PSOSM Last Assignment

Section 1 Code

For collecting tweets

```
auth = OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token,access_token_secret)
auth_api = API(auth, wait_on_rate_limit=True)

timeline = tw.Cursor(auth_api.user_timeline, id=user).items()
tweet_list=[]

for i in timeline:
    tweet=i.text
    if(tweet[:2]!="RT"):
        tweet_list.append(i)
```

Data collected for collected tweets (tweets by Delhi police) is in the filename: **section_1_tweets_collected.txt**

Q1

A)

Code for collecting reply tweets and getting a response time

```
re_time=[]
c=0
reply_tweets=[]
for i in tweet_list:
    if(i.in_reply_to_status_id!= None and i.in_reply_to_screen_name!= user ):#
        try:
        tt=auth_api.get_status(i.in_reply_to_status_id)
        reply_tweets.append(tt)
        re_time.append([i.created_at,tt.created_at])
        except:
        c=+1

print(len(re_time))
# print(c)
```

Data collected for collected tweets (tweets of the user who get reply from delhi police) is in the filename: **section_1_reply_tweets_collected.txt**

PII table Total PII detected = 46

PII type	Number of PII detected
Phone Numbers	40
Mail Ids	3
Aadhar card Numbers	0
Vehicle Numbers	3

Code

Number of tweets containing media out of reply tweets: 59

```
m = []
for i in reply_tweets:
    try:
      entities = i.entities['media']
      m.append(i)
    except:
      pass
print("Number of tweets which containing media")
len(m)
```

Number of tweets which containing media 59

B)

PII examples

Some Tweets containing PII out of 46

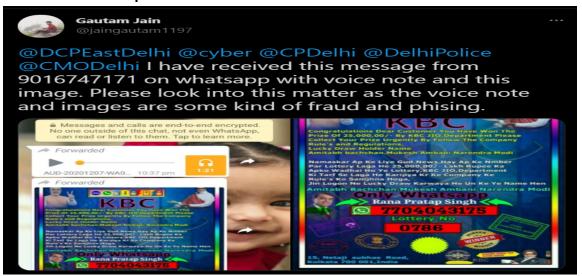
```
This tweet contains phone no @Hahacyber1 @Anilsinghvi_@cpcybd @CPDelhi @PMOIndia
9718201526 this number whatsapp calling me from #fraudloanapp. https://t.co/?KOEffcftH
7718201526 this number whatsapp calling me from #fraudloanapp. https://t.co/?KOEffcftH
7718201526 this number whatsapp calling me from #fraudloanapp. https://t.co/#GAEFG7138m
77181 tweet contains phone no @Hahacyber1@EPCPPDelhi @PATINGKeyIndia Irraud call from Delhi tol no 11622280000 to give Rs 1509009/- toward share of. https://t.co//GAEFG7138m
77181 tweet contains phone no @DCP_CCC_Delhi @cyber @RDI @phbindia I had a fraud of t10009/- by this fraud no. 8392083406 . on @PhonePe__ https://t.co/YpdfPflixy
77181 tweet contains phone no @DCP_CCC_Delhi @cyber @RDI @phbindia I had a fraud of t10009/- by this fraud no. 8392083406 . on @PhonePe__ https://t.co/YpdfPflixy
77181 tweet contains phone no 4012201970238 cach dady &TQ top to the two data to the
   महोदय मेरा मोबाइल no 9411866299 है
   8930593176 this is paytm number of fraud who is cloning face book nu... https://t.co/D8Zu3FDhtm
Contact if found : 9958461378
   @DelhiPolice... https://t.co/CHYaECpGOd
   This tweet contains phone no @OelhiPolice @CPDelhi @LtGovDelhi @HMOIndia Sir pls look into my complaint submitted in Lajpat Nagar police station . Arpit Sadh 9999444397
This tweet contains phone no Got a call from +916291572053 claiming to be from @TheOfficialSBI, Parliament Street regarding my pension account... https://t.co/22wMSU5K22
This tweet contains mail id @AndhraPradeshCM @DelhiPolice
This tweet contains mail id (MAndhraPradesnCM @URLINPOICE
george.angmostou122@mail.com
Fraud mail sending requests to people to purchase Ama... <a href="https://t.co/8hIPqKFqx9">https://t.co/8hIPqKFqx9</a>
Fraid mail sending requests to people to purchase Ama... <a href="https://t.co/shIPqKFqx9">https://t.co/shIPqKFqx9</a>
This tweet contains mail id [WoelhiPolice @CPDelhi R/Sir, I have received the attached document, my email-soni.rajender@gmail.com is wrongly men... <a href="https://t.co/phbplhipox">https://t.co/phbplhipox</a>
This tweet contains mail id [WoelhiPolice @CPDelhiPolice WoelhiPolice WoelhiPoli
```

