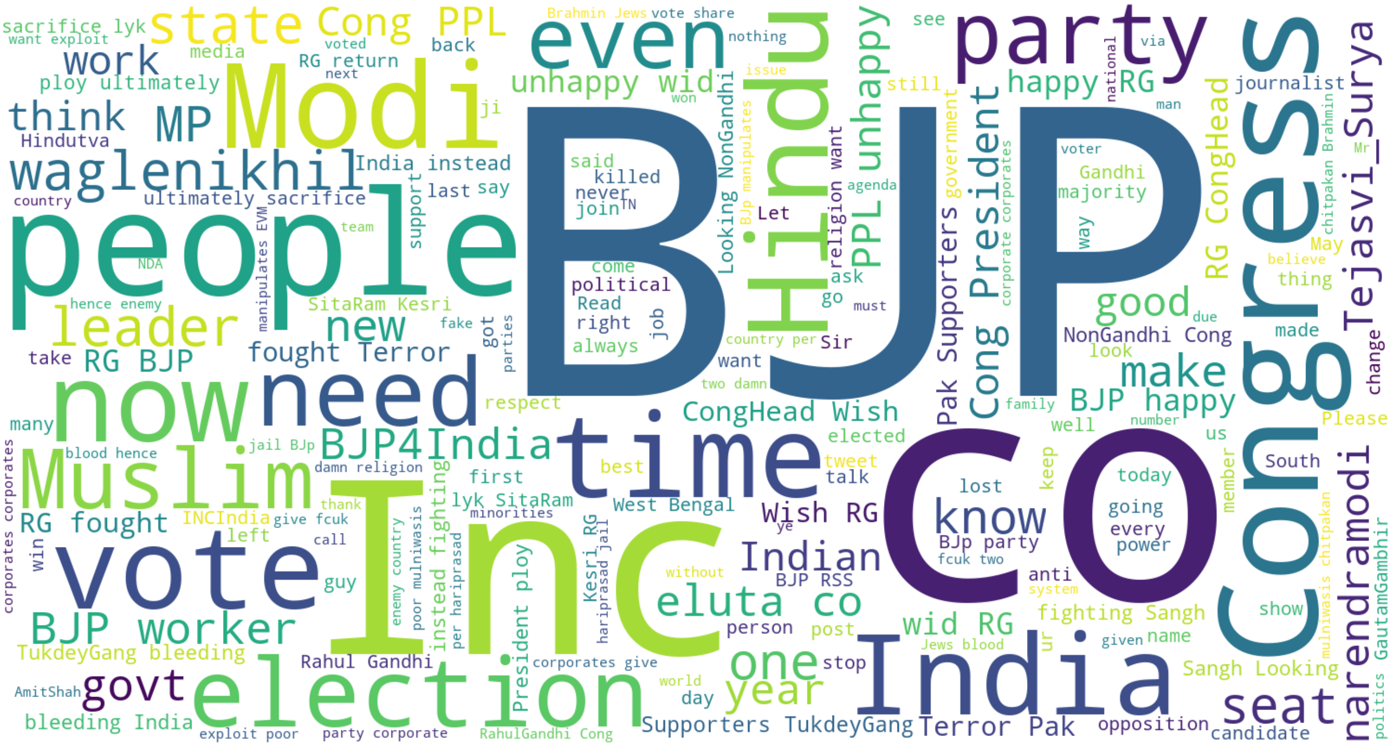
TWITTER SENTIMENT ANALYSIS REPORT

Recently while learning I came across WordCloud Python package for Visualization. For those who don't know about WordCloud, it is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. Significant textual data points can be highlighted using a word cloud. Word clouds are widely used for analyzing data from social network websites.

I collected around 32,000 tweets using Twitter API on what peoples were tweeting about the political parties and their leaders. After gathering the tweets and cleaning them, I visualized the tweets and the following was the result.



After visualising the tweets I wasn't able to conclude that the tweets were positive or negative. So, I visualized all the positive words said by the people and the following was the result.

And all the negative words said by the people in their tweets were the following.

Still, I wasn't sure that positive tweets were more or negative tweets So, I read about scoring the tweets and getting the net score.

```
        if strip_punctuation(i) in positive_words:
            c += 1
            p_list.append(strip_punctuation(i))
    return c,p_list
```

In [72]:
```
# lists of words to use
positive_words = []
with open("positive-words.txt") as pos_f:
    for lin in pos_f:
        if lin[0] != ';' and lin[0] != '\n':
            positive_words.append(lin.strip())

negative_words = []
with open("negative-words.txt") as pos_f:
    for lin in pos_f:
        if lin[0] != ';' and lin[0] != '\n':
            negative_words.append(lin.strip())
```

In [73]:
```
punctuation_chars = ["'", '"', ",", ".", "!", ":", ";", '#', '@']
p_score,p_list = get_pos(words)
n_score,n_list = get_neg(words)
net_score = p_score - n_score
f_str = 'Postive score = {}, Negative score = {}, Net score = {}'.format(p_score,n_score,net_score)
```

In [74]:
```
f_str
```

Out[74]:
```
'Postive score = 910, Negative score = 1000, Net score = -90'
```

Scoring of tweets on the basis of Positive words and Negative words

Thus I can conclude that the sample of around 32000 tweets contains more negativity than positivity. Hope you find it interesting. Suggestions are always welcome. Thank You!

Github link: https://github.com/Reactor11/Twitter-Sentiment-Analysis