Social-Media Data Analysis

Parler Data

The data is in the form of tables with following column details:-

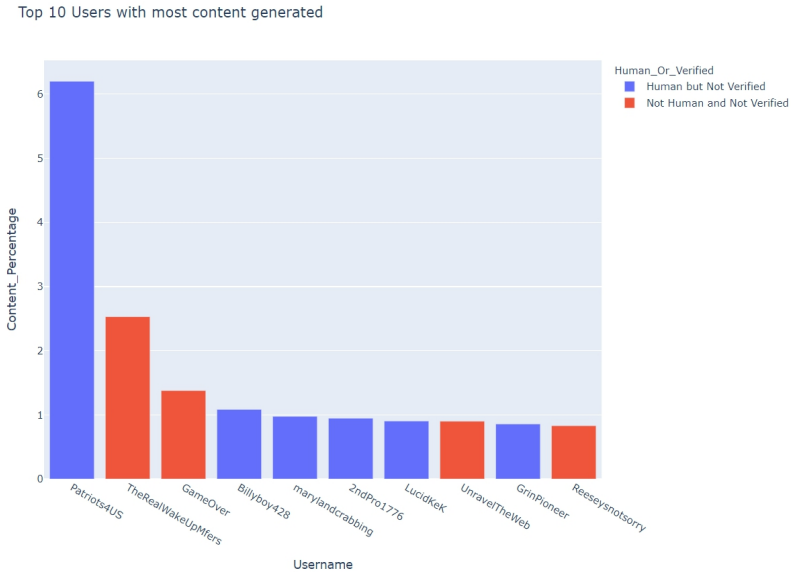
PARLER POSTS DATASET

Impressions	# of Impressions on the post.
Id	Id of the post.
Upvotes	# of Upvotes on the post.
At	Set of users mentioned in the post.
Comments	# of Comments on the post.
Reposts	# of Reposts of the post.
CreatedAt	Time of creation of post in 'YYYYMMDDHHMMSS' format
Body	Text within the post.
Creator	Id of the User who created the post.

PARLER USERS DATASET

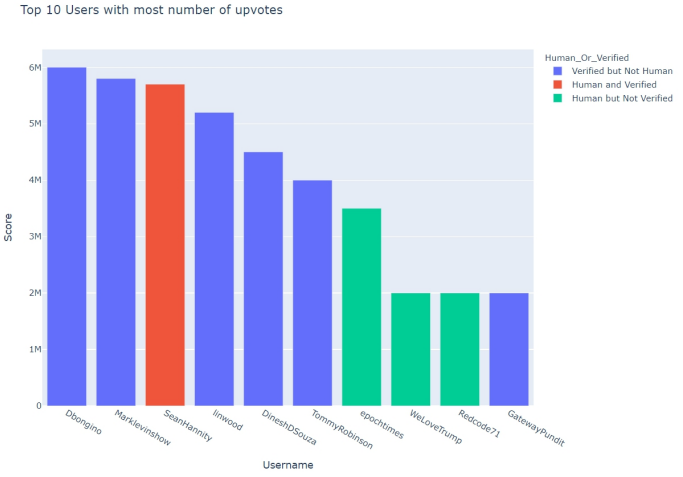
Name	Name as used by the user.
Username	Username of the user on Parler.
Score	Sum of upvotes of the user.
Id	User Id.
Bio	Bio of the user.
Joined	Date on which the user account was created (in 'YYYYMMDDHHMMSS' format).
Interactions	# of interactions of the user.
Human	Whether the user has been verified by Parler to be a human.
Verified	Whether the user is a prominent personality and has been verified by Parler.

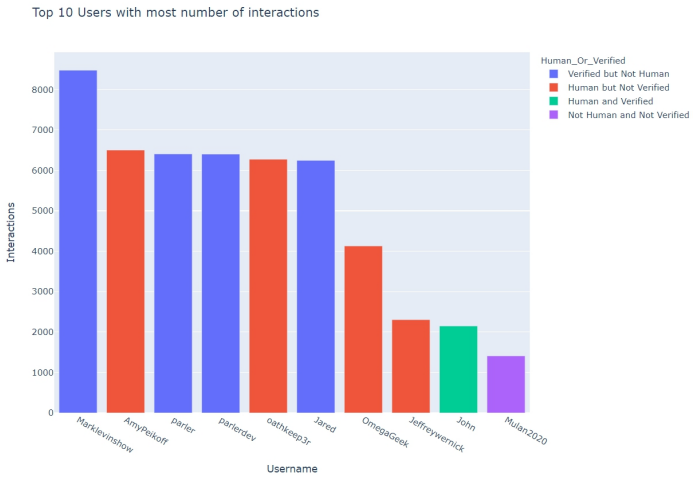
- Total Percentage of Content generated by the top 10 users is **16.60601304625043**.