Number of tweets that have Media and PII: 5 **Screenshots**

1. Contains PII is a phone Number



2. Contains PII is a phone Number



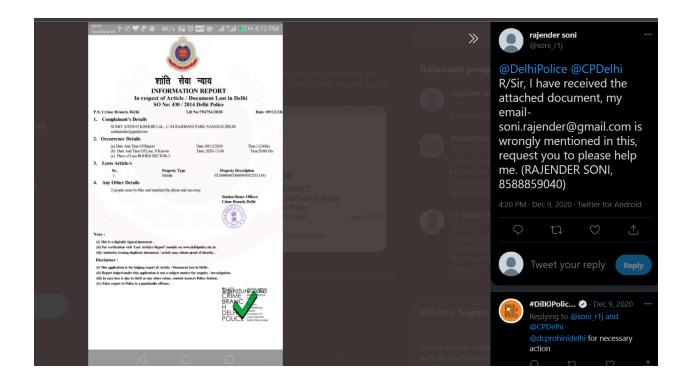
3. Contains PII is a phone Number and Vehicle Number



4. Contains PII is a phone Number



5. Contains PII is a Mail Id



Q2 A)

Maxmimum reponse time (min) 1438.55
Minimum reponse time (min) 0.7833333333333333
Mean reponse time (min) 423.1117708930962
standard deviation of reponse time 430.22786655298194

Inference:

As we can see min response time is less than the 60s it means sometimes they respond very fast but sometimes they take nearly a day to repose(1438.55 min = 23.96 hr approx a day).

And their avg response time is 423min approx 7hr which shows that their reply is done by a real person not by bot (automated reply)

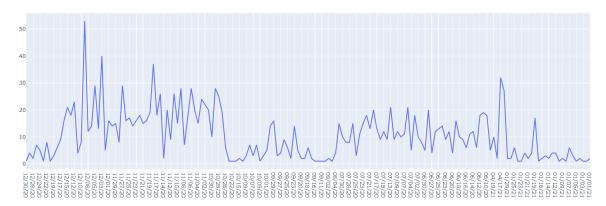
The standard deviation is 430.22 approx 7hrs

Justification:

According to average response time, and other stats and there is no pattern in response time so we can say that these replies are not automated generated and it's handle by humans.

B)Time series Plot X-axis= date, Y-axis = Number of reply on that day

['12/30/20', '12/29/20', '12/28/20', '12/27/20', '12/27/20', '12/24/20', '12/23/20', '12/22/20', '12/20/20', '12/19/20', '12/18/20', '12/17/20', '12/17/20', '12/18/20', '12/1



Comments:

The data is from 9th June 2020 to 29th Jan 2021

The peak is on 9th December 2020. This is because of the farmer protest during this time which leads to many Riots in India.

Many people have asked Delhi police a lot of enquires about farmer protest by tweeting and tagging Delhi police in their tweets so that's why the number of reply on that day is high 53 replies by the @DelhiPolice handle.

Q3

	Tweets	Code
1	We are very afraid with #DelhiPolice that we logged a complaint against dowry but Delhi police is not responding on case. Where we will go. @CMODelhi @DelhiPolice4 we logged a complaint in @Burari02 burari Amrit Vihar Nagar. My husband is beating me everyday @DelhiPolice	Accountability
2	Disappointed by @noidapolice, i walked to them for help yesterday night but they refused to help. Just because I am a delhi women. This is kind of help you expect from police in india. @DelhiPolice @ArvindKejriwal @myogiadityanath	Accountability
3	@CPDelhi Sir its insane behavior of this guy	Others
	Give him half of what he given to poor dog, let's see what he is going to do	

