Social Media and Network Analysis COSC2671

Assignment 2

Analysis on Twitter data related to Marijuana in USA

Team Members – Srujan Basavaraj(s3856311@student.rmit.edu.au)

Sunny Hirani(s3856019@student.rmit.edu.au)

Nitish Agrawal (S3881575 @student.rmit.edu.au)

Contents

Introduction..... 4

Data Collection..... 5

Data Pre-processing and Cleaning..... 8

Initial Insights
Analysis 1.....8
Tweet location Analysis 9
Analysis on data collected and state laws 10
Topic analysis 11
Twitter data flow analysis.....11
Hashtag analysis.....12
Tweet user mention analysis..... 13
Wordcloud 15
Network analysis 16

Analysis 2..... 17
Tweet location analysis 17
Analysis on data collected and state laws..... 18
Twitter dataflow analysis.....19
Hashtag analysis.....20
Twitter user mentions analysis.....21
Twitter data sentiment analysis.....23
Topic analysis.....24
Network analysis.....25
Analysis of before and after Biden's Announcement.....26
Tweet mention analysis.....26
Hashtag analysis(Before vs After).....28
Network analysis(Before vs After).....29
Conclusion from Analysis 2.....30
Intl impact.....30

Analysis 3..... 30
Hashtag analysis 30
Mention Analysis..... 31
Sentiment Analysis 32
Topic Analysis 34
Network Analysis 35

Introduction -

Social media has become a major part of our life in this day-to-day world and Twitter is a micro-blogging site that has lots of information to study analyze and capture the data from which we can find out inferences about the trends in the current world, we could analyze the moment we can analyze the changes happening in the world and, we can analyze what's gonna happen from these moments.

We were very much interested in this topic because it's been a hot topic for many of these years like legalizing marijuana making it illegal eyes or how this can be used for medical marijuana purposes.

there was a major change that came to this moment of legalizing marijuana for medical usage with the announcement coming from Mr job right and the United States president announcing that they will be decriminalizing the possession of marijuana for a small quantity because of many of the people in us. This change majorly came from the president because they analyze trends of people of different colours getting arrested or getting into legal proceedings because of carrying or possessing a small amount of marijuana showing this announcement came as a huge wave and we could see there are many tweets they were many reports and news articles which came out on the same day so we try to analyze the trend on the same and we are also trying to analyze how the moment of legalizing marijuana happened in us for last 10 years with the help of influential users tweets and we are also trying to analyze how the tweets were pro related to medical marijuana usage.

The USA acts as an epicentre of the world because it's one of the major developed nations and changes happening in us can be seen as a reflection in many of the other countries, so we are trying to analyze the data of marijuana-related tweets and trying to show the reach on Twitter related to marijuana have been carried out in last 6 months and in some of our analysis we have even taken data for last 10 years.

.

Data Collection –

The goal was to obtain information about a specific topic trending in the USA, so getting tweets through Twitter was initially seen as a first part to solve this problem.

Initially, we used Twitter API with elevated access, to fetch the data from Twitter. But as elevated access has some restrictions on how we collect the data like geo-location, user-data, user profile data, user, likes, or followers, no of retweets etc.

Later, we used **Tweepy** library to access the Twitter API using Python. But the issue for us in using this library is the number of tweets that it can fetch and the geo location which we need, but it needs higher access from twitter to fetch the user's geo-location. Number of tweets can be bypassed by using paginator method but still we had geo-location issue. So for that, in later stages, we stopped using the Tweepy.

Later, we started using **SNScrape** which is a scrapper for social networking sites. This library is so resourceful as it can fetch data not only from Twitter, but also from Instagram, Facebook, Telegram, Reddit, etc. This library can scrape user-profiles, geo-location, hashtags, no of followers, no of likes that a tweet has, similarly many other important items that we are needed.

So we created a word dictionary on marijuana, cannabis word terms with the help of Google. We collected words related to recreational and/or medicinal use of marijuana.

Below is our word dictionary used for this analysis –

wordlist=['medicinalcannabis','medicinalmarijuana','endocannabinoids','Cannabigerols','Cannabichromenes','Cannabinol','Cannabinodiol','CB1/CB2','sinsemilla','THC','cannabis','marijuana','bhang','dope', 'hashish', 'hemp', 'herb', 'tea', 'ganja', 'hash', 'joint', 'reefer', 'roach', 'weed', 'doobie', 'maryjane', 'pot', 'grass', 'bammer', 'bomb', 'blunt', 'dope','CBD']

Analysis 1 data collection -

So Biden announced the decriminalizing act on Oct 6th on Twitter, so we were interested in fetching the data, before and after the announcement so we collected from 15th June to 15th Oct 2022. We fetched around 10k of tweets related to both recreational and medicinal marijuana with the help of above wordlist and below query.

"Wordlist dates = ' lang:en until:2022-10-15 since:2022-06-15 -filter:replies min_faves:2'"

So, we were interested in tweets originating from USA, so we had put the condition that fetch only tweets originating from US mainland.

"if tweet.place != None and tweet.place.countryCode == 'US'"

So once, we fetched the data and analysed it, we could see that there were many bots which were spamming which would make us difficult for us to analyze the data. So, we put another condition to filter out the bots that were spamming the tweets.

"if tweet.place != None and tweet.place.countryCode == 'US' and tweet.user.followersCount > 1 and tweet.user.friendsCount > 1"

So, the condition was should have more than 1 follower and the friend count should also be greater than 1. So this will filter out the bots' tweets for us.

So with the help of SNScrape, we fetched the data related to:

Date, Username, Tweet, Hashtag used, mentioned user in the tweet, Retweets count, Followers count, Location, Keyword used(Wordlist).

“tweets.append([tweet.date,tweet.user.username,tweet.content,tweet.hashtags,tweet.user.location,tweet.mentionedUsers,word,tweet.replyCount,tweet.likeCount,tweet.retweetCount,tweet.user.followersCount])”

Analysis 2 data collection -

As we are doing analysis on medical marijuana related tweets, we collected data related to medical marijuana using wordlist containing only medical terms related to marijuana.

wordlist=['medicinalcannabis','medicinalmarijuana','endocannabinoids','Cannabigerols','Cannabichromenes','Cannabinol','Cannabinodiol','CB1/CB2','sinsemilla','THC']

Similarly, we used these medicinal words for data collection as above.

Analysis 3 data collection -

We collected influential person’s data for last 10 years, using a bit different method but with the help of SNScrape.

namelist = ['BernieSanders','morgan_freeman','SnoopDogg','WoodyHarrelson']

We selected this three influential people because this people are quite active on twitter and are celebrities having major fan-following base. And are quite in marijuana related movements.

Bernie Sanders is the US Senator; he was also the US representative for state at-large congressional district from 1991-2007. He was the longest serving independent in US congressional in history.

SnoopDog is an American Rapper and has sold over 35 million albums worldwide and also is a very avid user of marijuana.

Morgan Freeman is an actor, narrator, director who has won many awards like Golden Globe, Academy award etc.

So all these 3 users are the US citizens, so the wordlist which we used earlier. So we collected around 10k tweets for last 10 years with the help of below query. We filtered out replies and retweets and collected only tweets tweeted in English language. So we collected data similarly as we had done earlier.

"**username wordlist' lang:en until:2022-10-15 since:2012-09-15 -filter:replies -is:retweet'**"

Data cleaning and pre-processing -

After collecting the data, we started with the cleaning process. And for that we used Numpy, Pandas libraries.

The process that we did in data cleaning on tweets was removing the duplicate tweets, removing empty string, removing the punctuations, special characters, and emojis. And in tweets there were links containing http, https or any raw data which were also removed.

As we fetched the tweets with the keywords, there were many duplicate tweets in the dataframe, so we removed those duplicates from the dataset. Afterwards, we removed tweets related to few keywords which were quite misleading for future processing.

In the dataset, columns like hashtags, mentioned user had NaN values, so we changed those values to no hashtags and no mentions respectively.

In the location column, we had NaN values, so we set them to Washington, DC which is the capital of the USA.

We converted the data in the data column to specific format, which is given below,

"**format= '%Y-%m-%d %H:%M:%S'**"

Cleaning on tweets -

We converted all the tweets to lower cases as it becomes easier for processing. We removed all the hashtag and mentions from tweets and removed the white space, punctuations, emojis, number, links. With the help of NLTK stopword corpus, we removed all the stopwords from tweets.

We lemmatized the tweets and created the new columns.

We did all the above-mentioned processes for all the three dataframes which we collected above.

Insights –

Analysis 1 -

We chose tweets from the USA during the previous four months that included marijuana and cannabis. After analysing the data for the previous four months, we will analyse the data for tweets sent both before and after Joe Biden's statement. On each of the three cases, we have conducted hash tag analysis, location analysis, mentioned users' analysis, subject analysis, and network analysis.

There are three sections to this analysis:

1. Analysis of all the information gathered.

2. Examining data gathered before to Biden's announcement.

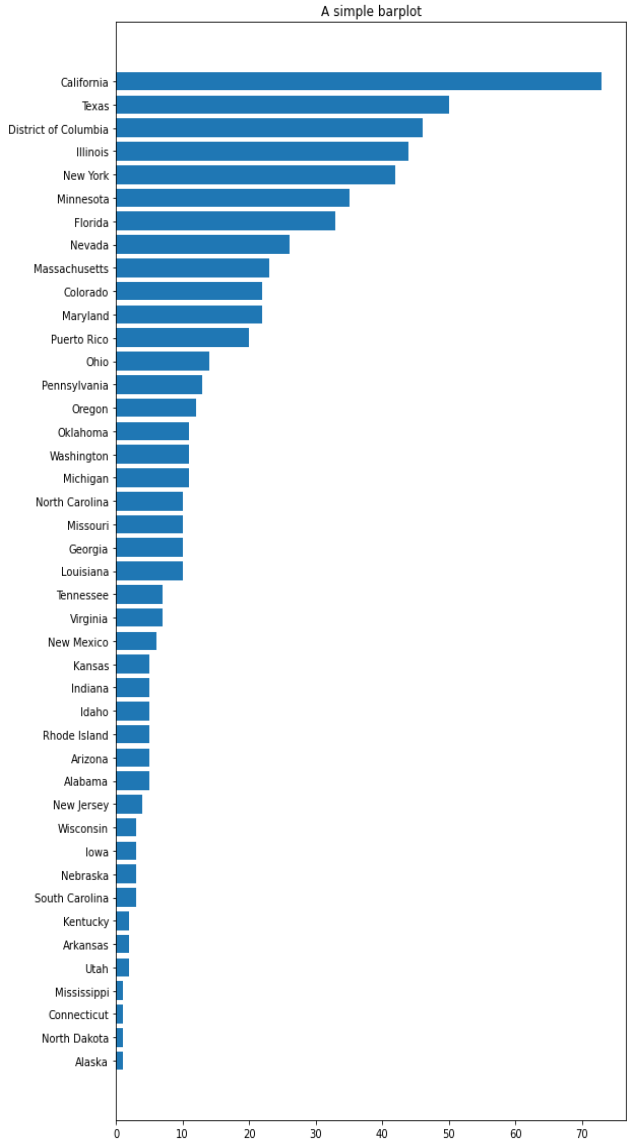
3. Data analysis on information gathered after Biden's announcement.

Tweet Location analysis:

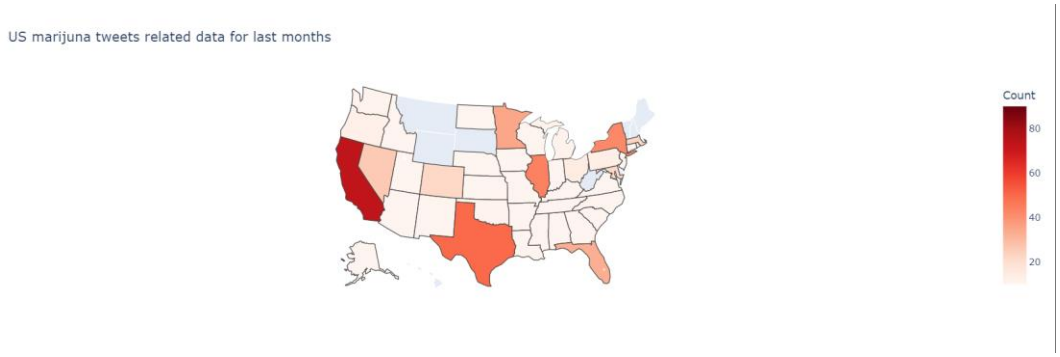
We gathered the information from American twitter users. Therefore, utilising the location of the tweets' origin, we can analyse how many tweets are coming from a specific area or state, which is useful for understanding the attitudes of individuals who use Twitter to tweet about medical marijuana or for any government agencies.

We can therefore observe from the bar plot, geopandas plot, and ploty plot that the most popular states for tweets were California, New York, Columbia, Illinois, NY, Minnesota, Florida.

The use of marijuana for medical purposes is now allowed in many US jurisdictions, and some have also legalised its use for other purposes.



