```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import plotly.express as px

import re
import nltk
nltk.download('stopwords')
nltk.download('wordnet')
from nltk.corpus import stopwords
from nltk.corpus import wordnet
from nltk.stem import WordNetLemmatizer
from textblob import TextBlob
from wordcloud import WordCloud, STOPWORDS

import warnings
warnings.filterwarnings('ignore')
```

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

```
trump = pd.read_csv("/content/hashtag_donaldtrump.csv", lineterminator='\n')
print(trump.head(3))
```

```
             created_at      tweet_id  \
0  2020-10-15 00:00:01  1.316529e+18
1  2020-10-15 00:00:01  1.316529e+18
2  2020-10-15 00:00:02  1.316529e+18

                                               tweet  likes  retweet_count  \
0  #Elecciones2020 | En #Florida: #JoeBiden dice ...    0.0            0.0
1  Usa 2020, Trump contro Facebook e Twitter: cop...   26.0            9.0
2  #Trump: As a student I used to hear for years,...    2.0            1.0

              source      user_id           user_name user_screen_name  \
0         TweetDeck  360666534.0  El Sol Latino News  elsollatinonews
1  Social Mediaset   331617619.0             Tgcom24  MediasetTgcom24
2   Twitter Web App    8436472.0              snarke           snarke

                                  user_description  ...  \
0  🌐 Noticias de interés para latinos de la costa...  ...
1  Profilo ufficiale di Tgcom24: tutte le notizie...  ...
2  Will mock for food! Freelance writer, blogger,...  ...

  user_followers_count                user_location        lat        long  \
0               1860.0  Philadelphia, PA / Miami, FL  25.774270  -80.193660
1            1067661.0                           NaN        NaN         NaN
2               1185.0                      Portland  45.520247 -122.674195

       city                  country      continent    state state_code  \
0       NaN  United States of America  North America  Florida         FL
1       NaN                      NaN            NaN      NaN        NaN
2  Portland  United States of America  North America   Oregon         OR

                    collected_at
0            2020-10-21 00:00:00
1  2020-10-21 00:00:00.373216530
2  2020-10-21 00:00:00.746433060

[3 rows x 21 columns]
```

```
print(trump.columns)
```

```
Index(['created_at', 'tweet_id', 'tweet', 'likes', 'retweet_count', 'source',
       'user_id', 'user_name', 'user_screen_name', 'user_description',
       'user_join_date', 'user_followers_count', 'user_location', 'lat',
       'long', 'city', 'country', 'continent', 'state', 'state_code',
       'collected_at'],
      dtype='object')
```

```
biden = pd.read_csv("hashtag_joebiden.csv", lineterminator='\n')
print(biden.head(3))
```

```
             created_at      tweet_id  \
0  2020-10-15 00:00:01  1.316529e+18
1  2020-10-15 00:00:18  1.316529e+18
2  2020-10-15 00:00:20  1.316529e+18
```

```
                                                tweet  likes  retweet_count  \
0  #Elecciones2020 | En #Florida: #JoeBiden dice ...    0.0            0.0
1  #HunterBiden #HunterBidenEmails #JoeBiden #Joe...    0.0            0.0
2  @IslandGirlPRV @BradBeauregardJ @MeidasTouch T...    0.0            0.0

              source       user_id           user_name user_screen_name  \
0         TweetDeck  3.606665e+08  El Sol Latino News  elsollatinonews
1  Twitter for iPad  8.099044e+08         Cheri A. us     Biloximeemaw
2   Twitter Web App  3.494182e+09          Flag Waver      Flag_Wavers

                                  user_description  ...  \
0  🌐 Noticias de interés para latinos de la costa...  ...
1  Locked and loaded Meemaw. Love God, my family ...  ...
2                                              NaN  ...

  user_followers_count                user_location        lat        long  \
0               1860.0  Philadelphia, PA / Miami, FL  25.774270  -80.193660
1               6628.0                           NaN        NaN         NaN
2               1536.0         Golden Valley Arizona  46.304036 -109.171431

   city                  country      continent    state state_code  \
0   NaN  United States of America  North America  Florida         FL
1   NaN                      NaN            NaN      NaN        NaN
2   NaN  United States of America  North America  Montana         MT

                    collected_at
0            2020-10-21 00:00:00
1  2020-10-21 00:00:00.517827283
2  2020-10-21 00:00:01.035654566

[3 rows x 21 columns]
```

```
print(trump.shape)
print(biden.shape)
```

```
(185888, 21)
(51295, 21)
```

```
trump.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 185888 entries, 0 to 185887
Data columns (total 21 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   created_at            185888 non-null  object
