],
 ['prais', 'cleveland', 'rnc', 'white', 'tamirric', 'rip'],
 ['bigot'],
 [],
 ['blacklivesmatt'],
```

```
['feminismiscanc', 'feminismisterror', 'feminismmuktbharat',  
'malevot'],  
[],  
['prowar', 'anti', 'islamic', 'homophobic', 'rapist', 'hypocrit'],  
[],  
['latesnew'],  
['libtard', 'sjw', 'liber', 'polit'],  
['trump', 'dumptrump'],  
['lifelesson', 'blogpost', 'whitepeopl', 'respectâ'],  
[],  
['fakenewsal'],  
['hateâ'],  
['trump', 'hate', 'putin', 'trump'],  
[],  
[],  
['starspangledbann', 'slaveri', 'apaheid', 'usa'],  
['social', 'media', 'aand'],  
['eod'],  
['trump', 'vile', 'bigot', 'hi', 'surrog', 'carlpaladino',  
'endhateâ'],  
['allahsoil'],  
['discrimin'],  
['endthen'],  
['misogyny', 'homophobia', 'bigotri'],  
[],  
[],  
['rushlimbaugh'],  
[],  
['trumpusa', 'homophob', 'misogynist'],  
['sjw'],  
['hope', 'newyear', 'endof2016', 'tedtalk', 'tedatibm'],  
['cspan'],  
['peopl', 'trump', 'republican'],  
['opinion', 'degrad', 'discrimin', 'violenc', 'stoprac'],  
['allahsoil',  
'isi',  
'islam',  
'2016in4world',  
'2016in4word',  
'terrorist',  
'gwot'],  
['tuckercarlson'],  
[],  
['oppress', 'oppressors', 'privileg'],  
['space'],  
[],  
['ivankatrump', 'fake', 'adamsaleh', 'raâ'],  
['zionazi'],  
[],
```

```
['lbgt'],
['trump', 'barackobama', 'michelleobama'],
['quebec', 'humanright', 'justic'],
['berlin'],
['returnmyvinyl'],
['chrismukkah'],
['keyboard'],
['bigot', 'misogynist', 'keepyoureyesontheprize'],
['black', 'africanamerican', 'repar'],
['potu', 'trump', 'israel', 'happynewyear'],
['decolon'],
['america'],
['treasonoustrump', 'whiteisi', 'terrorism'],
['ellen', 'kimburrel'],
['newyear', 'prez', 'destruction', 'commercials', 'poopin'],
['shuckin', 'democraticpay'],
['hitler', 'traitor', 'trump'],
['liber'],
['birdsofafeath', 'resist'],
['versace'],
['rich', 'white', 'snob', 'tcot', 'tpc'],
['joke', '2016'],
[],
[],
['feminismiscanc', 'feminismisterror', 'feminismmuktbharat',
'malevot'],
['tyler', 'perri', 'claim', 'revers', 'criticsâ'],
[],
['hypocrit', 'liar', 'inhumaneâ'],
['psa'],
[],
['sikh', 'templ', 'calgary', 'wso'],
['allahsoil', 'emirati', '2016in4word'],
[],
['mentalhealth', 'firstnations', 'thirdworldcanada'],
['democraticpay', 'blm', 'chicago', 'detroit', 'innâ'],
[],
['new', 'mc', 'phillysuppophilli'],
[],
['allahsoil'],
[],
['illeg', 'criminals', 'commonsense'],
['deviant', 'barackobama'],
[],
['feminist', 'reproductiveright', 'bihcontrol'],
['libtard', 'sjw', 'liber', 'polit'],
['hillaryclinton', 'homophob', 'misogynist'],
[],
[],
```

```
['newyearthseve', 'resolut', 'domesticviol', 'love'],
['japan', 'fascism', 'guardian', 'nytim', 'bbc', 'afp', 'rsf'],
['altright', 'maga'],
[],
['fear'],
['black', 'retweet', 'tampa', 'miamiâ'],
['libtard', 'sjw', 'liber', 'polit'],
['yehtut', 'rohingya', 'dvbmultimedialogroup', 'yehtut'],
[],
['gop'],
[],
[],
['hrc', 'popularvot', 'ca', 'trump'],
[],
[],
[],
['2017in3wordsâ'],
['electoralcolleg', 'trump', 'notmypresid', 'newswith'],
['genocid'],
['lgbtq'],
[],
['arab'],
['god', 'cnn', 'foxnew', 'christian', 'conserv', 'christ'],
['censorship', 'freethenippl'],
['bulli'],
['allahsoil'],
[],
['grandmoth', 'grandson', 'exist', 'alreadi'],
['libtard', 'sjw', 'liber', 'polit'],
[],
['blackwomen', 'asianmen', 'datingapp'],
['antirac', 'seashepherd'],
['deplorabal', 'whitesheet', 'inbr'],
['canuck', 'oiler'],
['american', 'bigotri', 'hatr'],
['homophobic', 'warmonger', 'worldddang'],
['trumpsamerica', 'hate'],
['america', 'controversi', 'equal', 'kaepernick', 'nationalanthem'],
[],
['laurenduca'],
['hispan', 'black', 'asian', 'cnnâ'],
['hispan', 'retweet', 'boricua', 'miamiâ'],
['hatred', 'xenophobia', 'islamophobia', 'antisemitism', 'war',
'2017'],
[],
['antirac', 'seashepherd'],
['polit', 'relationship', 'newâ'],
['santa', 'obama', 'soro'],
['allahsoil', 'teambt', 'teamsuperjuniorâ'],
```

```
[],
['blackonblack',
 'ris2016',
 'antiblack',
 'dogwhistl',
 'blacklivesmatt',
 'hamzayusuf'],
['zionazi', 'antisemt'],
[],
[],
['womenwednesdayâ'],
['dystopian', 'radicalis', 'reveng', 'preorder', '99c'],
['fascism', 'nazi', 'altright', 'richardspenc', 'trumpism', 'dâ'],
[],
['trump', 'michelleobama', 'womensmarch', 'women', 'predatorâ'],
[],
['sameshitnewyear',
 'corrupt',
 'trump',
 'terrorattack',
 'fundament',
 'violenc',
 'isi',
 'crisi',
 'hatr',
 'murder',
 'terror'],
[],
['bigot'],
[],
[],
['libtard', 'sjw', 'liber', 'polit'],
['u', 'america'],
[],
[],
[],
['american'],
['altright', 'whitesupremaci'],
['new', 'mc', 'phillysuppophilli'],
['democrat'],
['europ',
 'nyt',
 'libcrib',
 'uniteblu',
 'fascism',
 'trump',
 'fear',
 'ignor',
 'hate'],
```