Analysis on all the data collected :



States laws on medicinal marijuana:

State	Alzheimer's disease	Autism	Cancer	Crohn's disease	Epilepsy	Glaucoma	HIV/AIDS	Multiple sclerosis	Parkinson's disease	PTSD	Seizures	Wasting syndrome
Alaska[213]	No	No	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Arizona[214][215]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Arkansas[216]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
California[217]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colorado[218][219]	No	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Connecticut[220]	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Delaware[221]	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
District of Columbia[222]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Florida[223]	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Guam[224]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hawaii[225]	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Illinois[226]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Louisiana[227]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Maine[228]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Maryland[229]	No	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Massachusetts[230]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Michigan[231]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Minnesota[232]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Mississippi[233][234]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Missouri[235]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Montana[236]	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Nevada[237][238]	No	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
New Hampshire[239][240]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
New Jersey[241]	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No
New Mexico[242][243]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
New York[244]	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
North Dakota[245]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
N. Mariana Islands[246]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ohio[247][248]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Oklahoma[249]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Oregon[250][251]	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Pennsylvania[252]	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Puerto Rico[253][254]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rhode Island[255]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
U.S. Virgin Islands[256]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Utah[257][258]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Vermont[259]	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Washington[260][261]	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes

Source(Wikipedia)

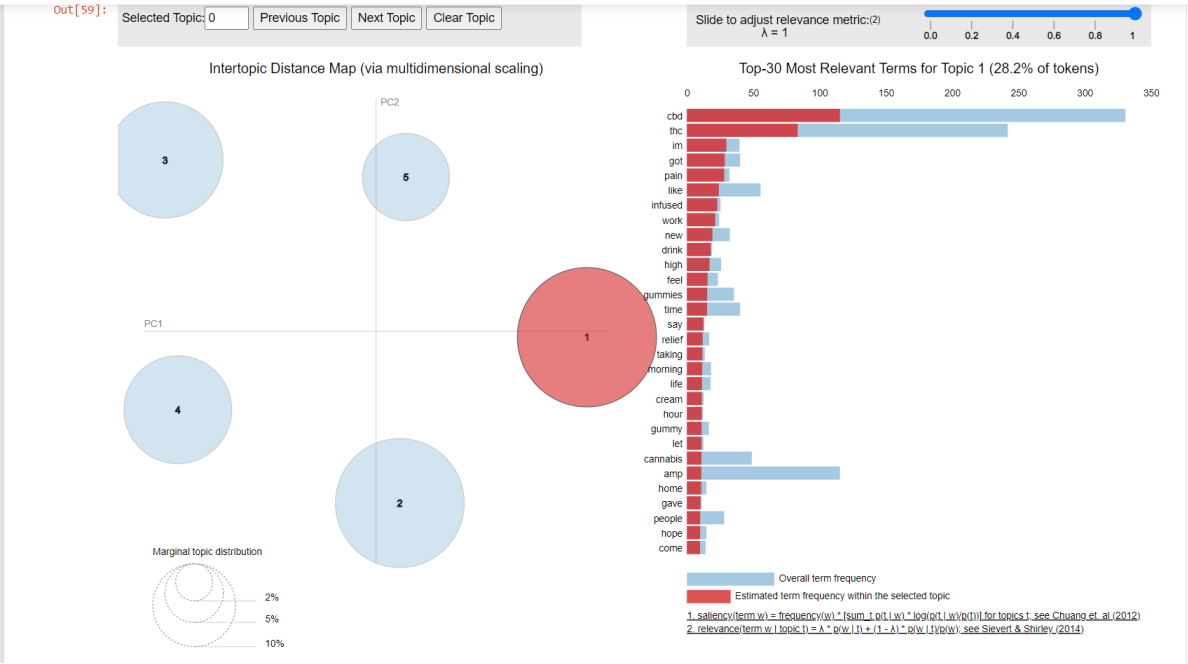
Above image also shows that which states allows the usage of medicinal marijuana and 37 agreed on the usage of it to some extend, but as the population of every state is huge they have some restrictions over it as well.

Topic Analysis:

The features of the tweets were extracted using the count vectorizer and the LDA were fed the features to get the topic related to the marijuana.

```
Topic 0:
thc cbd amp good smoke game coffee giving week high like age old tonight hope
Topic 1:
cbd free dog firework steel treat pt code pain weekend happy pet seltzer cat promo
Topic 2:
cbd thc amp day like today product great cannabis sleep weed night got time help
Topic 3:
thc amp cbd infused stereo miss live dont mark calendar available let app love use
Topic 4:
cbd thc new ve im gummies oil drink help gave store win going right anxiety
```

Above image, shows the different topics but related to marijuana.



The above diagram shows some of the most frequent unigrams occurring in the topics where people are involving or talking about cbd, or thc, got, or pan.

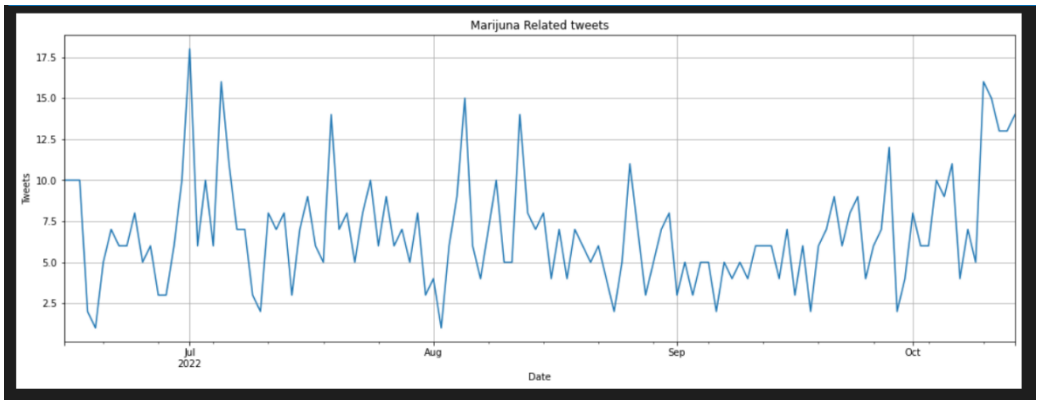
Twitter data flow analysis(Part of sentiment analysis):

If we examine the trends in the twitter data that we have amassed over time, we may determine whether there has been, is being, or will be a significant movement, change, or announcement.

So, after aggregating the data for daily use, we created a time series graph using the matplot toolkit. The graph shows that there was a significant influx of tweets on marijuana at the beginning of

October. When we analysed the data, we discovered that the spike occurred on October 6, the day Joe Biden unveiled the Decriminalization Law.

As a result, on that day, more people responded and tweeted about the subject.



Hash tag analysis:

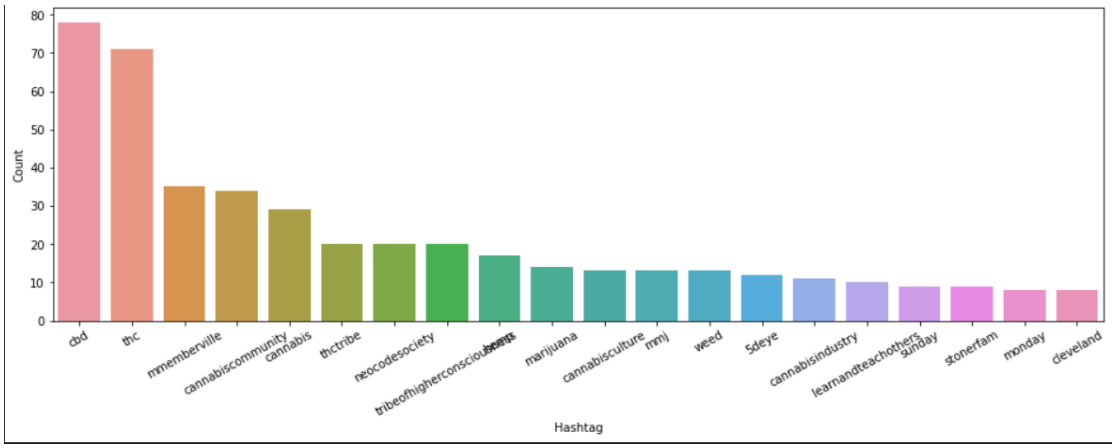
We will talk about and examine each hashtag that individuals used in their tweets. The use of hash tag analysis greatly aids in determining the trend that users want their tweets to endorse. Twitter's primary tool is the hashtag, which is used to collect the words of all users and to start a huge trend on the social media platform.

So, we took out top 10 hashtags which were used by the users during the data collection period we can see the major hash tags are cbd, mmemberville, thc, cannibiscommunity, cannabis and thctribe. We can see that all these hashtags belong to pro cannabis movement.

Few insights on hash tags:

Mmemberville: this is run by a Sibusiso Sithole who a musician and tweets on these belong to virtual stoner(cannabis smokers) community.

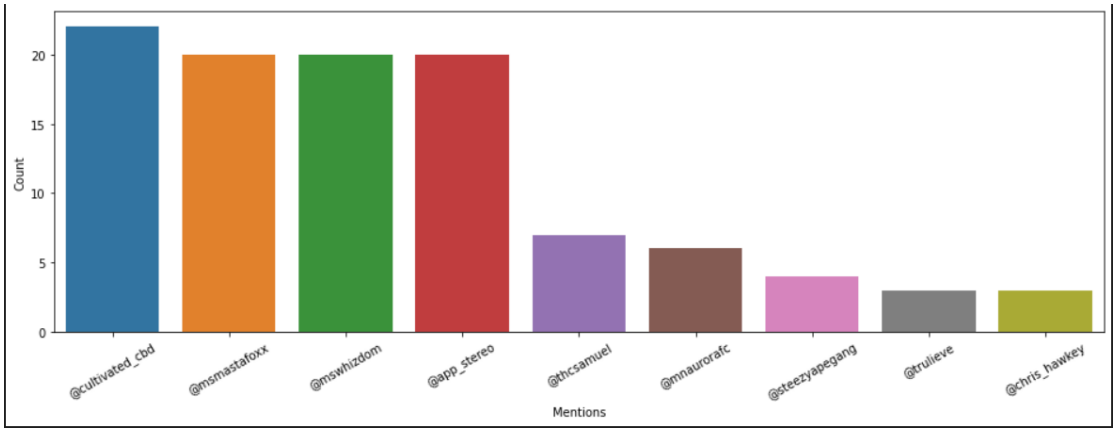
Cannibiscommunity: as the name says these are tweets belong to people who intake marijuana and discuss on the topic of marijuana.



Tweet User mentions analysis:

We can determine the type of communication taking place, analyse the subjects they are debating, and determine the significance of a specific tagged person by looking at who the Twitter users are tagging.

So, we plotted a graph for 10 most tagged user for the data collected and we can see the same below



@Cultivated_cbd,@msmastafox,@mswhizdom,@app_stereo,@thcsameul,@steezyapegang were few prominent users who were tagged in Cannabis related tweets from the bar graph.

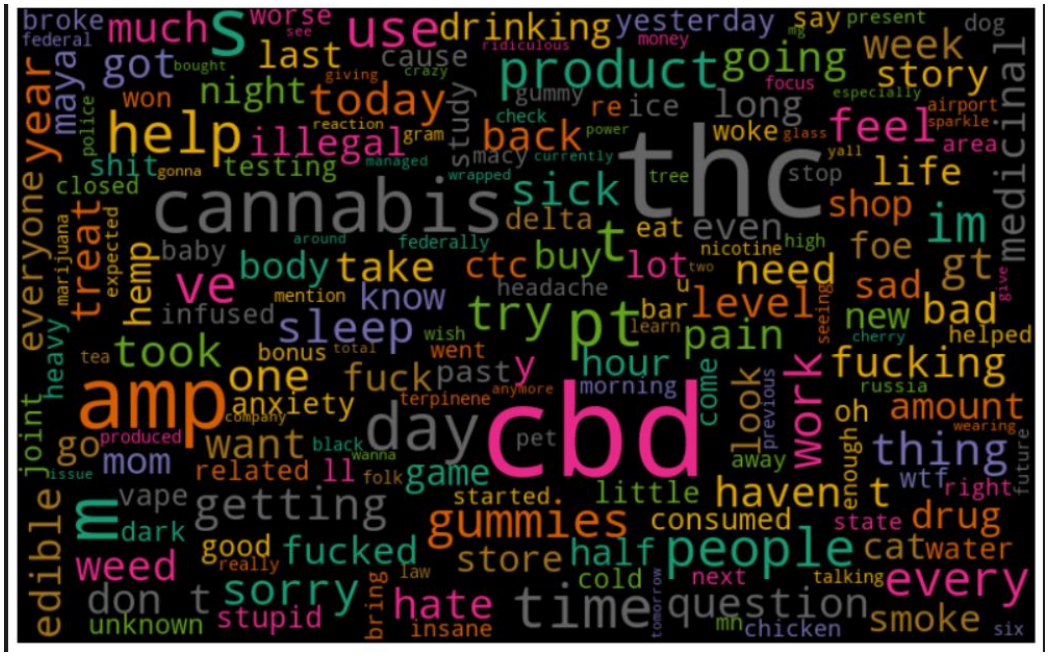
@Cultivated_cbd: is organic and vegan cbd sellers

@msmastafox: spreads message about business, mental health, spiritual advisor and also talks about cannabis and edible products of cannabis.

