

Section 1

- code and data attached

Answer 1

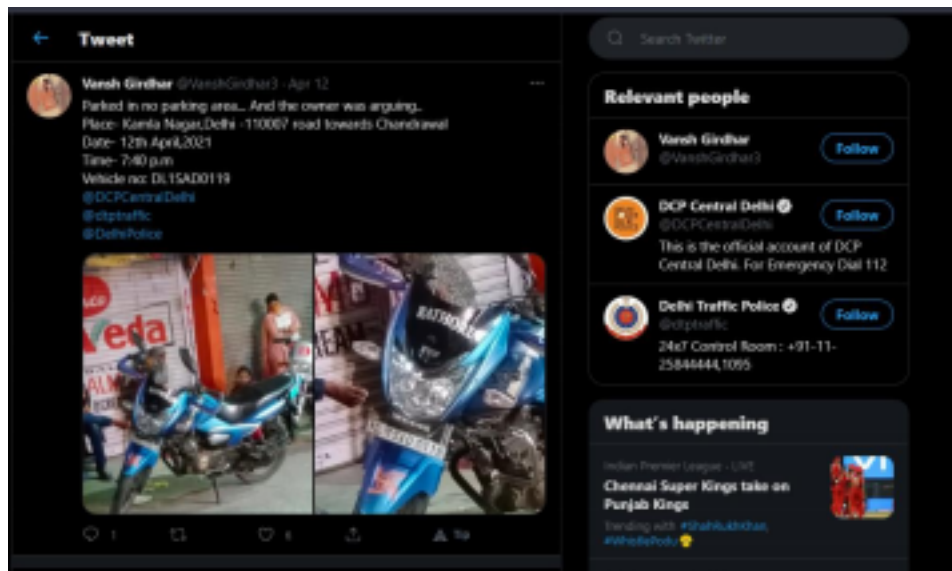
Twitter handle used: "dtptraffic"

Number of tweets collected: 3250

A.

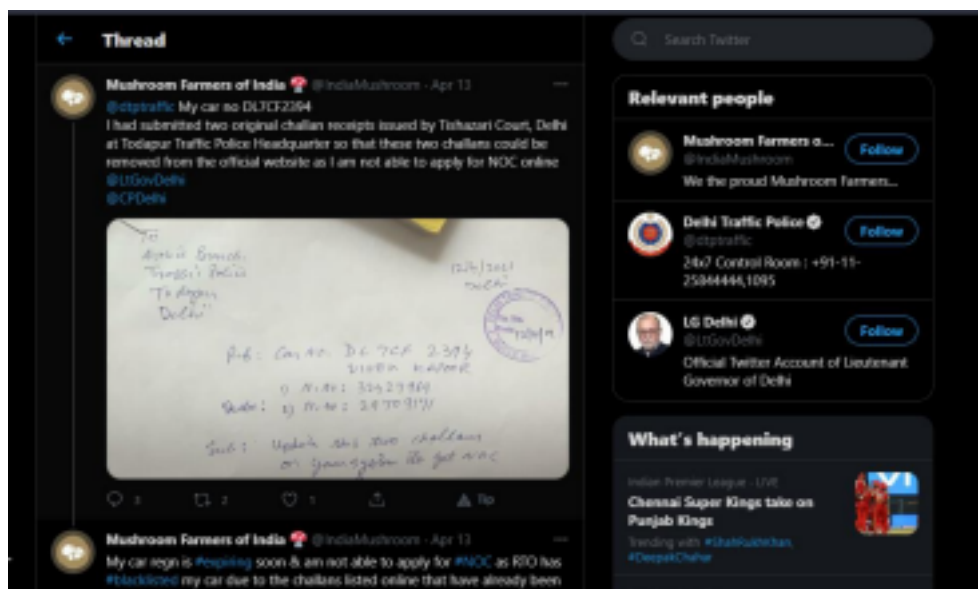
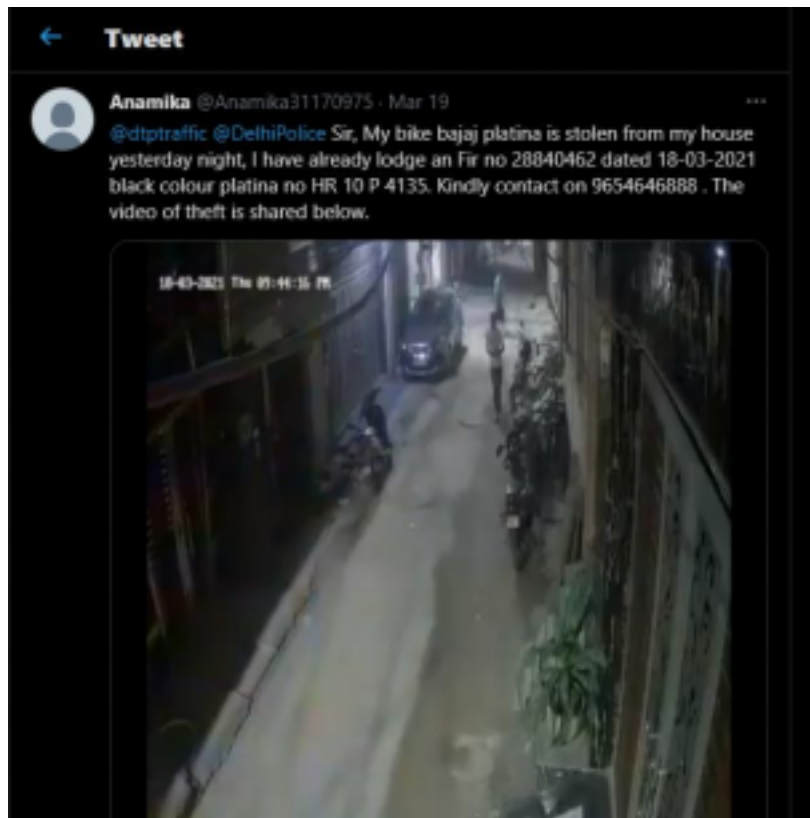
```
Phone numbers: 11
Adhar numbers: 0
Emails: 3
Pin Codes: 56
Vehicle Numbers: 141
credit cards: 0
```





