

**VISHWAKARMA INSTITUTE
OF INFORMATION
TECHNOLOGY
KONDHWA BK. , PUNE**

**BUSINESS INTELLIGENCE
AND DATA ANALYTICS
MINI-PROJECT**

Twitter based Election Analysis

<u>NAME</u>	<u>ROLL NO</u>	<u>GR NO</u>	<u>BATCH</u>
Raj Thakkar	321052	21810934	A2
Aadarsh Shetty	321048	21810861	A2
Amlan Nanda	321032	21810954	A2
Pranav Sarda	321047	21810773	A2

PROBLEM STATEMENT :

Try to predict the results of a local Indian Election through Sentiment Analysis of Tweets gathered from TwitterAPI. Compare the results of those Elections against the predicted results and the possible factors that may have affected the results. Also, try to analyze the Indian Political Scenario through general Twitter Sentiment Analysis. Also, try to analyze the results of the US Political Elections 2020 based on the available Tweets, and derive meaningful inferences from the same.

OBJECTIVE:

- Predict the results of an Indian State Election by collecting Twitter Data for the same, and compare it with actual results.
- Twitter Sentiment Analysis of Election related tweets through Python.
- To analyze the 2020 US Elections with respect to Tweets, and draw an analogy between the Results and the Tweet Trends.
- Create Dashboards for effective Visualizations of the Analyses done.

THEORY:

1.) RAPIDMINER

RapidMiner is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics. It is used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results visualization, model validation and optimization. RapidMiner is developed on an open core model.

2.) TABLEAU

It is the fastest growing data visualization and data analytics tool that aims to help people see and understand data. In order to transform the way people use data to solve problems, tableau software ensures to meet strict requirements. In other words, it simply converts raw data into a very easily understandable format. It provides top class interactive data visualization with the purpose to help organizations solve their data problems.

3.) TABLEAU DESKTOP:

With the help of tableau desktop anybody can create the reports and dashboards on the Tableau Server and work with by blending it into the system browser. Creating reports, charts and more is all done through tableau desktop. It allows you to connect with various types of data and converts it into simple understandable tableau reporting. Tableau dashboard and tableau data visualization can be accessed even through the mobile devices. The product of desktop product helps to create a great visualization, data stories, data analysis formation of workbooks etc.

4.) TABLEAU ONLINE:

The administrator tableau server handles the access to servers provides permissions to the different data sources and protects data. The data is stored on servers hosted in the cloud which are maintained by the Tableau group. It also helps you manage security, team worker and execution of Tableau products. In other words, it allows you to publish Tableau Dashboards on the Tableau Online and share the data discoveries with anybody and can access and explore the various features like interactive visualizations, explore hidden opportunities and so on.

5.)TABLEAU PUBLIC:

This version of Tableau is particularly built for the cost-effective users. The workbooks created are not saved locally in this version. In fact, it is saved to the tableau's public cloud which can be sighted and accessed easily by anyone. Anyone can download and access it effortlessly. Basically, there is nothing like privacy to these documents saved to the cloud. People who want to learn tableau and wish to share their information with the public can use this version.

6.)TWITTER API:

The Twitter API lets you read and write Twitter data. Thus, you can use it to compose tweets, read profiles, and access your followers' data and a high volume of tweets on particular subjects in specific locations.

API stands for [Application Programming Interface](#). This software provides "middleman services" between two applications that want to communicate with each other. Any requests you make go to the server first and the response given comes through the same route.

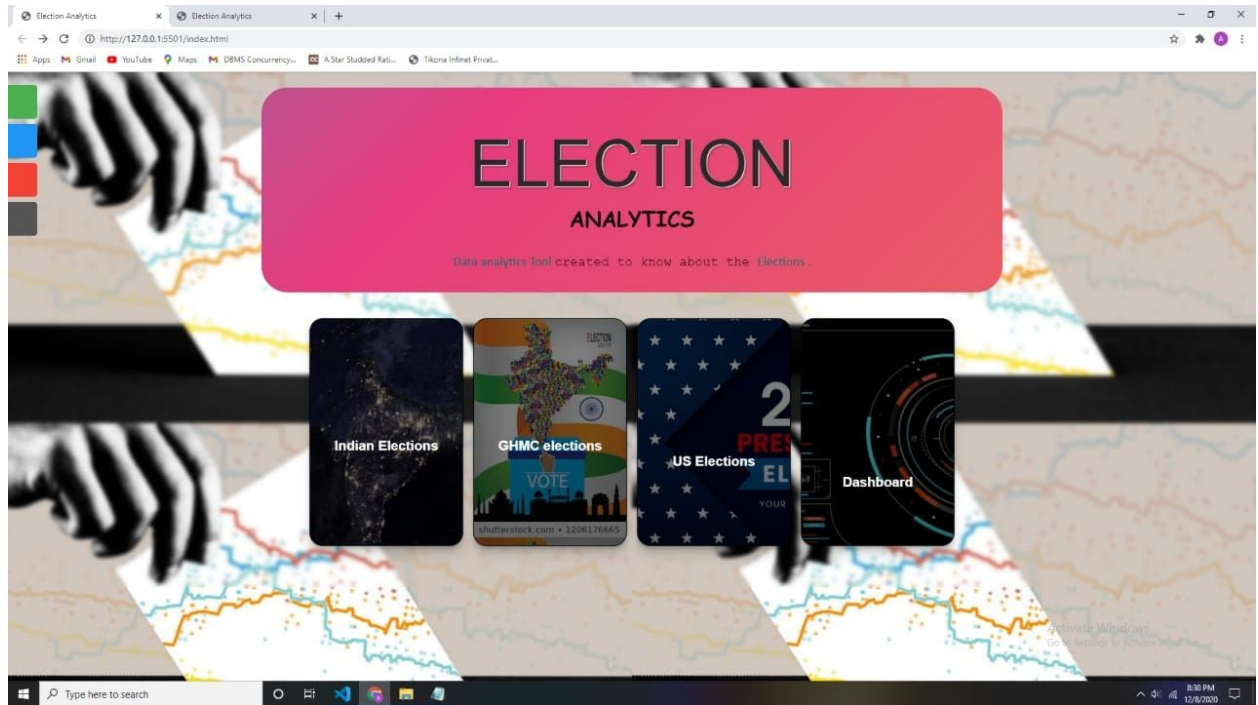
Datasets:

- **US Election 2020 Tweets | Kaggle**
- **Indian Election Tweets through Twitter API**
- **GHMC Election Tweets through Twitter API**

Tools Used:

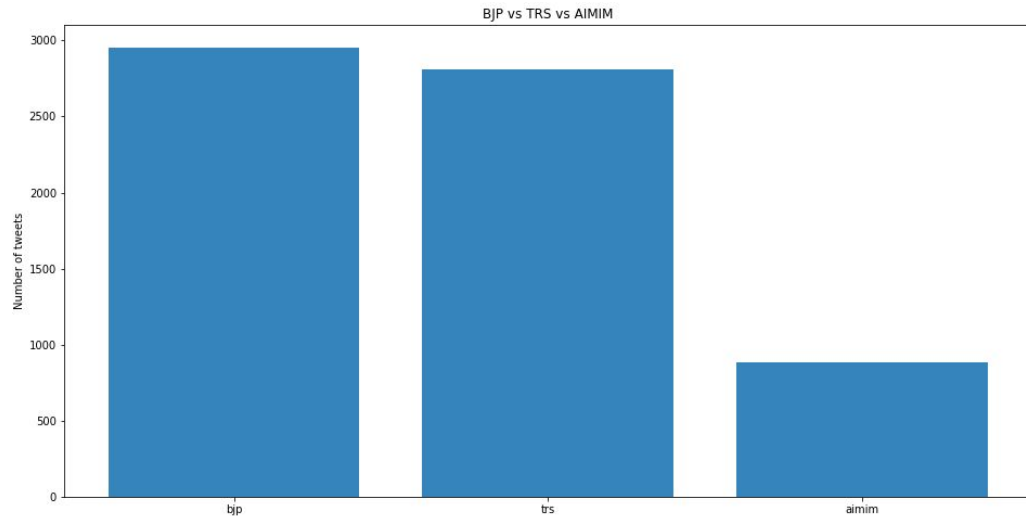
- **Rapidminer**
- **Tableau Desktop**
- **Tableau Online**
- **Tableau Public**
- **Python**
- **Twitter API**
- **Git & Github**

WEBSITE:



GHMC Elections:

1. Number of Tweets collected :

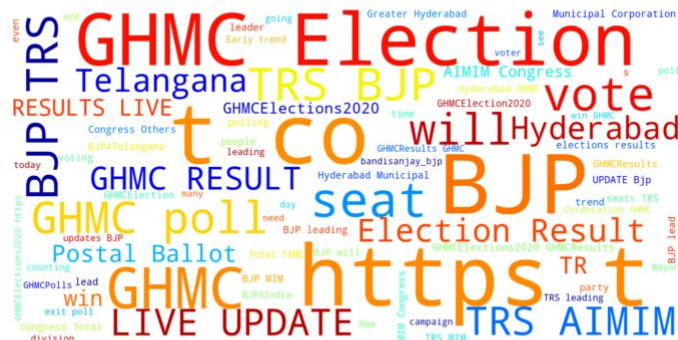


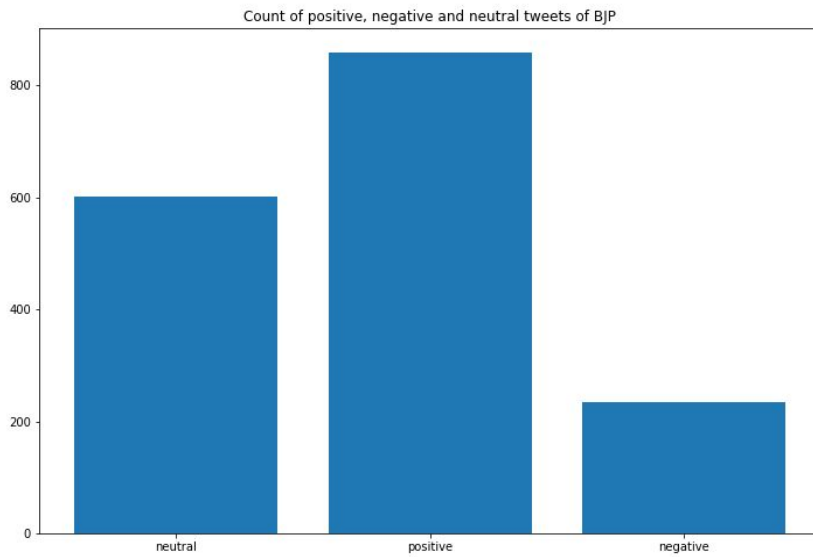
Number of Tweets collected were the highest for BJP at 3000, followed closely by TRS & AIMIM's tweet count was almost 1/3rd of the other 2 parties.

Given the tweet limit of 3000 at the time of collection of tweets, we can see that BJP's tweets must've been well over 3000 while TRS barely reached the 3000 threshold. AIMIM's tweets seem low, but seeing that it is a very young party it is a good tweet count.