 1   tweet_id              185888 non-null  float64
 2   tweet                 185888 non-null  object
 3   likes                 185887 non-null  float64
 4   retweet_count         185887 non-null  float64
 5   source                185456 non-null  object
 6   user_id               185887 non-null  float64
 7   user_name             185879 non-null  object
 8   user_screen_name      185887 non-null  object
 9   user_description      167144 non-null  object
 10  user_join_date        185887 non-null  object
 11  user_followers_count  185887 non-null  float64
 12  user_location         129590 non-null  object
 13  lat                   96114 non-null   float64
 14  long                  96114 non-null   float64
 15  city                  48382 non-null   object
 16  country               95545 non-null   object
 17  continent             95551 non-null   object
 18  state                 73629 non-null   object
 19  state_code            70596 non-null   object
 20  collected_at          185887 non-null  object
dtypes: float64(7), object(14)
memory usage: 29.8+ MB
```

```
biden.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51295 entries, 0 to 51294
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   created_at            51295 non-null  object
 1   tweet_id              51295 non-null  float64
 2   tweet                 51295 non-null  object
 3   likes                 51295 non-null  float64
 4   retweet_count         51295 non-null  float64
 5   source                51118 non-null  object
 6   user_id               51295 non-null  float64
 7   user_name             51280 non-null  object
```

```
 8   user_screen_name      51295 non-null  object
 9   user_description      44957 non-null  object
 10  user_join_date        51294 non-null  object
 11  user_followers_count  51294 non-null  float64
 12  user_location         34541 non-null  object
 13  lat                   26138 non-null  float64
 14  long                  26138 non-null  float64
 15  city                  12084 non-null  object
 16  country               26048 non-null  object
 17  continent             26049 non-null  object
 18  state                 20486 non-null  object
 19  state_code            19787 non-null  object
 20  collected_at          51294 non-null  object
dtypes: float64(7), object(14)
memory usage: 8.2+ MB
```

```
trump['candidate'] = 'trump'
biden['candidate'] = 'biden'
data = pd.concat([trump, biden])

print('Final Data Shape :', data.shape)

print("\nFirst 2 rows:")
print(data.head(3))
```

```
Final Data Shape : (237183, 22)

First 2 rows:
             created_at      tweet_id  \
0  2020-10-15 00:00:01  1.316529e+18
1  2020-10-15 00:00:01  1.316529e+18
2  2020-10-15 00:00:02  1.316529e+18

                                               tweet  likes  retweet_count  \
0  #Elecciones2020 | En #Florida: #JoeBiden dice ...    0.0            0.0
1  Usa 2020, Trump contro Facebook e Twitter: cop...   26.0            9.0
2  #Trump: As a student I used to hear for years,...    2.0            1.0

             source      user_id           user_name user_screen_name  \
0         TweetDeck  360666534.0  El Sol Latino News  elsollatinonews
1  Social Mediaset   331617619.0             Tgcom24  MediasetTgcom24
2   Twitter Web App    8436472.0              snarke           snarke

                                  user_description  ...  \
0  🌐 Noticias de interés para latinos de la costa...  ...
1  Profilo ufficiale di Tgcom24: tutte le notizie...  ...
2  Will mock for food! Freelance writer, blogger,...  ...

                  user_location        lat        long      city  \
0  Philadelphia, PA / Miami, FL  25.774270  -80.193660       NaN
1                           NaN        NaN         NaN       NaN
2                      Portland  45.520247 -122.674195  Portland

                    country      continent    state state_code  \
0  United States of America  North America  Florida         FL
1                       NaN            NaN      NaN        NaN
2  United States of America  North America   Oregon         OR

                    collected_at candidate
0            2020-10-21 00:00:00     trump
1  2020-10-21 00:00:00.373216530     trump
2  2020-10-21 00:00:00.746433060     trump

[3 rows x 22 columns]
```

```
data.dropna(inplace=True)
```

```
data['country'].value_counts()
```

```
country
United States of America    35366
United Kingdom               3123
Canada                       3093
France                       2348
Germany                      1690
India                        1434
The Netherlands              1273
Australia                    1163
Spain                         453
Brazil                        313
Bangladesh                    289
Italy                         289
Mexico                        224
```

