



BrainStation

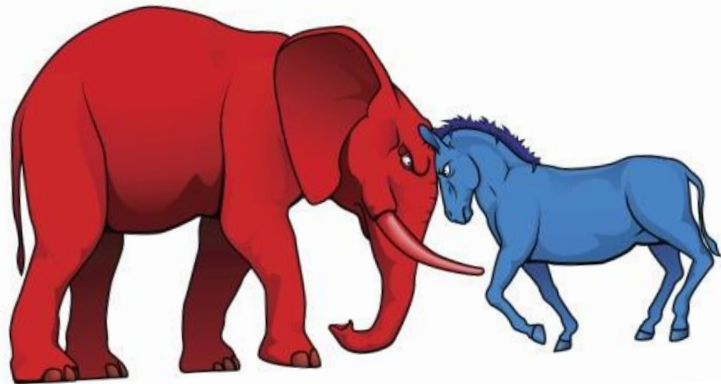
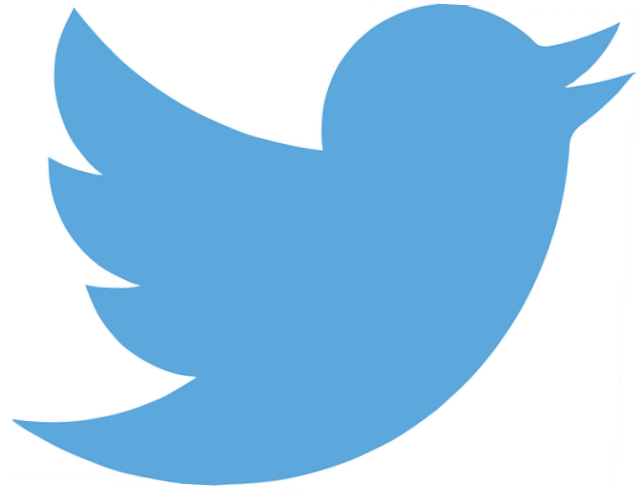
# ***Political Tweets***

## ***Classification***

Can we predict if a political tweet has a Democrat or Republican inclination?

