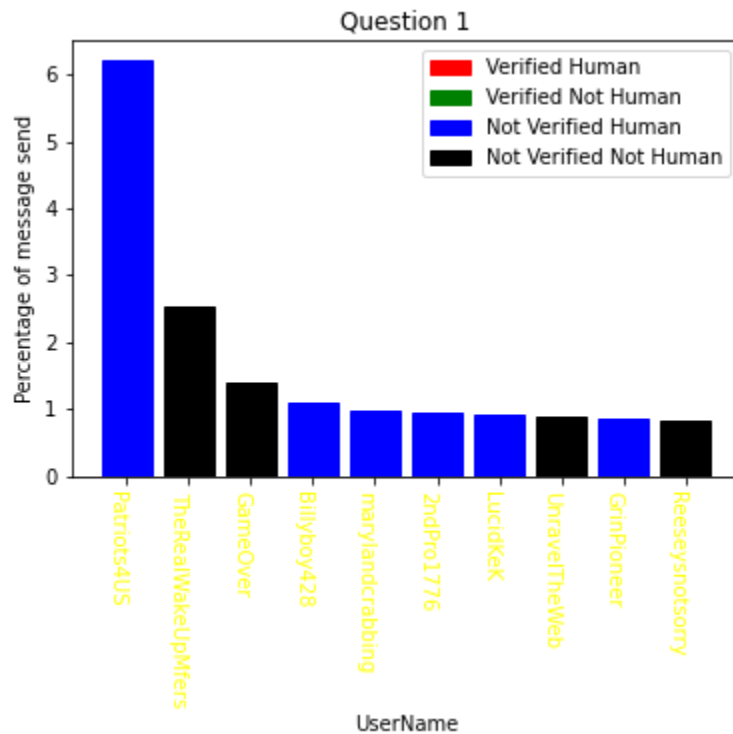


Q1 Top 10 users with most no of posts



Percentage of content contributed by each of top 10 user:

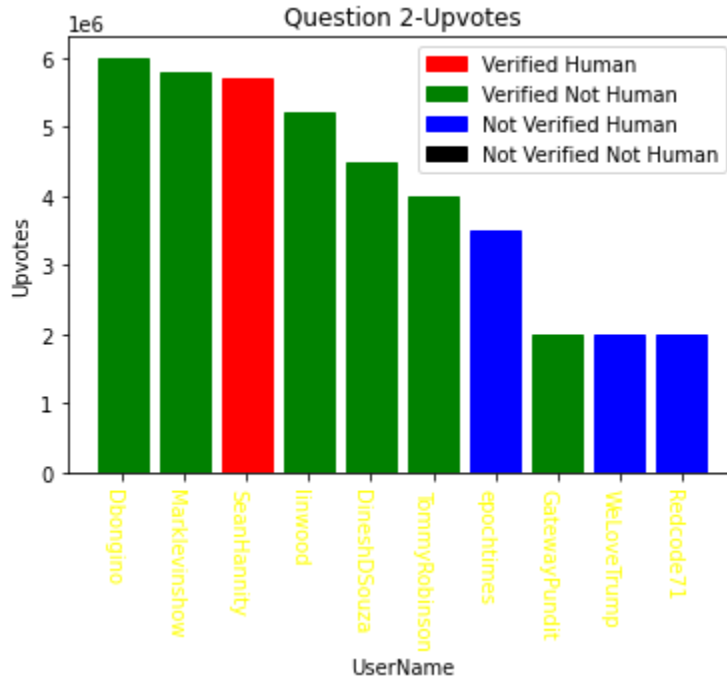
- Patriots4US 6.199421256559909
- TheRealWakeUpMfers 2.5307763990386976
- GameOver 1.3791750453676002
- Billyboy428 1.0839177988130855
- marylandcrabbing 0.9750355583893277
- 2ndPro1776 0.9475697680121634
- LucidKeK 0.9044092402766197
- UnravelTheWeb 0.8995046348521262
- GrinPioneer 0.8563441071165825
- Reeseysnotsorry 0.829859237824317

16.60601304625043 total content generated by top 10 cumulatively

Inference: The inference we could draw is that a large % of content is generated by a single user named as (Patriots4US 6.199421256559909) . Also, the top 10 user contributes 16% which is a large chunk when compared to the no of users in Userdata. These id's create a soft influence over other users since it is human tendency that when we see same content more no of times then we believe it to be true, it doesn't matter whether it is fake or real news. These ids create influence among other users and could change the perception of users.

