Ruling Political Party Sentiment Analysis and Hate speech Detection for Twitter

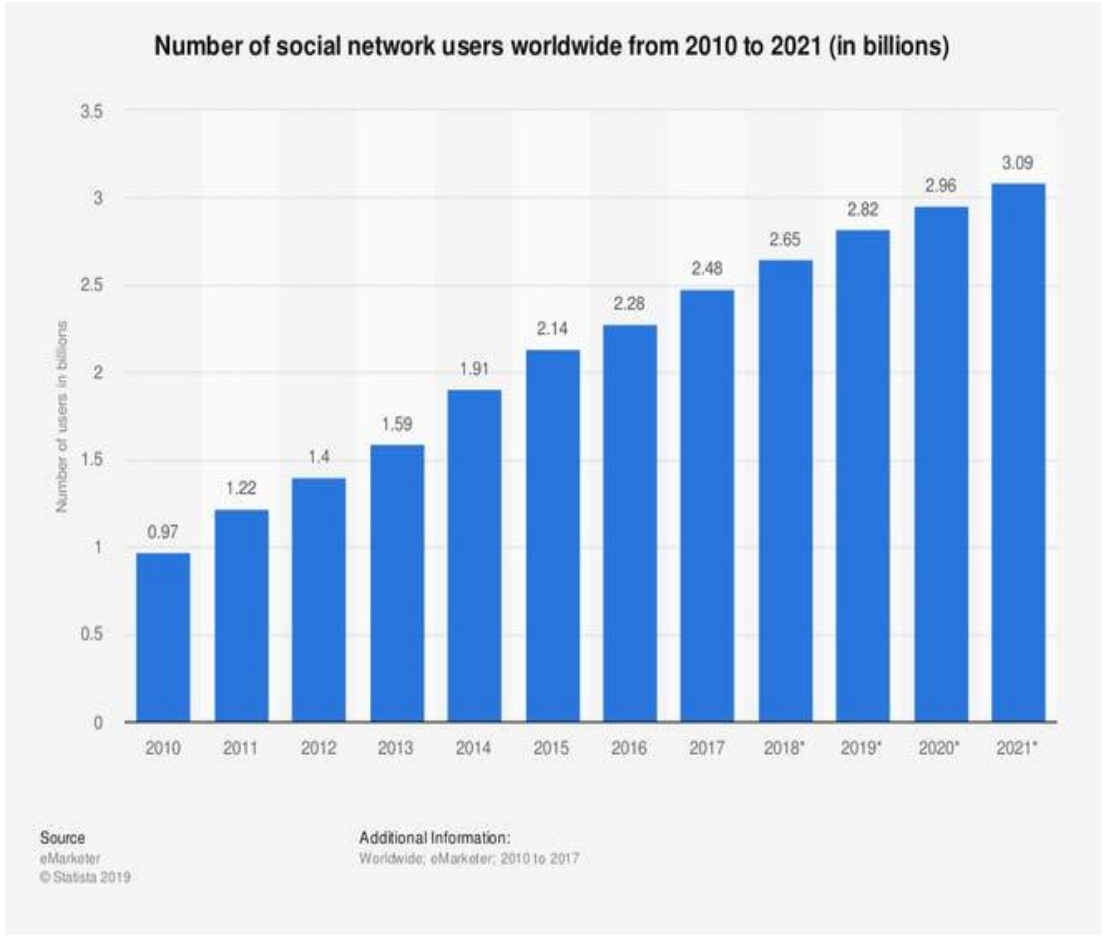
Submitted to:

Dr. Nitin Arvind Shelke

Submitted By:

Jiya Midha	**401703010**
Manan Vyas	**401703014**
Shourya Marwaha	**401708013**
Sumit Arora	**401708015**

INFOGRAPHICS



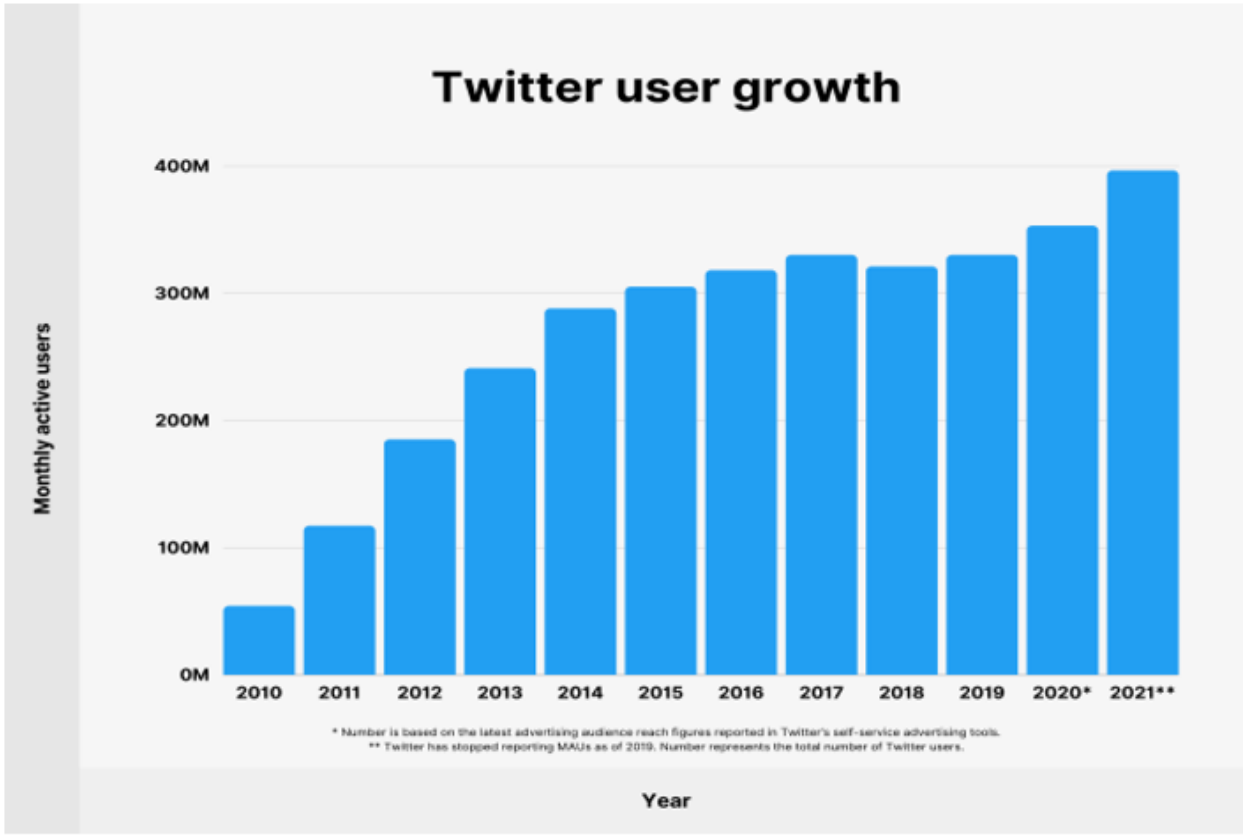
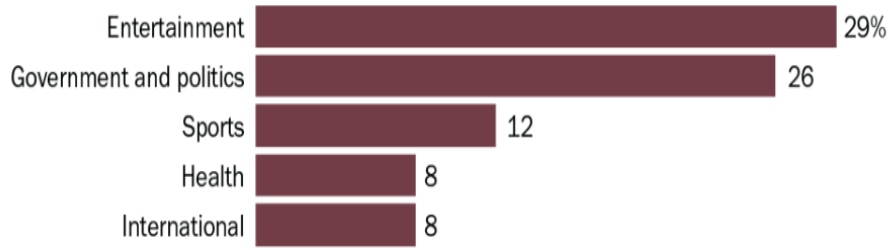