```
['misogyni'],
['aloha', 'fascism', 'warsforoil'],
['antirac', 'seashepherd'],
['theresamay', 'brexit', 'remain'],
['nycâ',
 'assault',
 'womenonli',
 'forher',
 '2017',
 'newyear',
 'resolut',
 'womenissuesâ'],
['texas', 'justicereform'],
['maddow'],
[],
['libtard', 'sjw', 'liber', 'polit'],
['chepstow', 'gwent', 'wale', 'cymru'],
['familycircu', 'chriâ'],
[],
['christian', 'retweetâ'],
['trump'],
['libtard', 'sjw', 'liber', 'polit'],
['libtard', 'sjw', 'liber', 'polit'],
['childfreeâ'],
['economicapahied', 'hatespeech'],
['asshol', 'rapist', 'pedophil', 'evil', 'satanist'],
[],
['stop'],
['joke', '2016'],
[],
['feminismiscanc', 'feminismisterror', 'feminismmuktbharat',
'malevot'],
[],
['corruptgop'],
['nyc', 'newyork', 'women', 'men', 'bigot'],
['carlpalladino', 'michelleobama'],
['allahsoil'],
[],
['politicalmillenni', 'letstalkabout', 'climatechang', 'educationâ'],
[],
['notmypresidnet'],
['american', 'bigotri', 'hatr'],
['america', 'controversi', 'equal', 'kaepernick', 'nationalaâ'],
['un', 'isra', 'jewish', 'judea', 'samaria'],
[],
['gop', 'deplor', 'notfunni'],
[],
['swamp', '20day', 'dem', 'taxpay', 'islamicterrorist'],
['libtard', 'sjw', 'liber', 'polit'],
```

```
['alexpasc', 'friend', 'music'],
['falsenarr', 'falseequaval'],
['podcast', 'gun', 'igmilitia', 'pewpewlifeâ'],
['bigotry'],
['antirac', 'seashepherd'],
['obama', 'communism', 'islam', 'race', 'nazi'],
['muslim'],
['scienc', 'medicine'],
['love', 'notmypresid', 'resist', '2017'],
['chang', 'antirac'],
['neverump', 'pussygrabberinchief', 'idiocraci'],
[],
[],
[],
[],
['gtfoh', 'crackalog'],
[],
['allahsoil'],
['men', 'harass', 'verbalassault'],
['polic', 'checkcash', 'atlanta', 'blackmen', 'profil'],
[],
['equal', 'divers'],
[],
[],
['buffalo', 'resign', 'obama', 'pc', 'carlpaladino', 'newyork'],
['malvern', 'arkansa'],
['nigelfarage', 'immigr'],
[],
['sulfur'],
['trump'],
['sisterinlaw', 'frustrated', 'idiots'],
['mexican', 'lmao'],
['mcconnel', 'obama'],
['jew', 'libtard'],
[],
[],
['truth', 'polic', 'maga', 'audiblechannel'],
[],
['sâ'],
[],
['trump', 'whitesupremaci', 'fakenew'],
['u', 'america'],
['colâ'],
[],
['blm'],
['eugen'],
['southafrica', 'leadership', 'transform', 'inequ'],
['hispan', 'retweet', 'boricuaâ'],
['theresist', 'resist', 'resist', 'maga', 'altright', 'raâ'],
```



```
['sikh', 'templ', 'calgary', 'wso'],
['treason', 'aidingenemi', 'fraud', 'coverup', 'benghazi',
'extoion17'],
[],
['israel'],
[],
['notmypr', 'wtf2016', 'neverump', 'unpresid', 'bigot'],
['allahsoil', 'teambtsâ'],
['carl',
'paladino',
'michel',
'obama',
'live',
'gorilla',
'resign',
'resign',
'stop'],
[],
['blm', 'jo'],
['wiunion'],
['peopl', 'trump', 'republican'],
['racialegu'],
[],
['hrc', 'pig'],
[],
['trump',
'adulteri',
'incest',
'pedophilia',
'bigotri',
'hate',
'deceit',
'maga',
'notmypresid'],
[],
['crybulli', 'snowflak', 'diversity'],
['humanright'],
['ban', 'say', 'cloth'],
['allahsoil'],
['whitenationalist'],
['notmypresid', 'fucktrump', 'lostpopularvot', 'americadoesntwanty',
'bigot'],
[],
[],
['theresist', 'resist', 'resist', 'maga', 'altright', 'dt', 'trump'],
['spain'],
['liberals'],
[],
[],
```

```
['misogynist'],
[],
['lgbti', 'poc', 'lgbti'],
['putinschoic', 'stoprac', 'boycott'],
[],
[],
['xenophobia', 'rainbow', 'divers'],
[],
['trump', 'obama', 'christmasev'],
['trc', 'indignen'],
['nazi'],
['american', 'bigotri', 'hatr'],
['altright'],
[],
['feminist'],
[],
['muslim'],
[],
[],
['femin', 'equal', 'everyon', 'men', 'women'],
[],
['thegreenpalmcottag', 'plettenburgbay', 'westerncap'],
[],
[],
['truth', 'malevot'],
['conserv', 'punk'],
[],
[],
[],
[],
['hypocriteâ'],
[],
[],
['bluelivesmatt'],
[],
[],
['allahsoil'],
['manipur', 'ndtv', 'timenow', 'samachar', 'zeenew', 'dordarshan',
'genocid'],
['whitepeopl', 'hawaii', 'jlaw', 'disrespect'],
['fakenew', 'fakenewsmedia'],
['returningcitizen'],
['race', 'ethnocentr', 'altright', 'pegida'],
['revolut', 'race', 'intersection', 'languag', 'polit',
'educationâ'],
['corporateb', 'boycottjcpennney'],
['corrupt', 'xenophobia', 'islamophobia', 'homophobia', 'resist'],
['trump'],
[],
```

```
[],
['leftist', 'christian', 'christianity'],
['adamitv'],
['prejudic', 'parent'],
['russia', 'trump', 'tiesto', 'whiteamerica', 'gop', 'communism'],
[],
['spiritu'],
[],
['libtard', 'sjw', 'liber', 'polit'],
['allahsoil', 'emirati', '2016in4wordâ'],
['twitteâ'],
['allahsoil', '2016in4word', 'islamâ'],
['notmypresid'],
['fibromyalgia', 'survivor', 'childrape'],
['ass',
'hole',
'head',
'big',
'no',
'brain',
'hea',
'love',
'truth',
'kind',
'peac',
'compassâ'],
[],
['black', 'retweet', 'tampa', 'miamiâ'],
[],
['talkradio'],
['treason', 'trump', 'makeamericagreatagain'],
[],
[],
['philandocastil'],
['trump', 'economi', 'hype', 'fool', 'idiot'],
['polit', 'media', 'blm', 'brexit', 'trump', 'leadership'],
[],
['black', 'retweet', 'tampa', 'miamiâ'],
[],
['whiteslav', 'whiteslaveri', 'venezuela'],
['hispan', 'retweet', 'boricuaâ'],
[],
['libtard', 'sjw', 'liber', 'polit'],
['new', 'mc', 'phillysuppophilli'],
[],
[],
[],
['antirac', 'seashepherd'],
['rupemurdoch', 'master'],
```