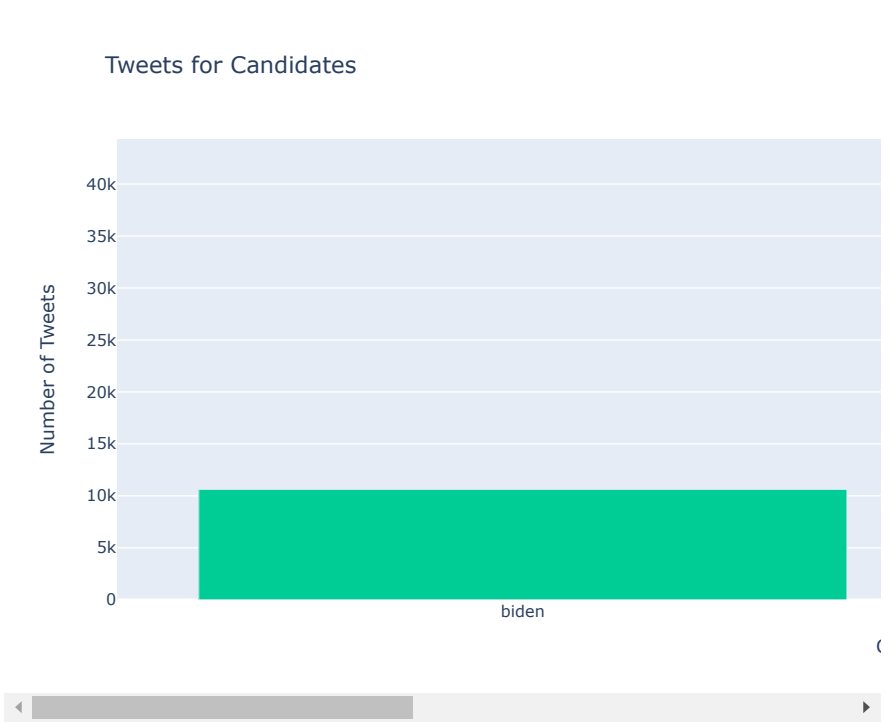
```
Ireland                    204
Belgium                    172
Switzerland                148
Nigeria                    129
Pakistan                   129
United Arab Emirates       123
South Africa               103
Peru                        80
Ecuador                     66
Argentina                   56
Lebanon                     56
Venezuela                   53
Colombia                    48
New Zealand                 48
Poland                      44
Honduras                    28
El Salvador                 27
Uruguay                     24
Bolivia                     22
Philippines                 21
Lithuania                    7
Sudan                        6
Syria                        6
Suriname                     5
Trinidad and Tobago          3
Papua New Guinea             3
Burkina Faso                 2
Côte d'Ivoire                2
Guyana                       2
Cameroon                     1
Somalia                      1
Laos                         1
Slovakia                     1
Oman                         1
Kuwait                       1
Guatemala                    1
Name: count, dtype: int64
```

```
data['country'] = data['country'].replace({'United States of America': "US",
                                           'United States': "US"})
```