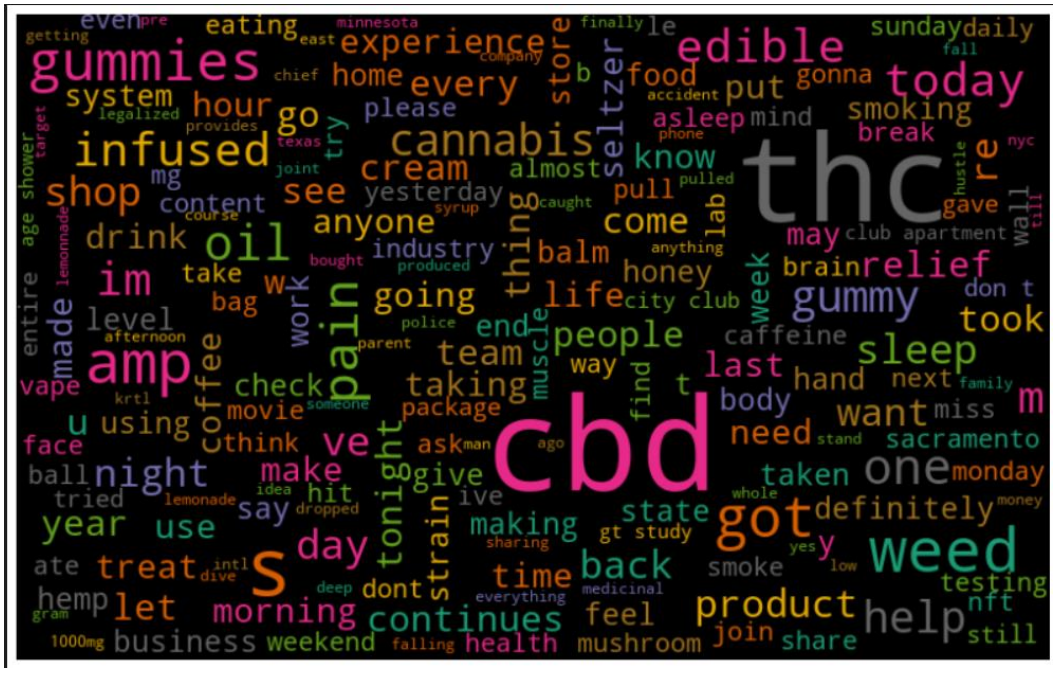
Wordcloud(part of sentiment analysis) –

Below are the wordcloud, for positive, negative, and neutral words in terms of sentiments expressed or shown by the people over Twitter.

As you can see from above image, that the positive sentiments in words are shown pretty highlighted like love, people, body, amazing, life,happy, etc.

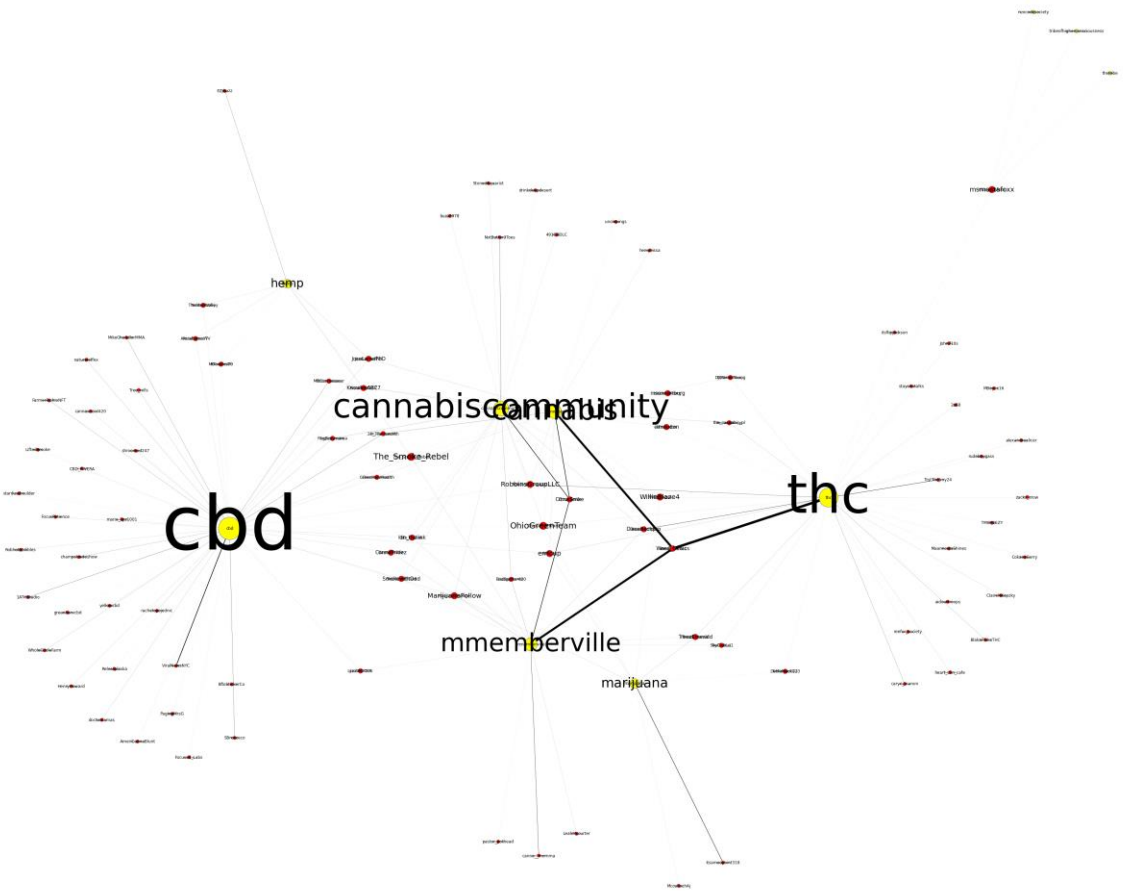


As you can see from above image, that it shows some negative sentiments in words like fuck, fucking, illegal, shit, drinking, bad, etc.



From above image, we can see that now the responses or the sentiments shown in the image are quite neutral like the words here are day, time, product, thing, come, cream, yesterday, etc.

Network Analysis –



Above image, shows the community of the various medicinal marijuana or mentions. It can be seen that cbd has the most nodes and is largely followed while marijuana in case of medicinal marijuana is followed lesser in comparison.

Analysis 2 -

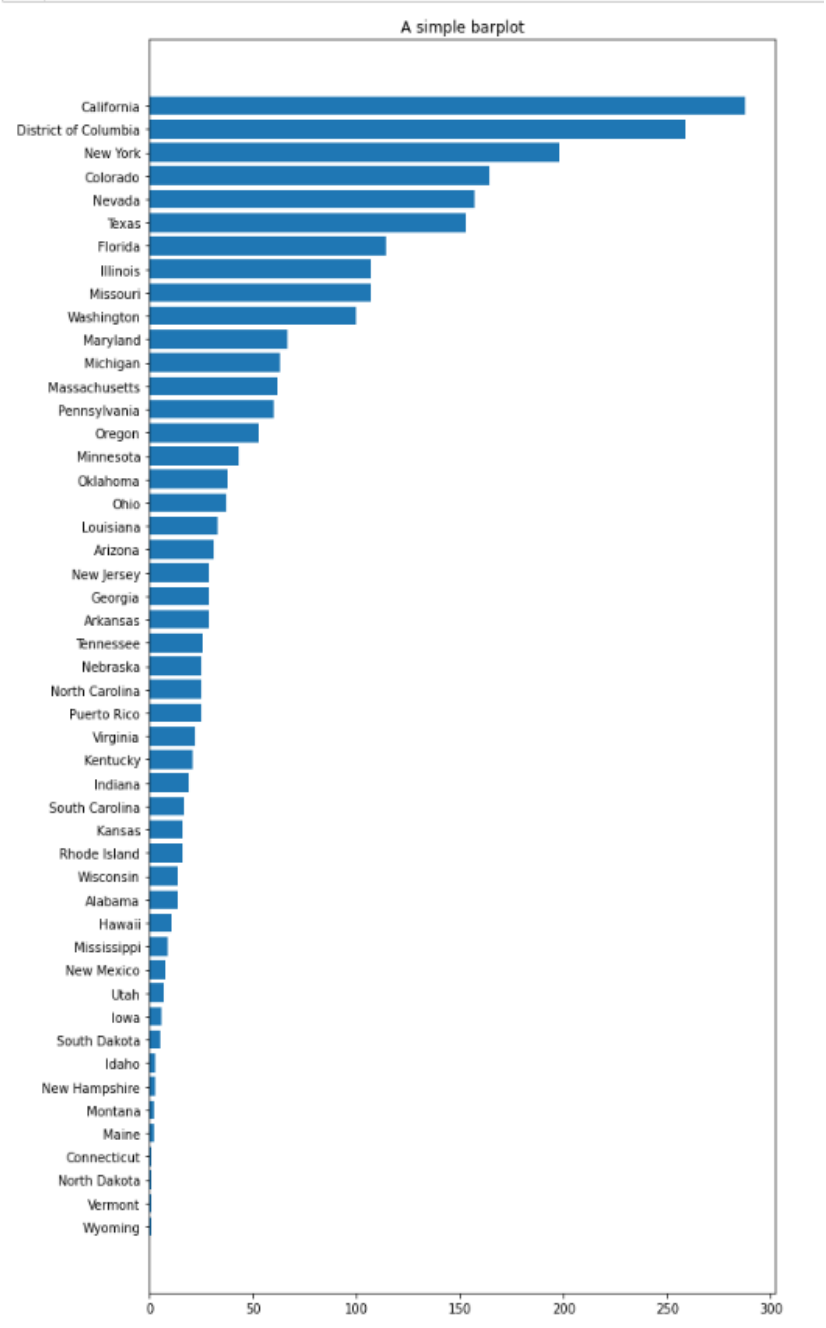
We selected tweets for last 4 months related to marijuana and cannabis originating from USA. And we will be analysing the data for last 4 months then we will be analysing the data for tweets generated before the Joe Biden announcement and after the announcement. We have done hash tag analysis , location analysis, mentioned users' analysis, topic analysis and network analysis on all three scenarios.

This analysis will be divided into three parts :

1. Analysis on all the data collected.
2. Analysis for data which was collected before Biden's announcement.
3. Analysis on data which was collected after Biden's announcement.

Tweet Location analysis:

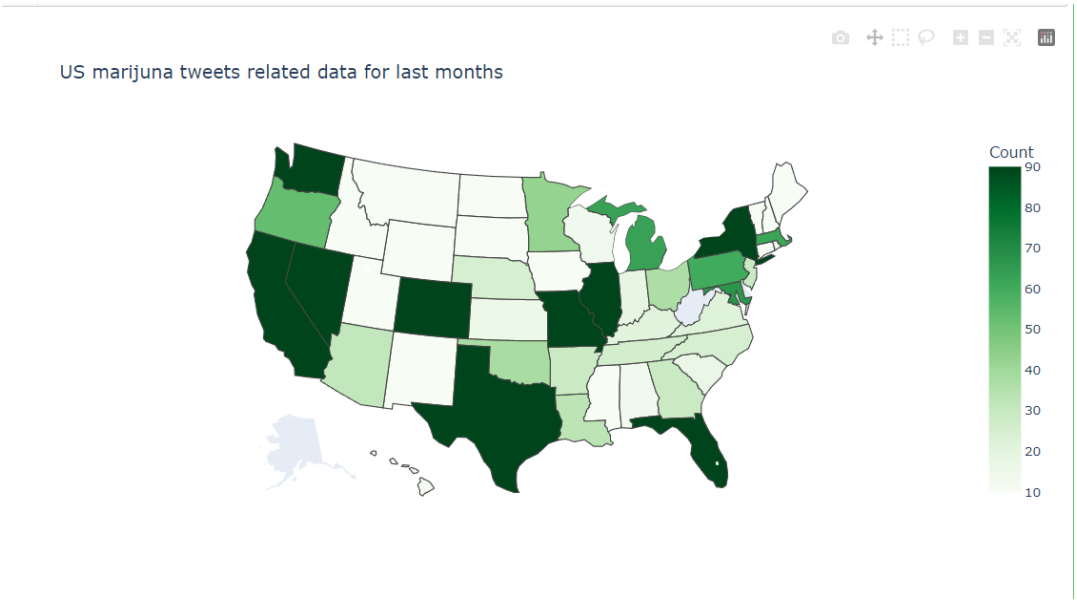
We had collected the data from the twitter users of USA. So, with the help of location where the tweets are originating, we can analyse how many tweets are generating from a particular place or state and this help in any government agencies or to understand the mindset of people using the twitter to tweet related to marijuana.



So, from the bar plot and geopandas, ploty plot we can see that major tweets were from the states of California, New York, Colorado, Nevada, Texas, Florida, Illinois ,Missouri, Washington.