Erick Rudelman | BrainStation

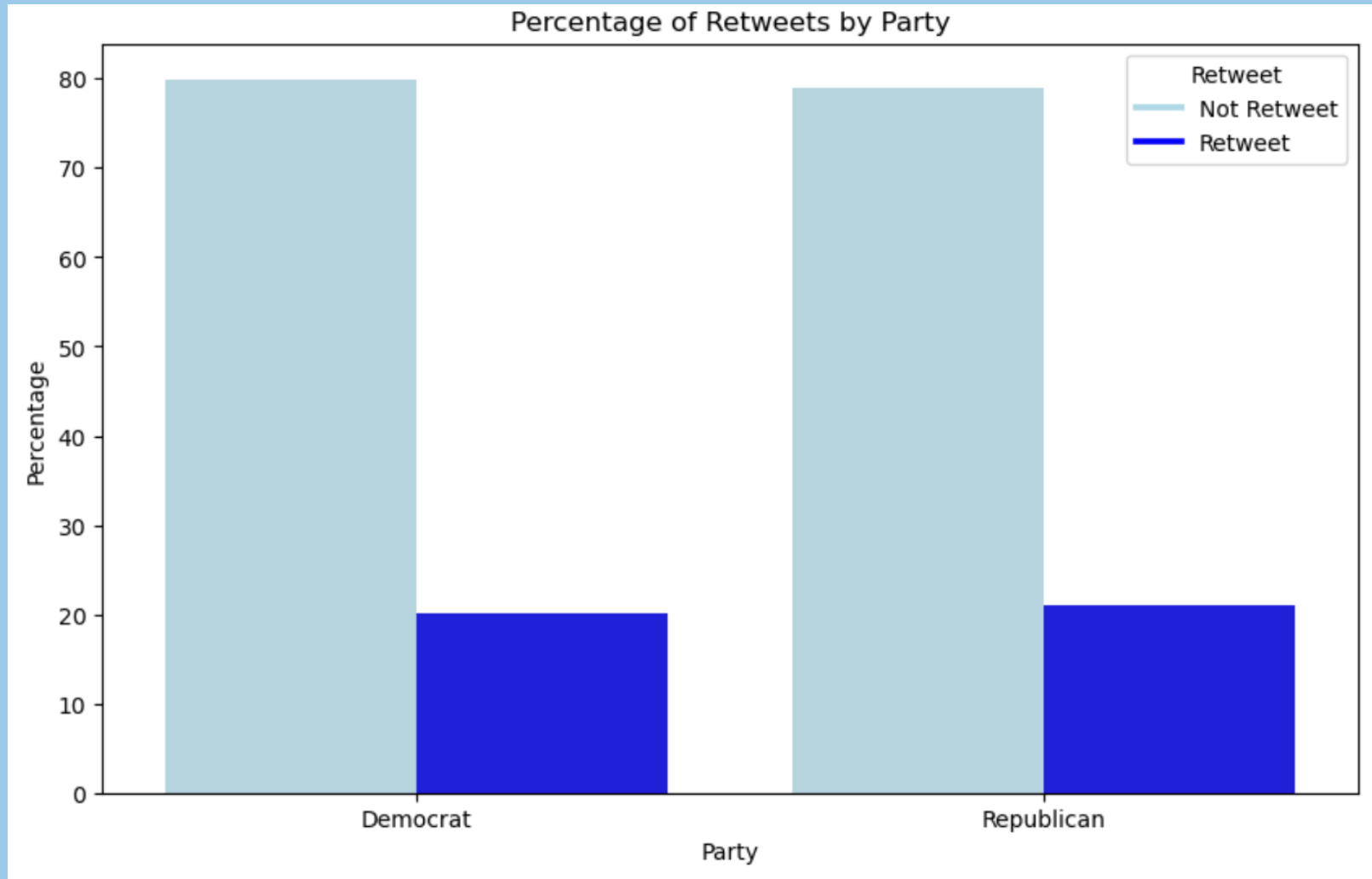
Sprint 2 – Jan 19, 2024





BrainStation

# Retweet Distribution by Party



# Hashtag Distribution

```
# Create two empty lists to store hashtags
republican_hashtags = []
democrat_hashtags = []

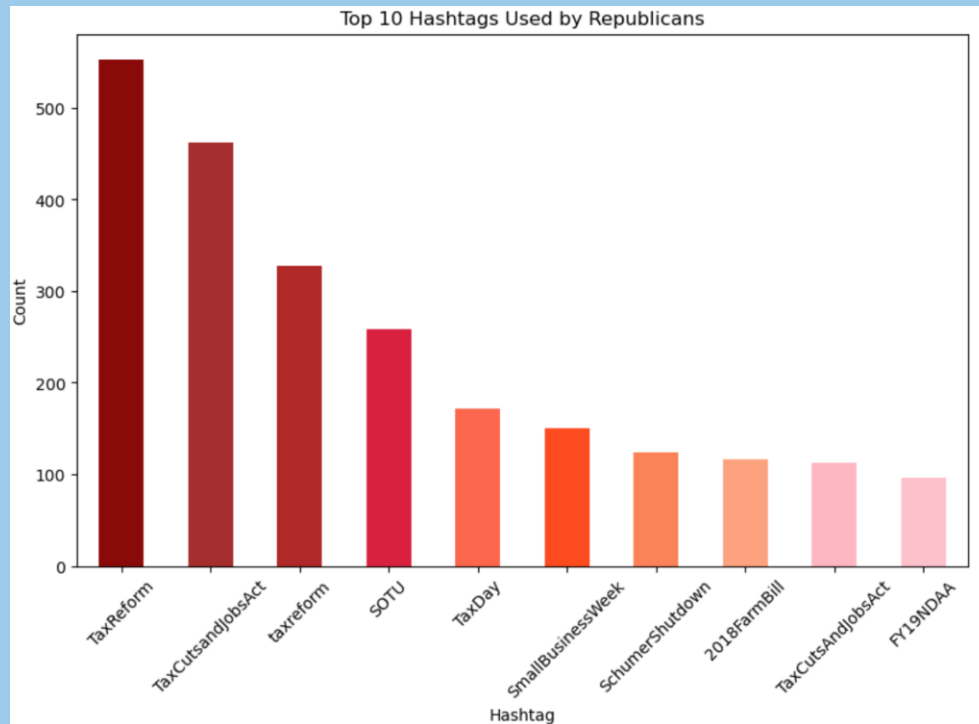
# Loop through each row to add hashtags to list
for index, row in df_cleaned.iterrows():
    if row['Party_binary'] == 1 and row['Hashtags'] is not None:
        democrat_hashtags.extend(row['Hashtags'])
    elif row['Hashtags'] is not None:
        republican_hashtags.extend(row['Hashtags'])
```