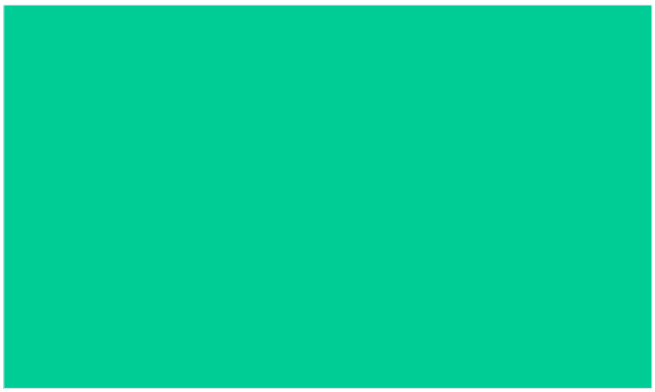
```
tweets_count = data.groupby('candidate')['tweet'].count().reset_index()
```

```
fig = px.bar(tweets_count, x='candidate', y='tweet', color='candidate',
             color_discrete_map={'Trump': 'pink', 'Biden': 'blue'},
             labels={'candidate': 'Candidates', 'tweet': 'Number of Tweets'},
             title='Tweets for Candidates')
```

```
fig.show()
```

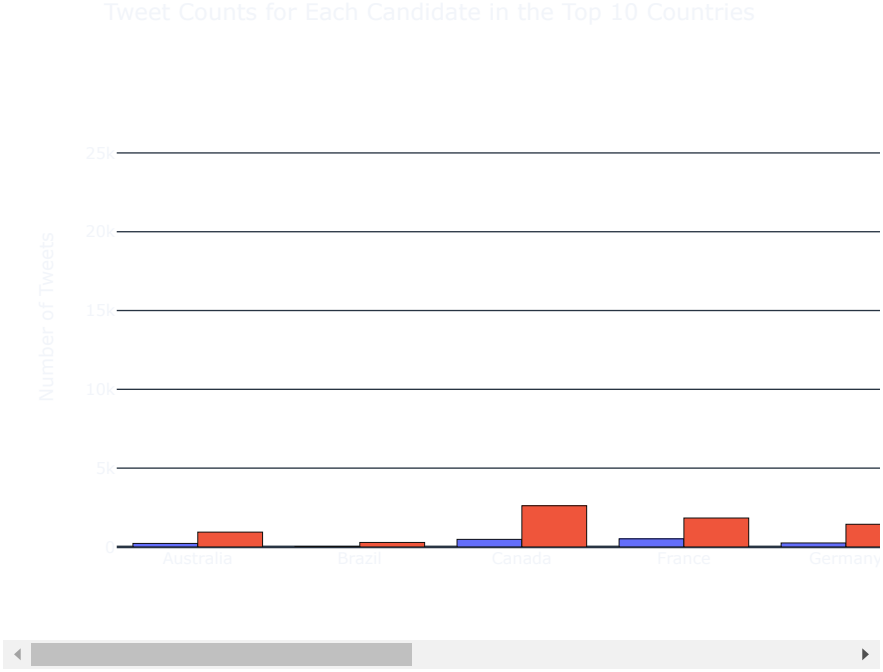


```
likes_comparison = data.groupby('candidate')['likes'].sum().reset_index()
fig = px.bar(likes_comparison, x='candidate', y='likes', color='candidate',
      color_discrete_map={'Trump': 'blue', 'Biden': 'green'},
      labels={'candidate': 'Candidate', 'likes': 'Total Likes'},
      title='Comparison of Likes')


fig.update_layout(plot_bgcolor='black',
        paper_bgcolor='black', font_color='white')


fig.show()
```



```
top10countries = data.groupby('country')['tweet'].count(
).sort_values(ascending=False).reset_index().head(10)



fig = px.bar(top10countries, x='country', y='tweet',
            template='plotly_dark',
            color_discrete_sequence=px.colors.qualitative.Dark24_r,
            title='Top10 Countrywise tweets Counts')


fig.show()
```

```
tweet_df = data.groupby(['country', 'candidate'])[
    'tweet'].count().reset_index()


tweeters = tweet_df[tweet_df['country'].isin(top10countries.country)]



fig = px.bar(tweeters, x='country', y='tweet', color='candidate',
            labels={'country': 'Country', 'tweet': 'Number of Tweets',
                    'candidate': 'Candidate'},
            title='Tweet Counts for Each Candidate in the Top 10 Countries',
            template='plotly_dark',
            barmode='group')


fig.show()
```




```
def clean(text):

  text = re.sub(r'https?://\S+|www\.\S+', '', str(text))
  text = text.lower()

  text = re.sub('[^a-z]', ' ', text)
  text = text.split()
  lm = WordNetLemmatizer()

  text = [lm.lemmatize(word) for word in text if word not in set(