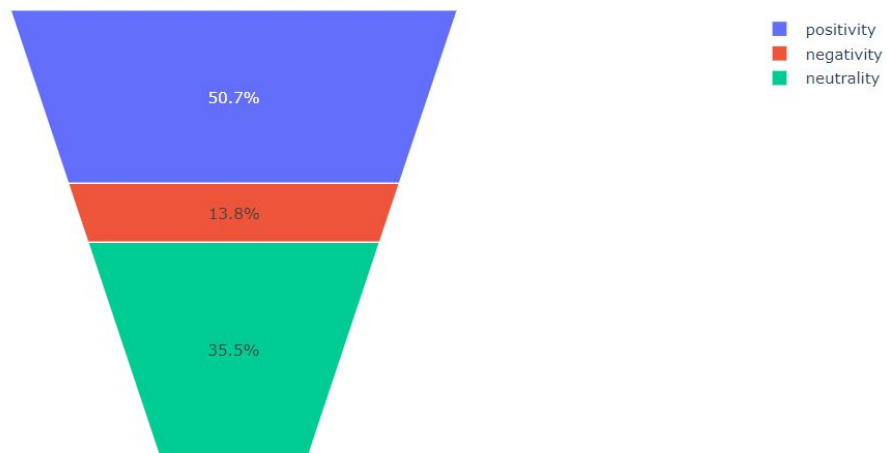
Also, we can assume that a good portion of BJP's tweets must've been non-local to Hyderabad GHMC Voters. On the other hand, most of TRS's tweets would be from the locals only. And, that of AIMIM would be even more so seeing the size of the party.

2. BJP Tweets Sentiment Analysis:





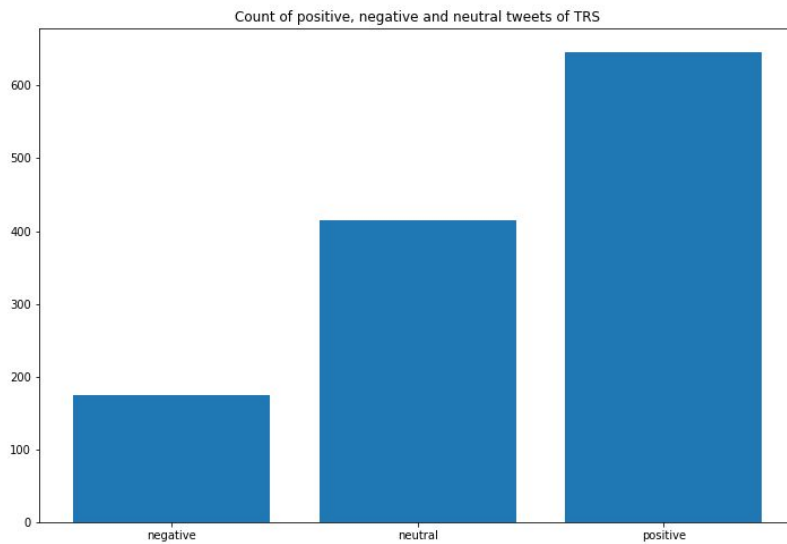
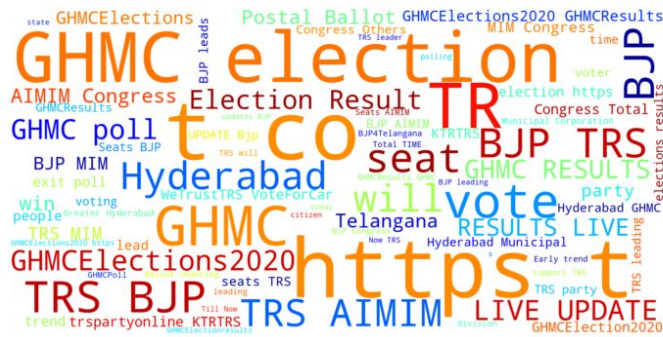
Sentiment analysis tweets BJP



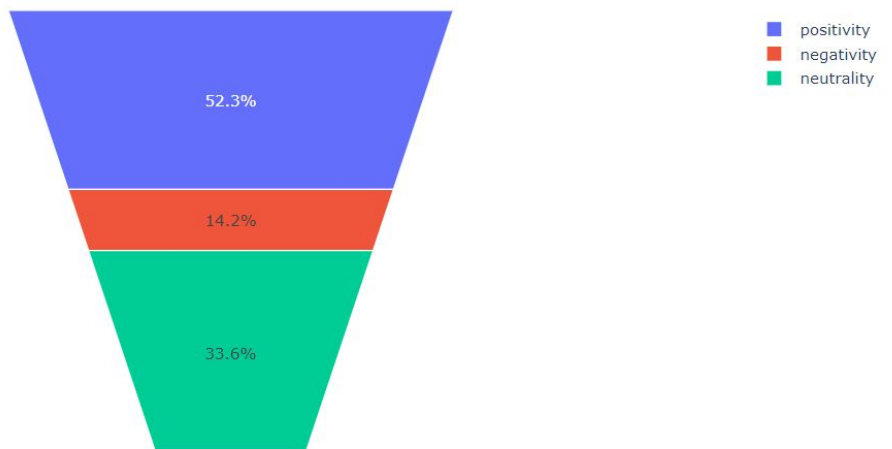
Positive Tweets : 859
Negative Tweets: 234
Neutral Tweets : 601

Positive/Negative : 3.67
Positive/Total : 50.71%

3. TRS Sentiment Analysis:



Sentiment analysis of tweets of TRS



Positive/Negative : 3.69
Positive/Total : 52.27%

A funnel chart illustrating the distribution of sentiment. The chart is divided into three horizontal segments: a top blue segment representing positivity at 49.6%, a middle red segment representing negativity at 15%, and a bottom green segment representing neutrality at 35.5%. A legend on the right side of the chart identifies the colors: blue for positivity, red for negativity, and green for neutrality.