There are many states in USA which have legalize the use of marijuana for the use of medical purposes and some have also legalized the use for recreational purposes

Analysis on the data collected:



State laws on use of marijuana:

State		Recreational	Medical	Transportation	Cultivation
California		Legal	Legal	Up to 1 oz (28 g)	Legal (six plants for personal use, or a commercial license)
New York		Legal	Legal	Up to 3 oz (85 g) of cannabis for recreational purposes or 24 grams of concentrated forms of the drug, such as oils.[107]	Up to three mature and three immature plants per person, maximum twelve per household.[108]
Colorado		Legal	Legal	Up to 2 oz (57 g)[23]	Legal (six plants for personal use, or commercially licensed[24])
Nevada		Legal	Legal	Legal	Adults at least 21 years old can grow in home (6 plants per household), or commercially licensed[85]
Texas		Illegal (De facto legal by refusal to arrest for less than 4 ounces in possession in Austin. "cite and release" in Houston, Dallas, San Antonio, Austin, and residents of Travis County.)	CBD oil (no more than 1% THC and no less than 10% CBD)	Not clearly stated	Illegal

Florida		Illegal	Legal	Medical use only	Medical use only
Illinois		Legal[49]	Legal	Up to 30 g (1.1 oz)	Five plants in home for medical use only, or commercially licensed for recreational[50]
Missouri	D	Decriminalized	Legal	Not clearly stated	Legal for medical use
Washington		Legal	Legal	Legal to possess up to 1 oz (28 g)	Legal with restrictions and commercial licensing
District of Columbia		Legal (no commercial sales)[172]	Legal (commercial sales)	Legal to carry up to 2 oz (57 g)	Legal to grow up to six plants (only three mature at a time) for recreational purposes; no provision for commercial recreational cultivation

Source(Wikipedia)

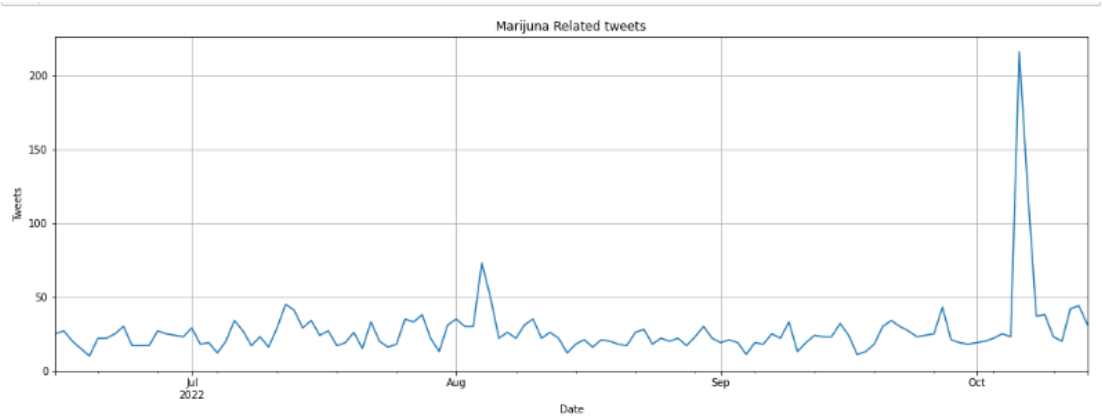
While analysisng the dtwitter data we sholud also consider the size of these states and their population these are very big states in USA and will be definetly having more users so the tweets might be more but we could see that 7 states have approved the recritional use of marijuna and also for medical purposes we could see that Texas is the only state mentiond has a very strict law on use and consumption of cannibis and its illegal in some cases for medical purposes as well.

Twitter data flow analysis:

If we analyse the twitter data which we have collected over a period of time we can see the trends and infer that a major movement or major change or annousment which might happen or going to happen or happened

So we plotted a time series graph using matplot library by grouping the data for everyday. From the graph we can see that in the beginning of october there was huge flow of tweets relating to marijuna. If analyize it we could see that the spike was seen on OCT 6th the day Joe biden announsed the De criminalization law

So people reacted and tweeted more about the topic on that day.



Hashtag analysis:

We will be discussing and analysing on all the hashtags which users used in their tweets. Hash tag analysis gives us a major help in finding out what is the trend people want their tweet to support. Hashtags are a major tool in twitter it is used to put a word from each user and to create a major trend on twitter by tagging the hashtag.

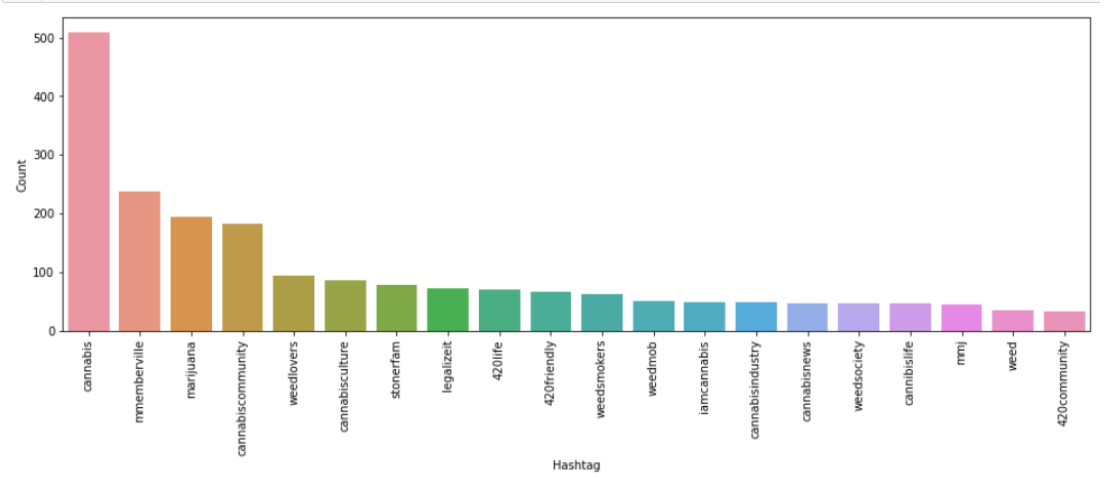
So, we took out top 10 hashtags which were used by the users during the data collection period we can see the major hash tags are Cannabis, mmemberville, marijuana, cannibiscommunity, stonefarm and legalizeit. We can see that all these hashtags belong to pro cannabis movement.

Few insights on hash tags:

Mmemberville: this is run by a Sibusiso Sithole who a musician and tweets on these belong to virtual stoner(cannabis smokers) community.

Cannibiscommunity: as the name says these are tweets belong to people who intake marijuana and discuss on the topic of marijuana.

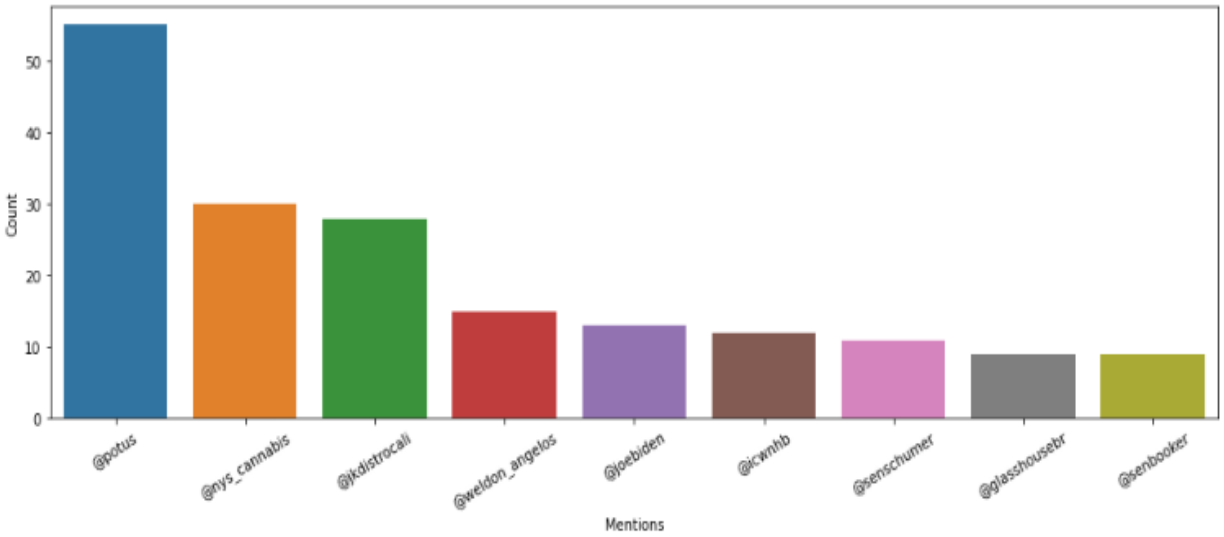
Legalizeit: here we can find tweets about people asking the government to legalize the consumption of cannabis for recreational and for medical purposes.



Tweet User mentions analysis:

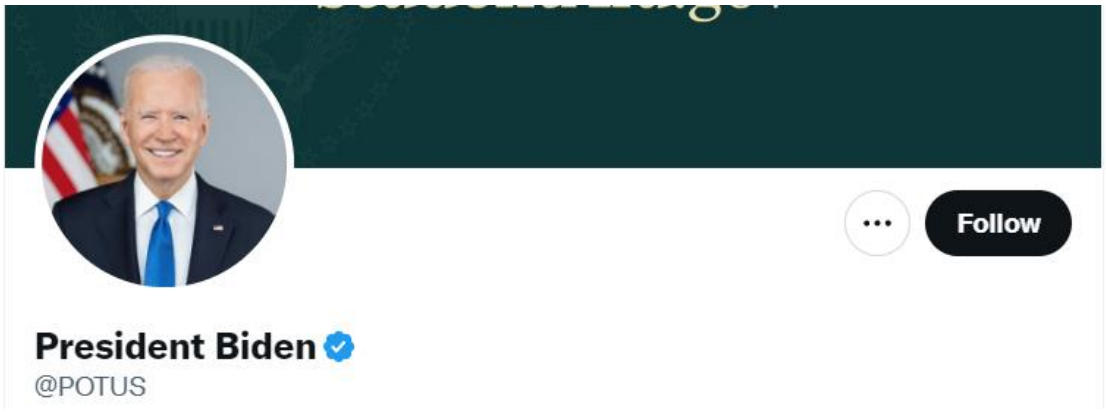
With the help of who the twitter users are tagging we can infer the kind of communication happening and analyse the topics they are discussing , and the importance of a particular tagged user

So, we plotted a graph for 10 most tagged user for the data collected and we can see the same below



@Potus,@nyc_cannabis,@jkdistrocali,@weldon_angelos,@joebide,@icwnhb,@senschumer were few prominent users who were tagged in Cannabis related tweets from the bar graph.

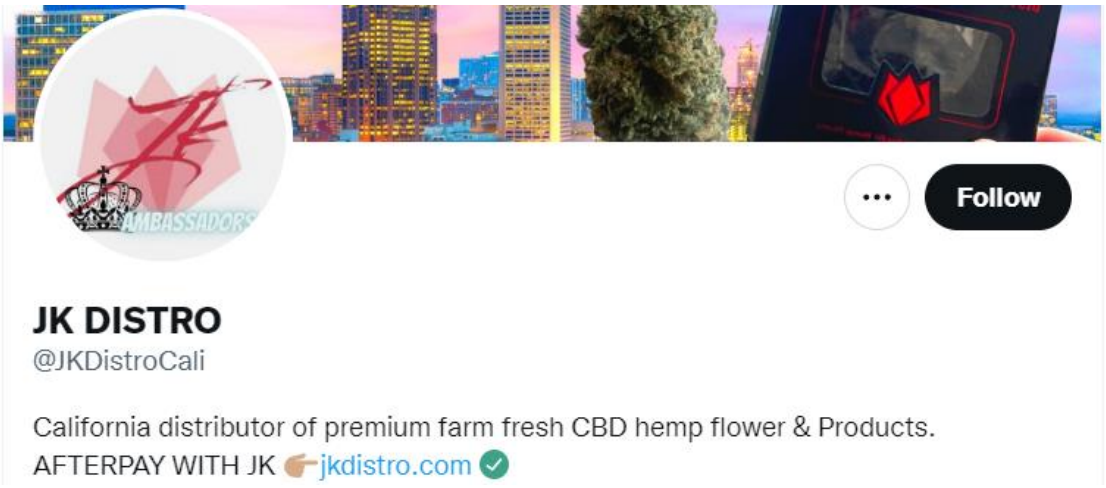
@POTUS is USA president twitter handle:



People were seen congratulating and accepting the move done by president for decimalization law

@nyc_cannabis : spreads message about cannabis and edible products of cannabis

@JKDistroCali : is a California based CBD hemp flower distributer



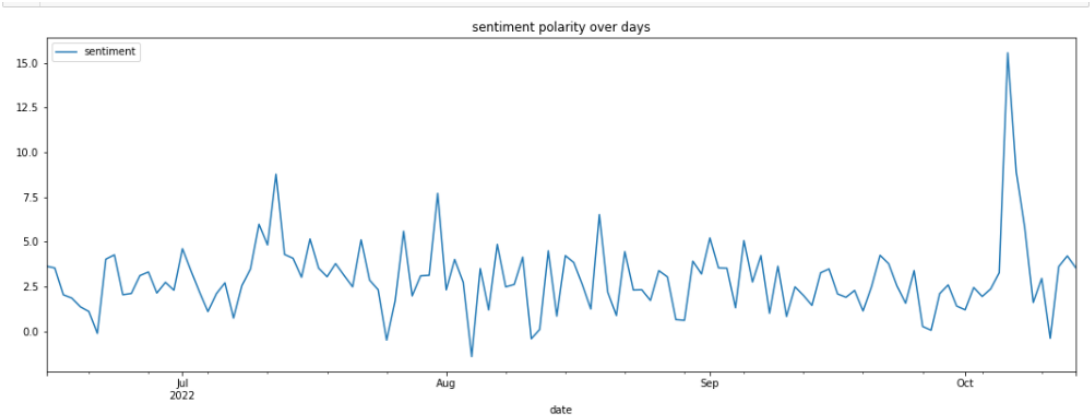
@weldon_angelos: Is a music producer and also a pro cannabis activist

@senschumer: is democratic leader which is currently in power in US and is a pro marijuana supporter

We could see that people are tagging and supporting their leaders move for decriminalizing law of marijuana possession and we can see that cannabis and marijuana products are quite famous on twitter and are quite proactive on cannabis movement.