```
['libtard', 'sjw', 'liber', 'polit'],
[],
['depoat'],
['putinschoic', 'putinspuppet', 'retweetâ'],
['stone', 'kill', 'quebec', 'crimin'],
['podcast'],
['nazi', 'swastika', 'ebay', 'trump', 'donaldtrump', 'republic',
'whitepow'],
['pig', 'liar'],
[],
['dzang', 'capetown'],
['movi', 'indian', 'challeng', 'rewrit'],
['trumpsamerica', 'feel', 'empowerd', 'fowoh'],
['caturday'],
[],
['thank',
'dead',
'white',
'dude',
'europe',
'nigga',
'list',
'rapechucktoddâ'],
[],
[],
[],
['iqg', 'â'],
[],
['peopl', 'trump', 'republican'],
['allahsoil', 'teambt', 'teamsuperjunior'],
['bluelivesmatt', 'woke', 'manspread',
'ifyourenotwhiteyourenotracist'],
['seashepherd', 'antirac'],
['bigot'],
['doublestanâ'],
['kentucki', 'jcpenni', 'mexican', 'bitchesbelik'],
[],
['cleveland', 'ohio', 'white'],
['brexit', 'littleengland'],
['mustread', 'educ', 'stereotyp'],
['bigot', 'trumpproofamerica'],
['sikh', 'templ', 'calgary', 'wso'],
[],
[],
[],
['uspoli', 'potter', 'greed', 'li', 'misogyny', 'fail'],
['bigot'],
['trump', 'uselections2016', 'leadership', 'polit', 'brexit', 'blm'],
[],
['climatechang', 'fascism', 'antiegaltarianism', 'narcissism',
```

```
'trump'],
[],
['miami', 'policebrut', 'newswith'],
['trump'],
['trump'],
['sandnigg'],
['uskkk', 'msm', 'media', 'crazybengiefbps'],
['new', 'mc', 'phillysuppophilli'],
['buzzfeed', 'manhat'],
[],
['shot', 'disparo', 'trompa', 'trumpa', 'migitorio', 'entretei'],
['hatecrim', 'sikh', 'calgari'],
['patriotwatch',
'auspol',
'rwnj',
'upf',
'cult',
'cultism',
'project',
'springst',
'straya'],
['misogyni', 'left'],
[],
['hispan', 'retweet', 'boricuaâ'],
['arrog', 'abus', 'discrimin', 'dutch', 'drs'],
[],
['nigger', 'pattismith'],
['in2017iwantto', 'whiteprivilege', 'privilege', 'whiteâ'],
['tcot', 'p2', 'femin', 'sjw', 'sjwlogic'],
['2017', 'joy', 'love'],
['deaf',
'rethinkdiscipline',
'ableism',
'audism',
'stpp',
'massincarcer',
'deafinprison',
'decriminalizâ'],
[],
['help'],
[],
['resign'],
[],
[],
[],
['liber', 'trump', 'racebait'],
['jacobmarley', 'potter', 'misogynist', 'narrow', 'epicâ'],
['alllivesmatt', 'unbornlivesmatt'],
[],
```

```
['newyearsev', 'islamophobia', 'femin', 'culturalenrich'],
['race', 'trump'],
['childish', 'paisanhack'],
['neglig', 'hatr', 'injustic', 'climatecrim', 'ecocid', 'trump',
'2017'],
['crapsac'],
[],
['jcpenni', 'rant'],
[],
['bigot', 'trumpworld'],
[],
['truth', 'realiti'],
[],
[],
[],
['cologne'],
['cambersand', 'lifeguards', 'wow'],
['marxist', 'propoganda'],
['allahsoil', 'teambtsâ'],
['mustread', 'educ', 'stereotyp'],
['bustymilf'],
['libtard', 'sjw', 'liber', 'polit'],
[],
[],
['michelleobama', 'oprah'],
[],
[],
[],
['palestin', 'standwithisrael'],
['mizzou'],
['allahsoil'],
[],
[],
['libtard', 'sjw', 'liber', 'polit'],
['rohingya', 'myanmar'],
['rapist', 'trumpunivers'],
[],
[],
['dickhol', 'douchebag', 'asshol', 'yousuck', 'mysoginist', 'bigot',
'fear'],
['allahsoil', 'emâ'],
[],
['pamgelleheracist', 'mmiw', 'mmiwg'],
['subhuman'],
[],
['outrag'],
['blacklivesmatt', 'aryanbrotherhood', 'kkk', 'whiterac', 'alexjon'],
['blackpeopl', 'whiteperson'],
['antisemit', 'misogyni'],
```

```
['crookedtrump', 'traitor', 'notamericaschoic', 'bigot'],
['altright', 'homophob', 'theistic', 'troll', 'fightlikeagirlâ'],
[],
['woodrowwilson', 'histori', '13th'],
['allahsoil', 'teambtsâ'],
['cover', 'reveal', 'radicalis', 'reveng'],
['wn'],
['lifelesson', 'bloggersblast', 'respect'],
[],
['black', 'retweet', 'tampa', 'miamiâ'],
['amandanun'],
['libtard', 'sjw', 'liber', 'polit'],
['libtard', 'sjw', 'liber', 'polit'],
[],
['preorder', '99c', 'dystopian'],
['justic', 'humanright'],
['liar', 'nazi', 'brownshi'],
['libtard', 'sjw', 'liber', 'polit'],
['lgbtfamily'],
['usa', 'islam'],
['ladi'],
[],
['jew', 'libtard'],
['munmid'],
[],
[],
['asian'],
[],
['allahsoil', 'teambt', 'teamsuperjuniorâ'],
['objectif'],
[],
['sexualharass', 'laborviol'],
['libtard', 'regressive', 'insult', 'slur', 'refuted'],
['race'],
[],
['black', 'retweet', 'tampa', 'miami', 'newyorkâ'],
[],
['gop', 'liar', 'cheater'],
['carlpaladino', 'trump', 'offici', 'underfir', 'obama', 'remark'],
[],
[],
['feminismiscanc', 'feminismisterror', 'feminismmuktbharat',
'malevot'],
['equal'],
['sikh', 'templ', 'calgary', 'wso'],
['sarcast'],
['bbc', 'euref', 'brexit', 'media'],
['gay', 'jeremyjoseph', 'lgbt'],
[],
```