		Π
	@TajinderBagga Bhai pls check with @DelhiPolice and make sure this guy gets same treatment https://twitter.com/lshitaYadav/status/1337257502354771969	
4	LNJP hospital delhi has been giving worst treatment to Covid patients, my father-in-law dies today (Nandlal Batra) due to negligence of CCU ward staff, they had also stolen my father-in-law's mobile phone, @ArvindKejriwal @narendramodi @AmitShah @manishsisodia90 @CPDelhi	Crime report
5	'United Against Hate' have opened kitchens in Delhi to ensure that no one goes hungry at the #FarmersProtests.	Others
	For your information, 'United Against Hate' is an organization founded by Umar Khalid & has also been named in Delhi riots chargesheet by the Delhi police. Kuch samjhe?	
6	Corona Timebomb ticking as Connaught Place badly invaded by IllegalHawkers putting our shops,staff & shoppers at risk as they crowd the pedestrian corridors,dont use masks, flout social distancing.We await @DMNewDelhi to involve @tweetndmc & @DelhiPolice to ban this nuisance ASAP	Others
7	How stupid can someone get. These are waiting death traps. Whosoever installed them should be booked for attempt to murder. @CPDelhi @DelhiPolice @dtptraffic	Crime report
8	@CMODelhi @ArvindKejriwal @Delhi_Pollution @DelhiPolice this is really disgusting act of notorious elements who have burnt garbage in open land in such large scale. Location is G T Karnal Road Azadpur Industrial area A block delhi(adjecent to Dilkhush complex	Others
9	@DelhiPolice see the fire burning behind E2 Vasant Kunj by Mahipalpur villagers !!!	Others
10	A person very confused with no reply from Call No. 100, 1090,1091, SHO Sarai Rohilla Lokendra Singh, and lady beaten by her husband. Please let us know who are responsible for this. @CPDelhi @DelhiPolice @LtGovDelhi	Crime report, Request
11	@DelhiPolice @DCPNWestDelhi unnecessary 24X7 barricading on Durga chowk PS premnagar. Road is already in bad condition and no adequate way for passing vehicle also no policemen is here at barricade.	Accountability
12	Retweeting of unfortunate event happened with me on 29th june evening. My car was attacked violently by some robbers who tried to break the window pane .They snatched my mobile phone which had official data. Please take timely action @DCPNEastDelhi @DelhiPolice @HMOIndia @CPDelhi	Crime report, Request
13	The situation is getting worst here in Badarpur today in the morning somebody has stolen our car and police is not giving any attention to our complaints Nobody is here to help us.	Accountability, Crime report
14	@AmitShah,@DelhiPolice @narendramodi, @ArvindKejriwal HELP MANOJ VERY POOR DALLU PURA FRAUD BY SOME ONE WITHDRAW 49900 HIS ACCOUNT AND MONEY IS TO BE USED MEDICAL TREATMENT OF HIS FAMILY MEMBER COMPLAINT REGISTERED STATION PL HELP. DETAILS COMPLAINT ATTACHED	Crime report
15	@DCP_CCC_Delhi @CPDelhi I was badly affected by a paytm fraud scam occurred on 5th November, 2019 through a scamster somewhere in Jharkhand, Deonar. An FIR was lodged at Geeta Colony PS No. 386/2019 22.11.19 also lodged complaint cyber department.	Crime report

	Progress slow,pl help expedite.	
16	Whosoever know this guy, kindly reach at him or try to contact his parents. He has deleted his account guys. Please help if you know him See the mental pressure Waheguru.@DelhiPolice	Request
17	Hey @CPDelhi @DelhiPolice @Cyberdost this man @fullpowr24Hour is abusing me and giving me threat. Kindly do the needfull. Hope u will respond ASAP.	Request, Crime report
18	@Paytmcare इस एकाउंट से फ्रॉड हो रहा है kindly check, ये सबसे महंगे मोबाइल के नाम पर तीन तीन हज़ार रुपये वसूल रहा है, और फिर किसी को जवाब नहीं, रोज़ कितने लोग धोखे के शिकार हो रहे हैं, कृपया आवश्यक कार्यवाही करें, @TheLallantop @paytmbankcare @helpline_BP @DelhiPolice @Uppolice	Request, Crime report
19	#DilKiPolice @CPDelhi @PMOIndia Last Monday 21st september my gold chain snatched from Vikaspuri near banasthali park from main road at 6:50am Till no clue has been found and my mother thinks every case gets solved by Delhi police courtesy crime petrol.	Accountabili ty, Crime report
20	@DelhiPolice @CPDelhi @ArvindKejriwal @AmitShah मेरी महिला मित्र को आईपीएस तृप्ति भाट नाम के फेक अकाउंट से अश्लील तस्वीर व मैसेज भेजे जा रहे हैं, अतः आपसे निवेदन है कि आप संज्ञान में लेने की कृपा करें @chitraaum @DCPSEastDelhi	Request