Sentiment	Percentage
positivity	49.6%
negativity	15%
neutrality	35.5%

Positive/Negative: 2.97
Positive/Total : 44.47%

TRS : 55
BJP : 48
AIMIM : 44

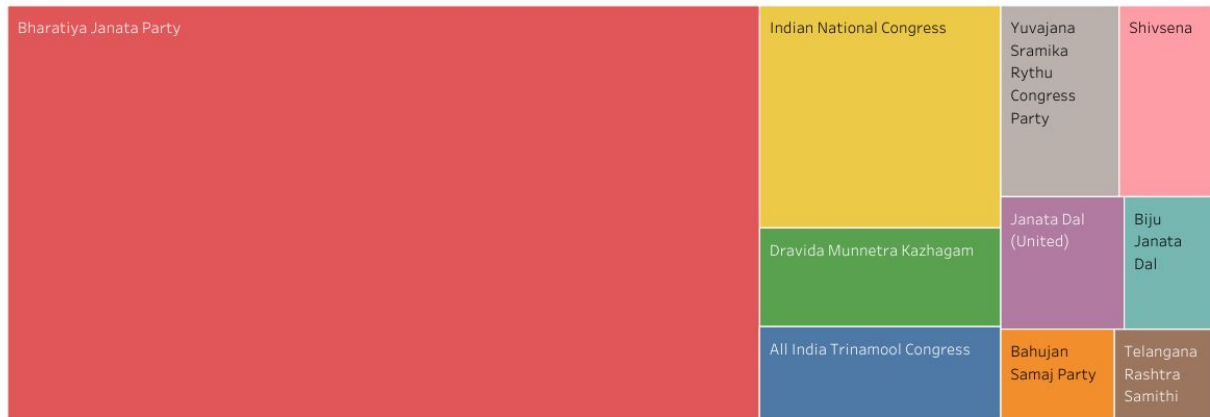
Therefore, we see that the Election Results were close for BJP & TRS. A little less close than should've been according to our Sentiment Analysis, but that is the gap introduced by the non-voter tweets for BJP.

Also, same way for AIMIM, it seems that AIMIM should've been a little less close to BJP, but again there were the non-voter tweets for BJP that in turn led to a rise in the Sentiment displayed on Twitter.

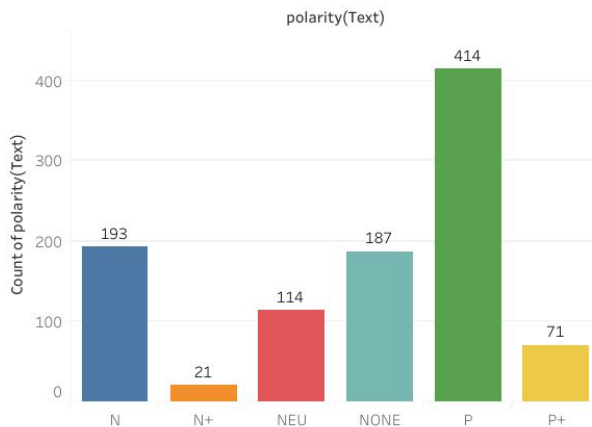
Indian Political Scenario:

Top 10 Parties of 2019 India Elections:

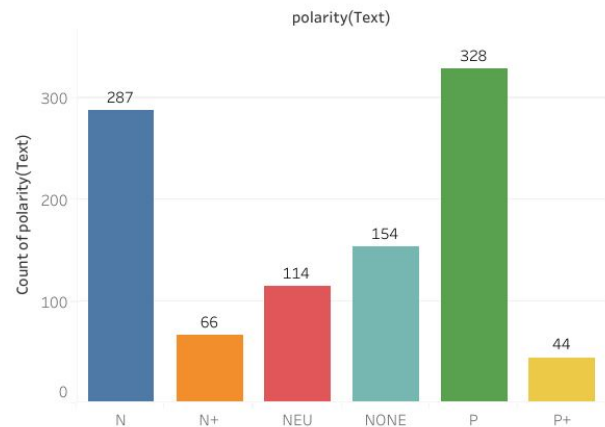
Top 10 Parties by Seats



1. BJP Sentiment



2. Congress Sentiment



Above, we see that BJP was the single largest party winning the 2019 Indian Elections with 303 seats and by a big lead above the second largest party, Congress having 131 seats.

The following are the Sentiments of the top 10 parties of 2019 elections, taken in 2020.

1. BJP :-

- a. Positive : 485
- b. Negative : 214
- c. Positive/Negative : 2.27

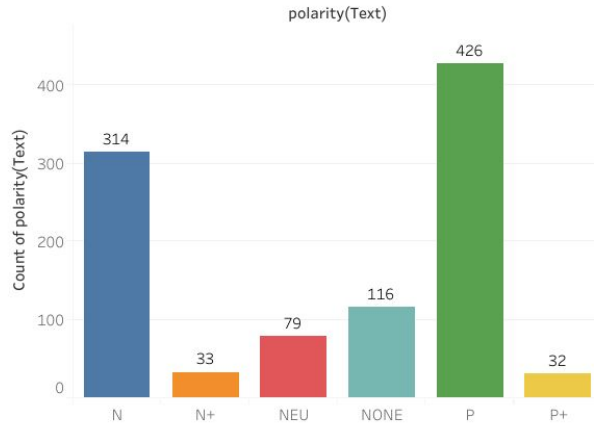
Hence, we see that BJP has a major Positive support even in 2020, just like 2019.