BrainStation

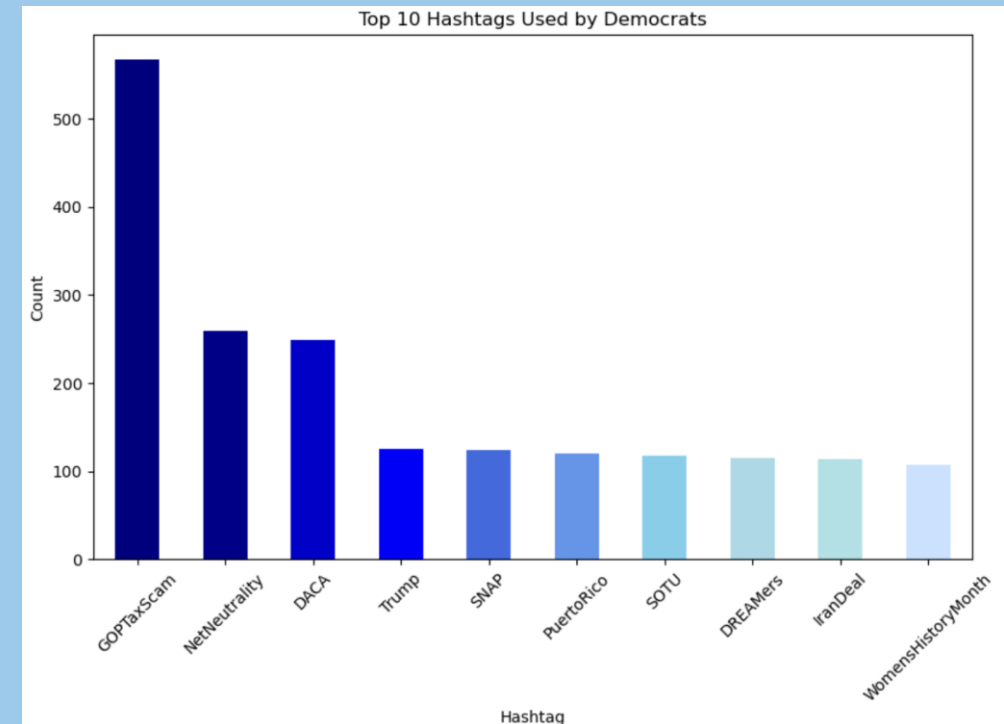
## TOP 10

### REPUBLICAN HASHTAGS



## TOP 10

### DEMOCRATIC HASHTAGS



