Assignment - 2
Arnav Tandon | 2018278

Section 1 :

```
In [15]: import tweepy,json
         from tweepy import *

         import pandas as pd
         import csv
         import re
         import string
         import preprocessor as p


         consumer_key= "dKFZqH88FytU2ebbFjQGFPmkr"
         consumer_secret= "wfj00vfMfAVFdYS6iy3rVYY56GGUDnddKhLgE5EbCjGkfTzF7kC"
         access_token= "3220731312-5lujZpUvkJ4Q0ltlVAD8ZwBPvT5QUOlJwP9EVR3"
         access_token_secret= "YJKquEf45wuZHyspdkie3Vt87H5VWkTK6aVWby5logt1A"

         auth = tweepy.OAuthHandler(consumer_key,consumer_secret)
         auth.set_access_token(access_token,access_token_secret)

         api = tweepy.API(auth,wait_on_rate_limit=True)

In [16]: from pandas import DataFrame

In [17]: import tweepy
         import datetime
         import re
         import tqdm


         number_of_tweets=5000
         userID="noidapolice"

         tweetList=[]
         tweetList = tweepy.Cursor(api.user_timeline, screen_name=userID, include_rts = False, tweet_mode ="extended").items(number_of_tw

         tweets_json = [tweet._json for tweet in tweetList]
         df_fin = pd.DataFrame.from_records(tweets_json)
```

Ans 1)

PII Table, Examples

```
        if(len(r2)>0):
            vehicle_plate_list.append(r2+[a])
        if(len(r3)>0):
            email_id_list.append(r3+[a])
        if(len(r4)>0):
            pan_card_list.append(r4+[a])
        if(len(r5)>0):
            aadhar_num_list.append(r5+[a])
        #if(len(r6)>0):
            pincode_list.append(r6+[a])
```

```
In [236]: print(phone_num_list)

[['9999419680', 'urgently need 2 units of plasma any blood group in noida please contact at 9999419680 @aajtak @dkumarchandel… httpstcox9sge0hi5c'], ['8285859901', '[illegible] 8285859901 [illegible] httpstcodpr8xpvjun'], ['8448830852', 'ye mobile phone chori ho gaya hai 8448830852 pidit ka mobile number emi number surveillance system par lagakar t… httpstcowlurm0pdbl']]
```

```
In [237]: print(vehicle_plate_list)

[]
```

```
In [238]: print(pan_card_list)

[]
```

```
In [239]: print(aadhar_num_list)

[]
```

```
In [241]: print(email_id_list)

[]
```

Here we can see the tweets having mobile numbers in the tweet text which is added with the PII obtained, here we can observe that when people ask for help/request they tend to share their personal information so that people can reach out to them with some solution or help which that person is asking.

of Tweets containing media

```
In [195]: count = 0
          media_links = []
          for ind,row in df_fin_reply_extend.iterrows():
              if ('media' in row.extended_entities):
                  count+=1
                  media_links.append(row.extended_entities['media'][0]['media_url'])

          print(count)
          #print(media_links)

53
```

```
In [187]: import requests

          listofimages=[]
          count = 0
          for url in media_links:
              filename = "images/"+"image"+str(count)+".jpg"
              count+=1
              result = requests.get(url, stream=True)
              if result.status_code == 200:
                  image = result.raw.read()
                  listofimages.append(filename)
                  open(filename,"wb").write(image)
```

Here I had 53 tweets which had some media in it, I downloaded all the media and analyzed each one of them to observe if I can get some useful images which contain PII