2. Congress :-

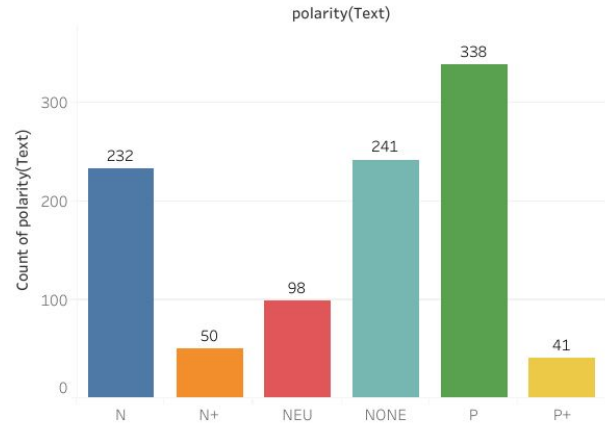
- a. Positive : 372
- b. Negative : 353
- c. Positive/Negative : 1.06

We see that Congress barely has a few more Positive Tweets than it has Negative Tweets. It might actually be performing worse than 2019.

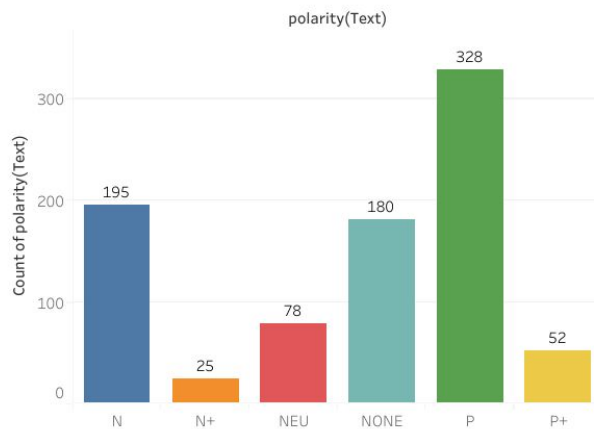
3.DMK Sentiment



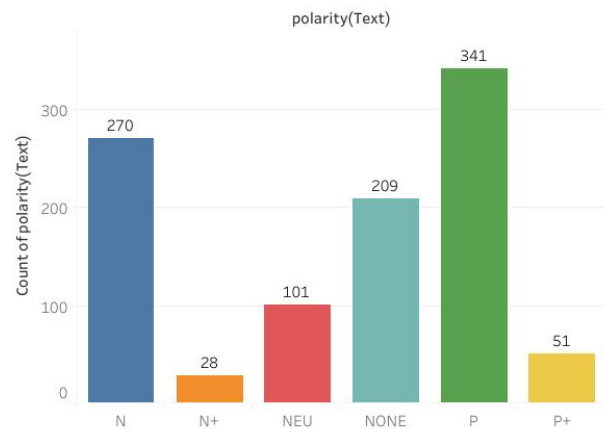
4.AITMC Sentiment



5.YSRCP Sentiment



6.Shiv Sena Sentiment



3. DMK:-

- Positive : 458
- Negative : 347
- Positive/Negative : 1.32

DMK has a good ratio of 1.32 given that it's a very new party yet.

4. AITMC:-

- Positive : 379
- Negative : 282
- Positive/Negative : 1.34

AITMC also has a ratio of 1.34, but you would expect an age old party to have a stronger **ratio than this**.

5. YSRCP:-

- Positive : 380
- Negative : 220
- Positive/Negative : 1.73

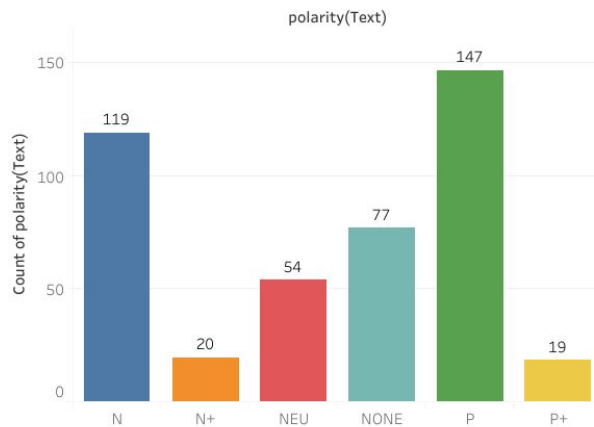
YSRCP has a good ratio of 1.73, given that it's not even 10 years old now.

6. Shiv Sena:-

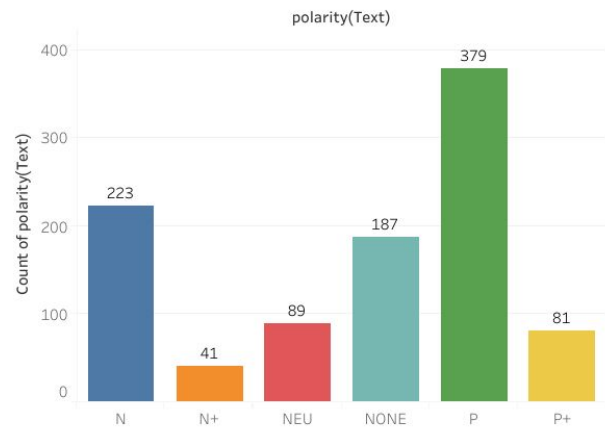
- Positive : 392
- Negative : 298
- Positive/Negative : 1.32

Shiv Sena has a ratio of just 1.32, performing a bit under the expectations since years.