B. out of 3250 tweets, 1639 contain media

```
Phone numbers: 2
Adhar numbers: 0
Emails: 0
Pin Codes: 33
Vehicle Numbers: 125
credit cards: 0
```



Answer 2

Min:1

Max:86190

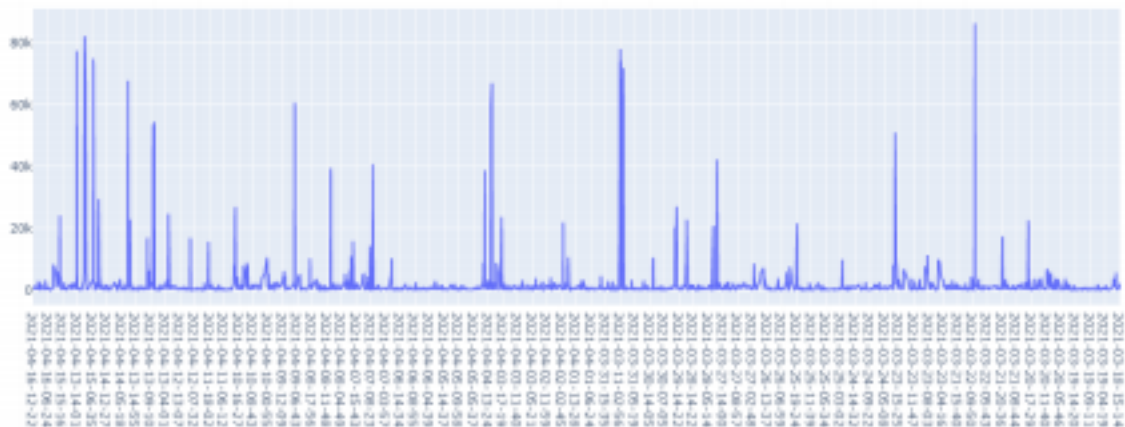
Average:1876

Standard deviation: 6437

Inferences:

1. As the minimum response is of 1 second, it indicates that this tweet must have been replied by a bot as it is not possible for police to resolve the query and reply to a tweet in one second.
2. Max response time is 1 day which indicated these replies must have been from a human.
3. The handle is controlled by humans but some tweets may require generic reply which may be done by a bot.
4. The average response time is coming out to be 30 minutes which is very efficient considering that police might be getting many such queries. Also, this is great way of comparing the efficiency police department with one another.
5. The maximum time taken is around 24 hours which shows that for some queries the police department might require more time to get to the solution for the query.

B.