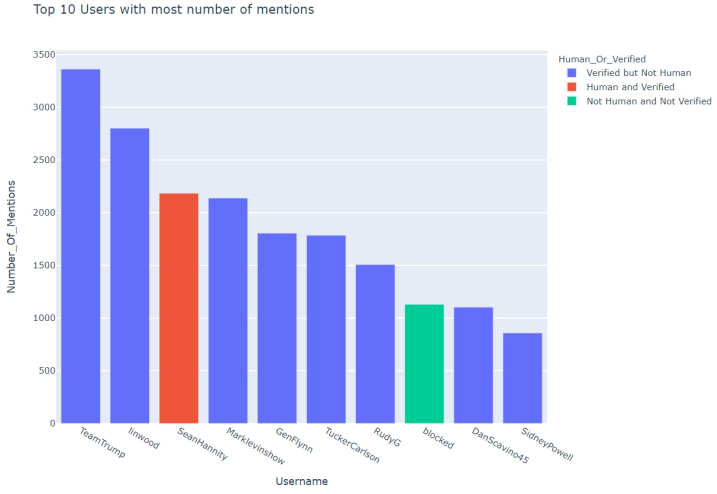


- Approximately 6% of the total content generated is contributed by the top user meaning this account is impacting a wide portion of the audience on parler. Another point to note is that it is not even a verified account.
- 6 users are **human but not verified(BLUE)** and 4 users are **neither human nor verified(RED)**.
- None of the top 10 users are verified user accounts and among them, 4 are bot accounts(Not Human). This means there is a high chance that the content shared by these accounts might not even be authentic.