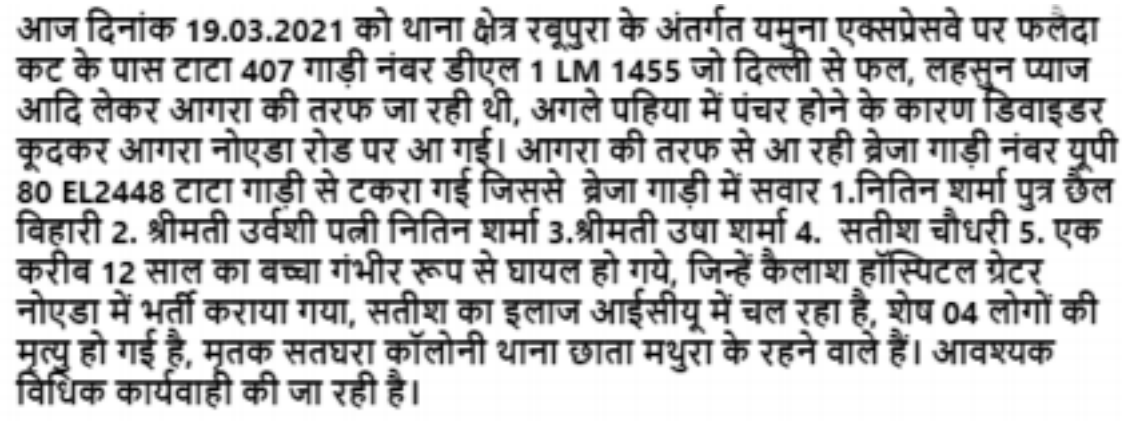
PII with screenshots

Below are some PII with screenshots I have taken different examples as in one of the media I

can observe it is a complaint with all the details containing address name and mobile number of the victim but posting these kind of information on twitter can be dangerous as people tend to take revenge when someone has filed a complaint against them

From the rest of the articles which I have posted below as examples contain the vehicle number along with the person name who did a crime by hitting the animals and it is a public information for telling the people about the assailant

Another articles displays the information about the car which had an accident on the main highway and details regarding the accident are shown in the media published.


आज दिनांक 19.03.2021 को थाना क्षेत्र रबूपुरा के अंतर्गत यमुना एक्सप्रेसवे पर फलैदा कट के पास टाटा 407 गाड़ी नंबर डीएल 1 LM 1455 जो दिल्ली से फल, लहसुन प्याज आदि लेकर आगरा की तरफ जा रही थी, अगले पहिया में पंचर होने के कारण डिवाइडर कूदकर आगरा नोएडा रोड पर आ गई। आगरा की तरफ से आ रही ब्रेजा गाड़ी नंबर यूपी 80 EL2448 टाटा गाड़ी से टकरा गई जिससे ब्रेजा गाड़ी में सवार 1.नितिन शर्मा पुत्र छैल बिहारी 2. श्रीमती उर्वशी पत्नी नितिन शर्मा 3.श्रीमती उषा शर्मा 4. सतीश चौधरी 5. एक करीब 12 साल का बच्चा गंभीर रूप से घायल हो गये, जिन्हें कैलाश हॉस्पिटल ग्रेटर नोएडा में भर्ती कराया गया, सतीश का इलाज आईसीयू में चल रहा है, शेष 04 लोगों की मृत्यु हो गई है, मृतक सतघरा कॉलोनी थाना छाता मथुरा के रहने वाले हैं। आवश्यक विधिक कार्यवाही की जा रही है।

Ans 2)

Response Time Statistics

```
In [246]: print(responsetimepolice.describe())

count    536.000000
mean       2.423950
std        4.955220
min        0.003333
25%        0.252778
50%        0.727083
75%        1.754306
max       59.961111
dtype: float64
```

Inferences, Justifications

Here we can observe that the average time taken by the police to reply on the tweets id close to 2.5 Hours which can be really dangerous in case of emergencies like fire,shooting, snatching which requires instantaneous reaction from the police but there can be cases which can be observed from the minm response time that police sees about the tweet and act to it but reply to it later.
25% tweets we responded within 15mins of tweeting whereas half of the tweets take more than an hour to get replied.