Inferences:

1. Most of the tweets are resolved in a reasonable amount of time.
2. Also, the response time has reduced in times on the current pandemic indicating police department are putting in more efforts for the betterment of the society.
3. Some tweets required more time to get resolved as compared to others. This may be due to the fact that some tweet may require steps that need time to get executed.
4. Also, by looking at the text of the tweet, it is noticed that the response time is significantly less if more info is provided like, phone number, vehicle number, address where the action is to be taken etc.
5. Also tweets that weren't clear or didn't have necessary info had larger response times.

Answer 3

	id	text	label
0	1.385030e+18	@dtptatic @dtptatic is Tikt Border open f...	assistance
1	1.382990e+18	@dtptatic @CPDelhi Dear Sir, intrastate tra...	assistance
2	1.385030e+18	@dtptatic sir they are not using mask and no...	report
3	1.382940e+18	@DelhiPolice are we allowed to travel from NCR...	assistance
4	1.382970e+18	@dtptatic @DCPNEastDelhi Sir I need to take...	assistance
6	1.382940e+18	@Virender9299 @dtptatic @ArvindKejriwal @UG...	report
6	1.382960e+18	@DelhiPolice @dtptatic Dear Sir, We are an ...	assistance
7	1.385020e+18	@dtptatic @ghazabadpolice @ArvindKejriwal C...	request
8	1.382930e+18	@dtptatic @CPDelhi Sir, what I have to do if...	assistance
9	1.385010e+18	@dtptatic one police man drive without helme...	report
10	1.382960e+18	@dtptatic Whether People working on contract...	assistance
11	1.382960e+18	@dtptatic @nada_authority is a private vehi...	assistance
12	1.382960e+18	@CPDelhi @dtptatic @DelhiPolice Dear Sir, PL...	request
13	1.385020e+18	@dtptatic Govt vehicle using for family mem...	report
14	1.385030e+18	@dtptatic stop line crossing, TV lower red l...	report

0s completed at 11:41 PM

A.

1. Request: containing keywords like please, can you, would you, could you, kindly etc. we can simply look through the sentences and label them as request if these phrases are present. This category of tweets will help the authorities gain trust of the people if they resolve the issues in these tweets in reasonable time.
2. Report: in these tweets, location of the incident, time, phone numbers, email id and other relevant data will be contained. Looking at the text of the data and checking for such details will help us label them as report. This category can also be used to calculate the crime rates in different areas. Areas having more tweets in report category would imply they are having high crime rate. This would help police department to make necessary arrangement in such areas.
3. Accountability: this category of tweet would contain text that would sort of put blame on police department for the conditions of the people. The text will show the disappointment of people on not keeping the expectations of people. These would also contain people complaining about the time taken by police department to respond and the outcomes of that late respond. This would force the department to work harder and reduce the number of tweets that fall in this category.
4. Appreciation: these would contain positive comments regarding the police department like thank you comments for solving the crimes. Looking for appraisal words like thank you, great work, keep it up, proud of police etc. can be used to tag under appreciation. These tweets will generally be as reply to tweets after the police solves the crime. This category of tweets will encourage department to work even harder.
5. Assistance: phrases like what, where, how, am I allowed, can I do this, etc. these are tweets asking the police department for some guidance on various issues like asking whether the traffic in some area is under control or not.
6. Others: this category would include tweets that are incomplete or lack necessary details. Also looking at these tweets manually, I can say that these tweets won't require the urgent

assistance from the police department.

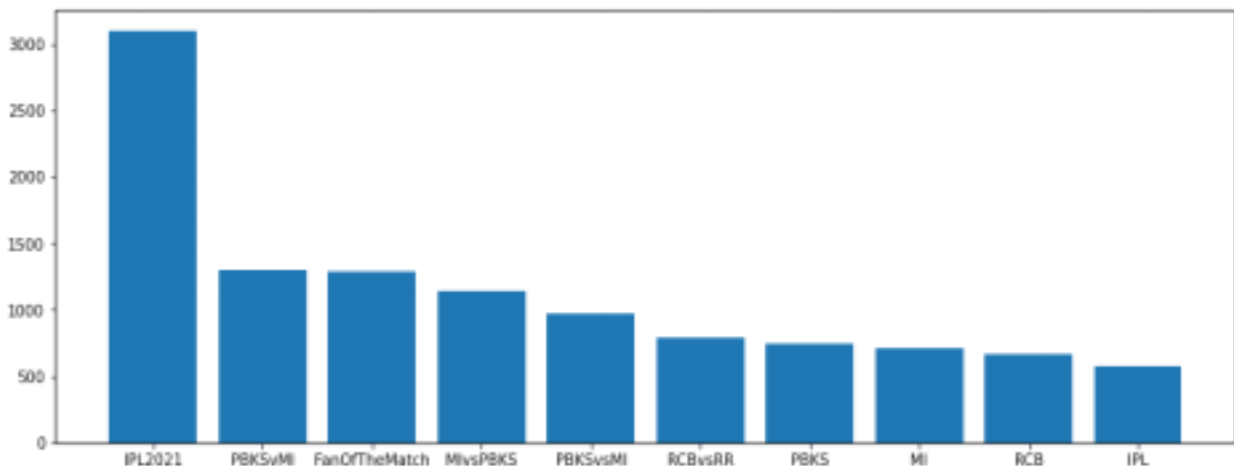
B.