TOP USERS BASED ON DIFFERENT CRITERIA





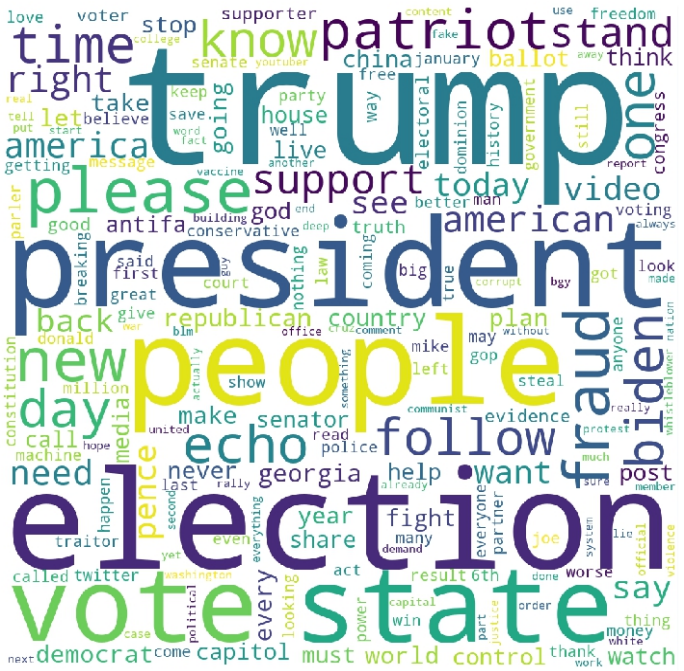


WORD CLOUDS

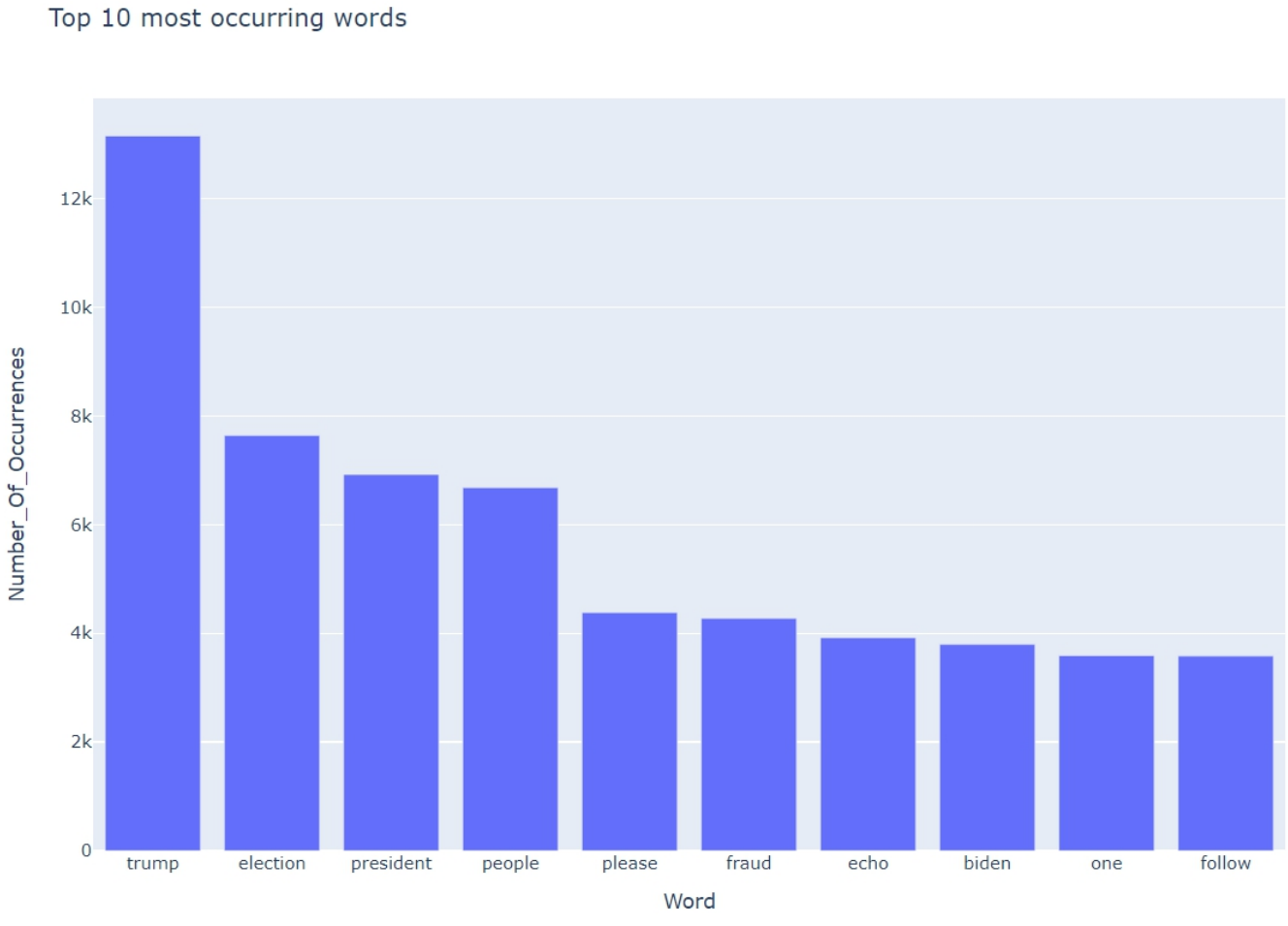
Bios of top 10 users with most number of interactions

All of the top 10 users are somehow related to parler, they either work for parler, promote parler in some way or they are verified non human accounts related to parler themselves. (Look at the words ‘parler’, ‘official’, ‘investor’, founder’, ’engineer’, statements’, ‘policy’, etc.)

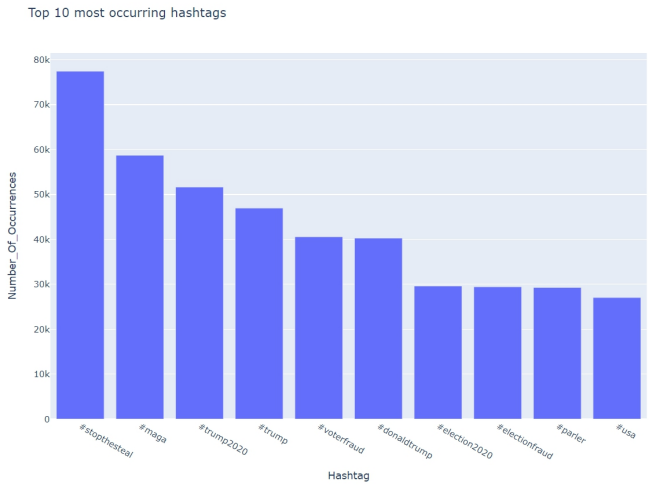
Major topics discussed in the posts

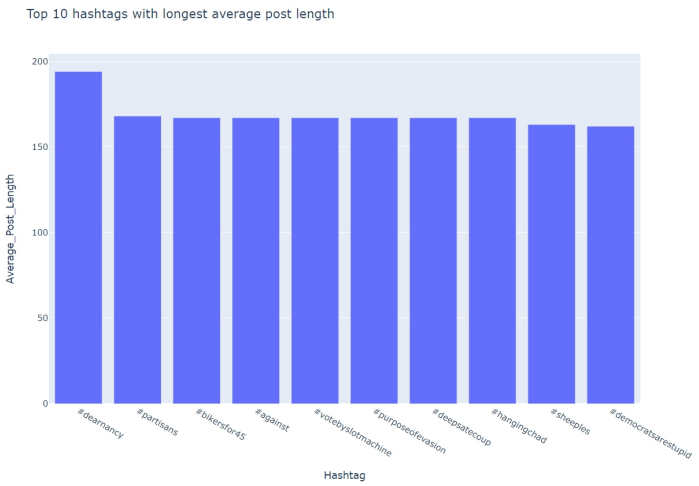


Most of the posts are about elections, people promoting for the leader they want to see elected as president (trump/biden), pleading for support, promoting patriotism, etc.