    stopwords.words('english'))]

  text = ' '.join(word for word in text)

  return text



def getpolarity(text):
    return TextBlob(text).sentiment.polarity

def getsubjectivity(text):
    return TextBlob(text).sentiment.subjectivity

def getAnalysis(score):
    if score < 0:
        return 'negative'
    elif score == 0:
        return 'neutral'
    else:
        return 'positive'



trump_tweets = data[data['candidate'] == 'trump']
trump_tweets = trump_tweets.loc[trump_tweets.country == 'US']
trump_tweets = trump_tweets[['tweet']]
print(trump_tweets.head())
```

```
                                                tweet
     2   #Trump: As a student I used to hear for years,...
     4   You get a tie! And you get a tie! #Trump ‘s ra...
     11  In 2020, #NYPost is being #censorship #CENSORE...
     12  #Trump #PresidentTrump #Trump2020LandslideVict...
     22  #Trump: Nobody likes to tell you this, but som...
```

```
trump_tweets['cleantext'] = trump_tweets['tweet'].apply(clean)
print(trump_tweets.head())
```

```
                                                tweet  \
     2   #Trump: As a student I used to hear for years,...
```

```
4   You get a tie! And you get a tie! #Trump ‘s ra...
11  In 2020, #NYPost is being #censorship #CENSORE...
12  #Trump #PresidentTrump #Trump2020LandslideVict...
22  #Trump: Nobody likes to tell you this, but som...

                                          cleantext
2   trump student used hear year ten year heard ch...
4                    get tie get tie trump rally iowa
11  nypost censorship censored twitter manipulate ...
12  trump presidenttrump trump landslidevictory tr...
22  trump nobody like tell farmer better way worki...
```

```
trump_tweets['subjectivity'] = trump_tweets['cleantext'].apply(getsubjectivity)
```

```
trump_tweets['polarity'] = trump_tweets['cleantext'].apply(getpolarity)
```

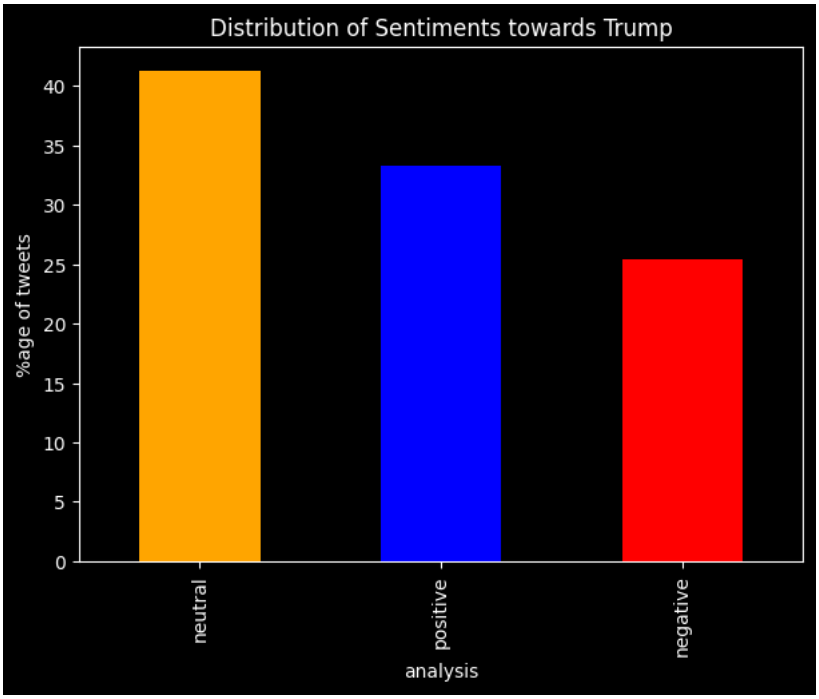
```
trump_tweets['analysis'] = trump_tweets['polarity'].apply(getAnalysis)
trump_tweets.head()
```