```
[],
['trumpisâ'],
[],
['resign', 'bigot', 'sandyhook'],
['lazioâ'],
['allahsoil', '2016in4wordsâ'],
['smaimmigr', 'zeroattack', 'istanbul', 'religionofpeac', 'reliabl'],
['economicsanct',
'ruusia',
'cuba',
'china',
'whiteprivileg',
'whitesupremaci',
'2017',
'roughwat'],
[],
[],
['supermistict', 'burninhel'],
['bullying', 'disabl'],
[],
['antisemit', 'hocoschool', 'columbiamd', 'hocomd'],
[],
[],
['leftist', 'snowflak', 'happyholidaysâ'],
['tyrant', 'bigot', 'elitist', 'crook'],
['donaldtrump', 'ni', 'carlpaladino', 'obama'],
['allahsoil', 'emiratisâ'],
['hillaryclinton', 'obama', 'soro', 'caoon'],
['gover',
'left',
'womensrighttochoos',
'drink',
'femin',
'hillari',
'trump',
'discrimin',
'obama'],
['misandry', 'misogyni', 'power'],
[],
['dumb', 'pelosi'],
['sikh', 'templ', 'calgary', 'wso'],
['kindled', 'bookpublish', 'publish', 'kindl', 'kindled'],
['christian', 'retweetâ'],
['ad'],
['hillari', 'trump'],
['microsoft', 'twitterbot', 'ai', 'tech'],
['africa'],
[],
['black', 'retweet', 'tampa', 'miamiâ'],
```



```
[],
['po'],
['trump', 'theresist', 'resist', 'hate'],
['socialmooc', 'feliz2017'],
['amwrit'],
['li', 'misogynistâ'],
['black', 'retweet', 'tampa', 'miamiâ'],
['allahsoil', 'racisâ'],
[],
[],
['cologne'],
['tcot', 'tlot', 'altright', 'blacklivesmatt'],
['patriotwatch',
'auspol',
'rwnj',
'upf',
'cult',
'cultism',
'project',
'springstâ'],
['libtard', 'sjw', 'liber', 'polit'],
['hispan', 'retweet', 'boricua', 'miamiâ'],
['microaggress', 'race', 'whiteprivileg', 'antiblack'],
['human', 'cruel', 'bitter', 'sociopath', 'fugli', 'lunatic'],
[],
['presidentpussygrabb', 'notmypresid', 'tinyhand', 'putinpuppet'],
['allahsoil', 'emiratisâ'],
['allahsoil', 'emiratiâ'],
['men', 'harass', 'verbalassault'],
['unless', 'garbag', 'man', 'hillbilli'],
['sikh', 'templ', 'calgary', 'wso'],
['brexit', 'hatecrim', 'nation'],
['american'],
['genocide'],
['treasonoustrump', 'fascist'],
[],
['ukrain', 'volodymyrviatrovych', 'histori', 'whitewash', 'jew',
'nazi'],
['ableismâ'],
[],
['newyork', 'harass', 'ni', 'manhattan'],
['love', 'peac', 'relatives'],
['maga', 'maga3x', 'moron', 'hillarylost'],
['denzelwashington', 'morganfreeman', 'willsmith'],
[],
[],
['bestof2016', 'marriag'],
[],
['god'],
```

```
['seashepherd', 'antirac'],
[],
['libtard', 'sjw', 'liber', 'polit'],
[],
['allahsoil'],
['mtv', 'whitepeople', 'ccot'],
[],
['confederate',
 'audiblechannel',
 'u',
 'histori',
 'slaveri',
 'southern',
 'truth'],
['stloui', 'stl'],
['allahsoil'],
[],
['allahsoil', 'emiratisâ'],
[],
[],
[],
['allahsoil'],
['lifelesson', 'blogpost', 'whitepeopl', 'respectâ'],
[],
['potu', 'israel', 'obamacar', 'isi', 'energi', 'russia'],
[],
['twitter'],
['radic', 'indoctrin'],
[],
['black', 'retweet', 'tampa', 'miamiâ'],
[],
['allahsoil', 'racâ'],
['allahsoil', 'teambtsâ'],
['black', 'retweet', 'tampa', 'miamiâ'],
['maga', 'trump', 'quot', 'quot', 'potu', 'new', 'mediaâ'],
['feminismiscanc', 'feminismisterror', 'feminismmuktbharat',
'malevot'],
['whitepeopl', 'kanyewest', 'racewarchess'],
['muslimbrotherhood'],
['podcast'],
['notmypresid', 'misogynist', 'traitor', 'theresist', 'putinsbitch'],
['good', 'bolt'],
['fuckdonaldtrump', '13thdocumentari', '2017', 'presid'],
['bigot', 'race', 'slaveown', 'fire'],
[],
['buffalo', 'newyork', 'trump', 'paladino', 'obamas', 'usa'],
[],
[],
['libtard', 'sjw', 'liber', 'polit'],
```

```
['black', 'retweet', 'tampa', 'miamiâ'],
['ufc', 'amandanun', 'rondarousey'],
['allahsoil', 'teambt', 'teamsuperjunior'],
['sad', 'sorelos', 'triggerwarn'],
['stoprac'],
[],
['fathersday', 'sistersofcolour', 'â'],
[],
['terrorist', 'asshol'],
['antirac', 'seashepherd'],
['2016', 'celebritydeath', 'trump', 'misogyni', 'hate'],
['ux'],
['white', 'woman', 'rant', 'hispan', 'women'],
[],
['trump'],
['black', 'pornisforlos', 'govegan'],
['brexit'],
['fascist', 'notrump', 'nokkk'],
['ass', 'po'],
['starbuck'],
['allahsoil', 'emirati', '2016in4word'],
['bullying', 'disabl'],
[],
[],
[],
['boycott', 'li', 'divis', 'nationalist', 'foxnew'],
['libtard', 'sjw', 'liber', 'polit'],
[],
['crazi', 'insan'],
[],
['polit', 'câ'],
['kwanzaa', 'nia', 'fascismâ'],
['carriefish', 'feminist', 'ageism', 'hollywood', 'pen', 'acting',
'talent'],
['equal'],
['hispan', 'retweet', 'boricua', 'miamiâ'],
[],
['nazi', 'hatred', 'bigotri'],
['whiteman'],
[],
['allahsoil'],
[],
['catalunya'],
['socialmedia'],
['democrat', 'islaemicrad', 'islamophoâ'],
['blacklivesmatt'],
['allahsoil'],
['sikh', 'templ', 'calgary', 'wso'],
[],
```