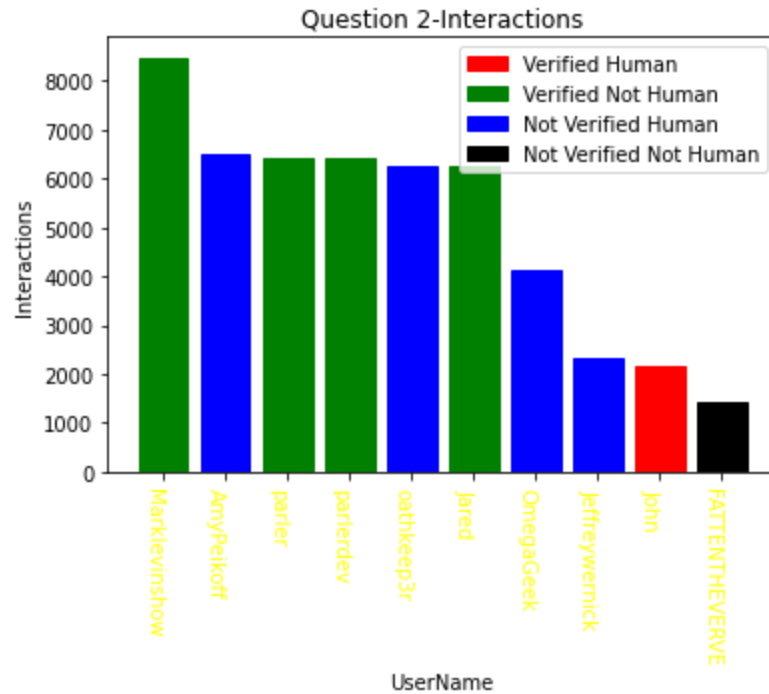
Q2:

A. (i) Top Ten usernames with most number of upvotes.



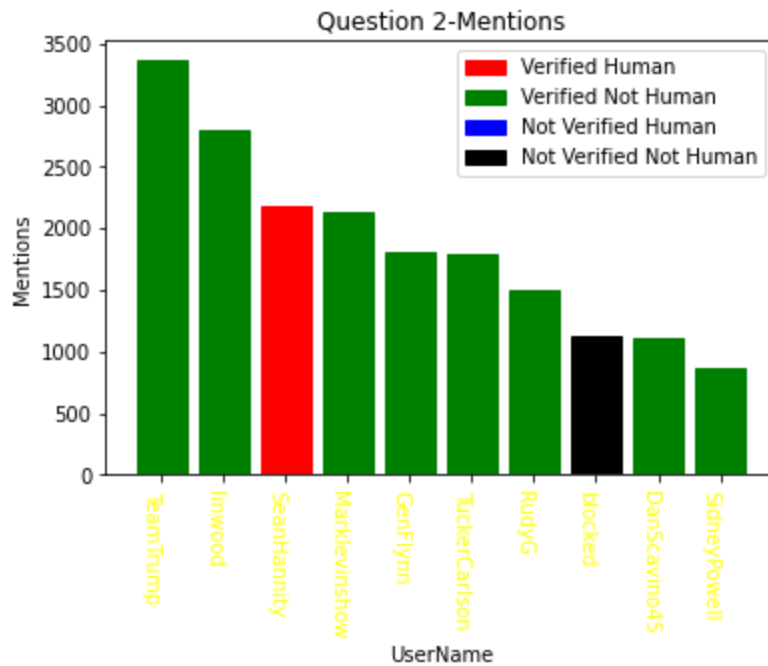
- Dbongino 6000000
- Marklevinshow 5800000
- SeanHannity 5700000
- linwood 5200000
- DineshDSouza 4500000
- TommyRobinson 4000000
- epochtimes 3500000
- GatewayPundit 2000000
- WeLoveTrump 2000000
- Redcode71 2000000

(ii) Top ten usernames with most number of interactions.



- Marklevinshow 8478
- AmyPeikoff 6503
- parler 6407
- parlerdev 6403
- oathkeep3r 6275
- Jared 6247
- OmegaGeek 4131
- Jeffreywernick 2307
- John 2151
- FATTENTHEVERVE 1410

(iii) Top ten usernames with most number of mentions.