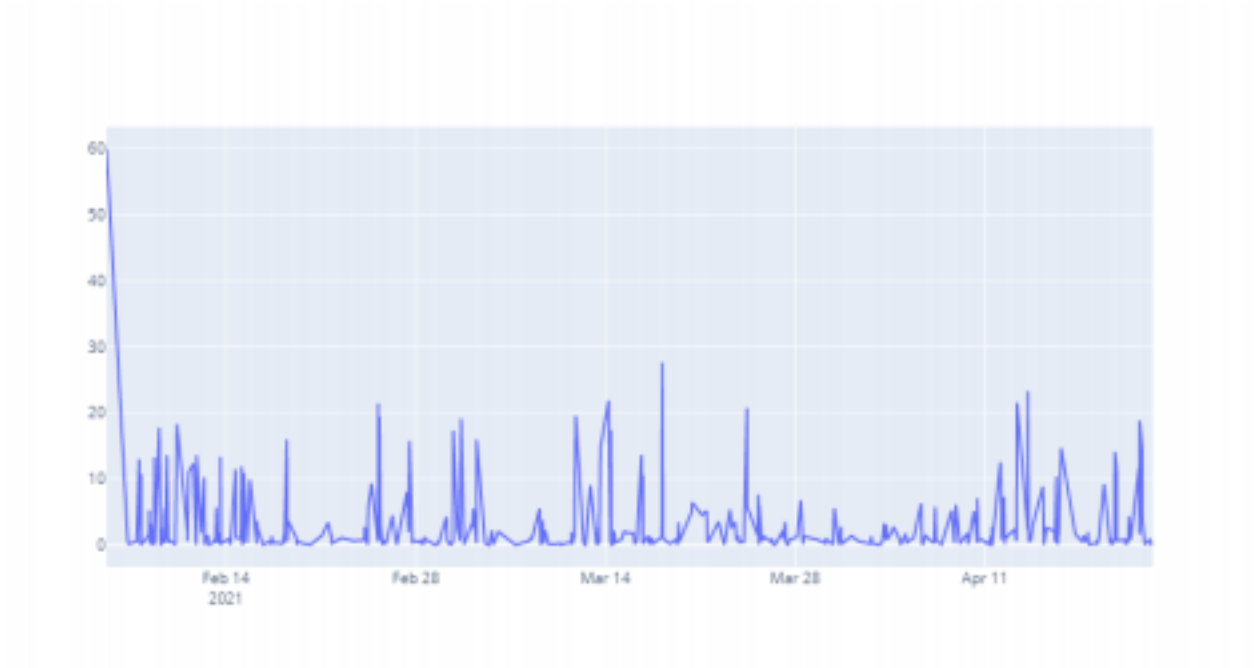
There can be cases that these 25% tweets can be crimes which require immediate action like fire,shooting,loot and the rest of the tweets which are replied late maybe for appreciation or accountability of the police which needs proper investigation before replying.

Methods can be adopted to reply back quickly based on segregating tweets and improving the enquiry time on twitter so that the mean time of replying can be brought down and the Noida police can be actively involved with the cases or reports done through twitter.

There can be efforts to make the twitter handle more active and responsive by having a dedicated team who checks the tweets regularly and reports the crime to the nearest police van near that location.

The quick response time can be also the case of reporting of road blocks during the farmers protest in the month of april and possible roadblocks in Noida as it acts as the border between Noida and UP

Time Series Plot



Comments

Here from the time series plot we can observe that the number of replies where way greater in the month of february and till mid march as lockdown restrictions were lifted and there were other mishaps and crimes being carried out in the city, then there is a sudden drop in end march and early April as most of the policemen were also deployed at various areas for handling covid and implementing the restrictions.

Recently mid april we can see spikes as due to the urgency in covid cases and scarcity of plasma and beds in the city, they are also trying their best to help the people and also educate them about the rules and lockdown guidelines so they are being replied to those tweets.