Twitter data Sentiment analysis:

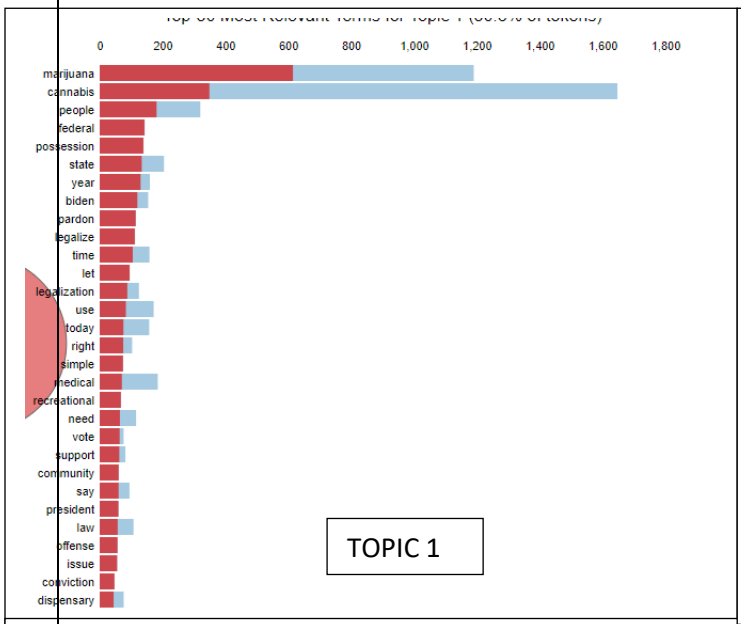
The study gathered and analysed the total data in sentiment analysis. The initial study shows that the sentiment peaked on the day of Biden's announcement, and we could see that there it was again at peak in the starting of August and in mid-July. The overall sentiment was positive and went below only very few times.



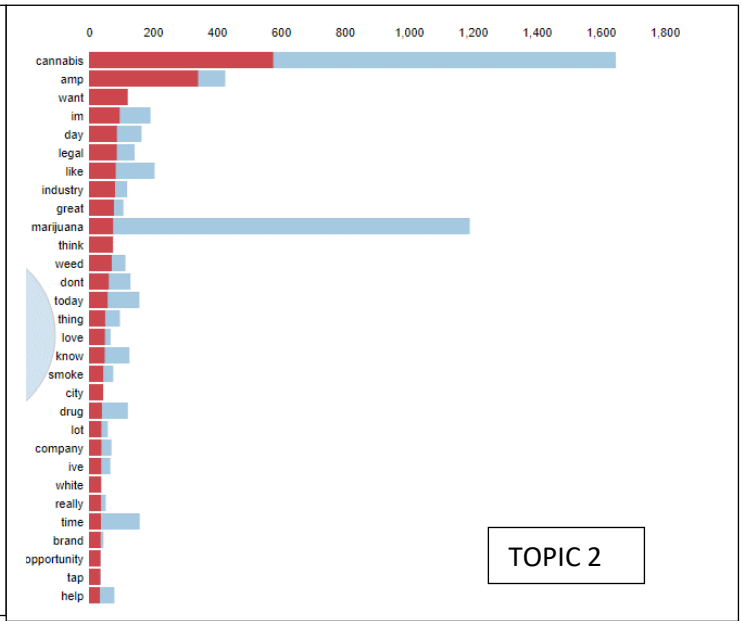
Topic Analysis of tweets:

Thematic content analysis was conducted on all the data. The tweets fell under these 4 themes

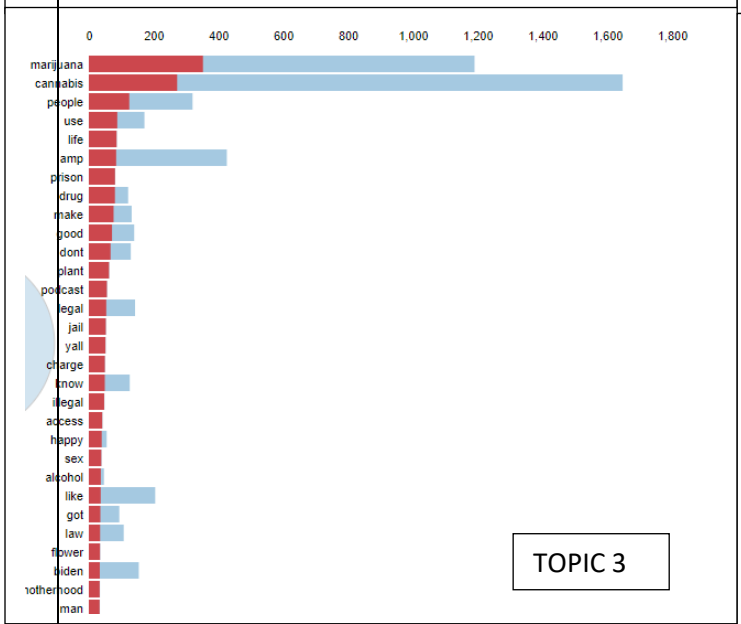
1. New work and job opportunities because of cannabis
2. Legalization of marijuana possession and Biden's new law
3. Improvements in medical marijuana because of legalization
4. Industrial growth in marijuana sector



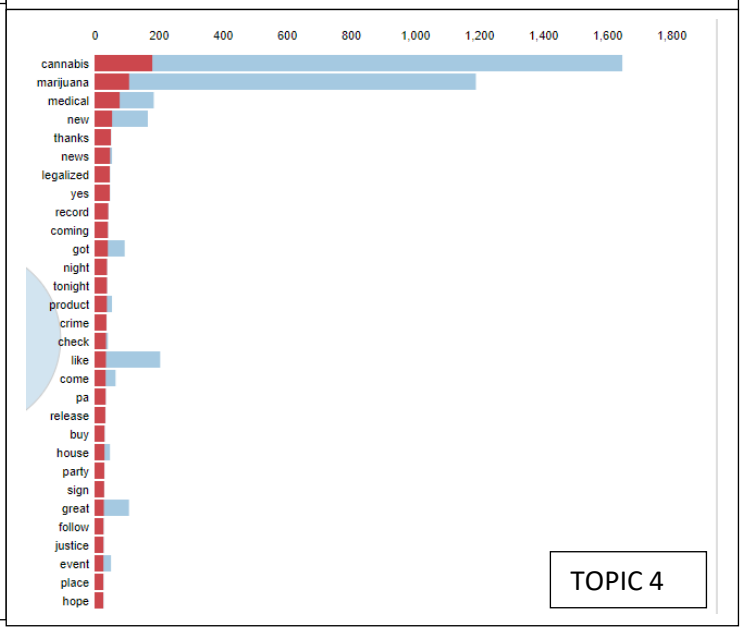
TOPIC 1



TOPIC 2



TOPIC 3



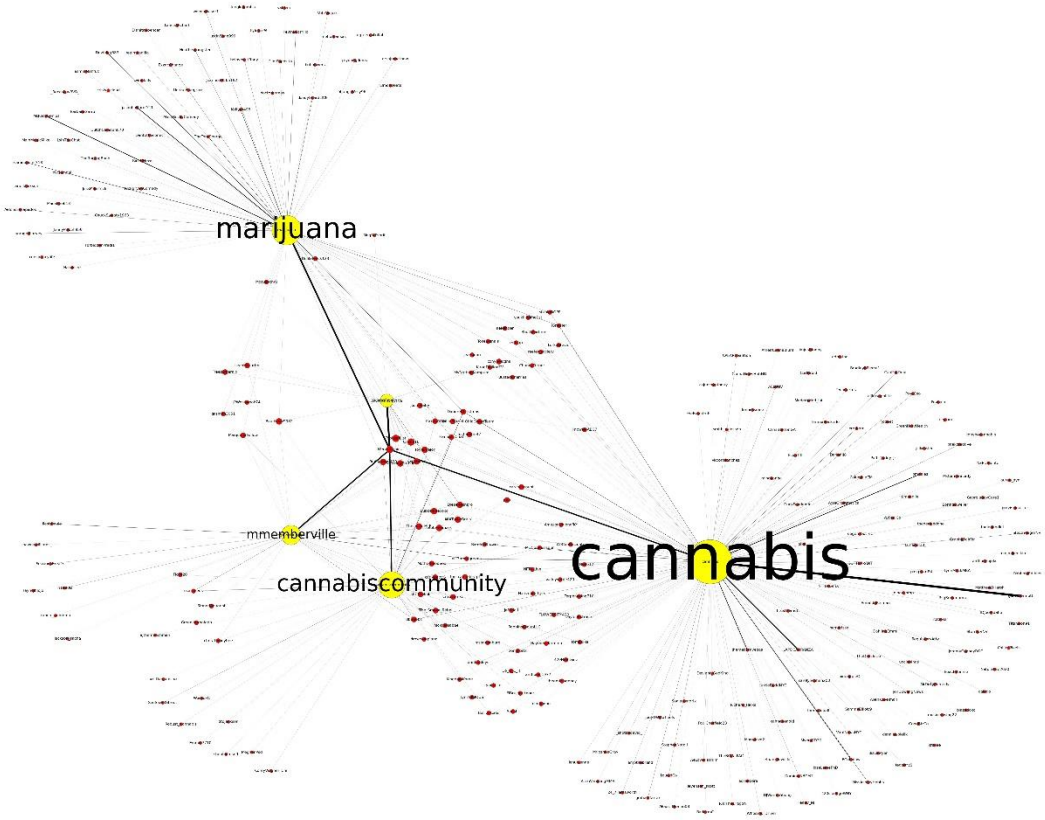
TOPIC 4

An important pattern is that people are thinking about the change and are ready to use the opportunity the marijuana here to provide if legalized people are speaking about the job and industrial opportunities marijuana is here to provide and are ready to accept the change.

People discussed on legalization, employment, medical advancements and many more.

Network analysis:

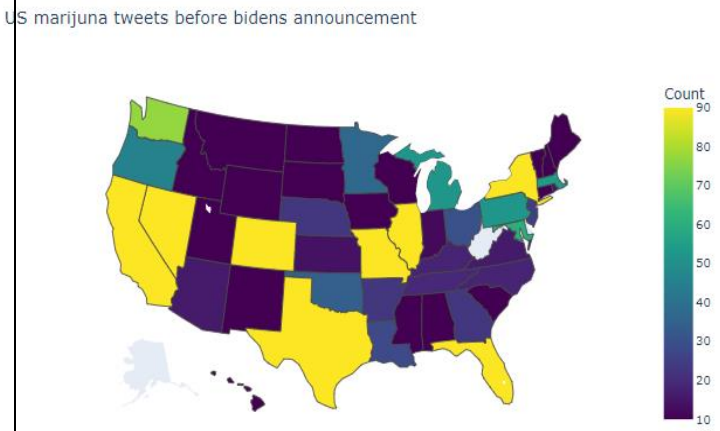
We did network analysis on twitter users selecting the hash tag while tweeting and the thickness of the edge is decided by the likes a particular user got when he tweeted by using the hash tags. We have taken only top 5 hash tags here because when we take top 10 the graph was not able to analyse because of density. The nodes in yellow are the hash tags and nodes in red are the users. We could see that many users used cannabis to tweet and it's the largest community formed and marijuana is next biggest community which was formed we can analyse from the below graph that 50 % users who used marijuana did not use any other hash tag where as in case of cannabis hashtag users more than 65 % users did not use any other hash tag in tweets their other users in both the community used the other hash tags like mmemberville, cannibiscommunity, weedlovers as their hash tag. We can see a user named @weedgenics used all the five hash tags regularly and has major likes on the tweets made

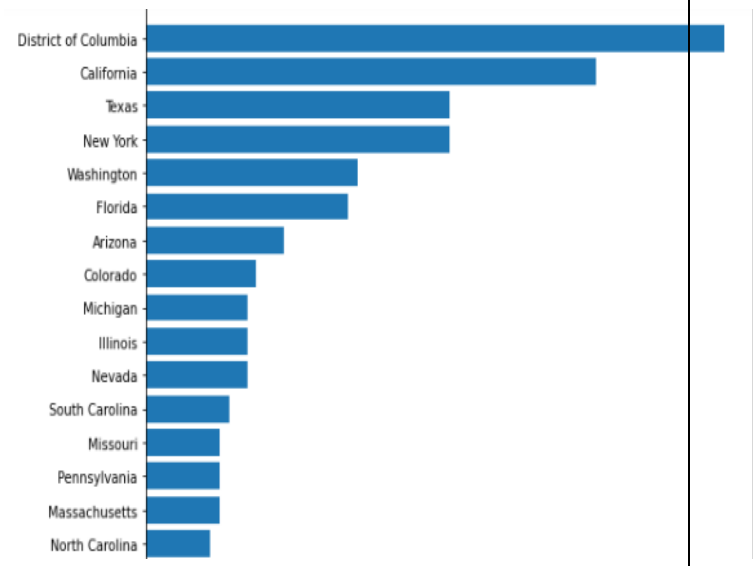


Analysis for data which was collected before Biden's announcement. And Analysis on data which was collected after Biden's announcement.

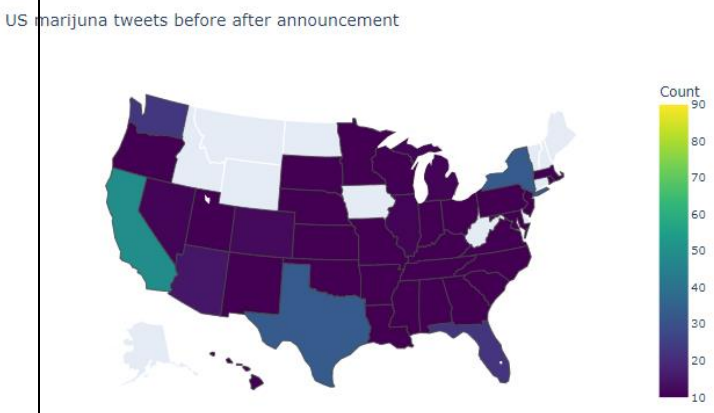
We will be comparing both data of before and after the Biden's announcement but the data before announcement is quite large compared to after announcement. So we will try to normalize this and compare the findings.