- TeamTrump 3363
- linwood 2802
- SeanHannity 2185
- Marklevinshow 2139
- GenFlynn 1805
- TuckerCarlson 1786
- RudyG 1507
- blocked 1130
- DanScavino45 1103
- SidneyPowell 859

B) Word Cloud

Note: Taken reference from GFG to plot word cloud.



Inferences:

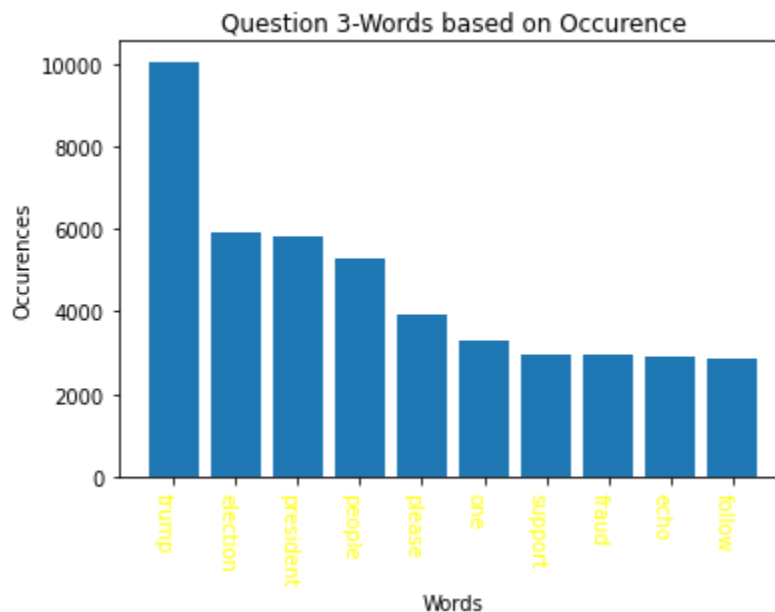
Since this is a bio, terms/professions like “investor”, “official”, “founder”, “podcast”, “advocate”, “lecturer”, “engineer”, “speaker” etc are highlighted. Since the app is parler so people tend to write “official parler account” that why these terms could also be found. Which also implies that they have a verified account, since only these types of accounts tend to write official in their bios. Some characteristics of the person could also be found like “friendly”, “team”, “committed”, “free”, “determine” etc these are the qualities that define that person.

Note: Taken reference from GFG to plot word cloud.

A)



Top 10 words based on occurrence



- trump 10068
- election 5934
- president 5806
- people 5294
- please 3937
- one 3267
- support 2964
- fraud 2949
- echo 2892
- follow 2868

Assumption: Data cleaning is done to remove data cleaning done to remove nan=no bio, websites, mentions, urls, short word of length=2 and hashtags. Written .org, .com specifically since the data preprocessing is done according to the current data set. Replaced “\n”, “\n\n” to “.”.

Also “now” and “will” are replaced with space, since these can be tenses used in sentences.

Inferences(word cloud): Since most of the users get activated during election time of US (done in Q4) the words like “trump”, “president”, “biden”, “elections”, “country”, “vote” etc are highlighted. Since the word “trump” could be seen many times and is also greater in size to “biden” this means most of the users were supporters of “trump” posting regarding him. These means most of the topic discussed there were related to these two personalities only.

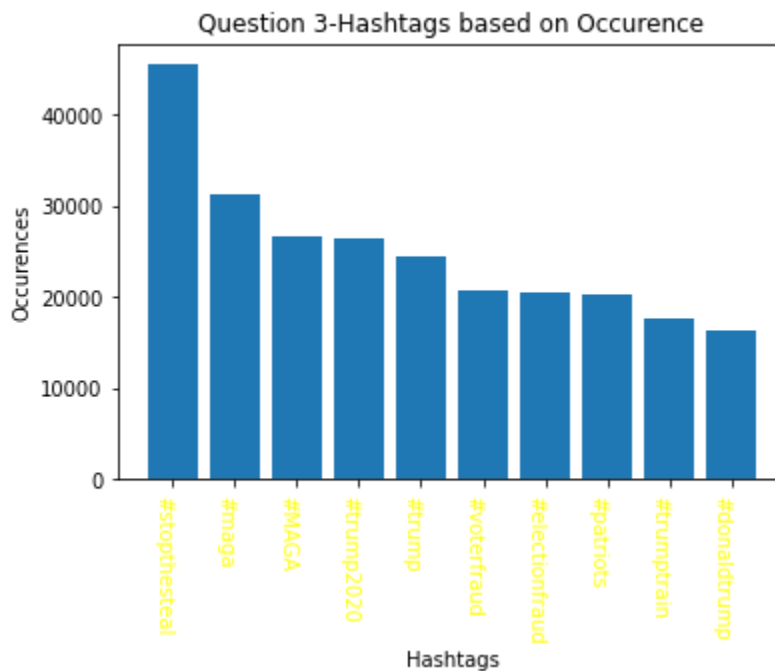
We could also see some words like “fraud”, “worse”, “re”, “traitor”, “evidence” etc making soft influence on people that the election was fraud/not done correctly, and there should be re-election and some are asking for proofs/evidence of Biden winning this election. Word like “great” could be seen which is used by Trump as a motto “Make America Great Again” is used in order to ask its followers to do something for America. Words like “going”, “video”, “protest” is showing their plan of protesting somewhere in America.