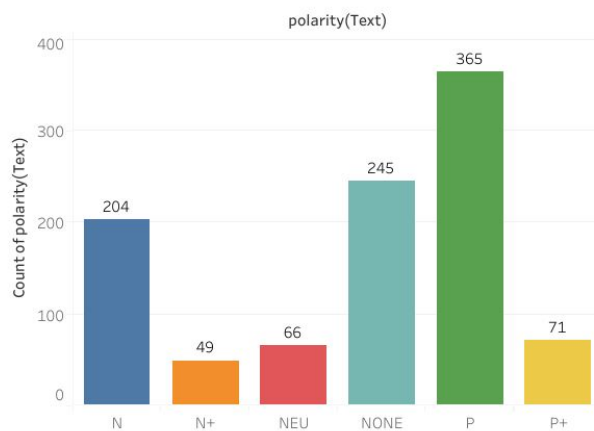
7.JDU Sentiment



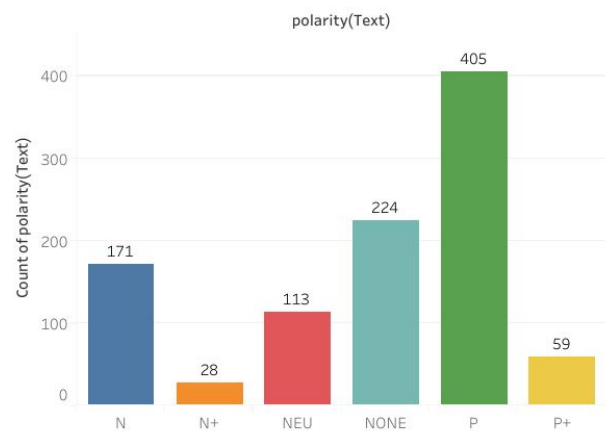
8.BJD Sentiment



9.BSP Sentiment



10.TRS Sentiment



7. JDU:-

- Positive : 166
- Negative : 139
- Positive/Negative : 1.19

JDU has performed badly in the Bihar elections too, evident from the 1.19 score.

8. BJD:-

- Positive : 460
- Negative : 264
- Positive/Negative : 1.74

BJD also is a very strong party with a 1.72 score, evident from the 2019 LA Elections.

9. BSP:-

- Positive : 436
- Negative : 253
- Positive/Negative : 1.72

BSP also seems to have a strong score despite its poor performance against BJP in 2019 UP Lok Sabha Elections.

10. TRS:-

- Positive : 464
- Negative : 199
- Positive/Negative : 2.33

TRS also seems to have a strong score of 2.33, winning against even BJP in GHMC.

Tabulated Results:-

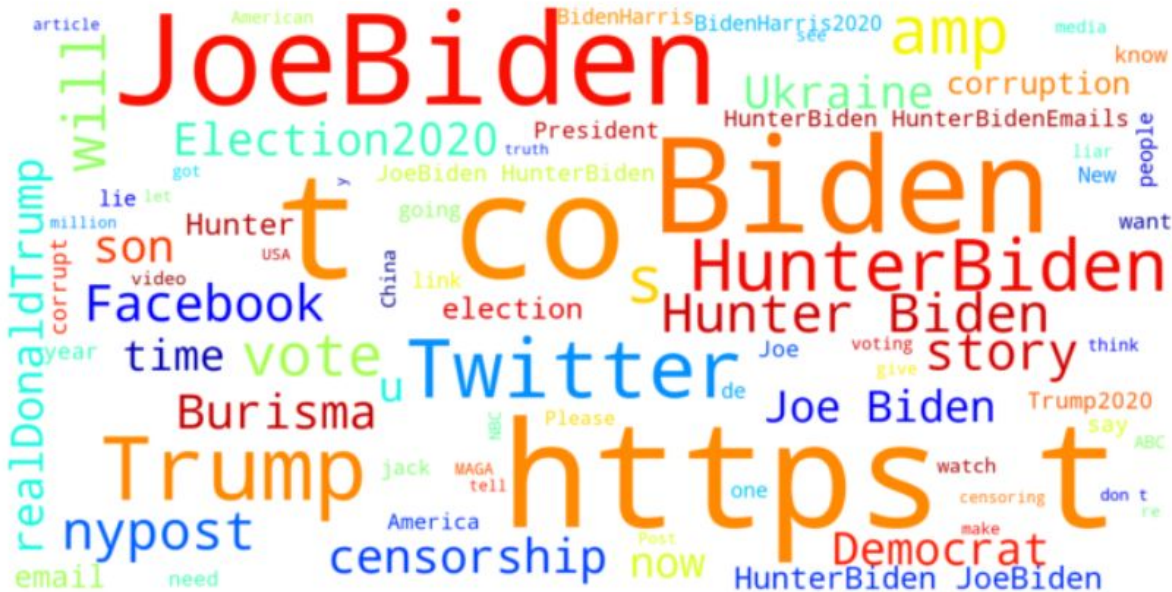
Party	Positive	Negative	Positive/Negative Score
TRS	464	199	2.33
BJP	485	214	2.27
BJD	460	264	1.74
YSRCP	380	220	1.73
BSP	436	253	1.72
AITMC	379	282	1.34
DMK	458	347	1.32
Shiv Sena	392	298	1.32
JDU	166	139	1.19
Congress	372	353	1.06

