Assignment-1

Q1 Top 10 usernames responsible for generating the most content.
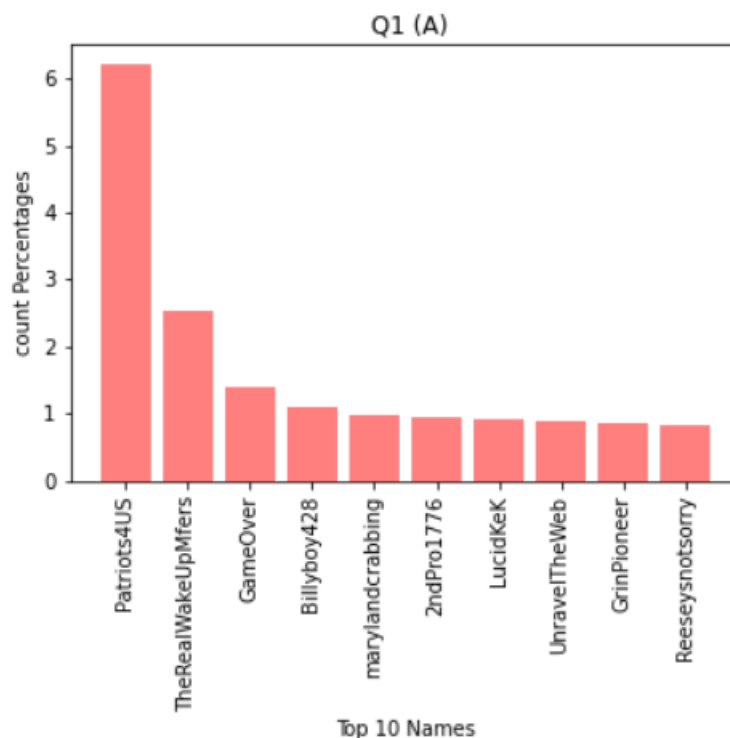
Inference: -

percentage of the total content is generated by the top 10 users cumulatively

16.60601304625043

Only 16.6 percent of people generating most top 10 content mean posting which is very less and showing only few users are posting.

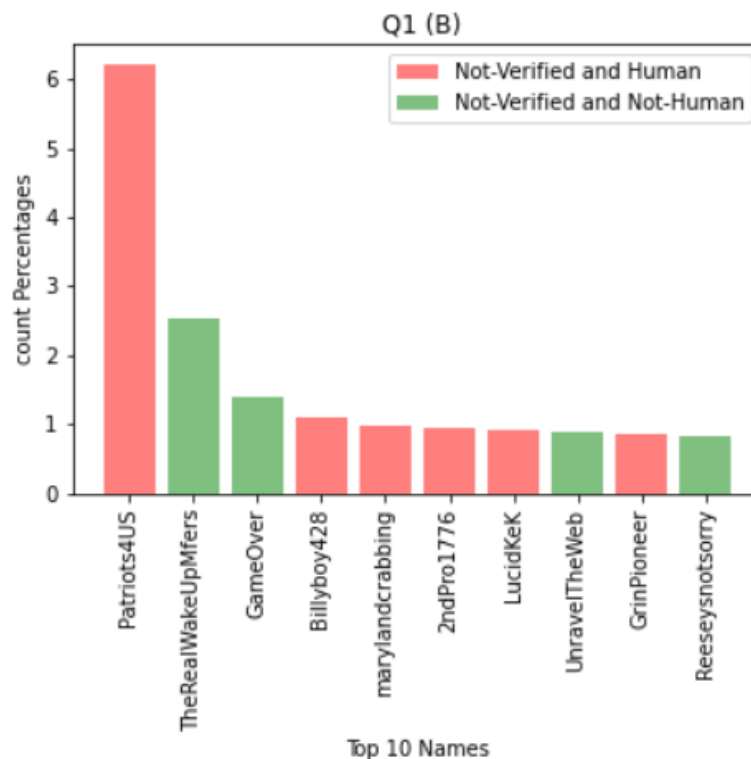Top 10 name you can see in the plot

Normal plot

Inference: -

As we can see in the plot Most content is generated in top 10 username #1 is Human but not verified so it may be an organization who is mass tweeting and the green ones (not verified and not human) may be bots who are misleading or spreading some agenda to people.