Inferences(top 10 words): All these words are related to elections taking support for Trump, influencing people that these elections are a fraud and we should follow Trumps instructions.

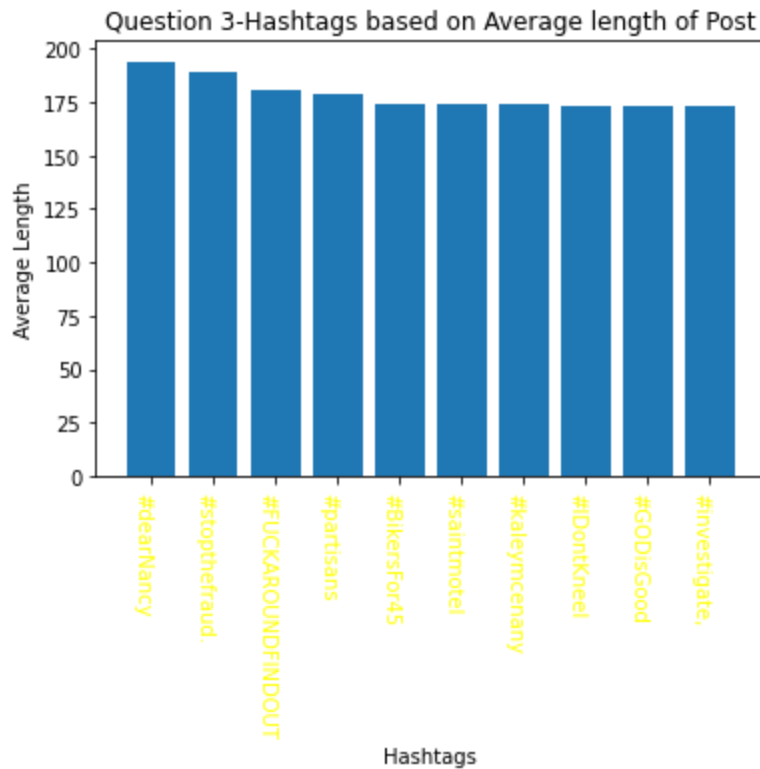
B)

Assumption: Assuming that all hashtags are different if there is a change in its case. For example “trump”, “Trump”, “TRUMP” will be treated differently. This is done according to Twitter which treats them differently when counting the no of tweets with a particular hashtag.

i) Top 10 hashtags based on number of occurrences.



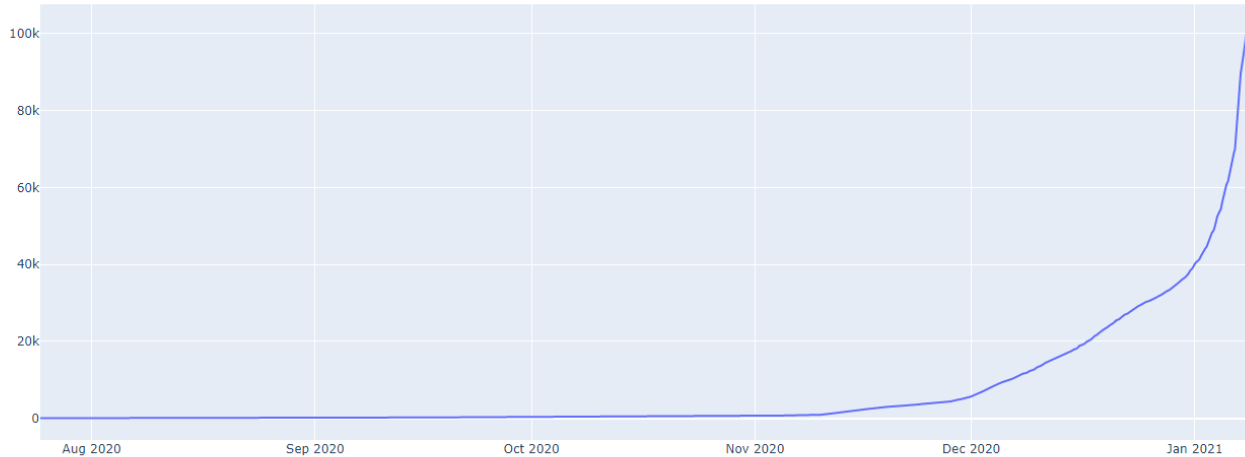
ii) Top 10 hashtags based on average post length