# Hashtag Word Cloud



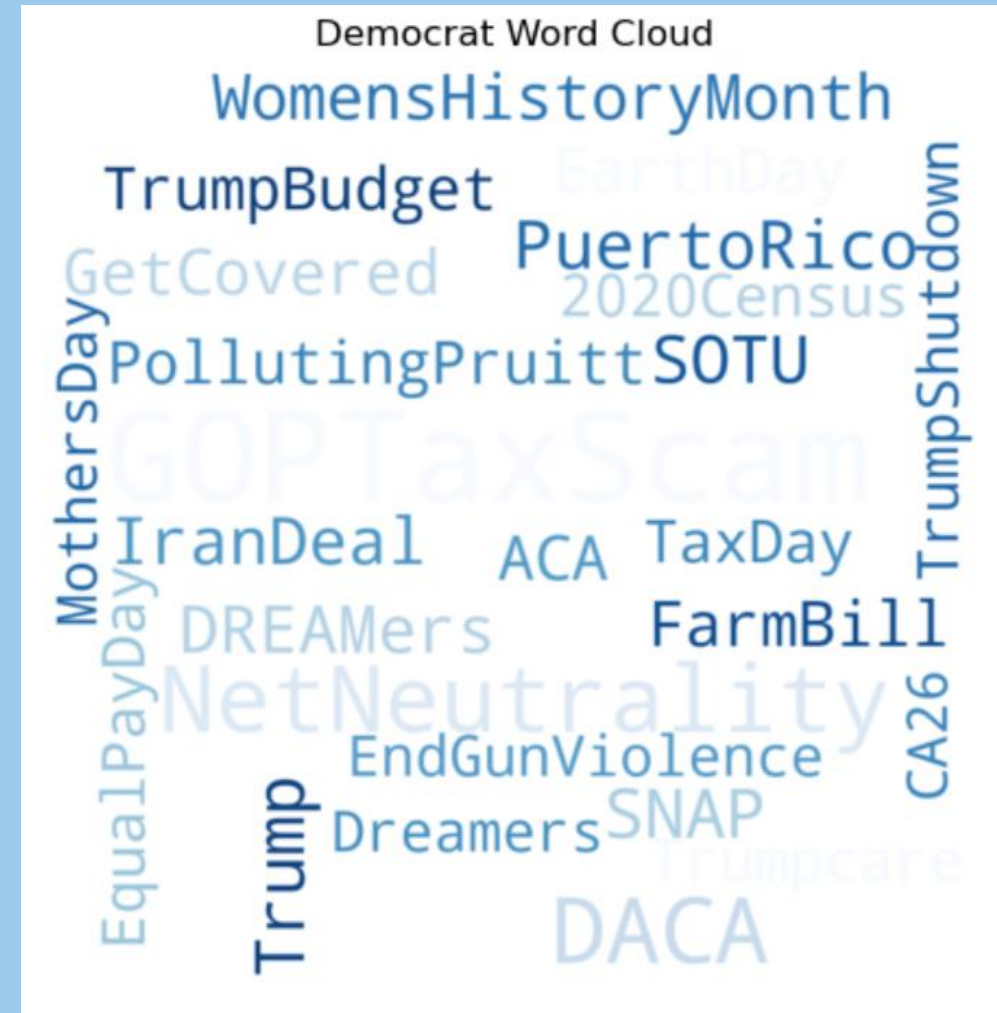
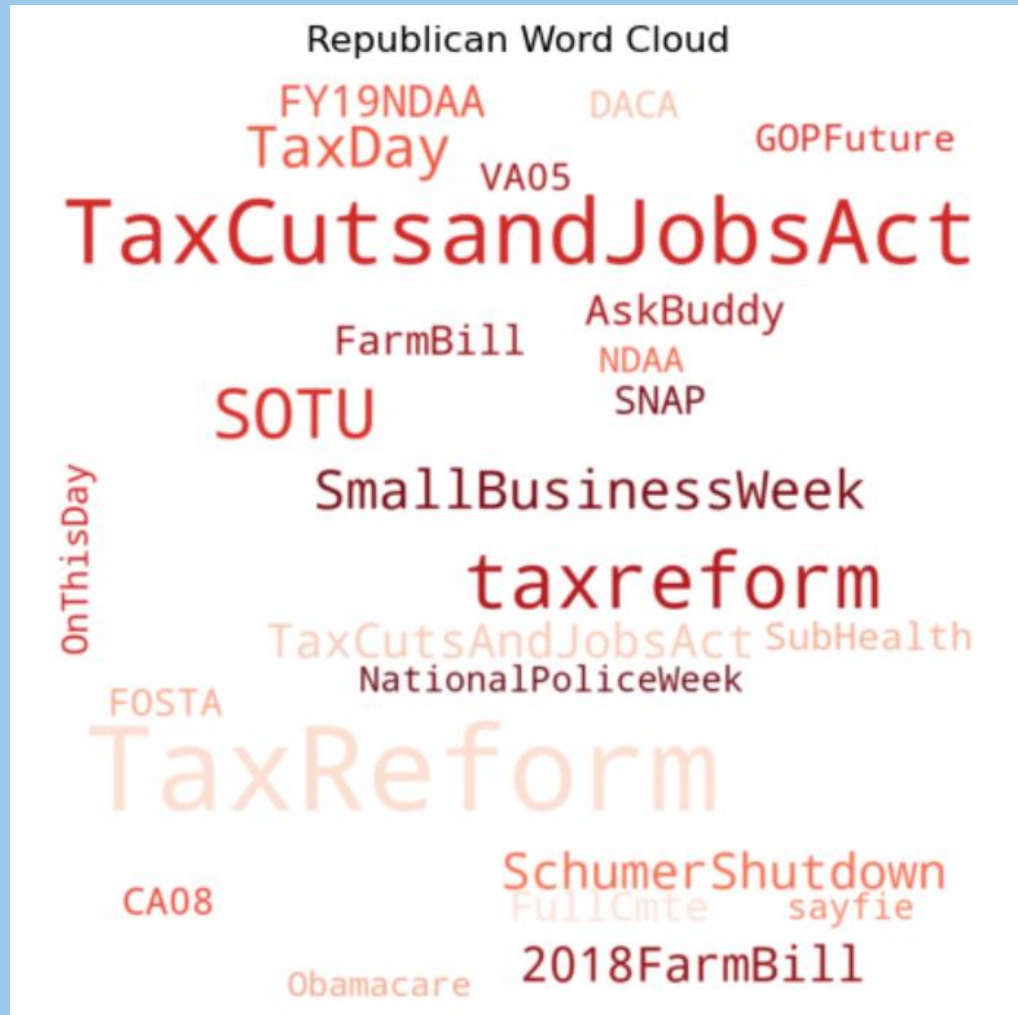
BrainStation