There is no verified user in top 10 means verified user post very less.

Verified/Human plot

```
Red color = Not-Verified and Human
Green color = Not-Verified and Not- Human
blue color = Verified and Not-Human
yellow color = Verified and Human
```


Q2 A 1)

top 10 usernames with most number of upvotes,
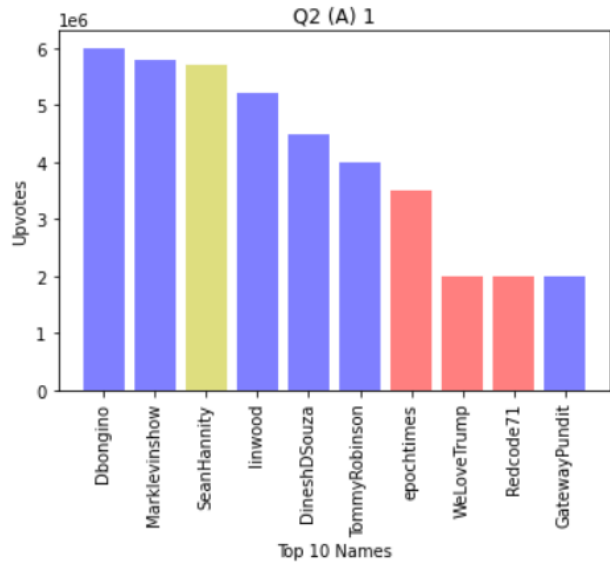
```
TOp 10 usernmae with most upvotes
['Dbongino' 'Marklevinshow' 'SeanHannity' 'linwood' 'DineshDSouza'
 'TommyRobinson' 'epochtimes' 'WeLoveTrump' 'Redcode71' 'GatewayPundit']
Red color = Not-Verified and Human
Green color = Not-Verified and Not- Human
blue color = Verified and Not-Human
yellow color = Verified and Human
```


A 2) top 10 usernames with most number of interactions

TOp 10 usernmae with most Interactions
['Marklevinshow' 'AmyPeikoff' 'parler' 'parlerdev' 'oathkeep3r' 'Jared'
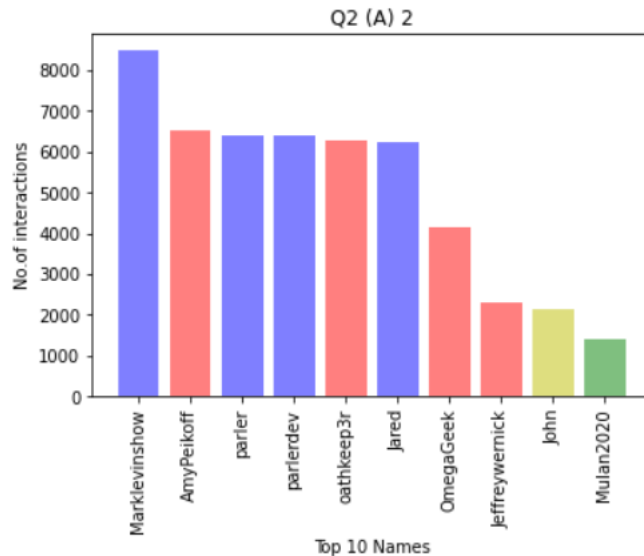 'OmegaGeek' 'Jeffreywernick' 'John' 'Mulan2020']
[8478 6503 6407 6403 6275 6247 4131 2307 2151 1410]
Red color = Not-Verified and Human
Green color = Not-Verified and Not- Human
blue color = Verified and Not-Human
yellow color = Verified and Human