1. After labelling the tweets it will become easier for police department to decide which tweets requires more urgent assistance. It would then help as the response time will become low and the efficiency will increase. This will generate more trust in people towards the police department thus reducing the communication gap between the department and the public.
2. more appreciation tweets will motivate the department to push its limit and increase its efficiency. This would imply more crimes will be solved and trust will increase and thus communication gap will decrease.
3. In accountable category people will complain about department taking more time to solve the crime or to respond which will indicate to the police that they need to work harder and respond early to gain trust of people.
4. Department will be able to better manage their time and devote more time to report, accountable, assistance category of tweets and thus will decrease the communication gap. 5. Also educating people on how to better use these resources for the betterment of the public can help in increasing the efficiency and decreasing the communication gap. Like not tweeting for small problems and tweeting for major problems with right necessary details so that proper action can be taken.

Section 2

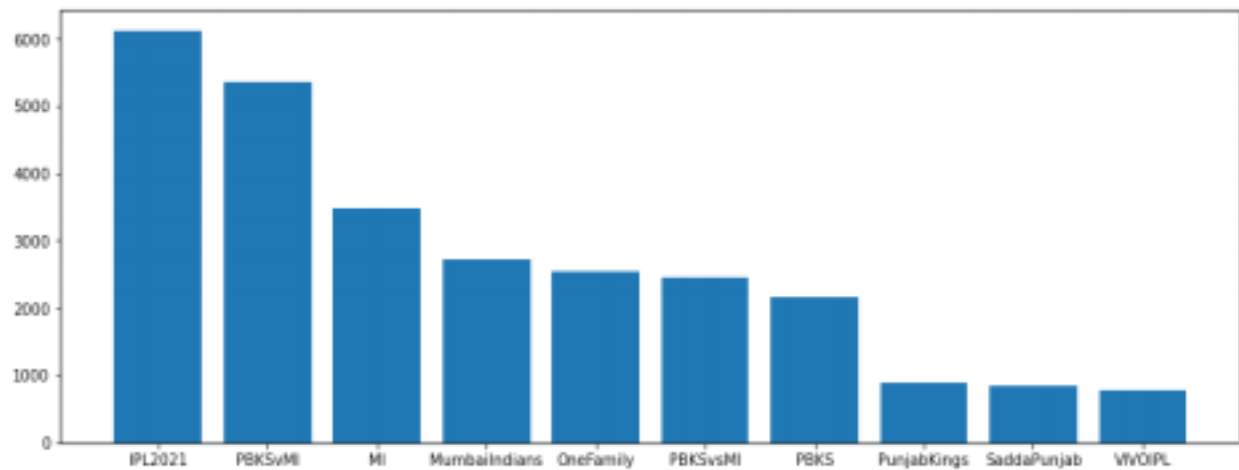
used-> ipl

➤ Based on number of occurrences



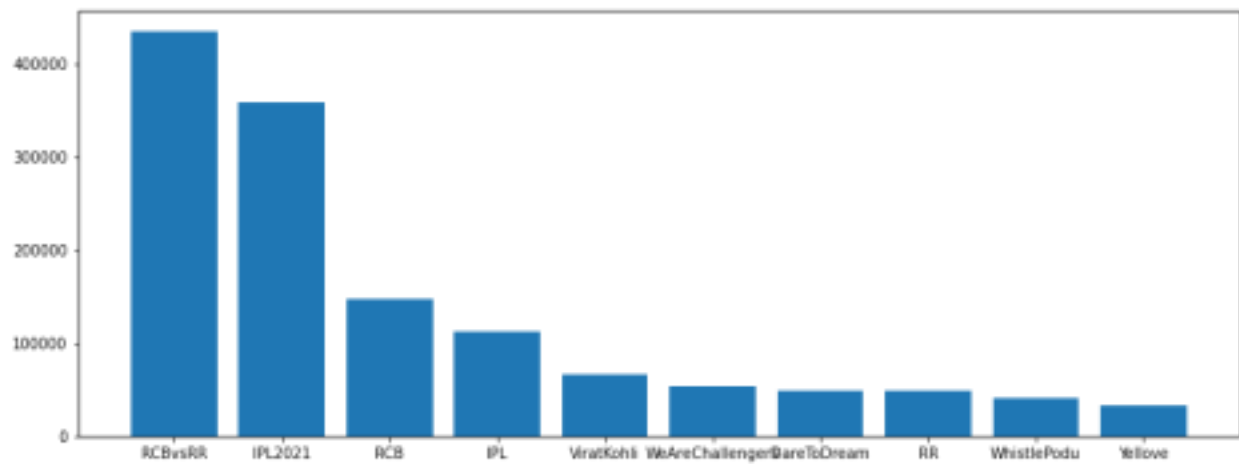
	Hashtag	R	F	U	CTM
0	IPL2021	40.897353	21.755972	2.249818	28.095525
1	PBKSvMI	35.990712	34.365325	2.153333	40.117730
2	FanOfTheMatch	2.321981	42.337461	3.800000	46.369659
3	MIvsPBKS	15.553603	45.342707	2.888325	49.786392
4	PBKSvsMI	32.507740	36.945304	1.697023	41.893101
5	RCBvsRR	92.484076	19.490446	1.256000	29.994854
6	PBKS	37.290503	40.363128	1.956284	46.048463
7	MI	55.947137	39.794420	1.830645	47.219779
8	RCB	85.779817	29.357798	1.406452	39.342231
9	IPL	62.807018	34.210526	1.496063	41.987291

➤ Based on number of likes



	HashTag	R	F	U	CTM
0	IPL2021	40.897353	21.755972	2.249818	28.095525
1	PBKSvMI	35.990712	34.365325	2.153333	40.117730
2	MI	55.947137	39.794420	1.830645	47.219779
3	MumbaiIndians	63.475177	64.539007	1.880000	72.766525