Figure 2 Number of active users on twitter

INFOGRAPHICS

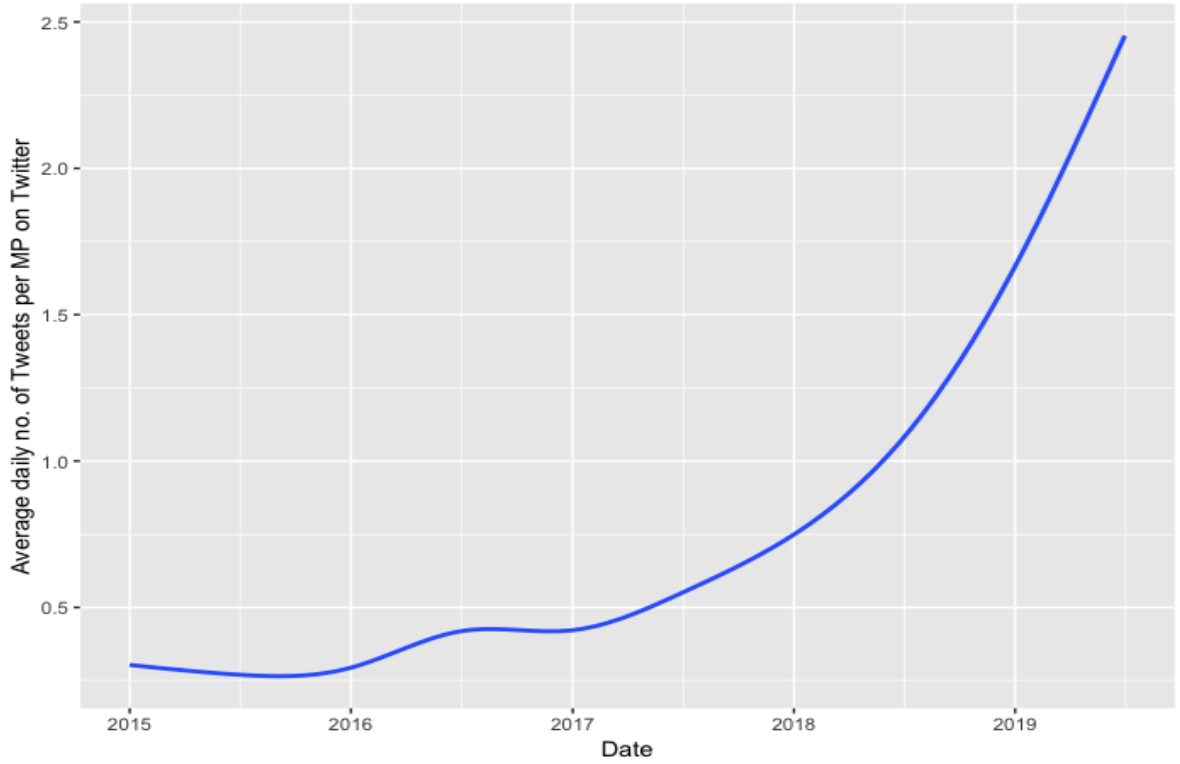
Entertainment, politics and sports topped the list of subject areas in Americans' news tweets in 2021

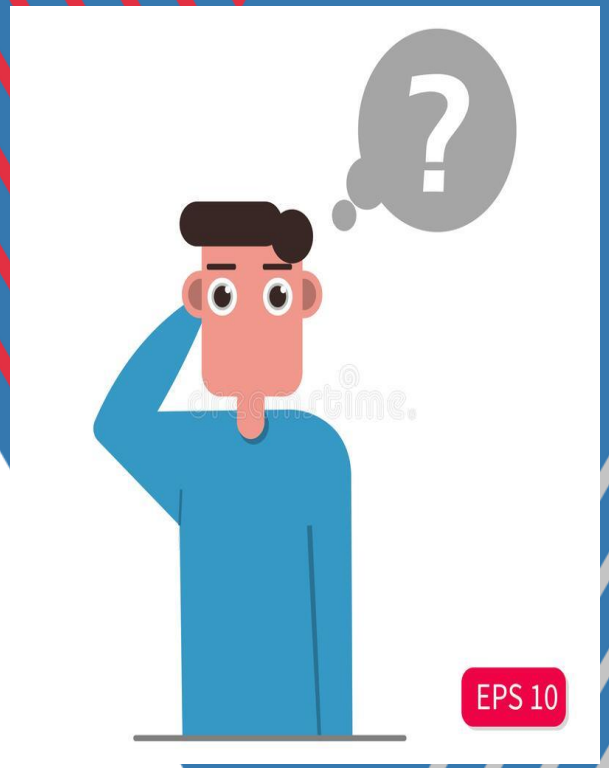
Among U.S. adult Twitter users who tweeted about news, the average share of a user's news-related tweets that were about each news topic