| | tweet | cleantext | subjectivity | polarity | analysis |
|---|---|---|---|---|---|
| **2** | #Trump: As a student I used to hear for years,... | trump student used hear year ten year heard ch... | 0.333333 | 0.333333 | positive |
| **4** | You get a tie! And you get a tie! #Trump ‘s ra... | get tie get tie trump rally iowa | 0.000000 | 0.000000 | neutral |
| | In 2020, #NYPost is being | nypost censorship | | | |

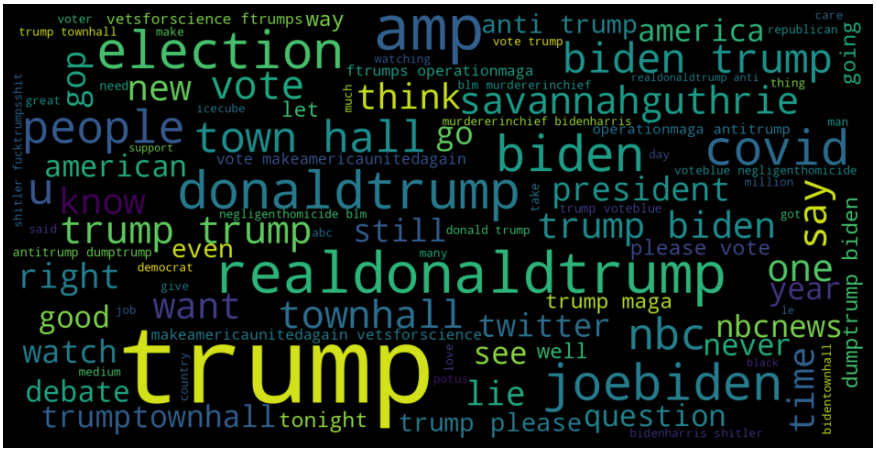
Next steps: Generate code with `trump_tweets` View recommended plots

```
plt.style.use('dark_background')
```

```
colors = ['orange', 'blue', 'red']

plt.figure(figsize=(7, 5))
(trump_tweets.analysis.value_counts(normalize=True) * 100).plot.bar(color=colors)
plt.ylabel("%age of tweets")
plt.title("Distribution of Sentiments towards Trump")
plt.show()
```



```
def word_cloud(wd_list):
    stopwords = set(STOPWORDS)
    all_words = ' '.join(wd_list)

    wordcloud = WordCloud(background_color='black',
                          stopwords=stopwords,
                          width=1600, height=800, max_words=100, max_font_size=200,
                          colormap="viridis").generate(all_words)
    plt.figure(figsize=(12, 10))
    plt.axis('off')
    plt.imshow(wordcloud)

word_cloud(trump_tweets['cleantext'][:5000])
```




```
biden_tweets = data[data['candidate'] == 'biden']
biden_tweets = biden_tweets.loc[biden_tweets.country == 'US']
biden_tweets = biden_tweets[['tweet']]
biden_tweets
```