Looking at the table we can say that :-

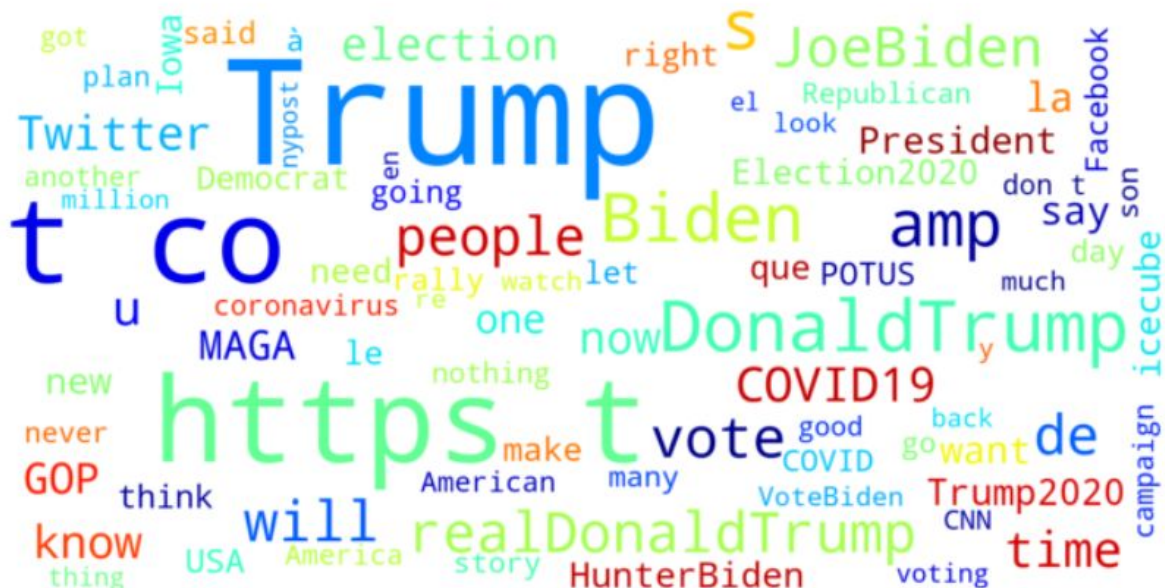
1. Parties having a score below 1.3 can be said to be performing bad.
2. Parties having a score of around 1.-1.4 can be considered good.
3. Parties having a score above around 1.7 score can be considered good.
4. Parties having a score above 2 can be considered really strong.
5. TRS has the highest ratio of 2.33 followed by BJP at 2.27. These 2 are very strong parties.
6. BJD, YSRCP & BSP seem to be performing well as all 3 are in the range of a score of 1.7 .
7. AITMC, DMK & Shiv Sena are barely performing okay with scores around 1.3 .
8. JDU has a bad score of 1.19 & Congress has the worst score of 1.06 from all parties evaluated.

US 2020 Elections:

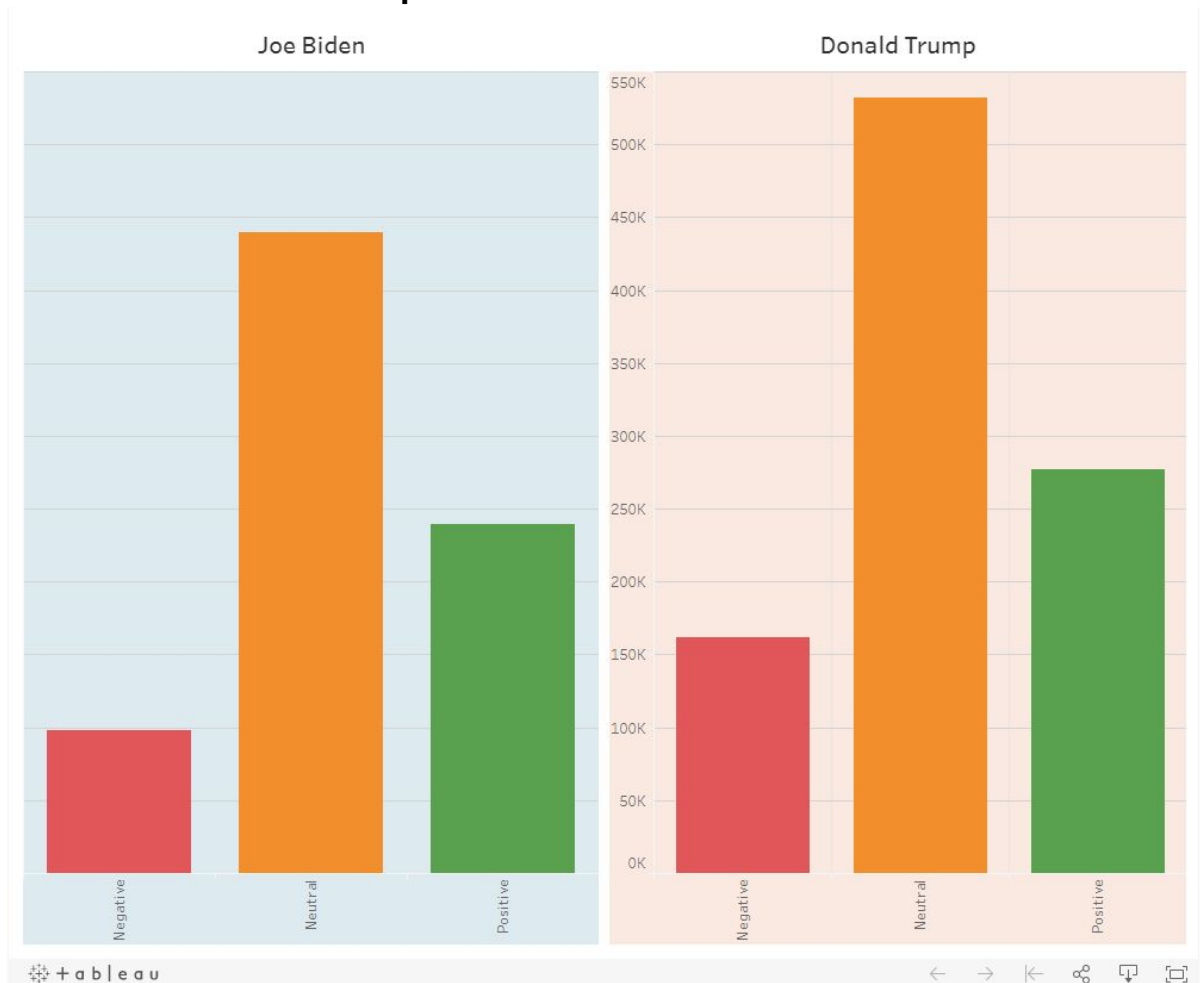
Wordcloud for Joe Biden Tweets:



Wordcloud for Donald Trump Tweets:



Joe Biden vs Donald Trump Tweets Sentiment:



Here, at first glance we see that Trump's Positive tweets are more than Biden's, but then again his negative tweets are more too.