ANALYSIS OF HASHTAGS

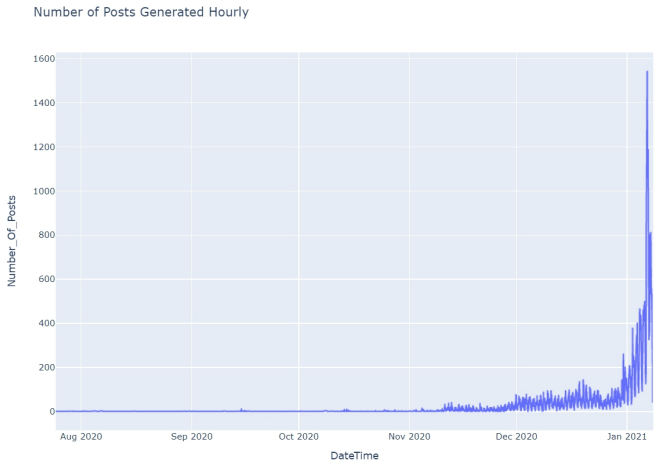




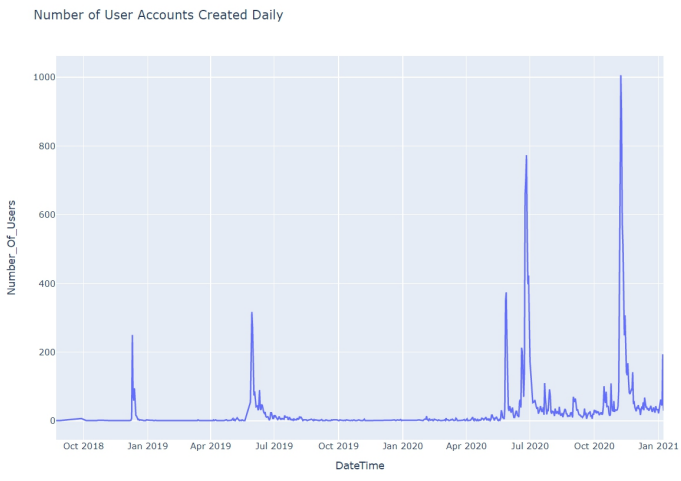
We can see from the above plots that all of the top 10 hashtags are related to 2020 presidential elections of US and most of them are related to trump supporting him in the elections.

TIME SERIES PLOTS TO STUDY GRANULAR PATTERNS

Content Generation on Parler



User Accounts creation on Parler



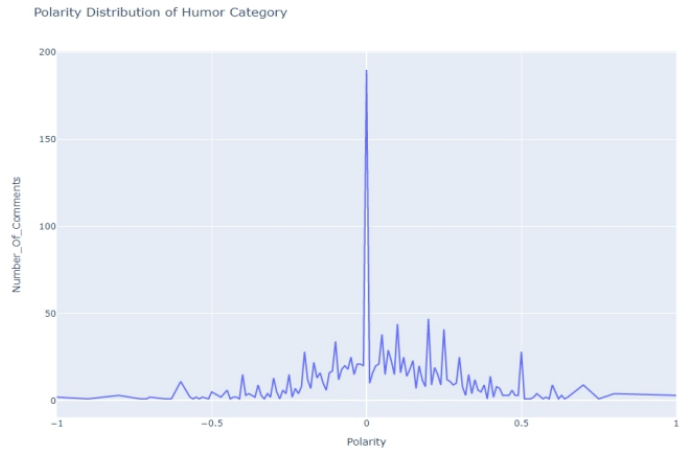
Number of posts created and number of users joined on parler increased exponentially near the end of 2020 which is the time when the US elections are concluded.

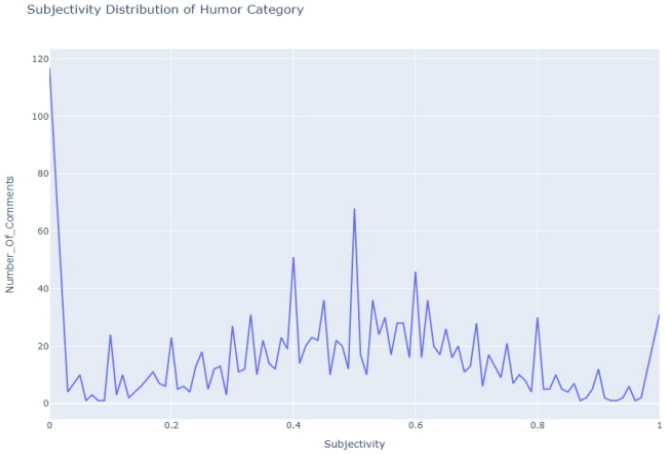
Reddit Data

The data is in the form a table consisting of two columns comments and their category (news/humor).