A 3) top 10 usernames with most number of Mentions
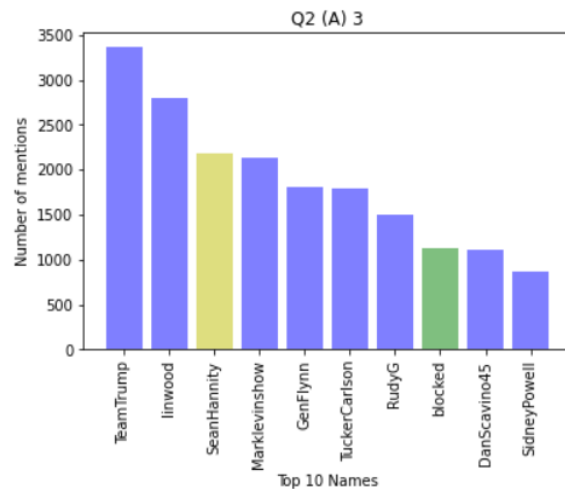
TOp 10 usernmae with most mentions
['TeamTrump', 'linwood', 'SeanHannity', 'Marklevinshow', 'GenFlynn', 'TuckerCarlson
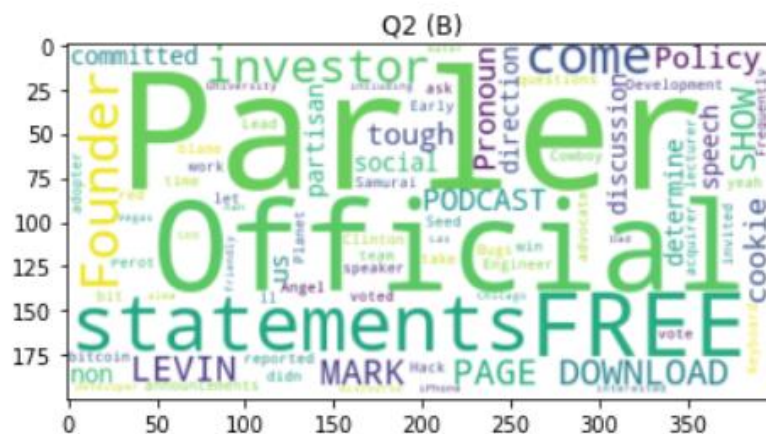[("'64ebc91ab5df4b768c37c21cc8a9e639'", 3363), ("'2ea66900ed69a5005fe51fc8b07a1711''
Red color = Not-Verified and Human
Green color = Not-Verified and Not- Human
blue color = Verified and Not-Human
yellow color = Verified and Human

Q2 B)

Inference: -

Basically top 10 users who have most interaction with others on paler or replying over almost every post. These user using these words in their bio mostly:- Parler, official , statement, Free, COME, Download , PAGE, Investor, Show, Podcast etc.

As we know from word cloud. bigger the size means greater the occurrence of that word so parler, official, statements and free are occurring mostly and maybe people are using **official and parler** in their bio for because they maybe official member of parler or having a official accounts and by looking at Plot of interaction (Q2 A) except mulan2020 all are human and 5 usernames are verified that's why they are using official and parler in their bio.

And using **Free** in their bio maybe showing they are advertising parler as parler is free and available to everyone and **Statement** is using in bio maybe they are declaring something in their bio.