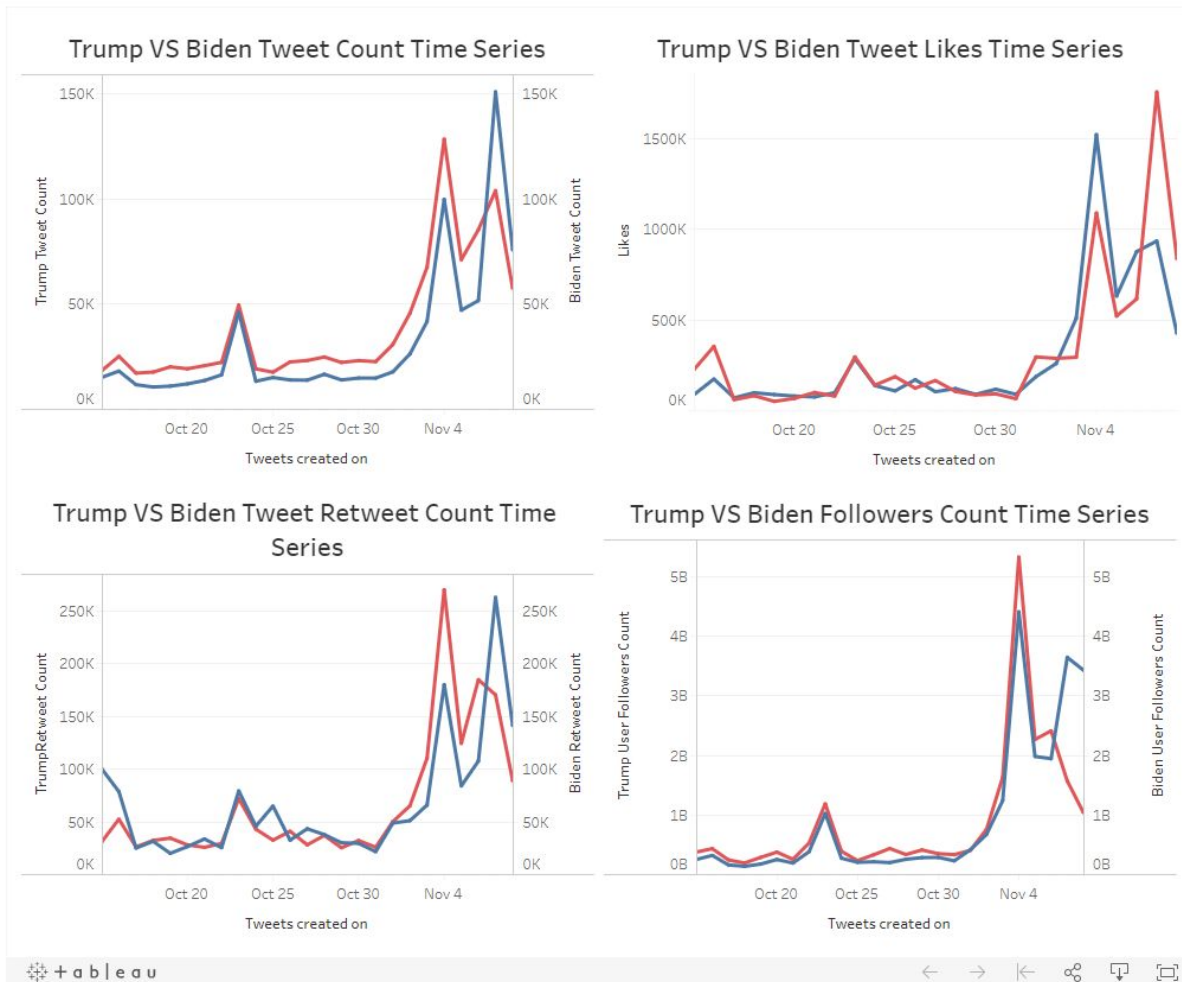
However if we look closely, the difference in Positive & Negative tweets of Biden is more than that of Trump. As Trump has approximately 115k(275k-160k) more Positive tweets than Negative. While, Biden has approximately 140k(240k-100k) more Positive Tweets than Negative.

So, we see that Biden is 25k Tweets stronger than Trump if we cancel out each of their negative tweets with same amount of Positives.

Also, Biden has 30.9% of Positive Tweets.

While, Trump has 28.7% of Positive Tweets.

The ratio of this comes out to be 108%(Biden/Trump). Therefore, Biden has 8% more support than Trump.



1. Tweet Count :-

We see that Trump tweets have always been above the number of Biden tweets. But, Biden tweets overtake Trump tweets at the final stages of counting as the polls and counts all favored Biden more.

2. Tweet Likes :-

In contrast to Tweet Count and every other trend, the Trump tweet likes actually go up in the end. The only reason for this however could be the heated frenzy of tweets from Trump's official Twitter account.

3. Retweet Count :-

Just like Tweet Count, Retweet Count of Biden tweets overtakes that of Trump tweets at the last stage of Counting.

4. Followers Count :-

The same way as other trends, Trump's followers fall down while that of Biden's keep on rising right at the final stages of the US Elections.

5. Overall :-

The Twitterati gets quite noisy about the US 2020 elections right in the last week of the elections and the whole Twitter is a heated battleground of Tweets from both sides trying to attack and defend tweets in the heated Twitter Presidential debates. Also, the general excitement of the US Elections all around the world in combination lead to an explosive rise in the Tweets made in the last week.