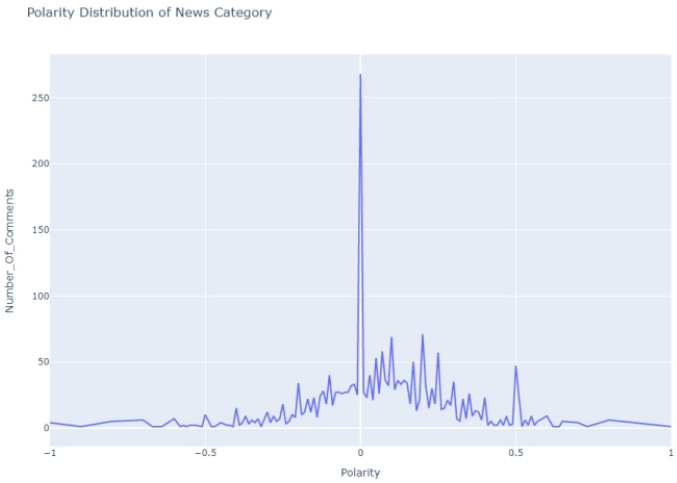
- Using Sentiment Analysis to extract the polarity and subjectivity of the comments:-

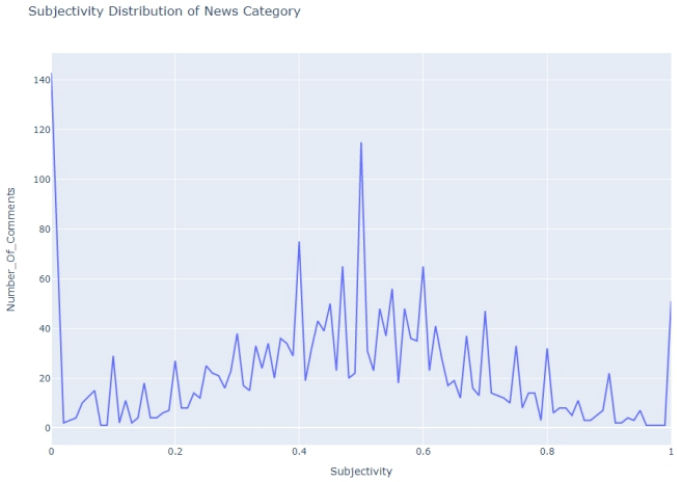
HUMOR CATEGORY





NEWS CATEGORY





Statistics:-

Humor Statistics

	polarity	subjectivity
mean	0.051587	0.471403
std	0.248748	0.242776

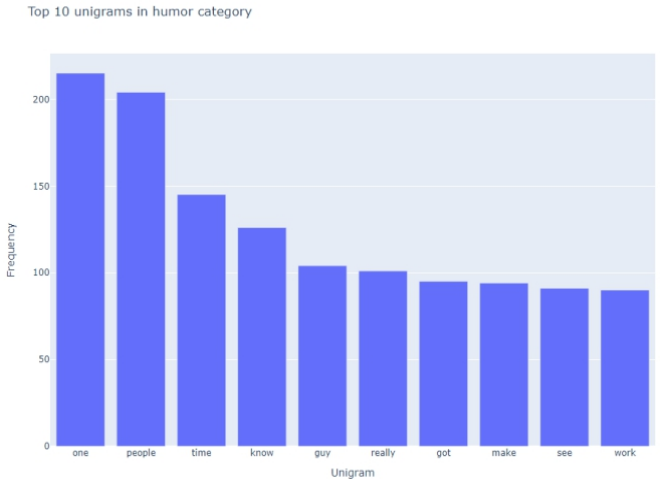
News Statistics

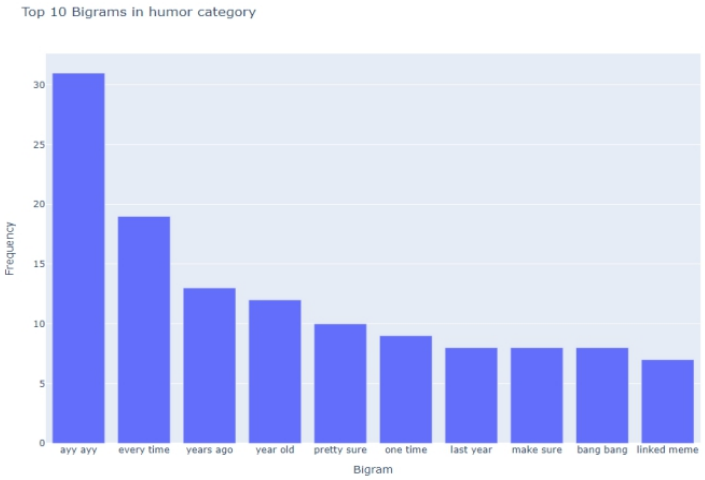
	polarity	subjectivity
mean	0.064049	0.470475
std	0.230468	0.232799

Most of comments whether they belong to Humor or News, have polarity close to **zero** meaning that they are more or less neutral in nature. Also, most of the comments whether they belong to Humor or News, have subjectivity values close to 0 and 0.5 which means the ones with value 0 are highly objective and the ones with value 0.5 are fairly subjective. We can also infer the same by looking at the mean and std statistics shown above. Mean polarity values for both Humor and News are close to 0 and mean subjectivity values for both of them are nearly 0.5. Also the standard deviation of the values are close to 0.2 which means they are not that much deviated from the mean value and are clustered around mean value.