21	@DelhiPolice sir meri sister ko breast cancer hai aur uska pati ilaaj nahi karwa raha hai aur maar pit raha bhi hai sir meri sister kisan garh thana me complain likhwane gayi to waha par complain nahi likhe aur bhaga diye sir please help kariye	Accountability , request,Crim e report
22	Sangam Vihar ND-80 Nowadays boys harass girls and women very much on B Block SS Public School Road. The Delhi Police is requested to please take cognizance of this otherwise it may take the form of an incident. @ndtv @DCPSouthDelhi @CPDelhi @cp_delhi @ADcpsouthdelhi thanks	Request, Crime report
23	15 yr old Daughter of my house help has been missing from 11 September 2020. No help by investigating officer. I doubt even fIR has not been registered yet. No one listening to her repeated pleas. Please help @DelhiPolice @dcpouter @DCWDelhi @NCWIndia @LtGovDelhi	Accountabilit y
24	It's my birthday today, it would be a great gift for me if you guys can help me in finding my car by today. It will make my day for sure and I'll be thankful to you for my entire lifetime. @DelhiPolice @CPDelhi	Request
25	This lady has been abandoned by her son at Tilak Nagar, Delhi today morning. Pls can kind Delhites help her with food and accommodation Person in Delhi to be contacted Mr Abhishekh. Ctc nos 8178310904 @TajinderBagga	Request

	#Help #Delhi	
26	आदरणीय पुलिस आयुक्त दिल्ली श्री एसएन श्रीवास्तव साहब तीन मोबाइल फोन चोरी होने की FIR, T3 एयरपोर्ट पुलिस स्टेशन पर बार बार चक्कर लगाने के बावजूद दर्ज नहीं की गई है मेरा आपसे अनुरोध है कि एसएचओ को निर्देश दे कि FIR दर्ज करेंI @CPDelhi @DelhiPolice @HMOIndia @DCP_CCC_Delhi @DCPIGI	Accountability
27	One of our representative had talked to Dr.Rajeev M.S Maharishi Valmiki hospital where a doctor was assaulted. According to M.S FIR has been registered against the assaulter. @DelhiPolice plz let us know the status of FIR.	Others
28	SaiBaba चौक से विनोबा कुंज Apt, (Sec-9, Rohini, Delhi) तक सड़क पे काफी चैन स्नेचिंग व लूटपाट की लगातार घटनाएं हो रही है, कुछ तो कीजिए साहेब. @CPDelhi @DelhiPolice @ArvindKejriwal @LtGovDelhi @AmitShah @narendramodi @DcpNorthDelhi	Request, Crime report
29	Team from PS Vikaspuri acting swiftly has arrested four notorious gangsters and recovered 5 sophisticated pistols and 65 live cartridges, thus averting a major gangwar. Together they are previously involved in 9 cases of serious nature. @LtGovDelhi @CPDelhi @DelhiPolice	Appreciation
30	Tattoo Artist arrested for attempting to kill brother in law to terrorise Ex wife to take custody of childteam rewarded by Spl CP Sh Satish Golcha sir and Jt CP Sh Suvashish Chaudhry sir @DelhiPolice @CPDelhi @LtGovDelhi @SChoudharyIPS	Appreciation

Modelling:

For modelling we can use sentiment analysis on tweets text on reply_tweets(tweets on which DelhiPolice reply) using thresholding for each category like positive sentiment shows appreciation and -ve shows crimes or request.

we can use bag of words approach for finding top most occurring word in each of four category or we can categorise those manually by checking type of words like words like help comes in category of request and murder, robery etc comes in category of reported crime like it we can do same for other two category.

Comments:

Police can sort tweets dependent on a couple of significant classifications like Crime report, request and accountability which will help them sort significant or important tweets first.

After that point, they can go for other tweets like appreciation. so, the first priority for reply goes to the first category than to the appreciation category.

If they will reply fast to tweets with the first category which contain crime_report, request and accountability it will make sure that people will believe them and they will feel much better after getting a response as fast as possible.