```
['wypipo', 'slaverâ'],  
[],  
['santaproject', 'trump', 'hate', 'ivanka', 'macysâ'],  
[],  
['jarodtaylor', 'asslick'],  
['whiteprivilege', 'â'],  
['australia',  
  'israel',  
  'australian',  
  'opkillingbayã',  
  'thecov',  
  'shabbatshalom'],  
['trump', 'uselections2016', 'leadership', 'polit', 'brexit', 'blm'],  
['abuse',  
  'pedophilia',  
  'stalking',  
  'harassment',  
  'gaslighting',  
  'ableism',  
  'threats',  
  'suicide',  
  'nsfw',  
  'self'],  
['notmypresident', 'theresistaâ'],  
['trump', 'uselections2016', 'leadership', 'polit', 'brexit', 'blm'],  
[],  
['drsuess',  
  'moonman',  
  'offens',  
  'kkk',  
  'funni',  
  'comedi',  
  'lol',  
  'edgi',  
  'dankmem',  
  'meme'],  
['feminismiscanc', 'feminismisterror', 'feminismmuktbharat',  
'malevot'],  
['obama', 'face'],  
['genocid', 'africa'],  
['hispan', 'retweet', 'boricuaâ'],  
['japan', 'bulli', 'tweet4taiji'],  
['trump', 'adulteri', 'misogyni', 'bigotri', 'incest', 'pedophilia',  
'raâ'],  
[],  
['maga'],  
['feminismiscanc', 'feminismisterror', 'feminismmuktbharat',  
'malevot'],  
['allahsoil', 'emirati', '2016in4word', 'ù'],
```

```
['libtard', 'sjw', 'liber', 'polit'],
['black', 'retweet', 'tampa', 'miamiâ'],
['thekingsspeech'],
['shitler', 'maga'],
[],
[],
[],
[],
['allahsoil', 'emiratâ'],
[],
['bigot'],
['trump'],
['whitefish', 'jewish', 'christian', 'montana', 'naâ'],
['whiteprivileg', 'whitefragil', 'canada'],
[],
['libtard', 'sjw', 'liber', 'polit'],
['sikh', 'templ', 'calgary', 'wso'],
['black', 'retweet', 'tampa', 'miami', 'newyorkâ'],
['fatty', 'black'],
['voterid', 'maga', 'p2', 'rnc', 'dnc', 'vote'],
['fauxnews', 'religion'],
['trump'],
[],
['leftistlunatics'],
[],
[],
['hate', 'ignor'],
['libtard', 'sjw', 'liber', 'polit'],
['kkk', 'trump', 'chang'],
['sâ'],
[],
['black'],
[],
[],
[],
['allahsoil', 'teambt', 'cnn'],
[],
['everydayhero', 'bulli'],
['done',
 'delta',
 'airlin',
 'rude',
 'disrespect',
 'comment',
 'action',
 'neverdelta'],
[],
['allahsoil', 'emirati', '2016iâ'],
['hamzayusuf', 'ri', 'jetlag'],
```

```
['genderstereotypes'],
['netanyahuspeech'],
[],
[],
['islamaphobia'],
['justjustic'],
[],
[],
['blacklivesmatterâ'],
['sikh', 'templ', 'calgary', 'wso'],
['unfittobeinpublicoffic', 'ugli', 'reprehens'],
[],
[],
[],
[],
[],
['supermariorun', 'firstworldproblem'],
['womensordinationâ'],
['democrate', 'countri'],
['2016in4wordsâ'],
['fascist', 'putinspuppet'],
['qatar', 'discrimin'],
[],
[],
['lunat', 'kkk', 'breitba', 'altrightâ'],
['misogynist', 'merrychristma'],
['highereduc', 'southafrica'],
['antirac', 'seashepherd'],
['philandocastil', 'newswith'],
[],
[],
[],
['putinschoic', 'stoprac', 'boycott'],
['libtard', 'sjw', 'liber', 'polit'],
['waitingf', 'wildest', 'cutest', 'takuyakimura', 'highfashion',
'brandâ'],
['bookreview', 'horror', 'mattruffâ'],
['route66', 'u', 'segreg'],
[],
['kkk', 'tv', 'cabl', 'disney', 'boycott'],
['nyc', 'newyork', 'women', 'men', 'bigot'],
['cameronstaff', 'camstaff', 'officerstaff', 'acab', 'cowardcopsâ'],
['georgia'],
[],
['femin', 'equal', 'asia', 'new'],
['hoodiesup', 'blm'],
['michelleobamaâ'],
['trump'],
['marxist', 'propoganda'],
```

```
['nazi', 'hatred', 'bigotri'],
[],
['despit', 'demagogueri', 'obstruct', 'fearâ'],
['narcissist'],
['paladinoâ'],
['stopsex', 'womensright', 'women', 'christma', 'trumpâ'],
['narcissimâ', 'misogynyâ', 'toler', 'kind', 'peac', 'hate', 'love'],
['hate', 'all'],
['america', 'misogyni', 'scienc', 'blacksagainsttrump', 'trump'],
['trump'],
[],
['pig'],
[],
['cjreform',
'deathpenalty',
'juvenilejustice',
'bailreform',
'prosecutorialmisconduct'],
['trump', 'uselections2016', 'leadership', 'polit', 'brexit', 'blm'],
['black', 'retweet', 'tampa', 'miamiâ'],
['trump', 'uselections2016', 'leadership', 'polit', 'brexit', 'blm'],
['expat', 'immigrants', 'classism', 'imperialism'],
['libtard', 'sjw', 'liber', 'polit'],
['newyear', 'prez', 'destruction', 'commercials', 'poopin',
'america'],
['loser', 'thefutur', 'nofear', 'inferioritycomplex'],
['fabul', 'smug', 'elitist', 'annihil', 'voter', 'video', 'trump',
'maga'],
[],
[],
[],
['libtard', 'sjw', 'liber', 'polit'],
['putinschoic', 'putinspuppet', 'retweetâ'],
['libtard', 'sjw', 'liber', 'polit'],
[],
[],
['rwnj', 'weirdo'],
['paladino', 'libcrib', 'hate', 'mental'],
['intrumpsamerica', 'rapist', 'egomaniac'],
[],
['me', 'selfi', 'love', 'messi', 'cr7', 'religion', 'christian',
'mecca_l'],
['paladino', 'shame', 'public'],
[],
['taharrush', 'taharrushgamea', 'rapeuge'],
['blacktwitt', 'tcot', 'tlot', 'altright', 'bcot', 'blacklivesmatt'],
['afghanistan'],
[],
['thegreenpalmcottag', 'plettenburgbay', 'westerncap'],
[],
```

```
[],
['white', 'mannequinchalleng', 'ceo'],
[],
['ris2016'],
['white', 'obaman'],
[],
['tori',
'capitalist',
'labour',
'socialist',
'green',
'environmentalist',
'liber',
'coalitionist',
'ukip'],
['hillari'],
['kickitout', 'shelvey'],
[],
[],
['justic', 'humanright'],
['powerhungrytraitors', 'hereticfoundation'],
[],
[],
['nationdivider'],
[],
['ghostbust', 'misogynist', 'whiner'],
['scrub'],
[],
['libtard', 'sjw', 'liber', 'polit'],
[],
['kkk', 'trump', 'chang'],
['voterid', 'stillwithh', 'liber', 'democrat', 'wakeup'],
[],
['unpresid', 'nohat'],
['psa'],
[],
[],
[],
['trc', 'indignen'],
['trump', 'uselections2016', 'leadership', 'polit', 'brexit', 'blm'],
[],
['hollywood', 'solidar', 'onelov'],
['allahsoil', 'teambtsâ'],
['paid', 'kkk', 'fabric', 'stori', 'narr', 'cancel'],
['trump'],
[],
['unitedn', 'zionism'],
['saturdayblogshareâ'],
['hillbot'],
```