- Common Words and phrases used in comments analyzed using unigrams and bigrams:-

HUMOR CATEGORY

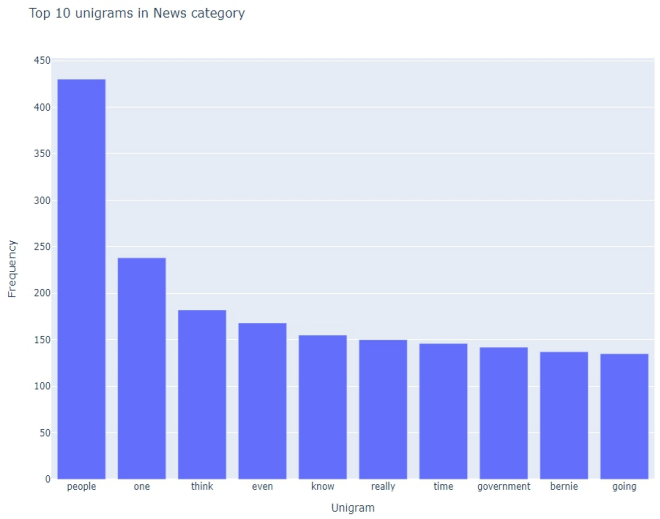


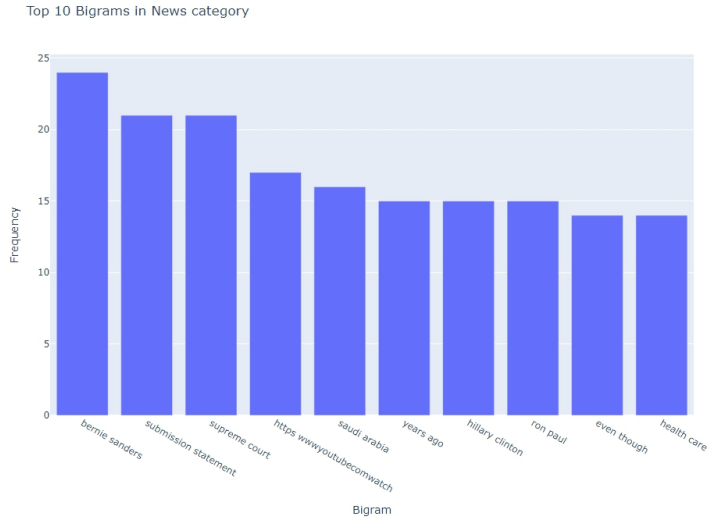


We see general words used in daily conversations, in the top 10 unigrams of humor category, with 'one' being the most commonly used (frequency = 215), then 'people' (frequency = 204) and so on.

The most common bi-gram in humor category is 'ayy ayy' (frequency = 31) which kind of a slang for greeting in informal language that is why, common in humorous comments. Other bi-grams similar to this, that are among the top 10 bi-grams are 'bang bang' (frequency = 8) and 'linked meme' (frequency = 7). Rest of them are just common bi-grams used in normal sentences.

NEWS CATEGORY





We see words like 'government', 'people' and 'bernie' (for **'bernie sanders'**, the **United States Senator**), 'going' and 'think' in the top 10 unigrams of News category. The above words indeed are commonly used in news headlines and statements.