Note: Shows average among 512 Twitter users in the sample who tweeted about news at least once during the time period studied.
Source: Content analysis of tweets of 512 American Trends Panel members, June 12-Aug. 31, 2021.

Pew Research Center

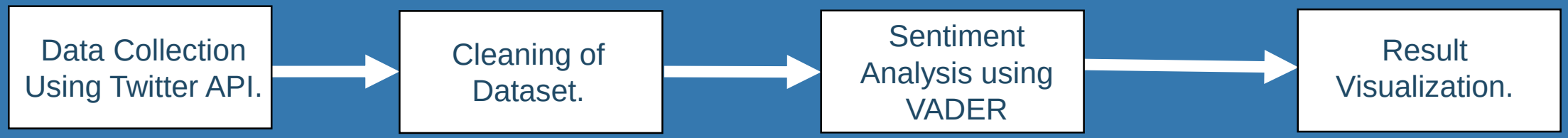




OVERVIEW

- On social media, people express themselves every day on issues that affect their lives.
- We collected Tweets concerned with the Ruling Party of United States of America in English Language.
- The dataset created for sentiment analysis was then divided into three categories: Positive, Negative and Neutral.
- The results were reported for the sentiment analysis in order to comment on the popularity of the ruling party.

Methodology



Data Collection using Twitter API

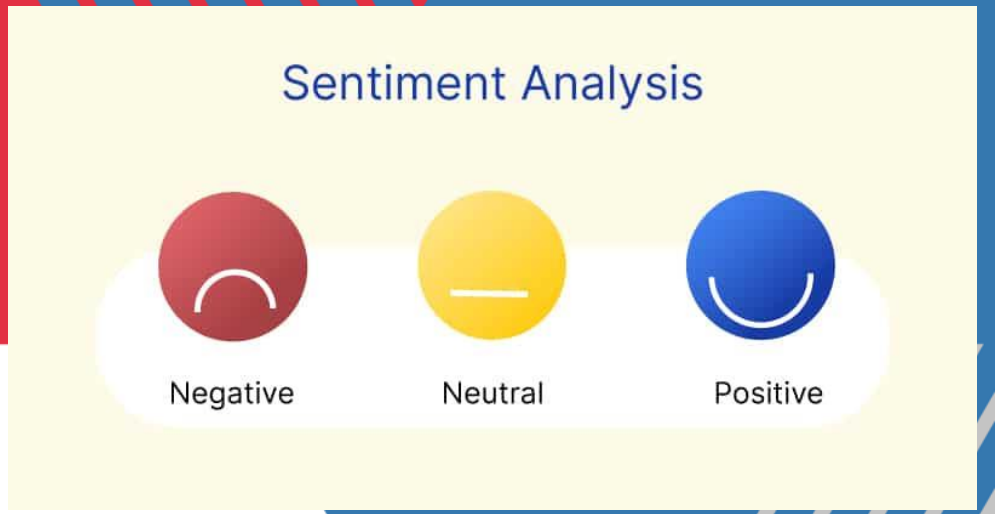
- **Tweepy**: Python library for accessing the Twitter API.
- Twitter Developer Account for Authorization and Accessing Data.
- Search Tweets based upon specific keywords or hashtags or '@' Handles.
- search_words = " biden, @joebiden, bidenharris, @kamalaharris, election, vote, #HunterBiden, #JoeBidenMustStepDown, trump, DemocraticParty ".

Cleaning of Dataset

- Remove unnecessary columns to clean the data.
- **Preprocessing**
  - Convert to lower case.
  - Remove punctuations, numbers, special characters, white spaces etc.
  - Tokenize
  - Remove Stop words
  - Normalize the text (Stemming / Lemmatization)

```
In [13]: #5 biden tweets after filtering
         biden_df.tweet.head()

Out[13]: 6     [nypost, censorship, censored, twitter, manipu...
         17    [comments, democrats, understand, ruthless, ch...
         22    [twitter, everything, help, democrats, win, el...
         25    [realjameswoods, bidencrimefamily, joebiden, h...
         29    [come, abc, please, right, thing, move, biden,...
         Name: tweet, dtype: object
```



Sentiment Analysis using VADER

VADER (Valence Aware Dictionary for Sentiment Reasoning) is a model used for text sentiment analysis that is sensitive to both polarity (positive/negative) and intensity (strength) of emotion. It can be applied directly to unlabeled text data.