Some are using **podcast** as maybe they are advertising or talking about something to people using podcast.
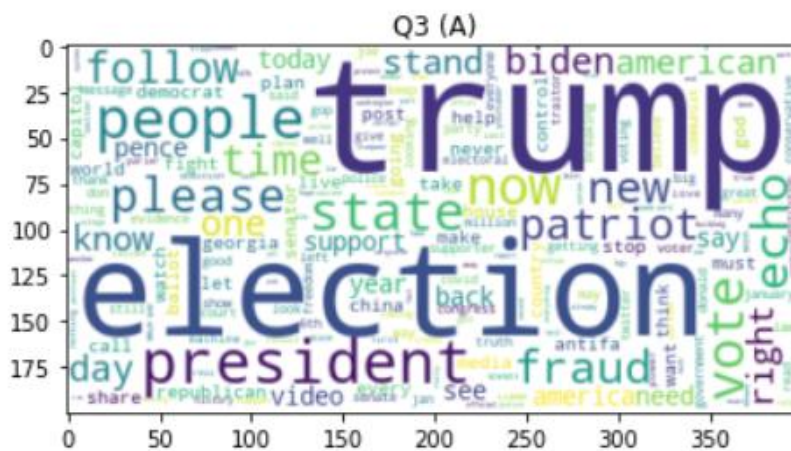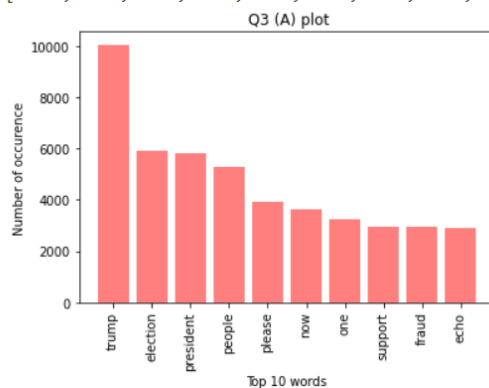
Q3 A)

What I have Observe?

Most used words are Trump, election, President etc. shows Mostly people are posting about us election of selecting US president and they are also posting about Trump. Trump is very famous as he is in most of the post (famous than biden) and many of the words are about US election and politics as people are mostly posting about election.
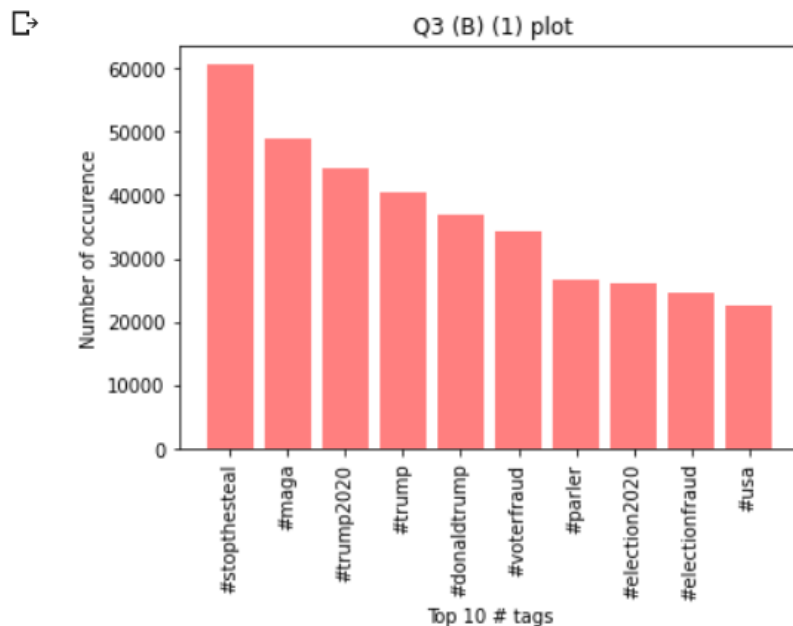


```
['trump', 'election', 'president', 'people', 'please', 'now', 'one', 'support', 'fraud', 'echo']
[10061, 5932, 5804, 5288, 3937, 3652, 3263, 2962, 2945, 2892]
```

Q3 B 1)