Ans 3)

Table of 30 Tweets with Codes

	id	text	Code
1	1.39E+18	Any lead for Plasma donor O+ blood group in Noida of a Coid positive person who recovered.@noidapolice @Uppolice @DeptGbn @CMOfficeUP	request
2	1.39E+18	Urgently need 2 units of plasma (any blood group) in Noida. Please contact at 9999419680. @aajtak @Dkamarchandelâ€¦ https://t.co/[illegible]	request
4	1.39E+18	@noidapolice pls help us ðŸ™ https://t.co/[illegible]	request
5	1.39E+18	Delhi, NCR - Request for plasma [illegible] @BabbarMridul https://t.co/cA2ncAafpx	request
9	1.39E+18	Om	request
13	1.39E+18	*Very	request
14	1.39E+18	@noidapolice please help https://t.co/[illegible]	other
15	1.39E+18	@vikashkatiyaris @nelovaoffice @DhirendraGBN @noidapolice @BloodDonorsIn @milking @PANKAJ PARASHAR_ @NoidaMediaClub Râ€¦ https://t.co/[illegible]	request
16	1.39E+18	Blood	request
21	1.39E+18	Urgently	request
26	1.38E+18	@Ashley_Ruskin One name, Ian Walker	other
29	1.38E+18	NEED	request
30	1.38E+18	#Noida	request
31	1.38E+18	@dingbn	request
35	1.38E+18	@PankajSinghBJP @vikhooshan @CMOfficeUP @112UttarPradesh @UPGovt @DeptGbn @CMHelpline1076 @CP_Noida @noidapolice Patâ€¦ https://t.co/[illegible]	other
41	1.38E+18	@Saurab	request
45	1.38E+18	Shooting incident in F block Sec-50 today morning. The person is in hospital. This is inspite of police patrolling.â€¦ https://t.co/[illegible]	accountability
49	1.38E+18	@PankajS	request
52	1.38E+18	@PrincessYoung Hi there [illegible]	other
65	1.38E+18	Bottas crashing has saved Hamilton so much	other
68	1.38E+18	@112UttarPradesh @noidapolice Babool our, Pasta Road, Sector 63A, https://t.co/TPM8pQ1PRj	other
84	1.38E+18	@ritaias2003 @ABPNews @noidapolice somewhere near Noida Extension and NH24 https://t.co/[illegible]	report
93	1.38E+18	RT @KatiAfterDark: Would you pull out or creampie me? [illegible] https://t.co/[illegible]	other
98	1.38E+18	ðŸ™	request
113	1.38E+18	@noidapolice Sir, I stay in Sec 78. Recently heard about few instances of looting, car snatching etc. in the area aâ€¦ https://t.co/[illegible]	report
118	1.38E+18	@UrvashiAnand @6 @CP_Noida @myogiadityanath @noidapolice Kindly look into this and take necessary action.	report
120	1.38E+18	@News18UP @noidapolice will find him, No worry.	request
121	1.38E+18	@Uppolice @noidapolice Open violation of night curfew . People dancing on main road on Noida Golf course . Is thisâ€¦ https://t.co/[illegible]	accountability

Modelling, Comments

Here I observed that tweets with words like violation, shooting, murder, crime, etc. are reporting or accountability tweets. Request tweets usually have words like please, action, etc. Using NLP too we can detect whether someone is asking for help or is holding police accountable for its actions.

@Uppolice @noidapolice Open violation of night curfew . People dancing on main road on Noida Golf course .: Accountability

Shooting incident in F block Sec-50 today morning. The person is in hospital. This is inspite of police patrolling. : Accountability

@noidapolice Sir, I stay in Sec 78. Recently heard about few instances of looting, car snatching etc. in the area : Report
I have so much Love for you but you keep Pushing me away and bringing me down.... wish I had luck with me fml.... : Other

Using this modelling, police can reduce its response times to incidents like fire, shooting if tweets are segregated by coding them through an algorithm. If someone holds the police accountable for something, they can take action on it and resolve the citizen’s grievance. All such steps help in improving efficiency and accountability of the police to the citizens and can bridge the communication gap that still exists between them.

Here we can use advanced technologies to get the keywords to report the importance of the tweet so that the police can react to it at the earliest and help avoid crime in the city.