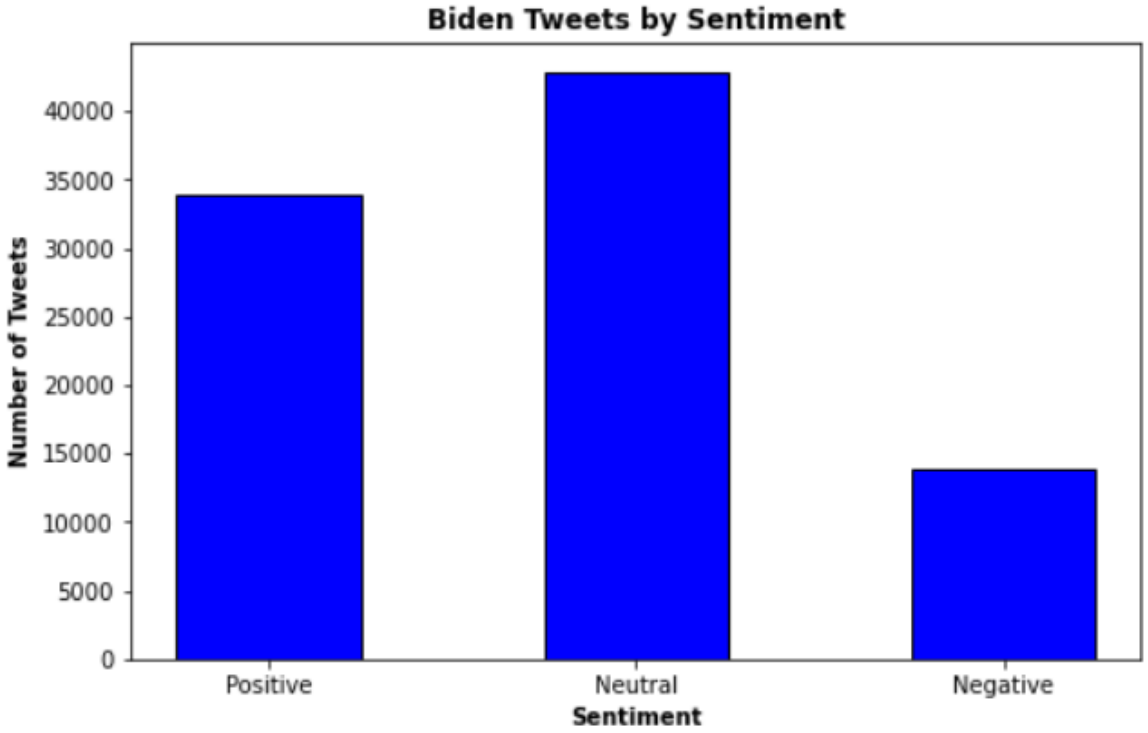
VADER sentimental analysis relies on a dictionary that maps lexical features to emotion intensities known as sentiment scores. The sentiment score of a text can be obtained by summing up the intensity of each word in the text.

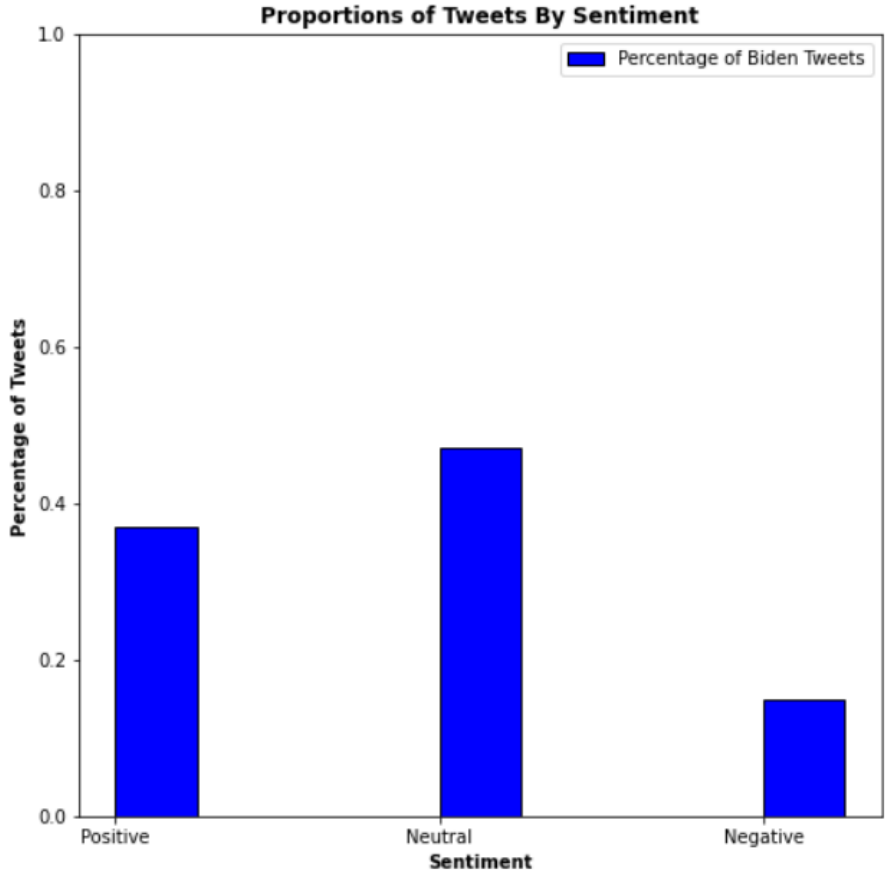
Polarity classification

We won't try to determine if a sentence is objective or subjective, fact or opinion. Rather, we care only if the text expresses a ***positive, negative*** **or** ***neutral*** opinion.