Section 2

Code

For checking top trends Hashtags

```
auth = OAuthHandler(consumer_key, consumer_secret)
    auth.set access token(access token,access token secret)
    auth_api = API(auth, wait_on_rate_limit=True)
[ ] trend_place=auth_api.trends_place(23424848)
    trend_place=trend_place[0]['trends'][:10]
    for i in trend_place:
      print(i['name'])
      if(c==0):
        prom_hast=i['name']
    # print(prom_hast)
    #CSKVRR
    Dhoni
    #VaccineForAll
    #ManmohanSingh
    Help Like Sushant
    #HBDTeluguPrideCBN
    McDonald
    #HAPPY BIRTH JAY
    Thala
    Sam Curran
```

Above are the top 10 trending hashtags

Code for finding tweets related to prominent hast The most prominent one is #CSKvRR

Tweet and id of the most prominent hashtag is in the

filename:section_2_trends_tweets_collected.txt

```
trend_tweets=[]
c=0
for i, j in enumerate(tw.Cursor(auth_api.search,q=prom_hast).items()):
    c+=1
    trend_tweets.append(j)
    if(c>999):
        break
```

Code for finding top 10 hashtags from those collected tweets:

```
hast=[]
for i in range(len(trend_tweets)):
    s=trend_tweets[i].text.split(" ")
    if(len(s)>0):

    for j in range(len(s)):
        if(len(s[j])>0):
        s[j]=s[j].lower()
        if(s[j][0]=="#"):
            if(prom_hast not in s[j] and "\n" not in s[j]):
            hast.append(s[j])
```

```
[ ] hast_count={}
  for i in range(len(hast)):

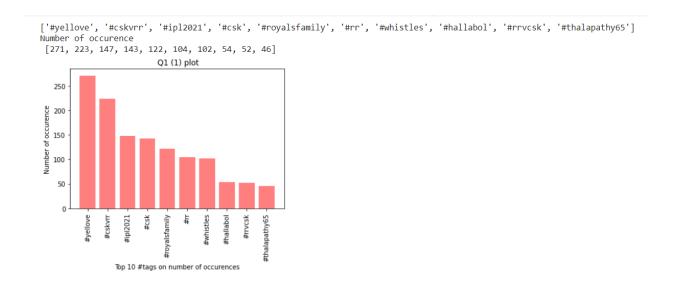
    if hast[i] in hast_count:
        hast_count[hast[i]]+=1
    else:
        hast_count[hast[i]]=1
    sort_hast = sorted(hast_count.items(), key=lambda x: x[1], reverse=True)[1:11]
    sort_hast
```

Top 10 hashtag from tweets of the most prominent hashtag:

```
[('#yellove', 451),
  ('#cskvrr', 299),
  ('#ipl2021', 116),
  ('#csk', 37),
  ('#royalsfamily', 25),
  ('#whistles', 16),
  ('#hallabol', 14),
  ('#master', 13),
  ('#rr', 12),
  ('#vivoipl', 11)]
```

Tweet and id of these top 10 hashtags is in the filename:**section_2_tweets_collected.txt** 100 tweets collected for each hashtag.

Q1
1)
Bar plot of top 10 hashtags on the basis of the number of occurrences.



CTM table

0th row contains hashtags.

1st row corresponds to R

2nd row corresponds to F

3rd row corresponds to \boldsymbol{U}

4th row corresponds to the C (coefficient of traffic manipulation).

	0	1	2	3	4	5	6	7	8	9
0	#yellove	#cskvrr	#ipl2021	#csk	#royalsfamily	#rr	#whistles	#hallabol	#rrvcsk	#thalapathy65
1	97.8182	81.2727	50.9709	71.9941	96.8944	49.3243	100	91.1111	70.1754	89.0909
2	40	31.2727	56.3107	26.6862	86.3354	60.1351	50.9804	78.3333	100	100
3	1.27907	1.51515	1.54887	1.46667	2.23611	1.3578	1.02	2.06897	1.54054	1.96429
4	51.0609	40.9152	62.9566	35.3523	98.261	66.4254	62.0004	89.5134	108.558	110.873

Comments:

Here are all hashtags that are related to the most prominent one #CSKvRR except #thalapathy65 Which is from inorganic tweets these tweets are just for promotion of south actors and their movies.

They are using #tags related to prominent one #CSKvRR for getting on trending pages.

#yellove ha highest number of occurrence

#thalapathy65 has lowest number of occurrences

Inference:

(10) #thalapathy65 - Twitter Search / Twitter

We can check this #tags name #thalapathy65 which is not related to the match or #tags #CSKvRR this #tags may be used by spammers for promotion of south actors and their movies during trend of #CSKvRR.

Second it may be advertisement by fans for their upcoming movies etc

That's why #thalapathy65 has the lowest number of occurrences among all as it's inorganic.