4	OneFamily	69.257951	70.671378	2.176923	79.774096
5	PBKSvsMI	32.507740	36.945304	1.697023	41.893101
6	PBKS	37.290503	40.363128	1.956284	46.048463
7	PunjabKings	57.819905	87.203791	2.740260	95.726042
8	SaddaPunjab	64.772727	89.772727	3.428571	99.678571
9	VIVOIPL	53.465347	87.128713	1.603175	94.078422

➤ Based on number of retweets



	HashTag	R	F	U	CTM
0	RCBvsRR	92.484076	19.490446	1.256000	29.994854
1	IPL2021	40.897353	21.755972	2.249818	28.095525
2	RCB	85.779817	29.357798	1.406452	39.342231
3	IPL	62.807018	34.210526	1.496063	41.987291
4	ViratKohli	91.205212	37.785016	1.273859	48.179396
5	WeAreChallengers	96.808511	71.276596	1.220779	82.178226
6	DareToDream	100.000000	89.189189	1.275862	100.465051
7	RR	84.012539	48.275862	1.483721	58.160837
8	WhistlePodu	92.727273	100.000000	2.115385	111.388112
9	Yellove	100.000000	100.000000	2.500000	112.500000

Inorganic or not:

I analyzed the number of tweets of a hashtag and total number of tweets; no hashtags appear to be inorganic as the number of tweets of a hashtag and total number of tweets are proportionate.

Inferences:

1. The plots are specific to a particular instant of time, as the topic chosen is ipl, which generates high tweets of volume everyday that too specific to match, the result will vary a lot.
2. In all three plots #IPL2021 appears due to the fact that people are using it because of the 2021th version of the IPL.
3. Along with the official franchise name, nicknames of the team are also widely used for example, "weAreChallengers", "whistlePodu".
4. I ran the code on 23rd of April, so the hashtags are related to the ipl franchises Mumbai Indians and Punjab kings. The hashtags also include franchise names from the match on 22nd April as the match was one day before only and had significant amount of tweets.
5. $f > 20$ denotes that major chunk of the content is generated by the top 50 users. 6. Also the list of hashtags does not include any unrelated hashtags. All the hashtags are related to each other. At first hashtags like "DareToDream", "OneFamily" may look unrelated but it is not the case, both these are used in context to IPL.