It was strange to see none of the appreciation posts as police workers also devote a good

amount of time for the people in the city and specially during these tough covid times.

Section 2 :

Code and Data Collected

In [1]:
```
import tweepy,json
from tweepy import *

import pandas as pd
import csv
import re
import string
import preprocessor as p


consumer_key= "dKFZqHB8FyNl2ebbFjQGFPmkr"
consumer_secret= "wFj0DvFMfAWYdYS6ly3rVYY546GUXnWKhLgtSEbCjGkFTzF7kC"
access_token= "3220731312-SlujZpUvKJ4Q0ltlVAD8ZWBPeTSQUOlJwPN5VRI"
access_token_secret= "Y3KquGfdSwuZHyspdkie7Vt87H5VwkTX6uVmby5logt1A"

auth = tweepy.OAuthHandler(consumer_key,consumer_secret)
auth.set_access_token(access_token,access_token_secret)

api = tweepy.API(auth,wait_on_rate_limit=True)
```

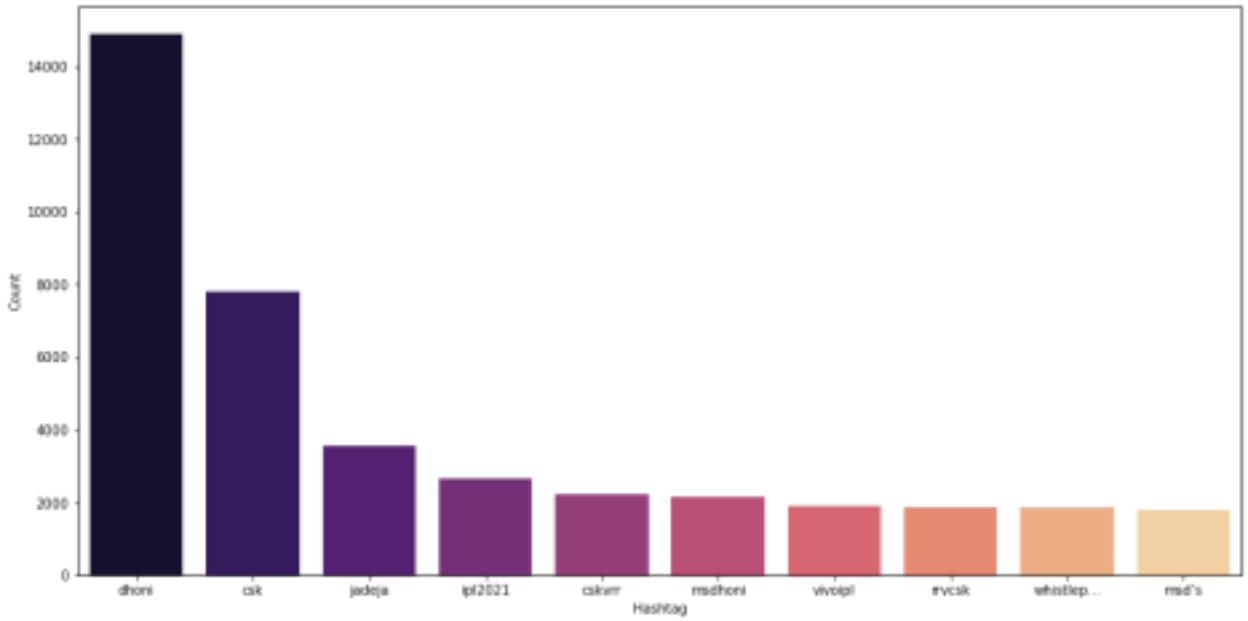
In [2]:
```
import tweepy
import datetime
import re
import tqdm


number_of_tweets=5000

tweetList=[]
tweetList = tweepy.Cursor(api.search, q='#dhoni', trim_user=True, tweet_mode = 'extended',lang='en').items(number_of_tweets)

tweets_json = [tweet._json for tweet in tweetList]
df_fin = pd.DataFrame.from_records(tweets_json)
```