| | tweet |
|---|---|
| 6 | In 2020, #NYPost is being #censorship #CENSORE... |
| 17 | Comments on this? "Do Democrats Understand how... |
| 25 | @RealJamesWoods #BidenCrimeFamily #JoeBiden #H... |
| 29 | Come on @ABC PLEASE DO THE RIGHT THING. Move t... |
| 34 | #realDonaldTrump addresses #JoeBiden and #Hunt... |
| ... | ... |
| 51270 | #JoeBidenLied when he claimed the Boilermakers... |
| 51273 | @Zigmanfreud Two huge issues I have - even if ... |
| 51276 | DID you know:\nS. 1338 (101st): #Biden-Roth-Co... |
| 51282 | @Michael68529417 @DineshDSouza @realDonaldTrum... |
| 51293 | Democrats Have Run Back To Back Criminals For ... |

7627 rows × 1 columns

Next steps: Generate code with biden_tweets View recommended plots

```
biden_tweets['cleantext']=biden_tweets['tweet'].apply(clean)
biden_tweets.head()
```

| | tweet | cleantext |
|---|---|---|
| 6 | In 2020, #NYPost is being #censorship #CENSORE... | nypost censorship censored twitter manipulate ... |
| 17 | Comments on this? "Do Democrats Understand how... | comment democrat understand ruthless china chi... |
| 25 | @RealJamesWoods #BidenCrimeFamily #JoeBiden #H... | realjameswoods bidencrimefamily joebiden hunte... |
| 29 | Come on @ABC PLEASE DO THE RIGHT | come abc please right thing move biden |

Next steps: Generate code with biden_tweets View recommended plots

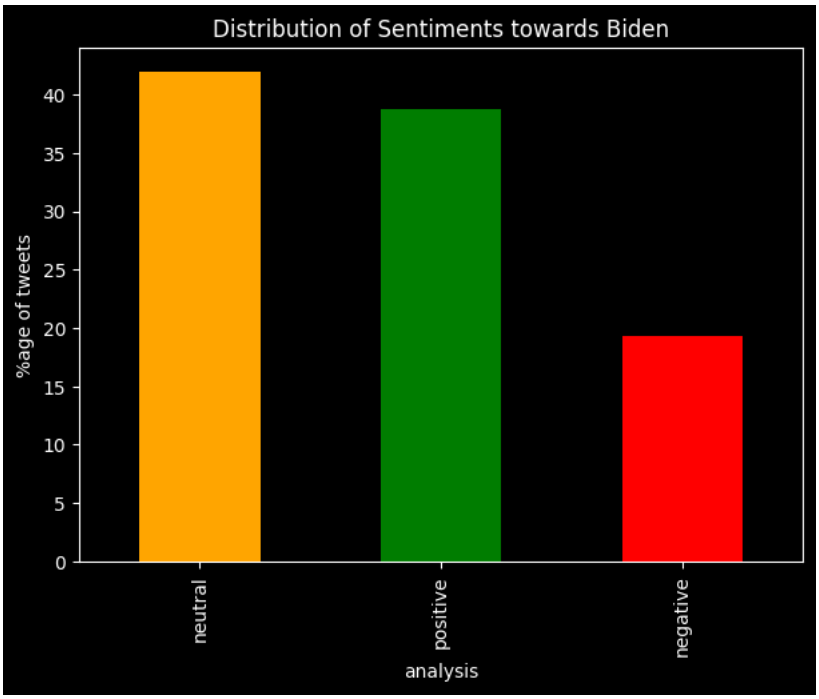
```
biden_tweets['subjectivity'] = biden_tweets['cleantext'].apply(getsubjectivity)
biden_tweets['polarity'] = biden_tweets['cleantext'].apply(getpolarity)
biden_tweets['analysis'] = biden_tweets['polarity'].apply(getAnalysis)
biden_tweets.head()
```

| | tweet | cleantext | subjectivity | polarity | analysis |
|---|---|---|---|---|---|
| 6 | In 2020, #NYPost is being #censorship #CENSORE... | nypost censorship censored twitter manipulate ... | 0.678571 | -0.148810 | negative |
| 17 | Comments on this? "Do Democrats Understand how... | comment democrat understand ruthless china chi... | 1.000000 | -1.000000 | negative |
| 25 | @RealJamesWoods #BidenCrimeFamily | realjameswoods bidencrimefamily | 0.000000 | 0.000000 | neutral |