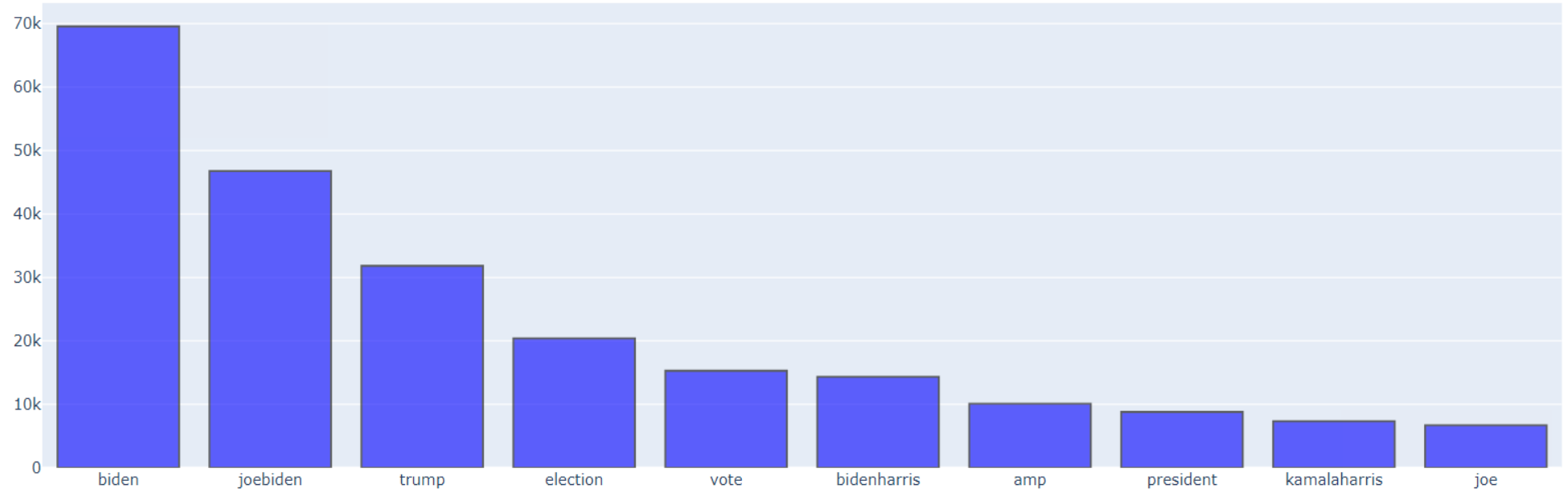
Result Visualization
&
Conclusion

Sentiment Analysis of Tweets for Joe Biden

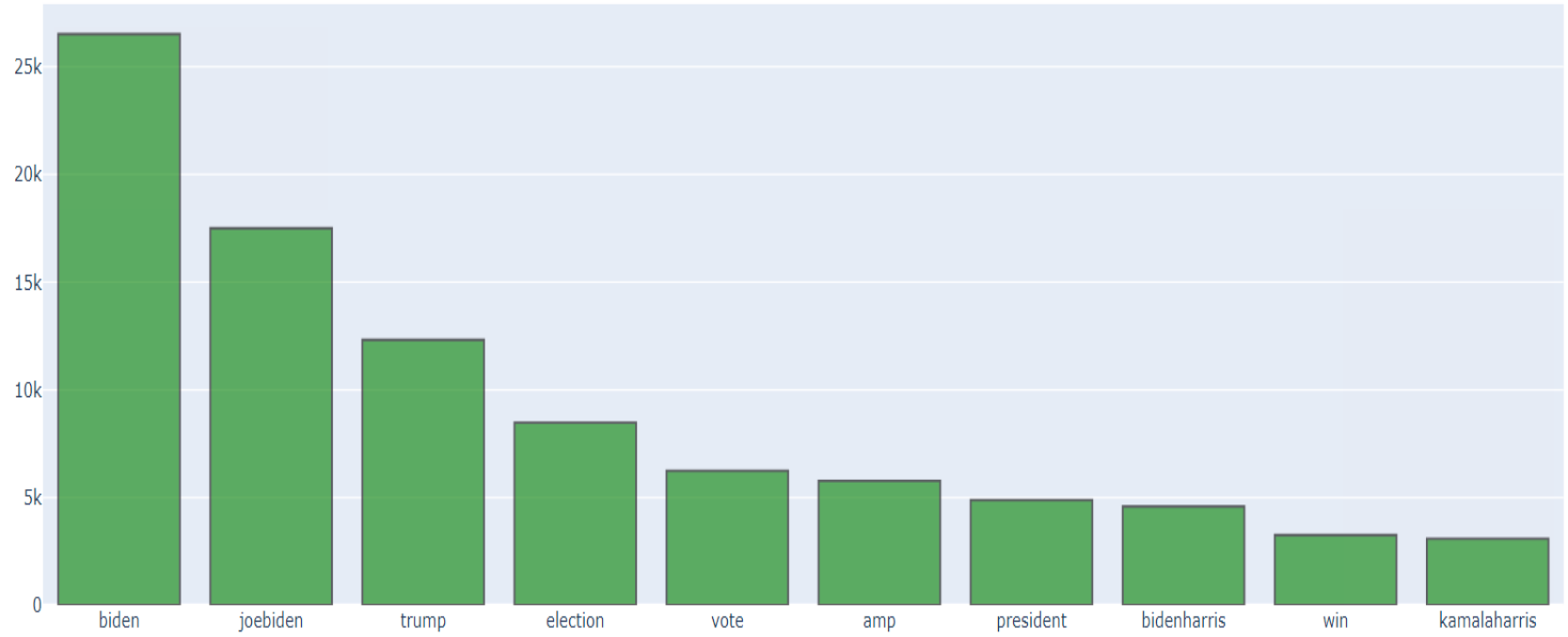




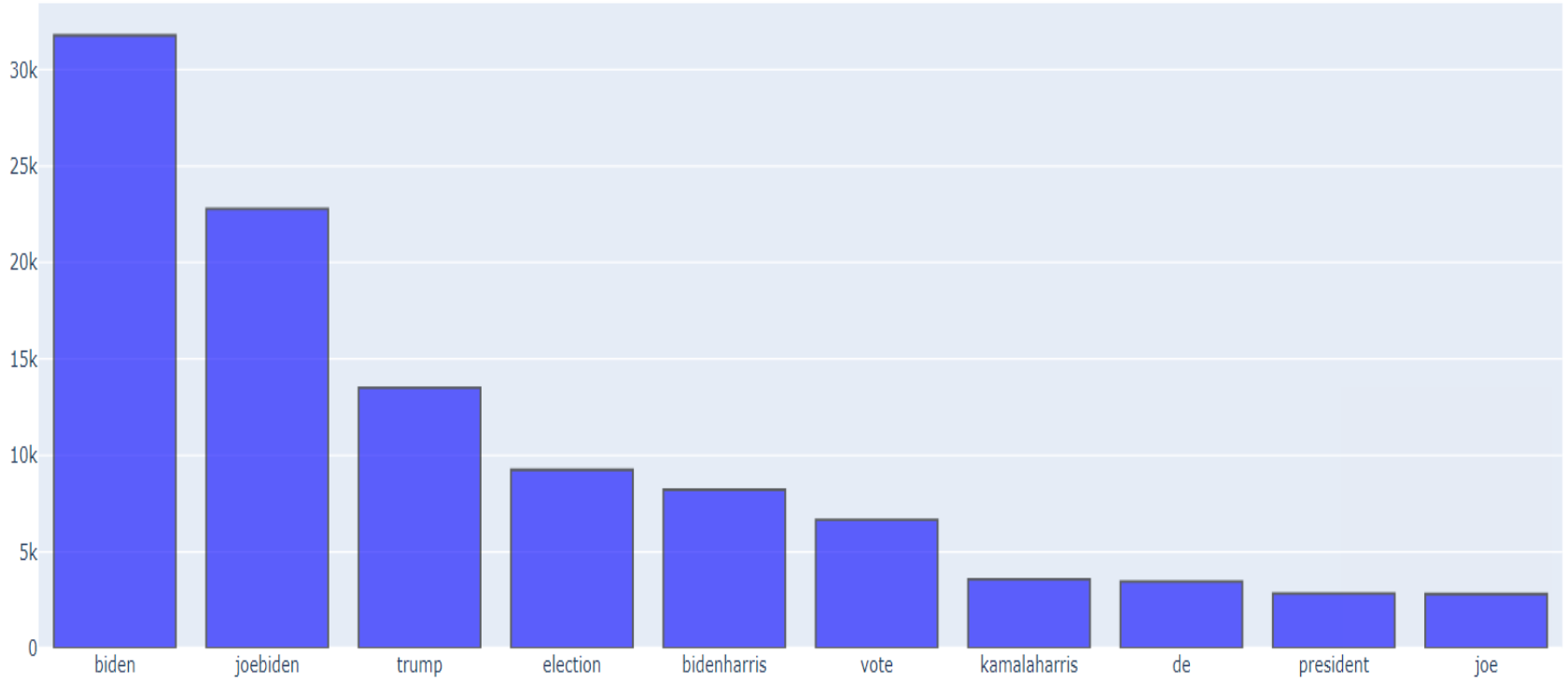
Frequent Words in Biden tweets



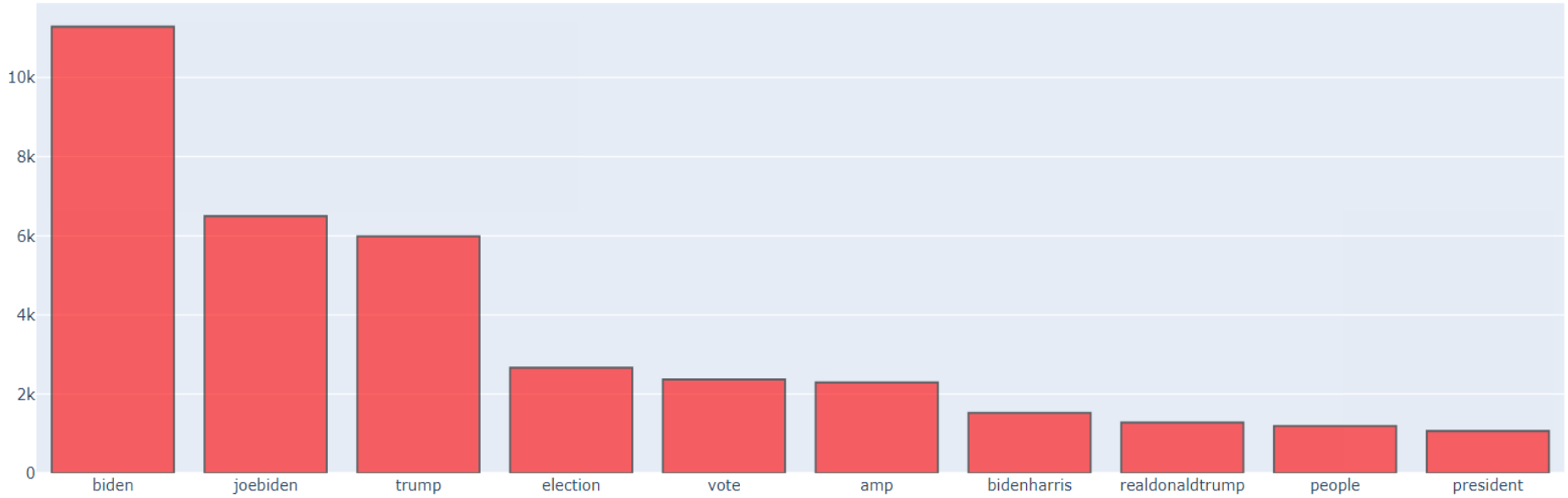
Frequent Words in Positive Biden Tweets

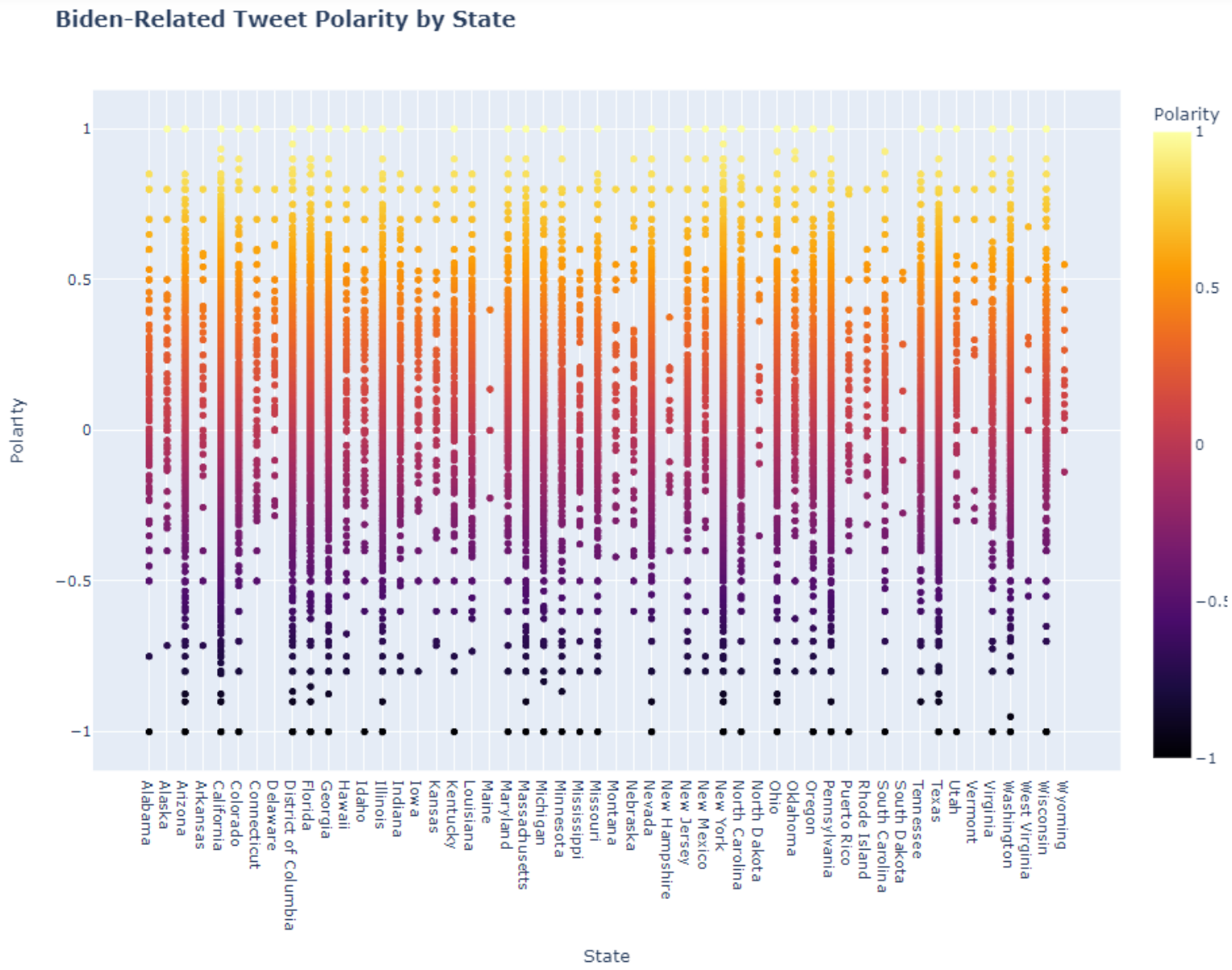