Ans 1)

3 Plots of Top 10 Hashtags

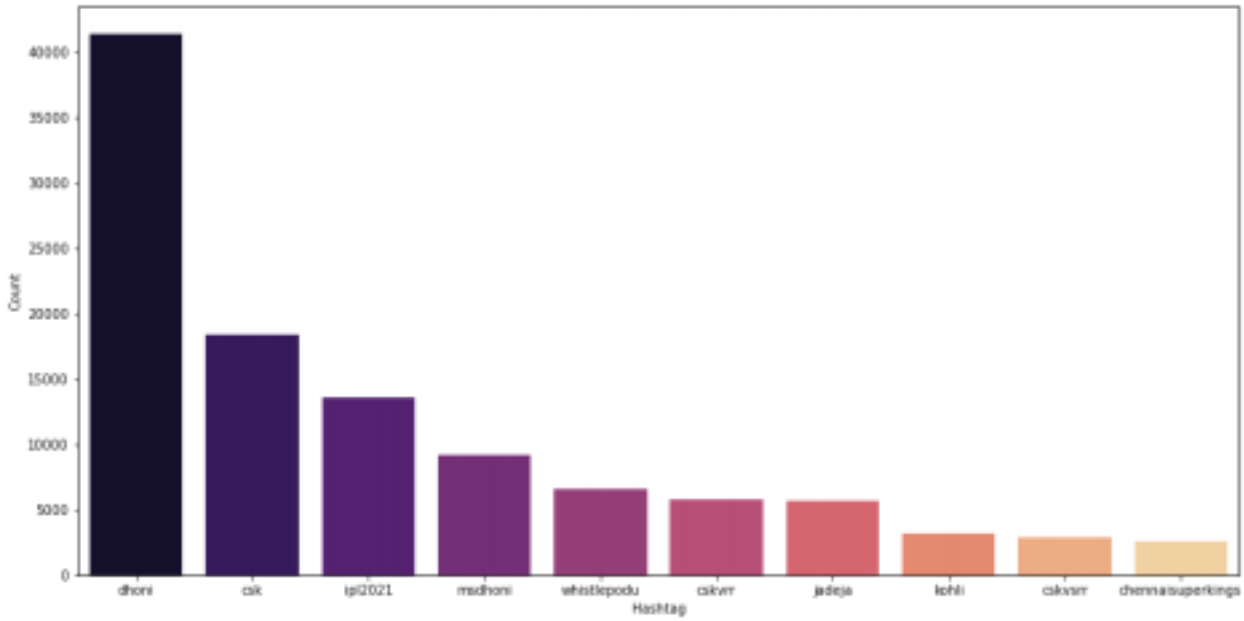
Based on occurrence

	Hashtag	Count
1	dhoni	14904
2	csk	7803
3	jadeja	3565
4	ipl2021	2649
5	cskvrr	2211
6	msdhoni	2138
7	vivoipl	1917
8	rrvcsk	1877
9	whistlep…	1861
10	msd’s	1808