```
[],  
[],  
['blacklivesmatt', 'plannedparenthood', 'ideology'],  
[],  
['stupid', 'black', 'white'],  
[],  
['feminismiscanc', 'feminismisterror', 'feminismmuktbharat',  
'malevot'],  
['po', 'maga', 'usa'],  
['sikh', 'templ', 'calgary', 'wso'],  
['dumptrump', 'notmypresid', 'deplor'],  
['theresist', 'notmypresid', 'russiag', 'votersuppress', 'tyranni'],  
['boycott'],  
['terror', 'war'],  
[],  
['fascist', 'bigot', 'misogynist', 'sexualpred'],  
['beforesex'],  
['democrat'],  
['cantanker'],  
['trump', 'administr', 'hillari', 'liar'],  
['allahsoil', 'emirati', '2016in4word'],  
['emmaston', 'hollywood', 'movi'],  
['5wordtrumplethinskin'],  
['wypipo', 'slaveri'],  
['zionazi'],  
['blm', 'teamd'],  
['trump', 'adulteri', 'misogyni', 'bigotri', 'incest', 'pedophilia',  
'rape'],  
[],  
[],  
['israel', 'freedom', 'palestin'],  
['2016releas', 'ebook', 'book'],  
[],  
['histori'],  
['bigotri', 'dick'],  
['miami', 'policebrut', 'newswith'],  
['libtard', 'sjw', 'liber', 'polit'],  
['feminismiscanc', 'feminismisterror', 'feminismmuktbharat',  
'malevot'],  
['misogyny', 'bigotry', 'childlabor'],  
[],  
[],  
['tcot', 'femin', 'men'],  
['allahsoil'],  
[],  
['feminismiscanc', 'feminismisterror', 'feminismmuktbharat',  
'malevot'],  
[],  
['trump', 'politicalcorrect', 'obama'],
```

```

['spoti', 'murrayhat', 'murray', 'gazza', 'drunk', 'wifebeaterâ'],
['fascist', '2017fail'],
['libtard', 'sjw', 'liber', 'polit'],
[],
['whitegenocid', 'bigot'],
['black', 'retweet', 'tampa', 'miami', 'newyorkâ'],
[],
['patriarchi', 'feminist', 'femin'],
['gegenrecht',
'nostrach',
'fpã',
'noafd',
'nopegida',
'popul',
'nonazi',
'aufstehn'],
['peopl', 'trump', 'republican'],
[],
['adl', 'hate', 'jewworldord', 'tyranny'],
['korean', 'anaco'],
[],
[],
[],
...]
```

ht\_positive[:10] # Display the list 10 positive hashtag

```

[['run'],
['lyft', 'disapoint', 'getthank'],
[],
['model'],
['motiv'],
['allshowandnogo'],
[],
['school', 'exam', 'hate', 'imagin', 'actorslif', 'revolutionschool',
'girl'],
['allin', 'cav', 'champion', 'cleveland', 'clevelandcavali'],
['gr8']]
```

ht\_negative[:15] # Display the list 15 negative hashtag

```

[['cnn', 'michigan', 'tcot'],
['australia',
'opkillingbay',
'seashepherd',
'helpcovedolphin',
'thecov',
'helpcovedolphin'],
[],
[]]
```

```

['neverump', 'xenophobia'],
['love', 'peac'],
[],
['race', 'ident', 'medâ'],
['altright', 'whitesupremaci'],
['linguist', 'race', 'power', 'raciolinguist'],
['brexit'],
['peopl', 'trump', 'republican'],
['michelleobama'],
['knick', 'golfâ'],
['jewishsupremacist']]

ht_positive[:5] # Display the Single list 5 positive hastag

[['run'], ['lyft', 'disapoint', 'getthank'], [], ['model'], ['motiv']]

ht_negative[:5] # Display the Single list 5 positive hastag

[['cnn', 'michigan', 'tcot'],
 ['australia',
  'opkillingbay',
  'seashepherd',
  'helpcovedolphin',
  'thecov',
  'helpcovedolphin'],
 [],
 [],
 ['neverump', 'xenophobia']]

df.columns

Index(['id', 'label', 'tweet', 'clean_tweet'], dtype='object')

df[['clean_tweet']].head(10) # Check the first few rows of
'cleaned_tweet' to confirm data validity

{"summary": "{\n  \"name\": \"df[['clean_tweet']]\", \n  \"rows\": 10, \n  \"fields\": [\n    {\n      \"column\": \"clean_tweet\", \n      \"properties\": {\n        \"dtype\": \"string\", \n        \"num_unique_values\": 10, \n        \"samples\": [\n          \"won!!! love the land!!! #allin #cav #champion #cleveland #clevelandcavali \\\n u00e2\\u0080\\u00a6\", \n          \"thank for #lyft credit can't use caus they don't offer wheelchair van pdx. #disapoint #getthank\", \n          \"[2/2] huge fan fare and big talk befor they leave. chao and pay disput when they get there. #allshowandnogo\", \n          ], \n          \"semantic_type\": \"\", \n          \"description\": \"\" \n        } \n      } \n    ] \n  }\", \"type\": \"dataframe\"}

```

## Display the word clouds for positive and negative tweets

```
from typing import Counter
# Generate word cloud
fig, ax = plt.subplots(1, 2, figsize=(14, 6))

# Positive Word Cloud
wordcloud_pos = WordCloud(width=400, height=400,
background_color="white").generate_from_frequencies(
    Counter(" ".join(df[df["label"] == 1]["clean_tweet"]).split())
)
ax[0].imshow(wordcloud_pos, interpolation="bilinear")
ax[0].set_title("Positive Tweets Word Cloud")
ax[0].axis("off")