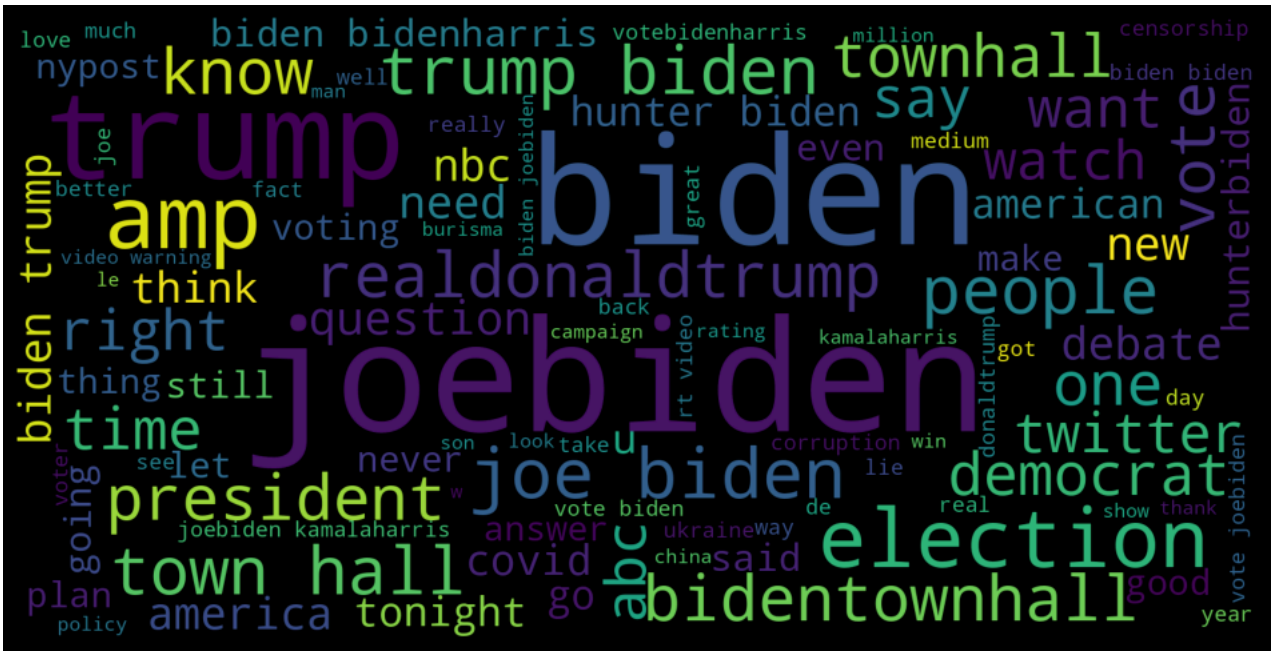
Next steps: Generate code with biden_tweets View recommended plots

```
plt.style.use('dark_background')
```

```
colors = ['orange', 'green', 'red']

plt.figure(figsize=(7, 5))
(biden_tweets.analysis.value_counts(normalize=True) * 100).plot.bar(color=colors)
plt.ylabel("%age of tweets")
plt.title("Distribution of Sentiments towards Biden")
plt.show()
```




```
word_cloud(biden_tweets['cleantext'][:5000])
```




```
trump_tweets.analysis.value_counts(normalize=True)*100
```

```
analysis
neutral     41.270413
positive    33.260031
negative    25.469555
Name: proportion, dtype: float64
```

```
biden_tweets.analysis.value_counts(normalize=True)*100
```

```
analysis
neutral     41.929986
positive    38.783270
negative    19.286744
Name: proportion, dtype: float64
```