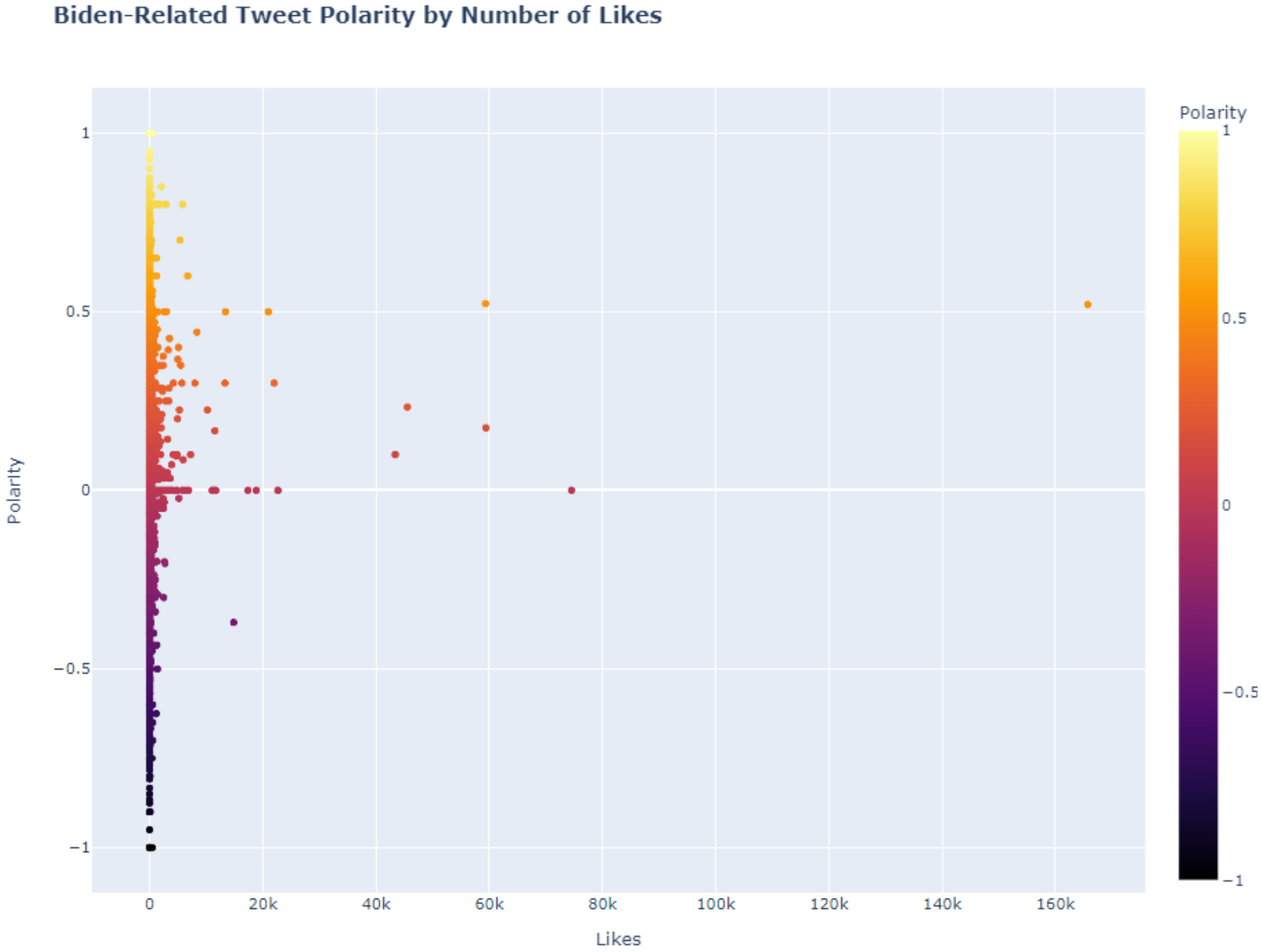
Frequent Words in Neutral Biden Tweets

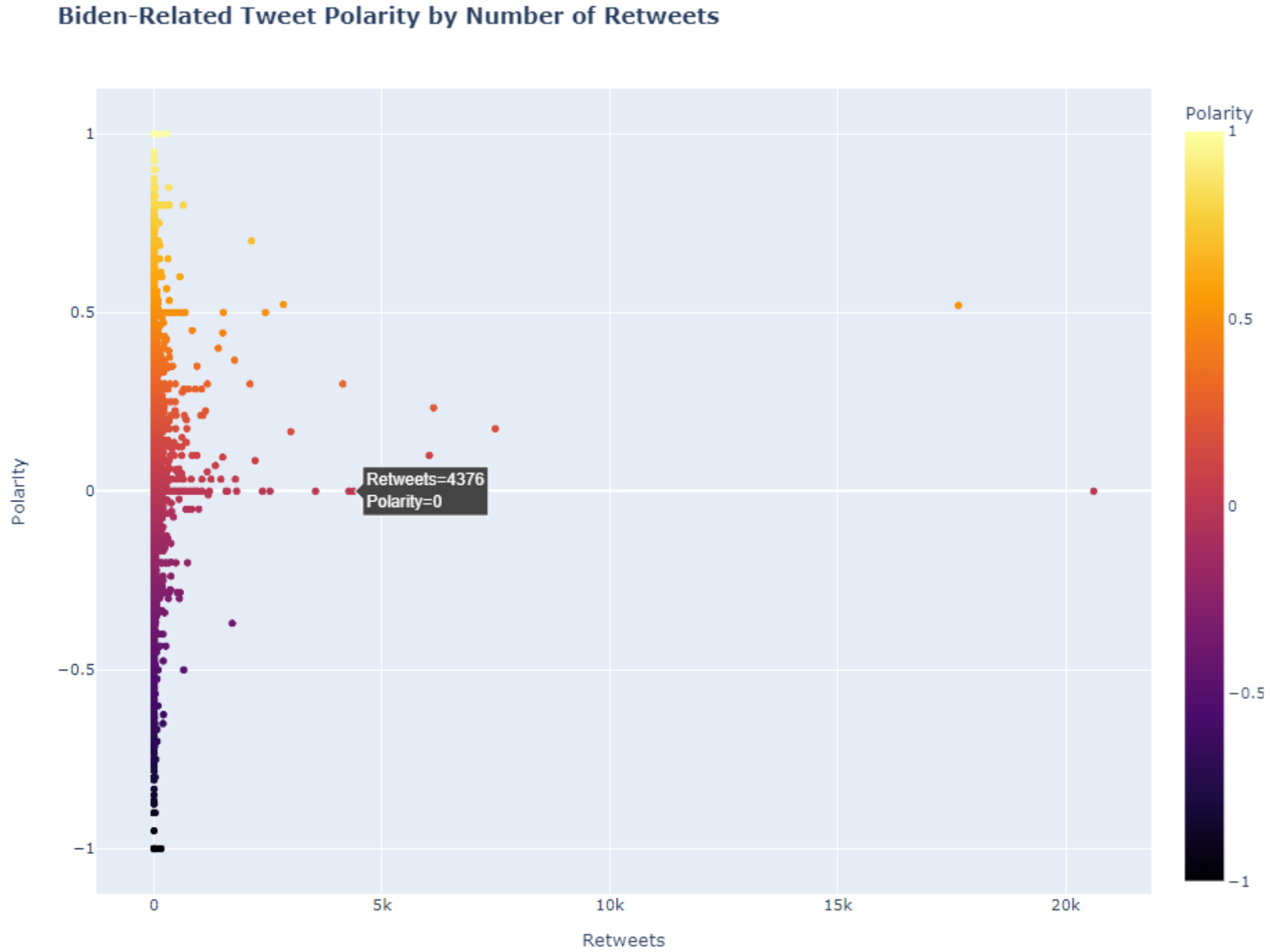


Frequent Words in Negative Biden Tweets




Biden-Related Tweet Polarity by State
Polarity
1
0.5
0
−0.5
−1
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Puerto Rico
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
State


Biden-Related Tweet Polarity by Number of Likes
Polarity
1
0.5
0
−0.5
−1
0
20k
40k
60k
80k
100k
120k
140k
160k
Likes
Polarity
1
0.5
0
−0.5
−1


Biden-Related Tweet Polarity by Number of Retweets
Polarity
1
0.5
0
−0.5
−1
Retweets=4376
Polarity=0
0
5k
10k
15k
20k
Retweets
Polarity
1
0.5
0
−0.5
−1

ACCURACY OF THE MODEL

VADER is a pre-trained model and has an accuracy of 96% which outperforms individual human raters (84% accuracy) at correctly labelling the sentiment of tweets into positive, neutral, or negative classes.

Challenges

Sentiment analysis is a hard task in NLP because even humans struggle to analyze sentiments accurately.

- Subjectivity & Tone
- Context & Polarity
- Irony & Sarcasm
- Comparisons
- Emojis
- Defining Neutral
- Human Annotator Accuracy

THANK YOU!