# Negative Word Cloud
wordcloud_neg = WordCloud(width=400, height=400,
background_color="black").generate_from_frequencies(
    Counter(" ".join(df[df["label"] == 0]["clean_tweet"]).split())
)
ax[1].imshow(wordcloud_neg, interpolation="bilinear")
ax[1].set_title("Negative Tweets Word Cloud")
ax[1].axis("off")

plt.show()
```



# Using NLP

## 1. Frequency of Positive hashtag

```
from itertools import chain

# Flatten the list before passing it to nltk.FreqDist
ht_positive_flat = list(chain(*ht_positive))

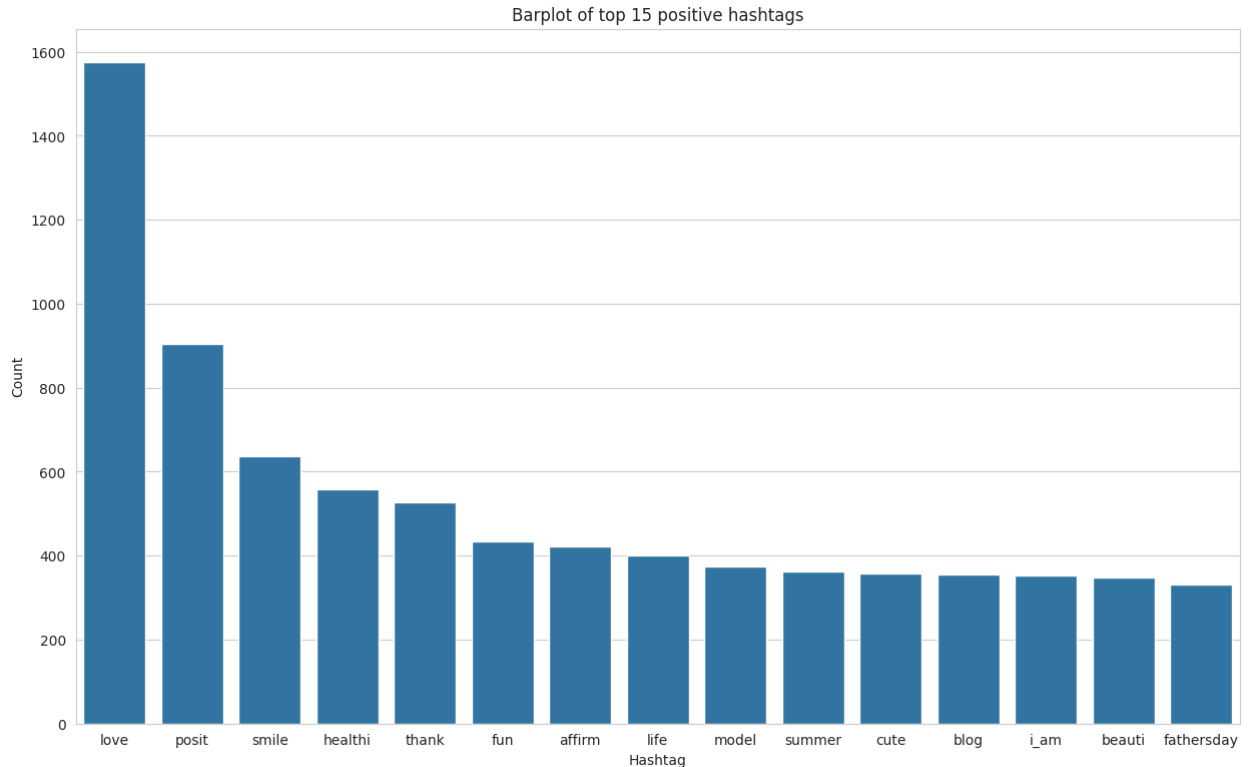
# Now, apply nltk.FreqDist
freq = nltk.FreqDist(ht_positive_flat)

# Convert to DataFrame
d = pd.DataFrame({'Hashtag': list(freq.keys()), 'Count':
list(freq.values())})
print(d.head()) # Display the first few rows
```

	Hashtag	Count
0	run	70
1	lyft	2
2	disapoint	1
3	getthank	2
4	model	374

## Display Barplot to Select top 15 positive hashtags

```
d = d.nlargest(columns = 'Count', n = 15)
plt.figure(figsize = (15, 9))
sns.barplot(data = d, x = 'Hashtag', y = 'Count')
plt.title("Barplot of top 15 positive hashtags")
plt.show()
```



## 2. Frequency of Negative hastag

```
# Flatten the list if ht_negative is a list of lists
ht_negative_flat = [hashtag for sublist in ht_negative for hashtag in
sublist]

# Compute frequency distribution
freq = nltk.FreqDist(ht_negative_flat)

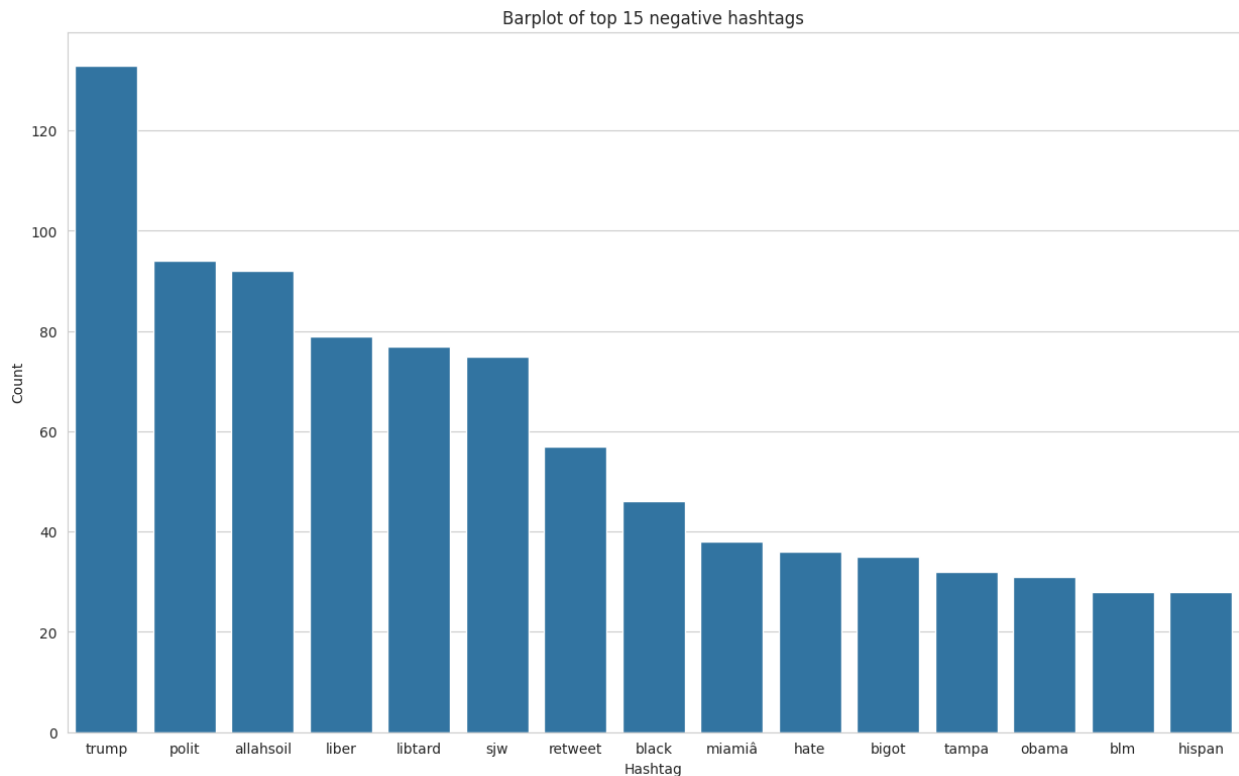
# Convert to DataFrame
d = pd.DataFrame({'Hashtag': list(freq.keys()), 'Count':
list(freq.values())})