## REPUBLICAN

## DEMOCRAT

Republican Word Cloud

Democrat Word Cloud



# Preprocessing



BrainStation

## Stopwords

- stopwords = stopwords.words('english')
- stopwords.append("rt") # retweets
- stopwords.append("u") # you
- stopwords.append("r") # are
- stopwords.append("amp") # ampersand
- stopwords.append("w") # with
- stopwords.append("th") #that tip
- stopwords.append("s")

### Original Tweet

Congrats to this month's #NevadaWomanInSTEM, Cristina Rocha Rodriguez!  
She is currently a student at @CSNCoyote & a... <https://t.co/r6GCiar0iM>

### Cleaned Tweet

congrats month cristina rocha rodriguez currently student

## Tokenizer

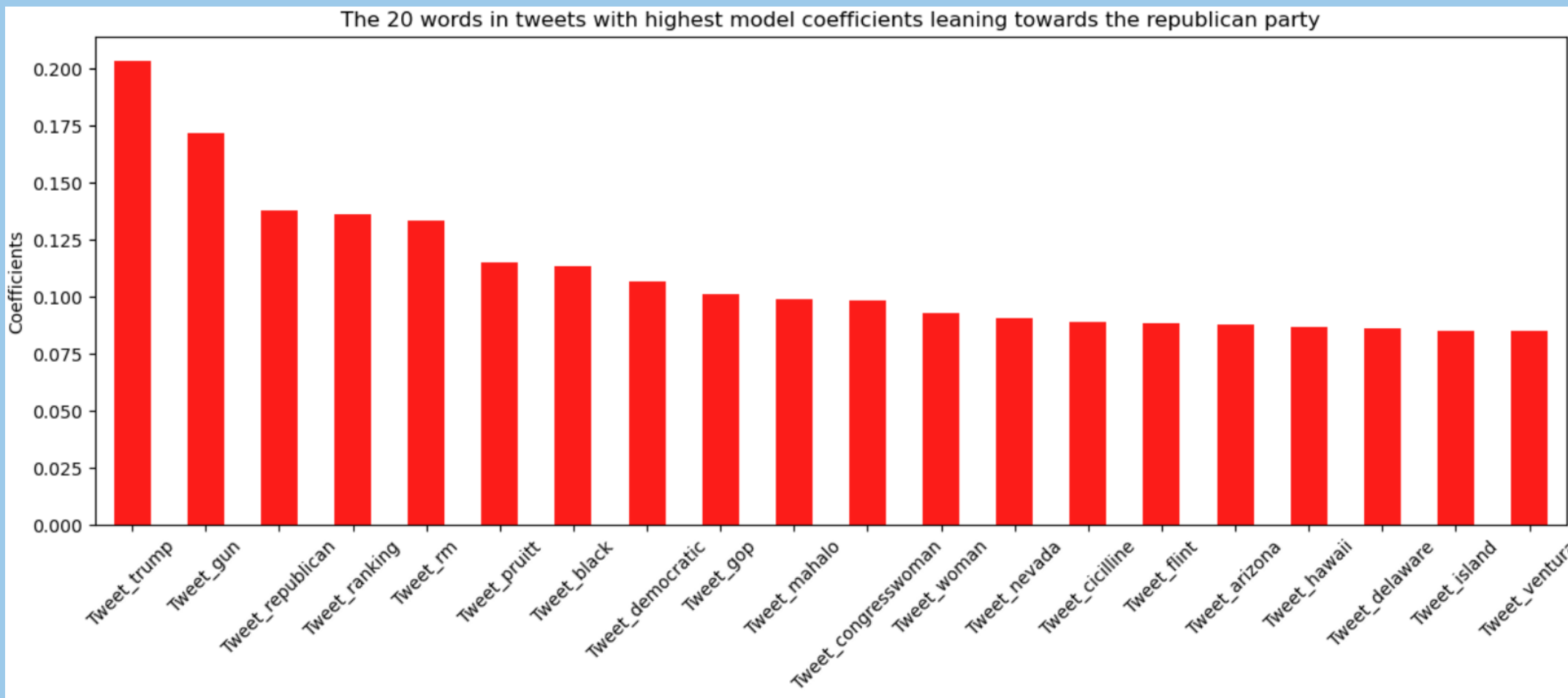
- tweet = re.sub("@[A-Za-z0-9]+", "", tweet)
- tweet = re.sub(r'#\w+', '', tweet)
- tweet = re.sub("[!@\$\$%^&\*()\_+\\|/?,.:;'`']+", "", tweet)
- tweet = re.sub(r'http\S+', '', tweet)
- tweet = tweet.lower()
- tweet = " ".join(tweet.split())
- tweet = ' '.join([word for word in tweet.split() if word.lower() not in stopwords])
- tokens = nltk.word\_tokenize(tweet)
- lemmatized\_tokens = [lemmatizer.lemmatize(token) for token in tokens]
- tweet = " ".join(lemmatized\_tokens)
- tweet = tweet.rstrip(' ', 1)[0]