As we can see in the plot #stopthesteal is most used tag and it's about US election basically all top 10 # tags are about us election or related and we can consider that some of # tags show popularity of Donald trump because those #tags are started or related to trump or his supporters.
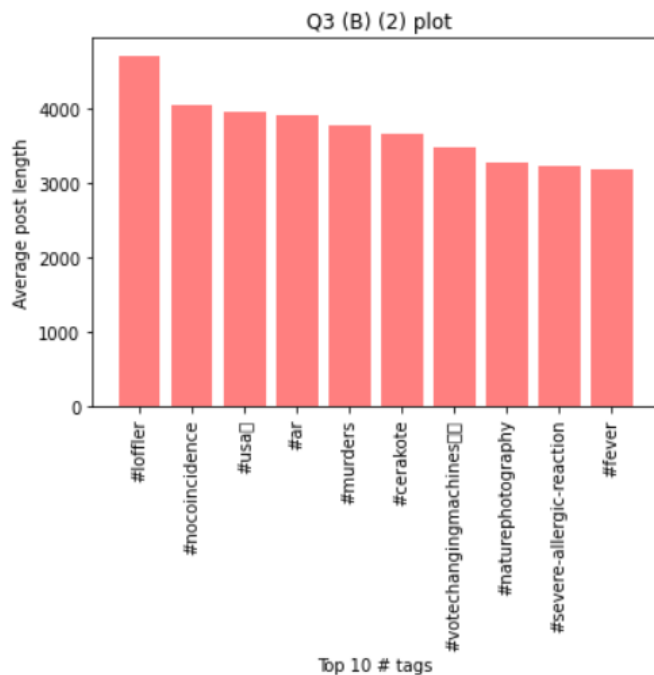
Mostly are maybe in Favour of trump or maybe criticizing or related to trump him that's why most #tags are related to trump

Q3 B 2)

In this plot basically these # tags are averagely used by people like #usa # murders are some # tags that we consider or maybe used by people mostly in a long or big post (post contains more text) people are very well describing their post that are related to these # tags given below in plot.

And the post with greatest Post length contains #loffler whose full name is Kelly Lynn Loeffler a American business man and politician. Peoples are posting or describing briefly and using #loffler

Q3 (B) (2) plot

Q4 A

Observations

The data is from 24 july 2020 to 8 jan 2021

And the day at which most of the post has created is 6 jan 2021 and no. Of post that created is 16.593k.

And most of the post that is generated is on month of dec and jan 2021 and mostly in jan 2021

Before 9 nov 2020 there are less than 100 post creation per day.

There was an incident name Capitol hill incident happens in us that's maybe a reason of most post creation at 6 jan 2021

['2020-07-24', '2020-07-25', '2020-07-27', '2020-07-28', '2020-07-29', '2020-07-30', '2020-07-31', '2020-08-01', '2020-08-02', '2020-08-03', '2020-08-04', '2020-08-05', '2020-08-06',
[3, 2, 1, 4, 5, 4, 1, 7, 4, 2, 7, 4, 6, 3, 4, 1, 3, 1, 6, 1, 3, 2, 1, 1, 1, 2, 3, 5, 3, 3, 2, 4, 1, 1, 2, 1, 1, 3, 4, 4, 8, 3, 2, 2, 14, 38, 19, 9, 3, 3, 2, 5, 4, 1, 1, 2, 4, 4, 1, 2,

Q4 B

Observations

The data is from 23 Aug. 2018 to 8 jan 2021

Ther are 5 big peeks in graph showing most of the account has created on these 5 days 10 dec 2018, 30 may 2019, 28 may 2020, 26 June 2020 and 8 nov 2020

In increasing order.

And on 8 nov 2020, 1006 accounts have been created (it is max)

It may be a special event or someone has dealing with some organization for creating fake account for creating fake tweets.

['2018-08-23', '2018-08-24', '2018-08-27', '2018-08-28', '2018-09-28', '2018-10-05', '2018-10-20', '2018-10-22', '2018-10-26', '2018-11-08', '2018-12-06', '2018-12-07', '2018-12-08',
[1, 1, 1, 1, 7, 1, 1, 2, 2, 1, 1, 1, 2, 8, 250, 73, 59, 95, 39, 16, 14, 10, 4, 5, 2, 3, 3, 2, 3, 2, 1, 1, 2, 1, 3, 1, 2, 2, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 2, 1, 1,