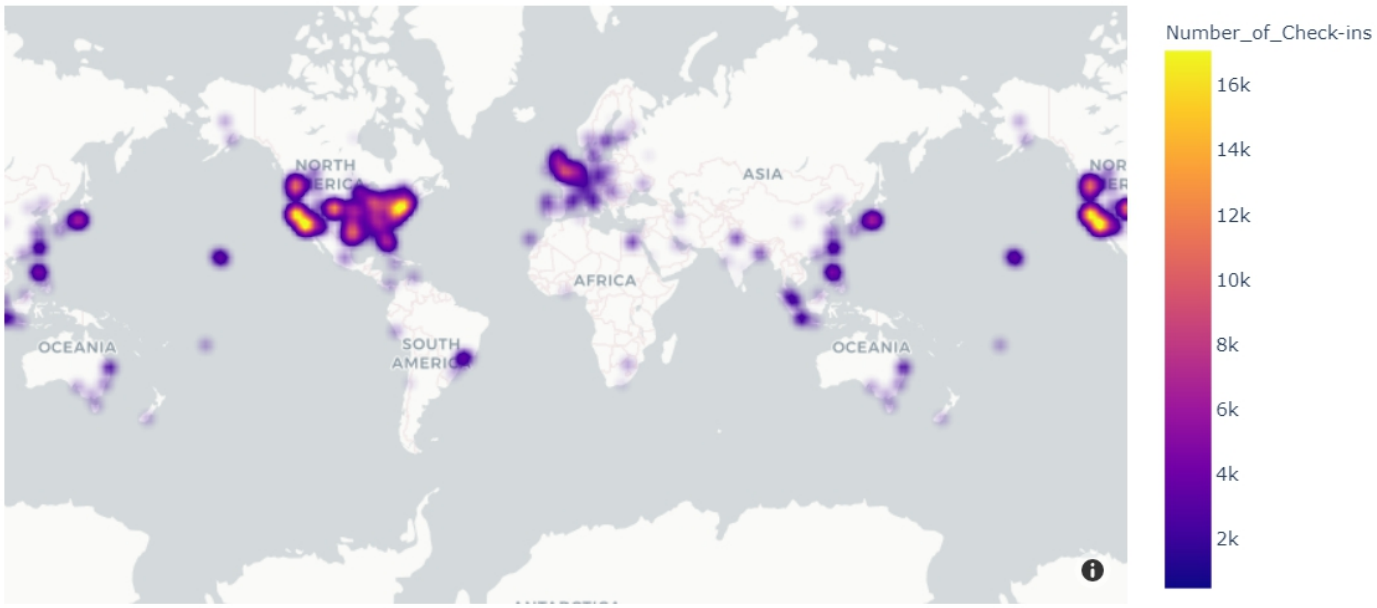
'bernie sanders' is the most commonly used bigram in the comments of News category meaning most of the comments are centered around him and others like **'hillary clinton', 'ron paul', 'supreme court', 'saudi arabia'** etc. Also, we see **'https wwwyoutubecomwatch'** in the top 10 bigrams which means that a youtube link has also been commonly shared among these comments.

Brightkite Data

The data is in the form of a table with columns and data types as shown in the example below:-

[user]	[check-in time]	[latitude]	[longitude]	[location id]
58186	2008-12-03T21:09:14Z	39.633321	-105.317215	ee8b88dea22411
58186	2008-11-30T22:30:12Z	39.633321	-105.317215	ee8b88dea22411
58186	2008-11-28T17:55:04Z	-13.158333	-72.531389	e6e86be2a22411
58186	2008-11-26T17:08:25Z	39.633321	-105.317215	ee8b88dea22411
58187	2008-08-14T21:23:55Z	41.257924	-95.938081	4c2af967eb5df8
58187	2008-08-14T07:09:38Z	41.257924	-95.938081	4c2af967eb5df8
58187	2008-08-14T07:08:59Z	41.295474	-95.999814	f3bb9560a2532e
58187	2008-08-14T06:54:21Z	41.295474	-95.999814	f3bb9560a2532e
58188	2010-04-06T06:45:19Z	46.521389	14.854444	ddaa40aaa22411
58188	2008-12-30T15:30:08Z	46.522621	14.849618	58e12bc0d67e11
58189	2009-04-08T07:36:46Z	46.554722	15.646667	ddaf9c4ea22411
58190	2009-04-08T07:01:28Z	46.421389	15.869722	dd793f96a22411

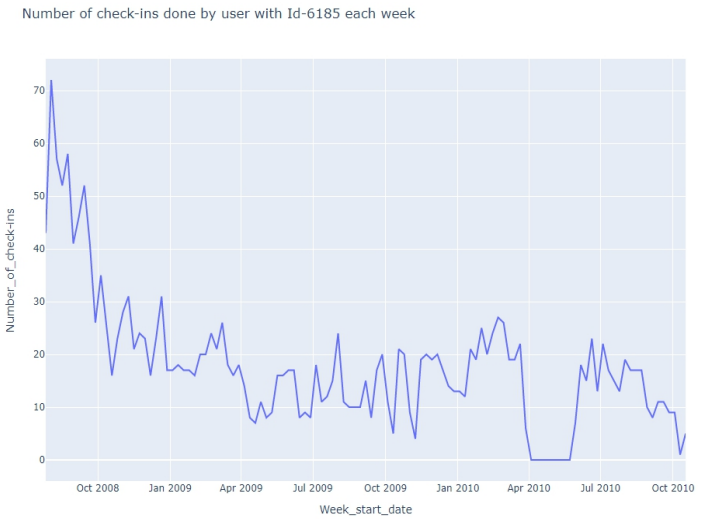
GEOGRAHPIC HEATMAP FOR ANALYSIS OF CHECK-IN LOCATIONS



We can clearly see that the concentration of check-ins in middle and southern part of North America, is the maximum, then in the western Europe. There are very few check-ins in northern part of North America, Africa, Oceania(Australia) and Northern Asia.There are no check-ins in Antarctica. This also shows that most of the check-ins are near the Tropic of Cancer.