# **Top Model Coefficient: Republican Words in Tweets**



BrainStation

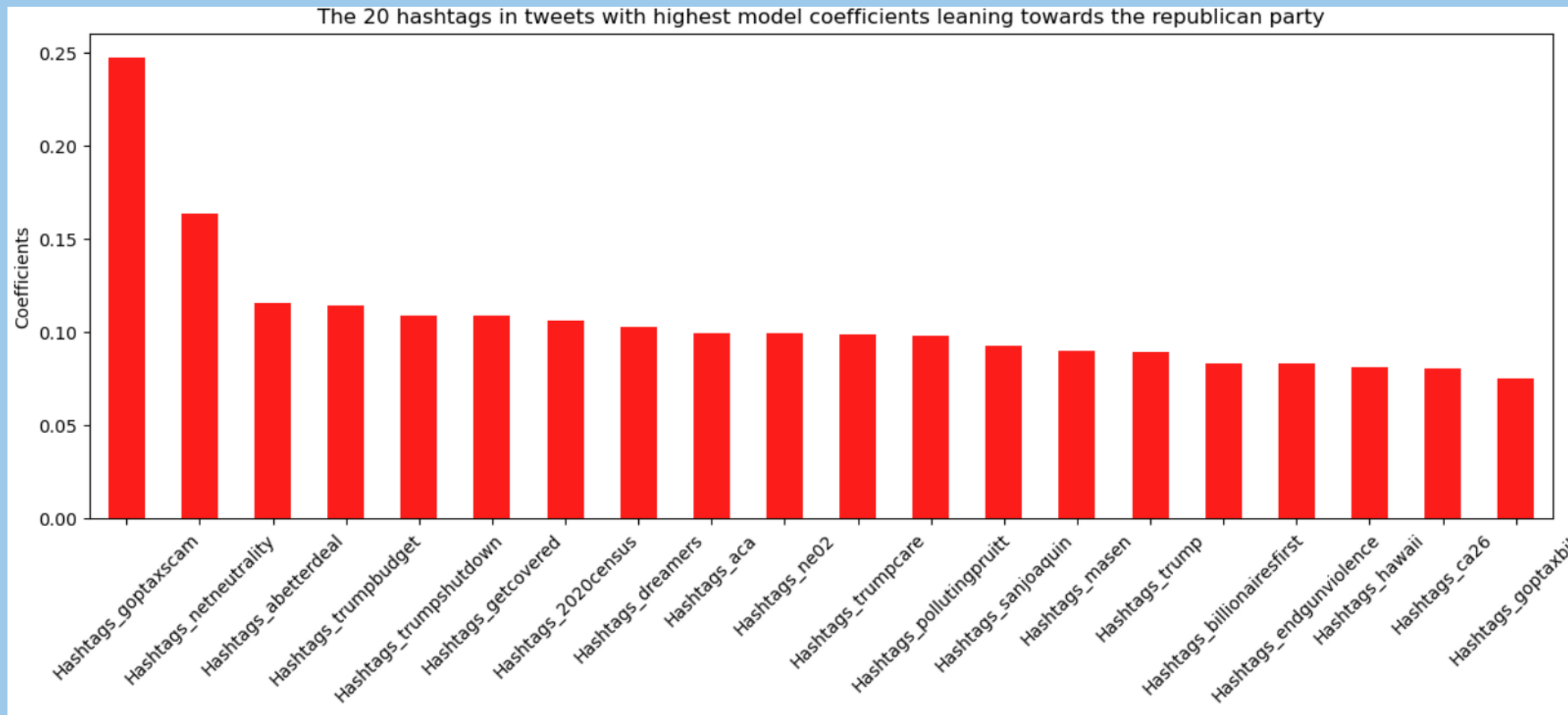


# Top Model Coefficient:

## Republican Hashtags in Tweets



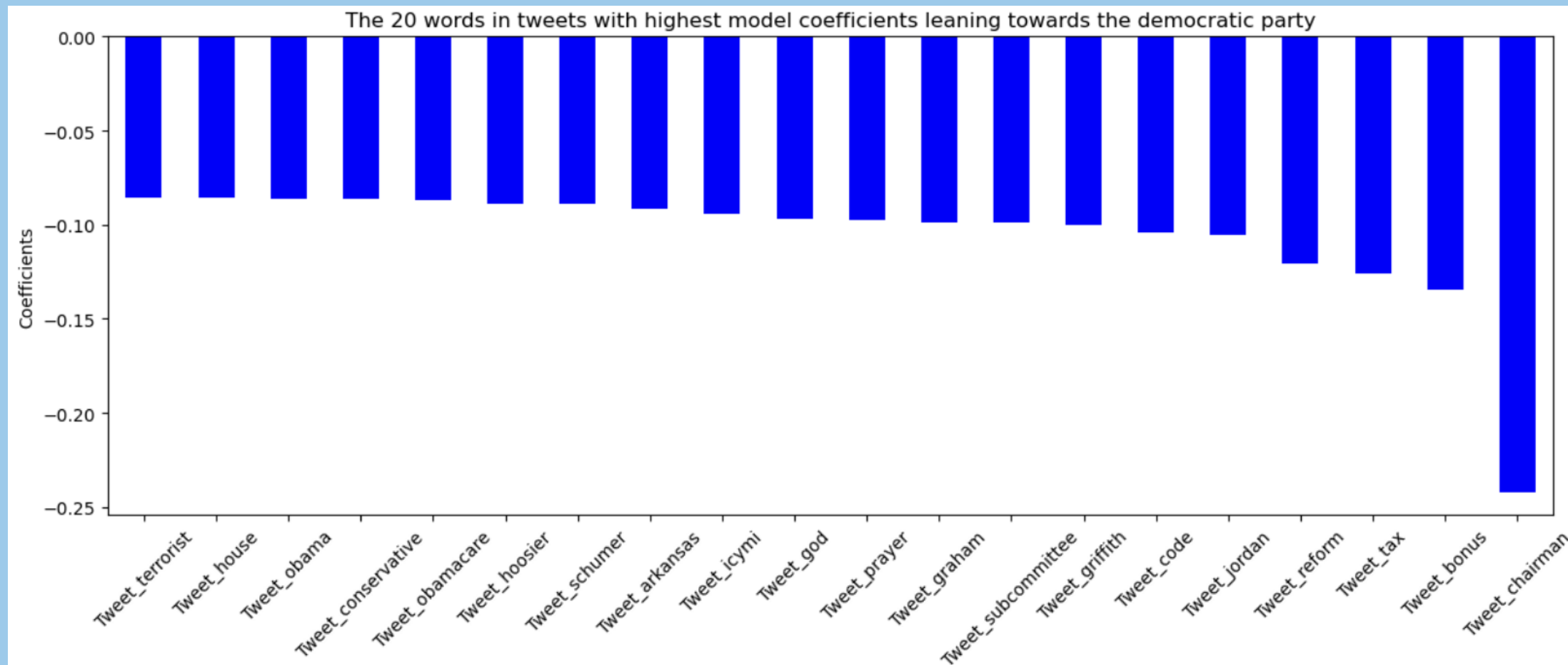
BrainStation