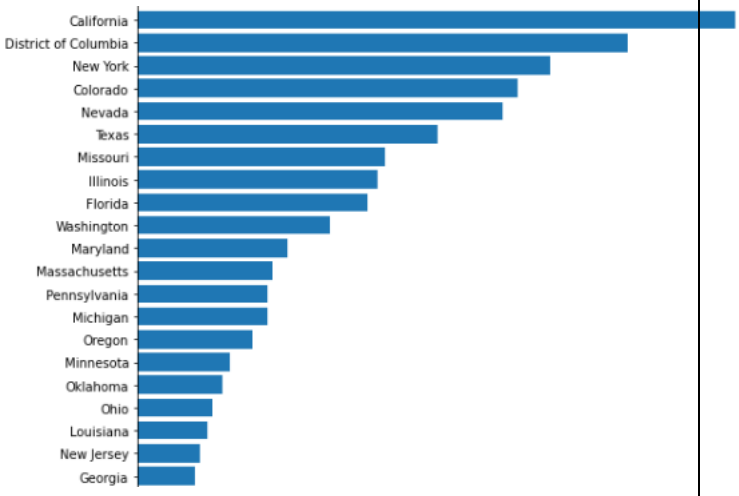
Twitter location data before announcement:





Twitter location data after announcement:

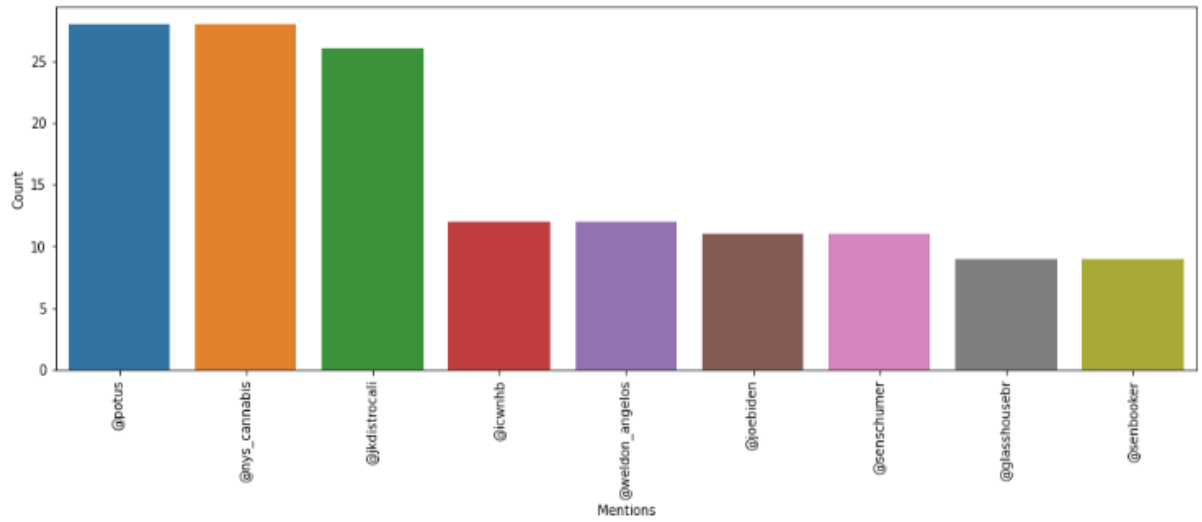




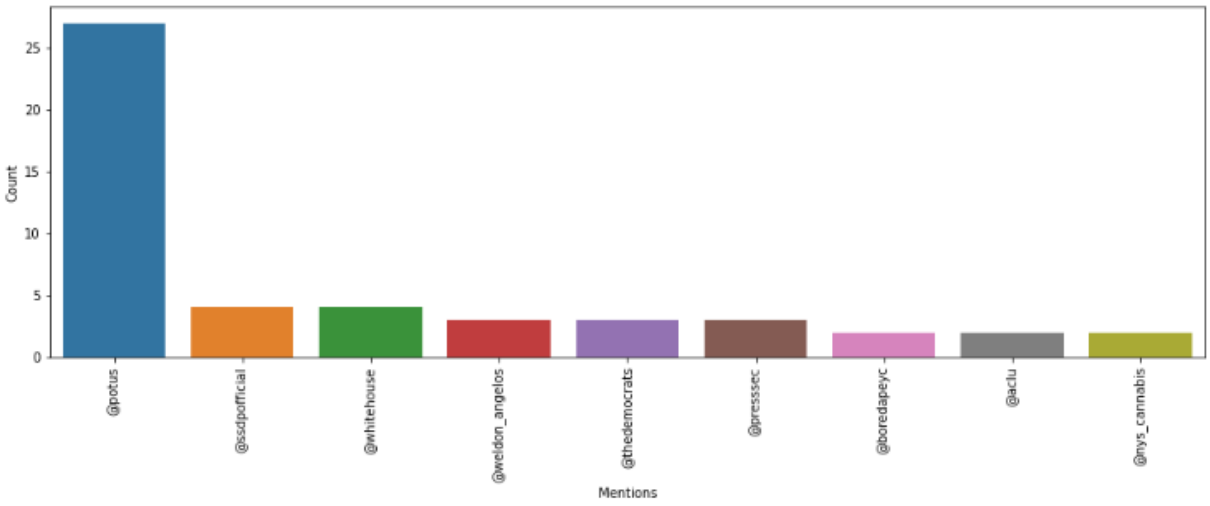
We could see some changes in the tweets made many of the states started to talk about the law change and joined the conversation on twitter. States on the east coast joined the conversation. And most of the central America had a equal participation.

Tweet mention analysis :

Before:



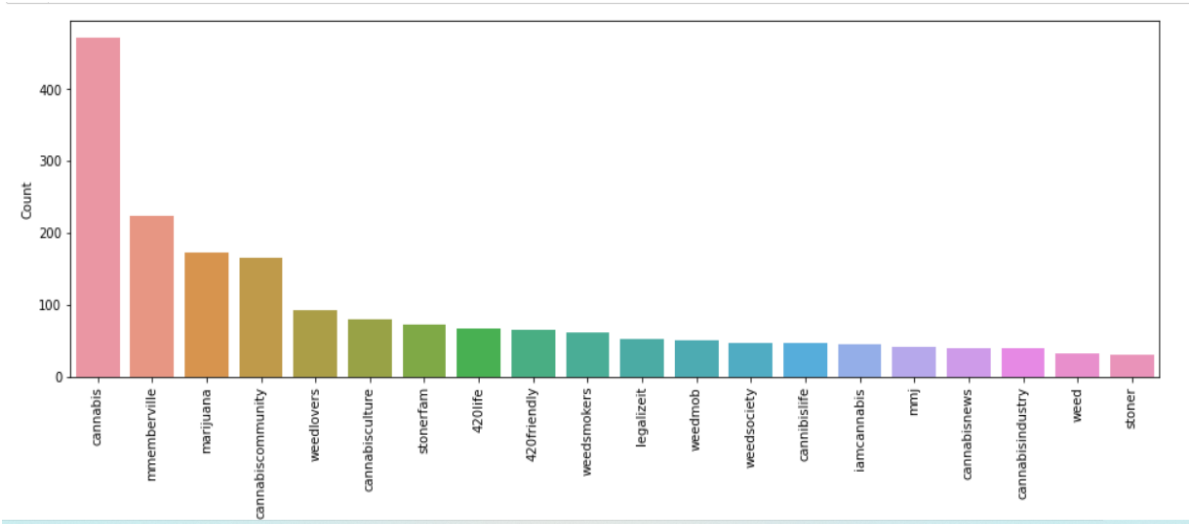
After:



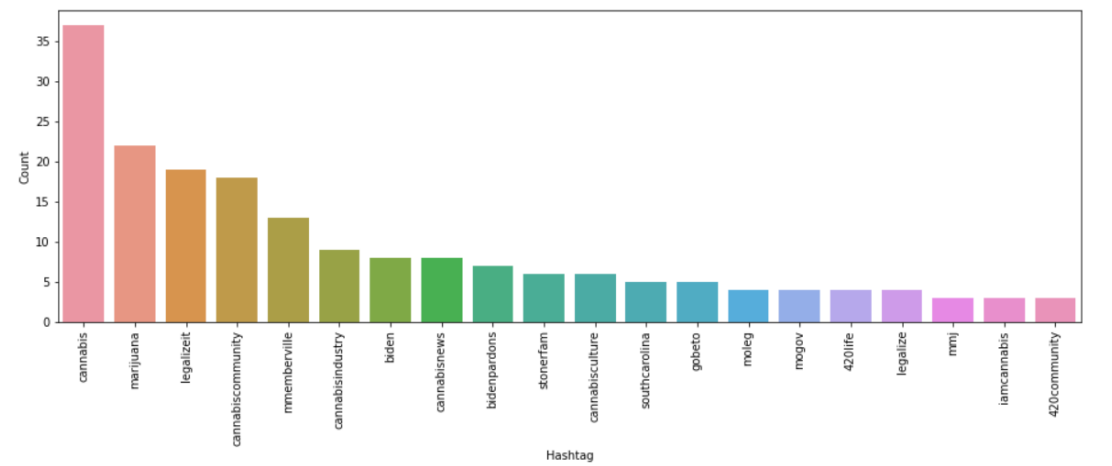
We could see many democrats' leaders mentions were found in the data we found after the Biden's announcement. @potus was the only mentioned twitter handle which still retained the number one position for the mentions

Hashtag Analysis:

Before:



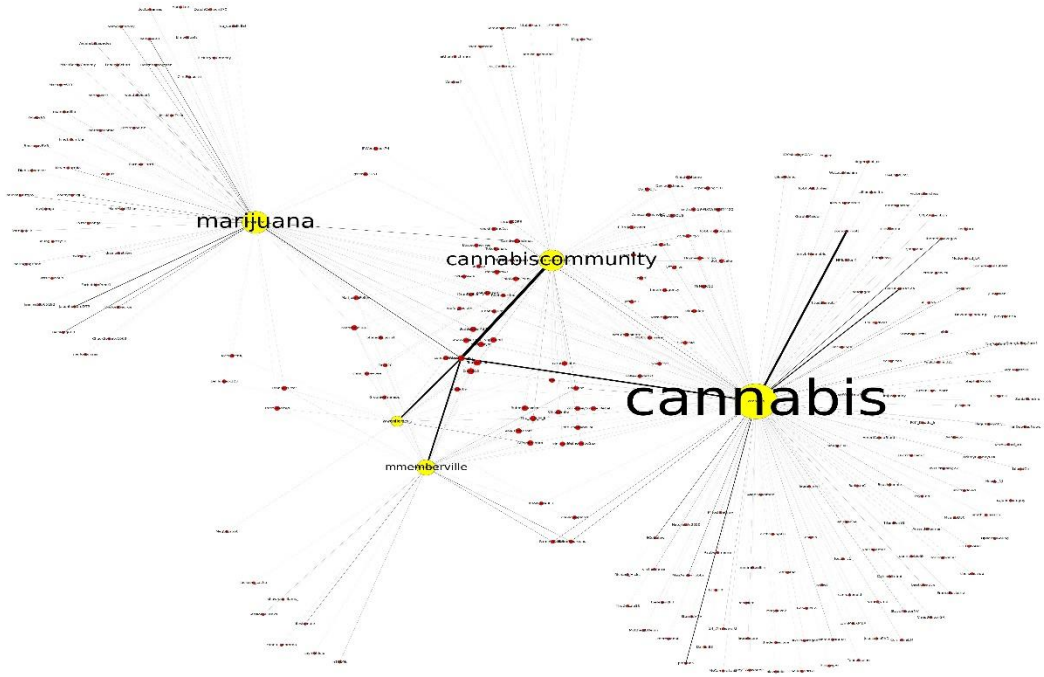
After:



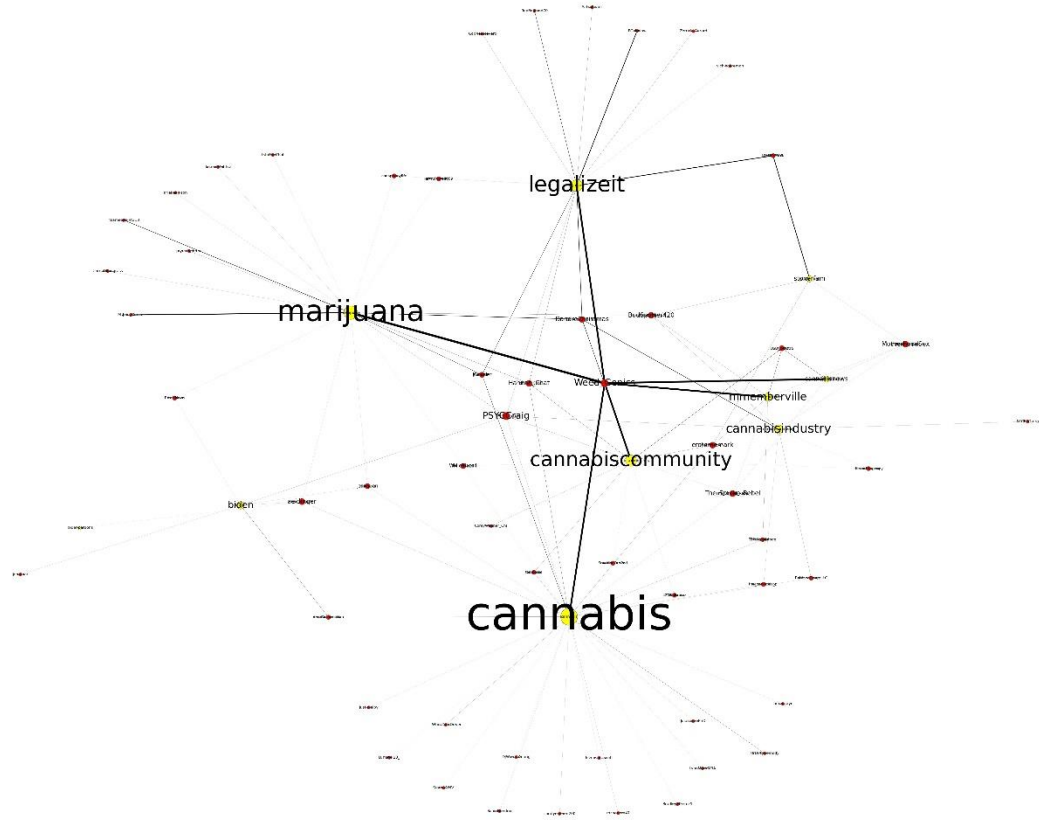
We could see a drastic change in the hash tag usage. Legalizeit and cannabisindustry hash tags were used more after Biden's announcement. Meaning people want more legal freedom relating to cannabis culture in USA

Network analysis:

Before:



After:



As data after the announcement was very less, we cannot conclude anything by comparing both the network graphs

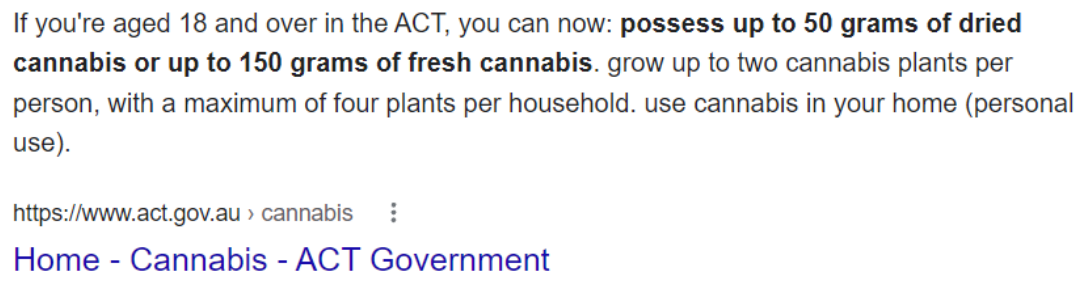
But before we could see that cannabis and marijuana formed a community but in the after graph, we could see that all the tweets were quite interlinked and was used in a equal manner but we could see partial communities were formed for each hashtag

Conclusion from analysis 1:

People are ready for change in a positive manner and are ready to accept the change marijuana is going to bring into industry, medical science, recreational activities. We could see tweets majority in favour of legalizing the marijuana and also welcoming the change of new Biden's take on marijuana.

International Impact:

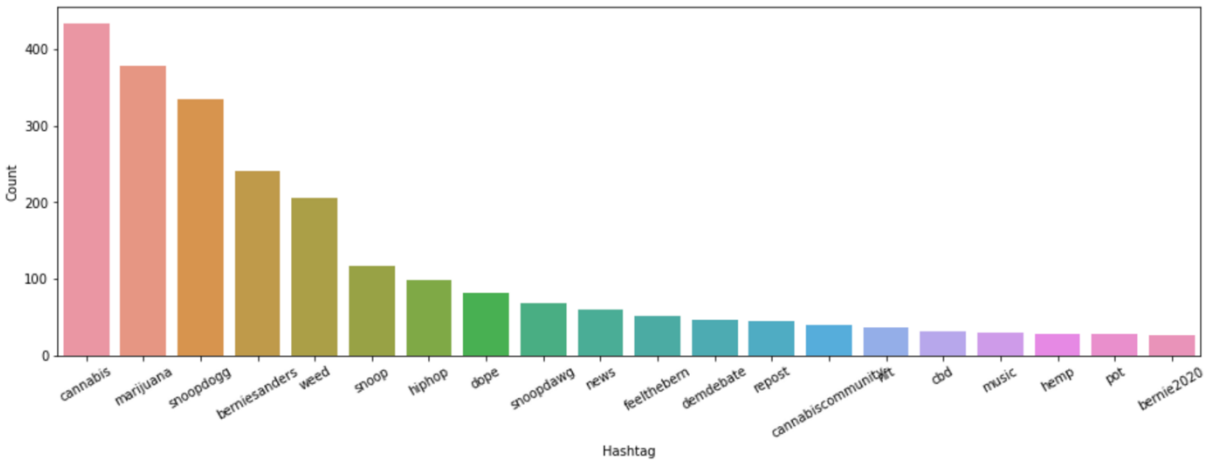
As we said above in introduction if USA changes it impacts other countries, we could a similar move in Australia Canberra has introduced a new law that states that it will also decriminalize the small position of marijuana with people only fines will be collected but not imprisonment



Analysis 3 -

Hashtag Analysis -

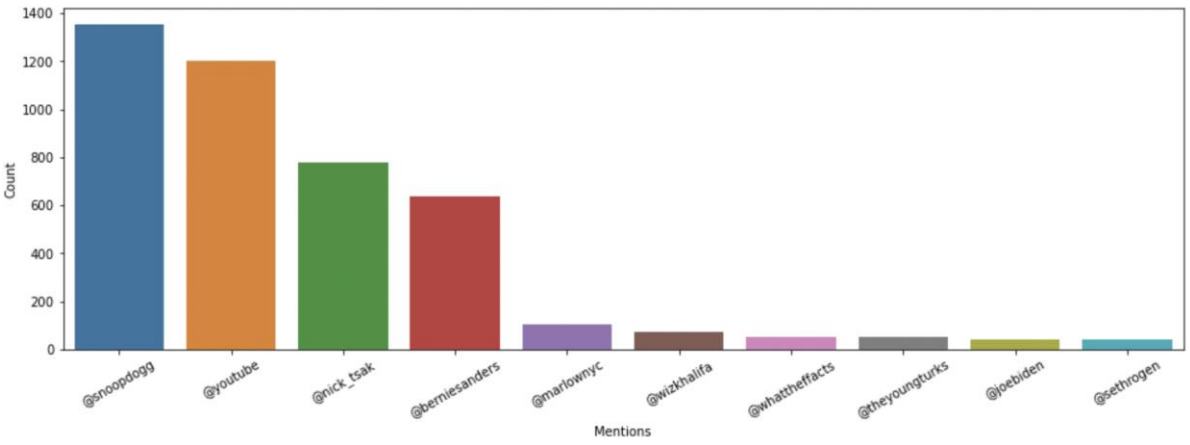
One of the significant tasks was to extract hashtags for the analysis. First the hashtags were extracted from the data frame. The top 10 hashtags were plotted to get the insights of the data.



The above figure gives us some insights of the data where cannabis was leading followed by marijuana and snoopdogg. Some of the other hashtags included Bernie sanders, weed, snoop, dope, news, cannabis community, cbd, pot. There were two major influencers in the top hashtag list which were Bernie sanders and the snoop dog because there are actively involved in the conversation related to marijuana. The data also suggest that people are more inclined towards marijuana and cannabis rather than weed and dope.

Mention Analysis -

The next major step is to identify the people that are engaged with the audience on this topic. It can be identified through plotting the histogram of mention.

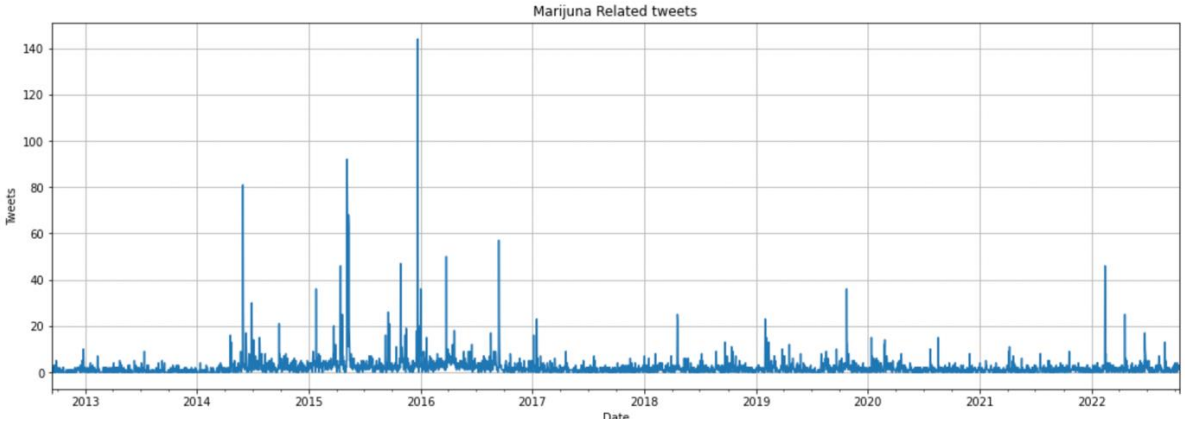


The most involved influencers were snoopdogg, Bernie Sanders, Nick Tsak, Wiz khalifa. People also

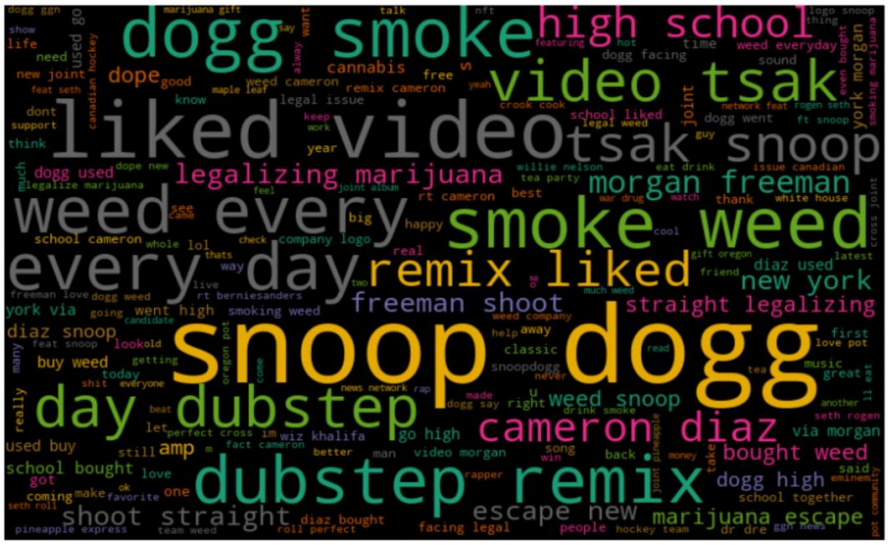
used youtube to bring attention about the drugs through the video. Joe Biden made entry to the list with the recent announcement about the laws regarding drugs.

Sentiment Analysis -

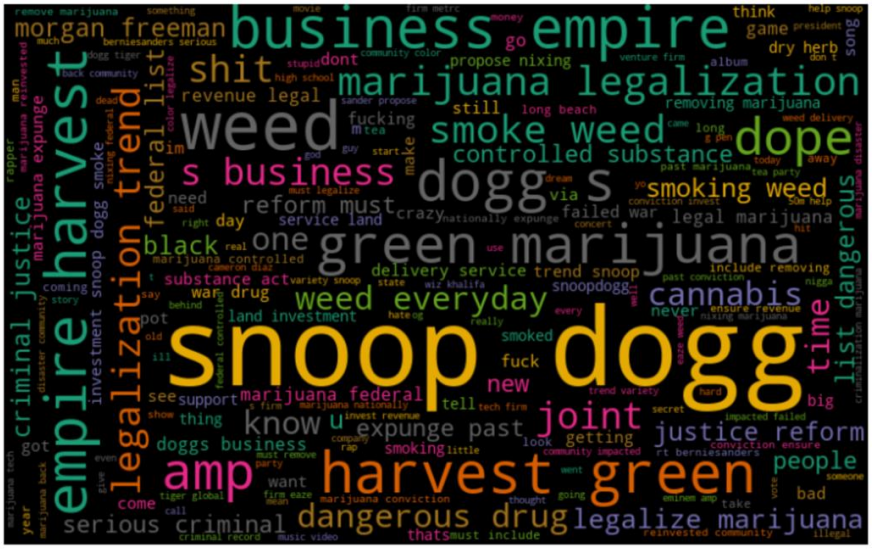
Every tweet was classified with the sentiments of positive, negative, and neutral. It was found that most of the tweets were positive followed by neutral and negative. The tweets were plotted in the span of 10 years from September 2012 to October 2022.