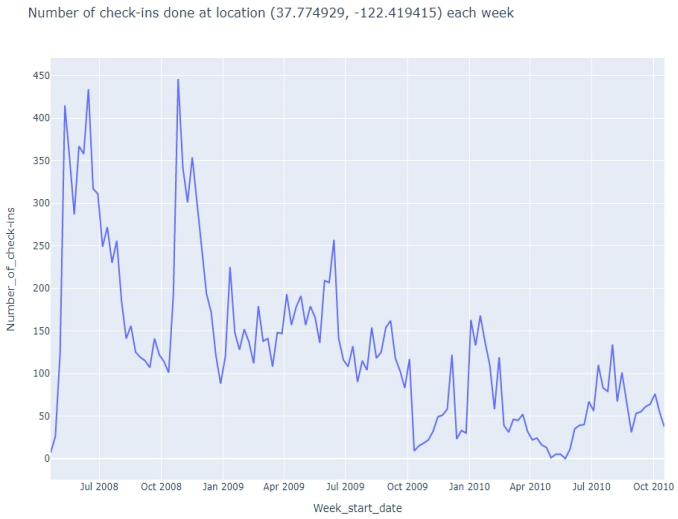
TEMPORAL ANALYSIS FROM TIME-SERIES PLOTS

User with most number of check-ins



We can see peaks in the weeks of July 2008, July 2009 and July 2010 and also near Jan 2009 and Jan 2010. Also, We see valleys in the weeks of April 2009 and April 2010. This clearly shows that people check-in more in July and less in April.

Location with most number of check-ins



We see a similar trend in the above graph as well, peaks in the weeks of July 2008, Jan 2009, July 2009, Jan 2010 and July 2010 and valleys in the weeks of Oct 2008, Oct 2009 and Oct 2010 i.e. people check-in more in July and Jan and less in Oct at that location.

Twitter Data

Tweets from various handles were collected for analysis using Tweepy wrapper for Twitter API.

- Tweets from @DCPNorthDelhi were collected to detect potential PIIs from their text data.

PIIs Detected

PII Type	Number of PIIs
emails	0
phone-no	6
pincode	11
aadhar-no	0
vehicle-no	8

Examples:-

@DelhiPolice @CPDelhi Dear Sir, My car White Suzuki Vitara Breeza has been stolen from Shakti Nagar, Delhi-110007 on yesterday 14/02/2021 @ 8 am. In this regards, when i visited my nearest PS Roop Nagar, they told me that we cant register your FIR as DP website is not working.

Location: rui mandi sadar bazaar delhi 110006
Social distancing, covid norms, crowding, encroachment on streets
Authorities are sleeping over it
@drharshvardhan @CMODelhi @msisodia @SatyendarJain @LtGovDelhi @DelhiPolice @DcpNorthDelhi @DCPNWestDelhi @CPDelhi @dtptraffic https://t.co/6jMhy7mw6d

@LtGovDelhi @DcpNorthDelhi @CPDelhi @PMOIndia @AmitShah
Rem 04 :Despite of all efforts made including meetings with SHO Burari, ACP Civil lines, ADCP, DCP/North & Jt CP Cen, nothing has so far been done wrt FIR No.1004/14.May resolve it now as it is already delayed. 8178082002.

@CPDelhi save my life subhash chand
9999877060 https://t.co/PHdJtAxagi

Delhi police please help me 8700034868 https://t.co/6afOVQwYIk

VIOLET THE ZEBRA CROSSING RULE
VEHICLE NO IS
DL9CAK3298
DL9COS7802 https://t.co/ObfKiAbfO1

PIIs Detected in media

REVENUE Department

GOVERNMENT OF NCT OF DELHI

ON COVID-19 DUTY

Name: Akshit Gulati

Type of Service:
General Provision stores

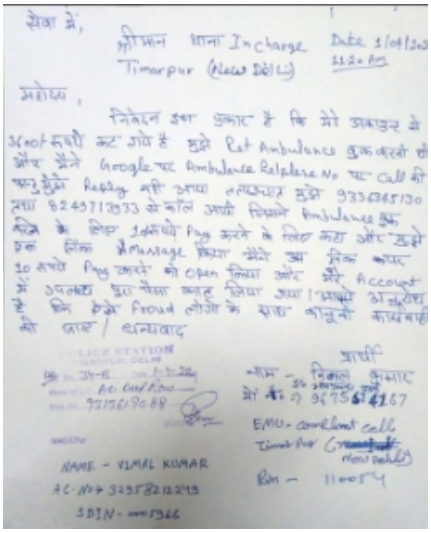
Office/Place of Engagement:
C-1/85,phase-4,

Date: From 10-04-20 to 14-04-20

Time: From 05 AM to 08 PM

ePass ID:
9SF6NZV

DM - South
Issuing Authority



- Tweets having "**#CovidIndia**" were collected for analysis.

Hashtags frequently used along with #CovidIndia