# Top Model Coefficient: Democrat Words in Tweets



BrainStation



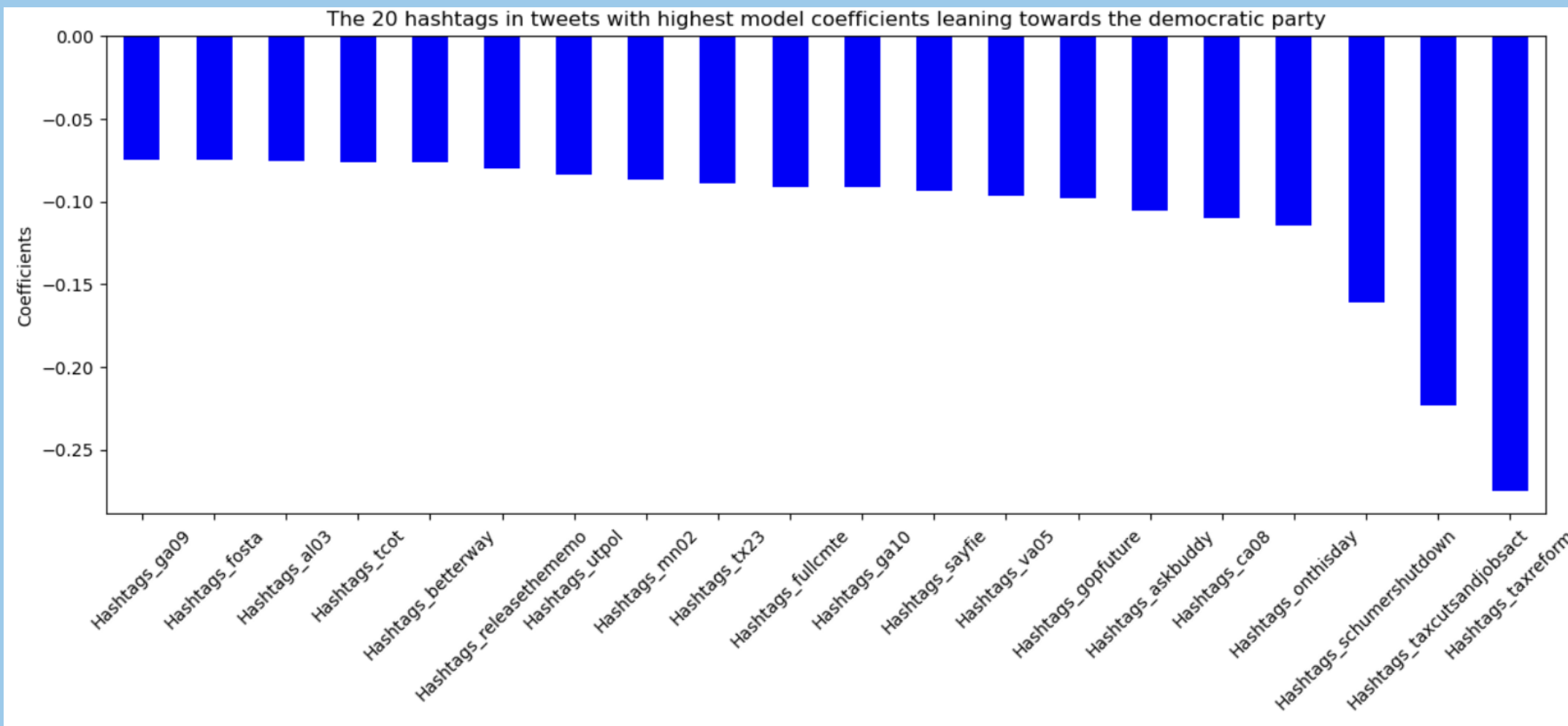


# Top Model Coefficient:

## Democrat Hashtags in Tweets



BrainStation





## ***Next Step***

Memory Allocation Optimization

Cross Validation for Logistic Regression

Pipeline Implementation for Naïve Bayes

Sentiment Analysis Implementation

Scraping New Dataset

Neural Network Application

Further EDA





**BrainStation**

***Thanks!***

Erick Rudelman | BrainStation