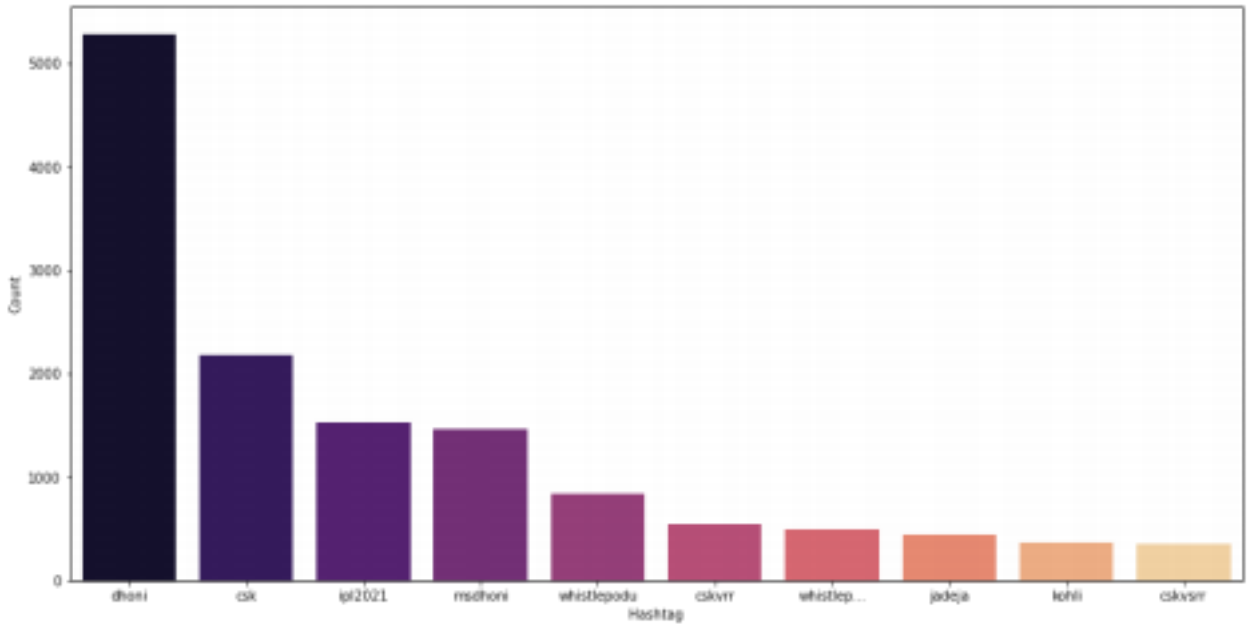
Based on likes

	Hashtag	Count
1	dhoni	41446
2	csk	18407
3	ipl2021	13598
4	msdhoni	9182
5	whistlepodu	6633
6	cskvrr	5754
7	jadeja	5719
8	kohli	3171
9	cskvsrr	2837
10	chennaisuperkings	2549



Based on retweets

	Hashtag	Count
1	dhoni	5286
2	csk	2185
3	ipl2021	1523
4	msdhoni	1464
5	whistlepodu	843
6	cskvrr	544
7	whistlep…	489
8	jadeja	444
9	kohli	370
10	cskvsrr	354



CTM Tables

	Hashtag_Retweet	ctm
0	(dhoni, 5286)	11.778050
1	(csk, 2185)	13.488348
2	(ipl2021, 1523)	18.235938
3	(msdhoni, 1464)	20.233863
4	(whistlepodu, 843)	25.015414
5	(cskvrr, 544)	22.179010
6	(whistlep..., 489)	22.902069
7	(jadeja, 444)	18.407499
8	(kohli, 370)	24.590442
9	(cskvsrr, 354)	23.360300

Inferences and Comments

Here, ‘rrvcsk’ and ‘whistlepodu’ have a very high CTM value suggesting that there might have been inorganic content generation with lots of retweets and lots of content by top 50 users to bring it into the trending section. None of the hashtags are unrelated to the topic (RR vs CSK IPL match in this case). Kohli is an unrelated hashtag since he’s a player of the RCB team.

Here from the other hashtags we can observe hashtags like kohli and jadeja, it maybe due to the fact that jadeja is being considered as the vice captain of the chennai super kings and kohli maybe due to the reason he is the captain of the royal challenger bangalore IPL team and also the captain of the current indian cricket team, so people often try to include Kohli in discussions related tot dhoni as people try to compare kohli and dhoni, dhoni being the one of the best cricket captains india ever witnessed because of winning the all the ICC tournaments.

Whistlepodu is the tagline of the chennai super kings so people tend to use this hashtag to support the CSK team and most other common hashtags include ipl2021 and ipl so ipl is considered as the cricket festival of india so people are very excited about it and so we can witness many of the posts containing these hashtags as currently ipl is going on and it can be a case that this will be dhoni's last ipl so people are more talking about ipl and dhoni.