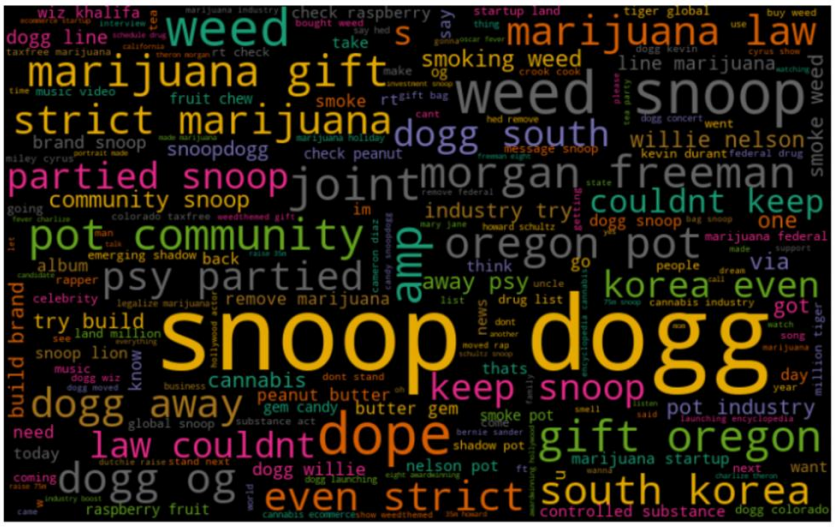
It was seen from the figure that major tweet spiked were observed during the year 2015-2016. Next the tweets were spiked due to the announcement of US president Joe Biden about the drugs law. The major reason for spike of tweets in the year 2015-2016 were because the death rate in US due to Drug overdose were significantly increased. As a result people were expressing concerns on the twitter about the drug related deaths. People expressed major of the sentiments during that year.



The above figure shows the positive sentiments expressed by people on the twitter. It can be observed that people are talking about smoke, weed, everyday consumption of the marijuana. One of the interesting facts is people are expressing their views on legalizing the marijuana in the country. Also, some of the people were expressing their views about the consumption of the drugs in the high school.



The above figure represents the negative sentiments about the marijuana. Snoop dogg one of the most influential persons is active on this. Some of the people are expressing their views on legalizing the drugs while other were concerned about the marijuana empire as a business. Some people were talking about reforming the justice in views of expressing their concerns over the marijuana while other were talking about weed consumption on daily basis.



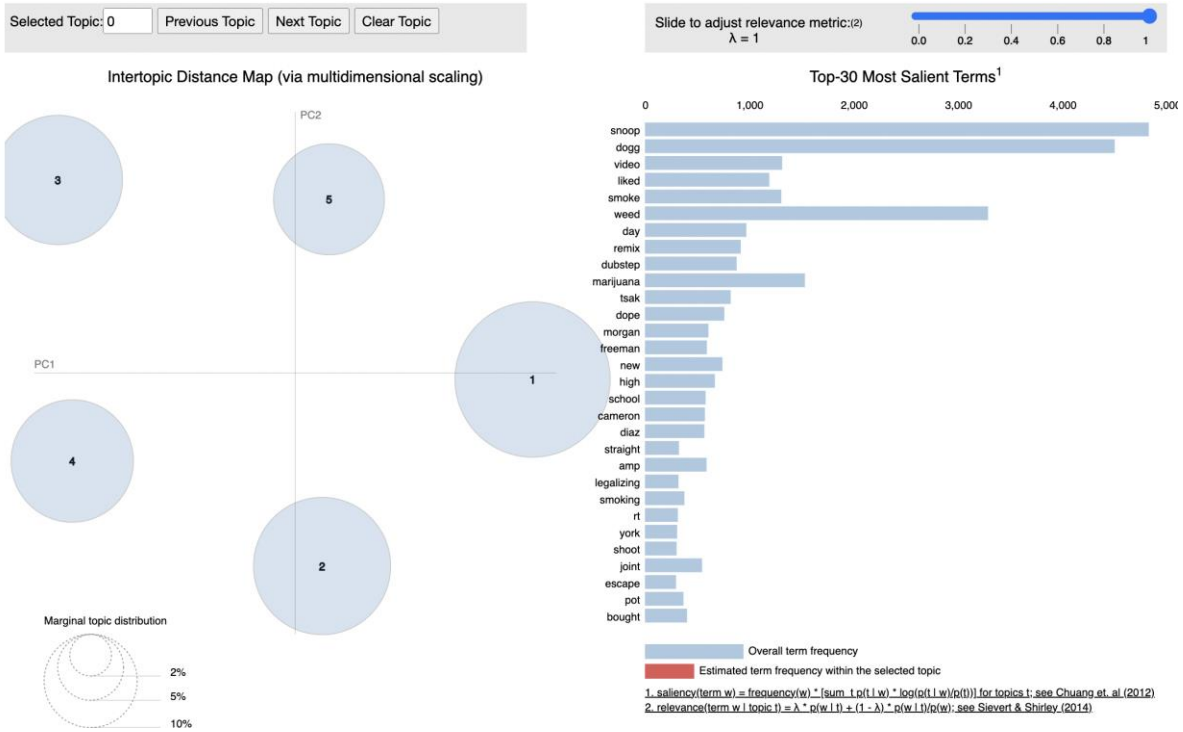
The above figure shows the neutral views about the marijuana from the people active on twitter. Some of them were involved with snoop dogg and Morgan freeman on this issue. A minor community of people were expressing the concerns about the pot community. Others were expressing to have strict marijuana rules.

Topic analysis -

The features of the tweets were extracted using the count vectorizer and the LDA were fed the features to get the topic related to the marijuana.

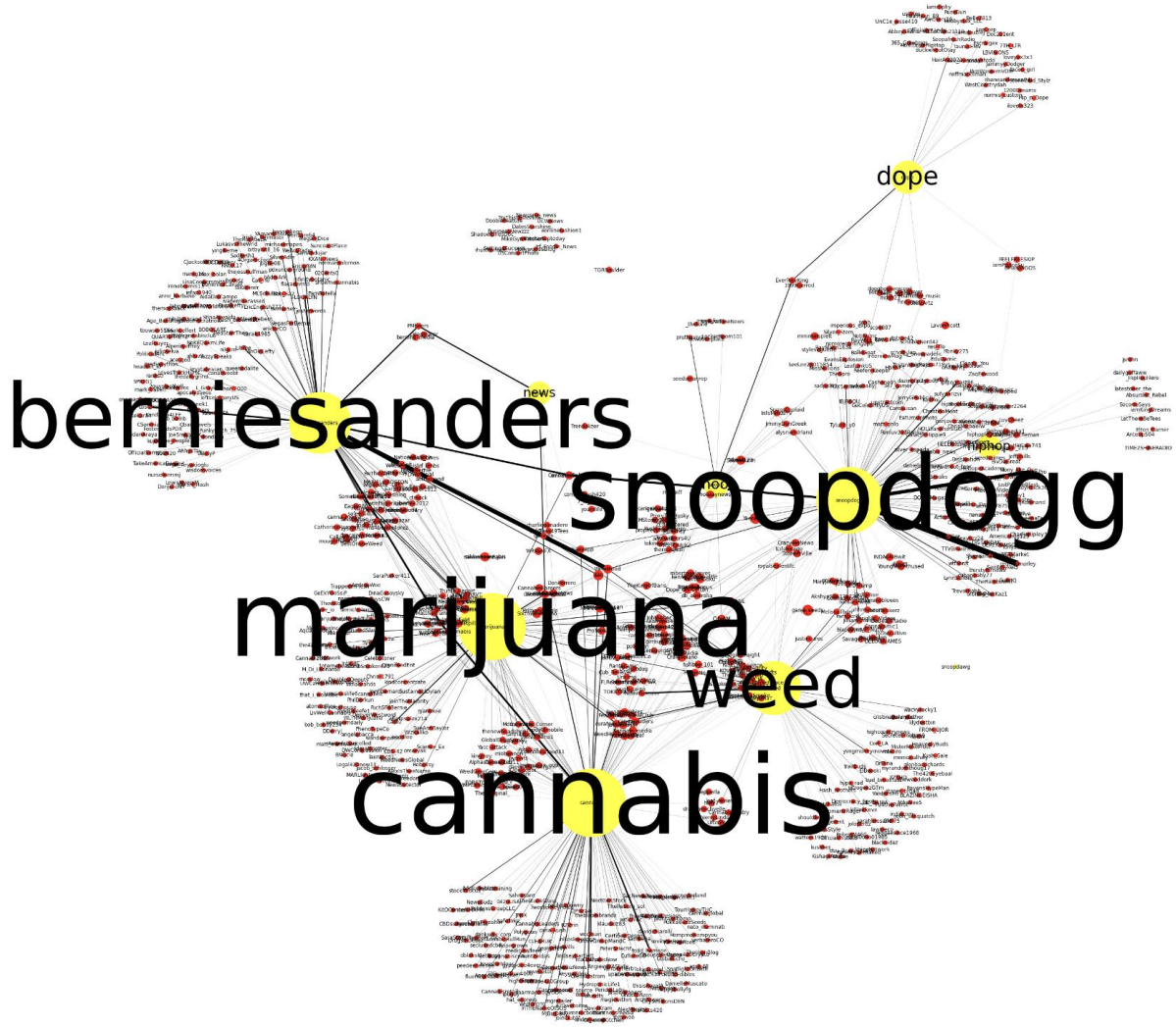
```
Topic 0:
marijuana rt community gift legal pot 5000 oregon drug away law business new big legalization
Topic 1:
dope joint amp love know like tea smoke weed world come shit got think need
Topic 2:
snoop dogg smoke weed cannabis marijuana check everyday wiz news south federal 420 khalifa list
Topic 3:
snoop dogg weed high school cameron diaz bought smoking used buy og amp joint marijuana
Topic 4:
video liked day remix dubstep tsak weed new morgan freeman marijuana straight legalizing york shoot
```

The above diagram shows the 5 different topics related to marijuana. People are talking about legalizing pot as it has been a business. Another topic involved consumption of dope and weed on everyday basis. Others are suggesting snoop dog to watch the news channel daily regarding the drug news due to the consumption of the marijuana in the high school. Others are talking to Morgan freeman about legalizing the weed.



The above figure represents some of the most frequent unigram occurring in the topics where people are involving Snoopdogg related to the news of high school marijuana consumption. While others were expressing the concern of legalizing it through Morgan freeman.

NETWORK ANALYSIS



The above figure represents the community of the various influencers and the marijuana. The marijuana community is very large followed by cannabis. While dope and weed has very small community. Majority of the community is revolving around Bernie sanders and Snoopdogg as there are actively involved about this matter.

The connection between the community of Bernie sanders and snoop dog is shown by first constructing the egonet of Bernie sanders and then randomly finding the connection between two different communities.

```
sEgoId = client.get_user(username="BernieSanders").data.id
sEgoName = client.get_user(username="BernieSanders").data.username
maxResults = 10
```

```
egoGraph = nx.DiGraph()
lUserFields = ["id","name","public_metrics"]

#
# Retrive followed (people who follows the ego user)
#
twitterResponse = client.get_users_following(id=sEgoId, max_results=maxResults, pagination_token=None, tweet_fields=None, user_fields=lUserField
# loop through the results from first page
for user in twitterResponse.data:
    sFollowedName = user.username
    egoGraph.add_node(sFollowedName, followerCount=user.public_metrics["followers_count"])
    egoGraph.add_edge(sEgoName, sFollowedName)
while True:
    try:
        twitterResponse = client.get_users_following(id=sEgoId, max_results=maxResults, pagination_token=twitterResponse.next_token, tweet_field
    except AttributeError:
        break
    for user in twitterResponse.data:
        sFollowedName = user.username
        egoGraph.add_node(sFollowedName, followerCount=user.public_metrics["followers_count"])
        egoGraph.add_edge(sEgoName, sFollowedName)
```

```
graphFile='ego.graphml'

inDeg = egoGraph.in_degree(sEgoName)
outDeg = egoGraph.out_degree(sEgoName)

print('in degree of ego = {}'.format(inDeg))
print('out degree of ego = {}'.format(outDeg))


lInNeighbours = [sNeigh for sNeigh in egoGraph.predecessors(sEgoName)]
lOutNeighbours = [sNeigh for sNeigh in egoGraph.successors(sEgoName)]
print('in neighbours of ego = {', end='')
print(*lInNeighbours, sep=', ', end='')
print('}')

print('out neighbours of ego = {', end='')
print(*lOutNeighbours, sep=', ', end='')
print('}')
with open(graphFile, 'wb') as fOut:
    nx.write_graphml(egoGraph, fOut)
```

```
in degree of ego = 0
out degree of ego = 10
in neighbours of ego = {}
out neighbours of ego = {NextGenAmerica, NikemaWilliams, shannaforPA, TeenVogue, donnaimamTX, Public_Citizen, _karaeastman, audreydenney, Kunkel4Congress,
onvention}
```

The above diagram shows the egonet of Bernie Sanders. It can be seen from the above diagram that the neighbours of Bernie sanders are NextGen America, public citizen, Nikema Williams, DemConvenction etc.

In [185]:
```
import random
def get_random_walk(node, walk_length):
    random_walk_length = [node]

    #loop over to get the nodes visited in a random walk
    for i in range(walk_length-1):
        # list of neighbors
        neighbors = list(G.neighbors(node))
        # if the same neighbors are present in ranom_walk_length list, then donot add them as new neighbors
        neighbors = list(set(neighbors) - set(random_walk_length))
        if len(neighbors) == 0:
            break
        # pick any one neighbor randomly from the neighbors list
        random_neighbor = random.choice(neighbors)
        # append that random_neighbor to the random_walk_length list
        random_walk_length.append(random_neighbor)
        node = random_neighbor

    return random_walk_length
```

In [192]:
```
get_random_walk('HEELAussie', 10)
```

Out[192]:
```
['HEELAussie', 'snoopdogg', 'MorveenaFacts']
```

From the above diagram the person HEELAussie belonging to the Bernie Sanders's community is connected to the Snoopdogg community via the MorveenFacts.