Inference: All these hashtags are trying to convey that Trump is our true president, these elections were fraud and they should support him. #stopthesteal implies that is stealing of votes from Trump to Biden. “#don’t kneel” saying not to believe these election results. “#patriots” were also there which means only people believing in Trump as their president is the true patriot and rest are not.

Q-4

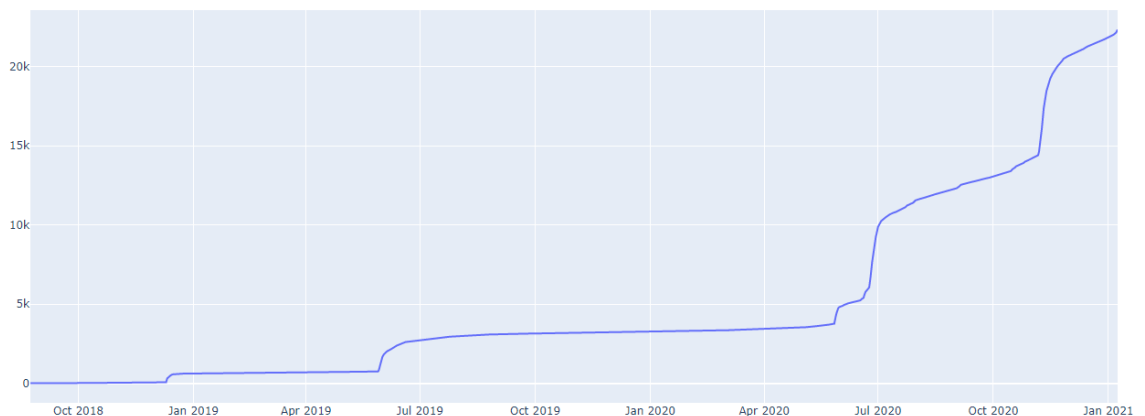
1. Content generation on parler



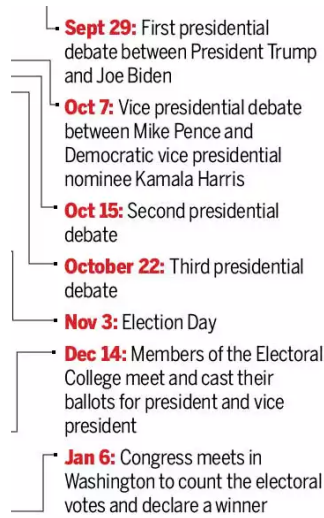
Assumption: I have taken cumulative sum due to mathematical advantage. Because of slope, we could calculate which date has the greatest no of activity.

Inference: See initially there was no activity on parler. The election date was 3rd November around that time only activity started increasing on parler. After that time the growth of activity increased exponentially. Jan 20 the winner was declared and as could be seen the activity was highest during that time. When merged with the body of post we could predict that something was about to happen.

2. User Account creation on Parler



Inference:



Reference:

<https://timesofindia.indiatimes.com/world/us/us-presidential-elections/us-presidential-election-2020-key-dates-and-events/articleshow/78634153.cms>

When we combine these two images we could conclude that during each important the user creation activity on parler increased. We could see the exponential increase during the last months since the results were during that time.

Sources/references:

- <https://plotly.com/python/time-series/>
- <https://www.geeksforgeeks.org/generating-word-cloud-python/>
- <https://towardsdatascience.com/how-to-group-data-by-different-time-intervals-using-python-pandas-eb7134f9b9b0>