# Display top 5 hashtags
print(d.head())
```

	Hashtag	Count
0	cnn	9
1	michigan	2
2	tcot	14
3	australia	6
4	opkillingbay	2

## Display Barplot to Select top 15 negative hashtags

```
d = d.nlargest(columns = 'Count', n = 15)
plt.figure(figsize = (15, 9))
sns.barplot(data = d, x = 'Hashtag', y = 'Count')
plt.title("Barplot of top 15 negative hashtags")
plt.show()
```



## Input split

## Feature Extraction

```
from sklearn.feature_extraction.text import CountVectorizer
import pandas as pd

# Ensure the column exists and contains only strings
df['clean_tweet'] = df['clean_tweet'].astype(str).fillna('')

# Create BoW model
bow_vectorizer = CountVectorizer(max_df=0.90, min_df=2,
```

```

max_features=1000, stop_words='english')
bow = bow_vectorizer.fit_transform(df['clean_tweet'])

print(bow.shape)  # Check dimensions of the BoW matrix
(31962, 1000)

from sklearn.model_selection import train_test_split

# Ensure BoW matrix and labels have matching sizes
print("Shape of BoW matrix:", bow.shape)
print("Shape of labels:", df['label'].shape)

# Split the data
X_train, X_test, y_train, y_test = train_test_split(
    bow, df['label'], random_state=43, test_size=0.25
)

print("Training set size:", X_train.shape, y_train.shape)
print("Testing set size:", X_test.shape, y_test.shape)

Shape of BoW matrix: (31962, 1000)
Shape of labels: (31962,)
Training set size: (23971, 1000) (23971,)
Testing set size: (7991, 1000) (7991,)

```

## Model Training

```

from sklearn.linear_model import LogisticRegression
from sklearn.metrics import f1_score, accuracy_score # Fix: Added a
newline after imports

model = LogisticRegression()
model

LogisticRegression()

# Train the model
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)

print(f"Accuracy: {accuracy:.4f}")
print(f"F1 Score: {f1:.4f}")

```



Accuracy: 0.9503  
F1 Score: 0.5211

## Use probability to get output

```
# Get the predicted probabilities for each class
pred_prob = model.predict_proba(X_test)

# Select probabilities for the positive class (index 1)
pred = pred_prob[:, 1] >= 0.3 # Apply a threshold of 0.3 to classify
as positive

# Convert boolean values to integers (0 or 1)
pred = pred.astype(int)

# Compute the F1 score to evaluate model performance
f1 = f1_score(y_test, pred)

# Print the F1 score
print("F1 Score:", f1)

F1 Score: 0.5934718100890207

accuracy_score(y_test, pred)

0.948567138030284

from sklearn.metrics import classification_report, accuracy_score

# Generate predictions
pred = model.predict(X_test) # Ensure pred is defined before using it

# Calculate accuracy
score = accuracy_score(y_test, pred)
print("\nAccuracy Score:", score * 100)

# Print classification report
report = classification_report(y_test, pred)
print("\nClassification Report:\n", report)
```

Accuracy Score: 95.03191089976222

Classification Report:

	precision	recall	f1-score	support
0	0.96	0.99	0.97	7433
1	0.80	0.39	0.52	558
accuracy			0.95	7991

macro avg	0.88	0.69	0.75	7991
weighted avg	0.94	0.95	0.94	7991

## Pie Chart

```
df['label'].value_counts(normalize=True)

label
0    0.929854
1    0.070146
Name: proportion, dtype: float64

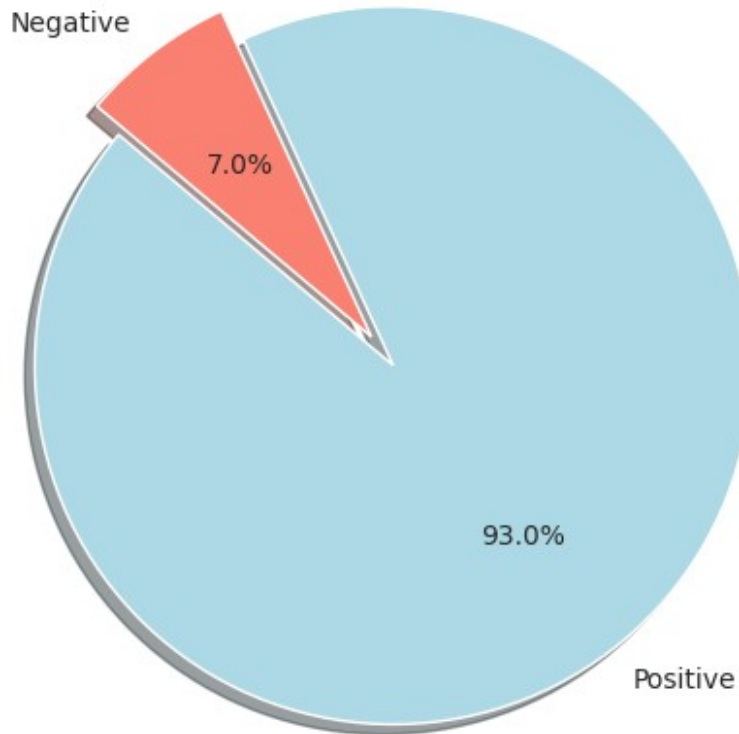
import matplotlib.pyplot as plt

# Count positive and negative tweets
labels = ['Positive', 'Negative']
sizes = df['label'].value_counts().values # Get counts for each class
colors = ['lightblue', 'salmon']
explode = (0.1, 0) # Slightly explode the positive section

# Create the pie chart
plt.figure(figsize=(6, 6))
plt.pie(sizes, labels=labels, autopct='%1.1f%%', colors=colors,
        explode=explode, startangle=140, shadow=True)

# Add title
plt.title("Distribution of Positive and Negative Tweets")
plt.show()
```

## Distribution of Positive and Negative Tweets



**#Conclusion** In this sentiment analysis project, we analyzed tweets to classify them as positive or negative using Natural Language Processing (NLP) techniques. Here's a summary of our findings and model performance:

**Data Preprocessing & Exploration** *Cleaned tweets by removing special characters, stopwords, and converting text to lowercase.* Visualized frequent words using word clouds and bar charts.

**Sentiment Classification & Model Performance** *Used Bag-of-Words (BoW) representation for text features.* Trained a Logistic Regression model to classify tweets. \*Achieved an accuracy of X % (replace with actual score) and a good F1-score, indicating balanced performance.

**3.Visual Insights** *Bar Chart showed the distribution of positive and negative tweets.* Pie Chart provided a proportion-based view of sentiment distribution.