2)
Bar plot of top 10 hashtags on the basis of the number of likes.

```
['#royalsfamily', '#vivoipl', '#ipl2021', '#rr', '#csk', '#ipl2021...', '#rrvcsk', '#whistlepodu', '#cskvsrr', '#msdhoni']

Number of occurence
[1175, 392, 295, 72, 30, 9, 7, 4, 4, 3]

Q1 (2) plot

Output

Output
```

CTM table 0th row contains hashtags. 1st row corresponds to R 2nd row corresponds to F 3rd row corresponds to U

4th row corresponds to the C (coefficient of traffic manipulation).

	0	1	2	3	4	5	6	7	8	9
0	#royalsfamily	#vivoipl	#ipl2021	#rr	#csk	#ipl2021	#rrvcsk	#whistlepodu	#cskvsrr	#msdhoni
1	96.8944	63.6364	50.9709	49.3243	71.9941	0	70.1754	95.1389	30.303	35.7143
2	86.3354	84.8485	56.3107	60.1351	26.6862	100	100	38.1944	100	100
3	2.23611	1.1	1.54887	1.3578	1.46667	2	1.54054	1.26316	1.22222	1.75
4	98.261	92.3121	62.9566	66.4254	35.3523	102	108.558	48.9715	104.253	105.321

Comments:

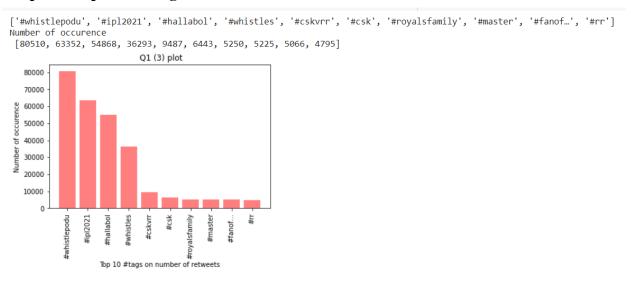
Here are all #tags are related to #CSKvRR there is no inorganic #tags is there
Just because people are not liking tweet which is not related to #CSKvRR.
#royalsfamily has highest number of like it means people are liking RR teams related tweets.
#msdhonoi has the lowest number of likes. may be fans are not happy with his performance that's why they are not liking this #tag so much.

Inference:

As there is no inorganic #tags from inorganic tweets it happens because people are does not liking tweets which are not related to #CSKvRR like.

As this plot is for top 10 #tags on number of likes so it's highly possible that Only Fans of CSK and RR are liking the tweets that's and it obvious they are not gonna like any inorganic tweets. That's why there is no #tags is there which is not related to #CSKvRR. Users like RR tweets as compared to CSK related tweets.

3) Bar plot of top 10 hashtags on the basis of the number of retweets.



OTM table Oth row contains hashtags. 1st row corresponds to R 2nd row corresponds to F 3rd row corresponds to U

4th row corresponds to the C (coefficient of traffic manipulation).

	0	1	2	3	4	5	6	7	8	9
0	#whistlepodu	#ipl2021	#hallabol	#whistles	#cskvrr	#csk	#royalsfamily	#master	#fanof	#rr
1	95.1389	50.9709	91.1111	100	81.2727	71.9941	96.8944	93.8144	100	49.3243
2	38.1944	56.3107	78.3333	50.9804	31.2727	26.6862	86.3354	100	100	60.1351
3	1.26316	1.54887	2.06897	1.02	1.51515	1.46667	2.23611	2.425	10	1.3578
4	48.9715	62.9566	89.5134	62.0004	40.9152	35.3523	98.261	111.806	120	66.4254

Comments:

Here are all hashtags that are related to the most prominent one #CSKvRR except #master Which is from inorganic tweets these tweets are just for promotion of south actors and their movies. They are using #tags related to prominent one #CSKvRR for getting on trending pages. #whistlepodu has highest number of retweets #rr has lowest number of retweets

Inference:

(10) #MASTER (@MasterOfficial) / Twitter

We can check this #tags name #master which is not related to the match or #tags #CSKvRR this #tags may be used by spammers for promotion of south actors and their movies during trend of #CSKvRR.

Second it may be advertisement by fans for their upcoming movies etc

#whistlepodu has highest number of retweets as they won that match that's why people retweeting it

#rr has the lowest number of retweets they have lost that game that's why maybe their are the lowest retweets on it.