

Section I

I analysed the data of DCPWestDelhi for this section. This data has been saved in dataHW2Section1.txt.

1. I used regex or regular expressions to extract the following data from the text corpus obtained from extracted data.

a. I made regex for the following:

- Aadhar Number
- Phone Number
- Email ID
- IP Address
- Pincode
- License Plate Number

The table looked like:

# of PII detected	
aadhar	6
phone number	13
email ID	0
ip address	0
pincode	0
license plate	2

Here are some examples:

License Plates: ['DL-11C-7351', 'DL09C3421']

Adhaar: ['6622-2350-9284', '2161 6729 3627', '8384-2795-9970', '6213-3631-4249', '7667-9750-5883']

Phone Numbers: ['9670032256', '+91 84505 55222', '5000004676', '+918135022555', '07635689895']

b. I used media extracted fro tweets which had them to find PIIs

<https://t.co/HA7Gb1RNbO>
<https://t.co/DQ0I4PD0Ey>
<https://t.co/6DZf70TpmB>
<https://t.co/wt1IilyG3L>
<https://t.co/TKlusM6vXA>
<https://t.co/6EIoyed0Ns>
<https://t.co/vfgcpXfL4Q>
<https://t.co/oZC75qu2GY>
<https://t.co/dsDp8WEL2U>
<https://t.co/uvqy9WiTse>
<https://t.co/JKkFrcUmNP>
<https://t.co/IRtsC6tCrT>
<https://t.co/FaEeHEku8f>
<https://t.co/WKnjOmsQeF>
<https://t.co/7zz1IMWNcE>
<https://t.co/KzpWEzjgev>
<https://t.co/Nk0Qt7ybUS>
<https://t.co/g772t1bwWr>
<https://t.co/GheAt6KOac>
<https://t.co/QNzGV4RYyp>
<https://t.co/28eJTQmsIT>

Examples:

CONTACT US

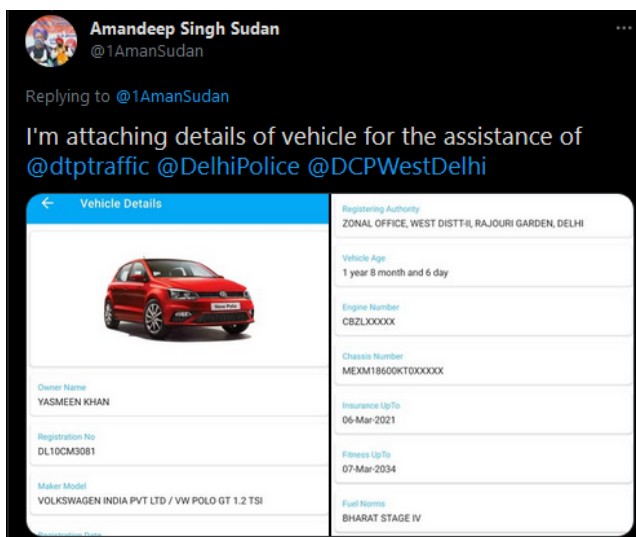
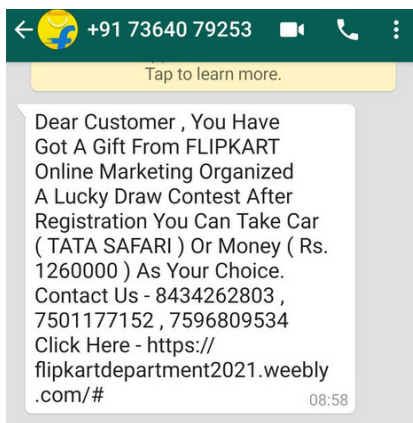
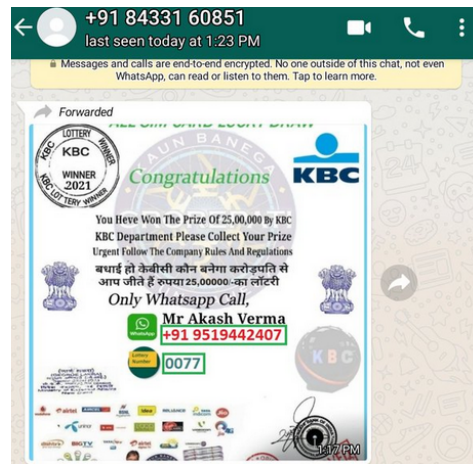
Email: support@espotshop.com

Call us: +916828512347, +916928545677

Business Hours: 9am-7pm (7 Days Open)

Office Address: Esportshop Shopping Private Limited, 52 . Gopal Bhavan, Kelkar Road, Ram Nagar, Dombivli(e), Mumbai, Maharashtra 421201

Note: Average Response time is 15 to 20 Minutes.



2.

a. Response time statistics:

- i. Minimum response time: 0 days, 0 hours, 0 minutes, 0 seconds. (Instant response)
- ii. Maximum response time: 11 days, 1 hour, 40 minutes, 24 seconds
- iii. Mean response time: 0 days, 19 hours, 14 minutes, 58.3 seconds
- iv. Standard deviation of response time: 1 day, 13 hours, 54 minutes, 59.1 seconds

```
print(min(difTimeArr))
```

```
0:00:00
```

```
print(max(difTimeArr))
```

```
11 days, 1:40:24
```

```
timeDifDf=pd.DataFrame(difTimeArr,columns=['time'])
```

```
timeDifDf['time'].mean()
```

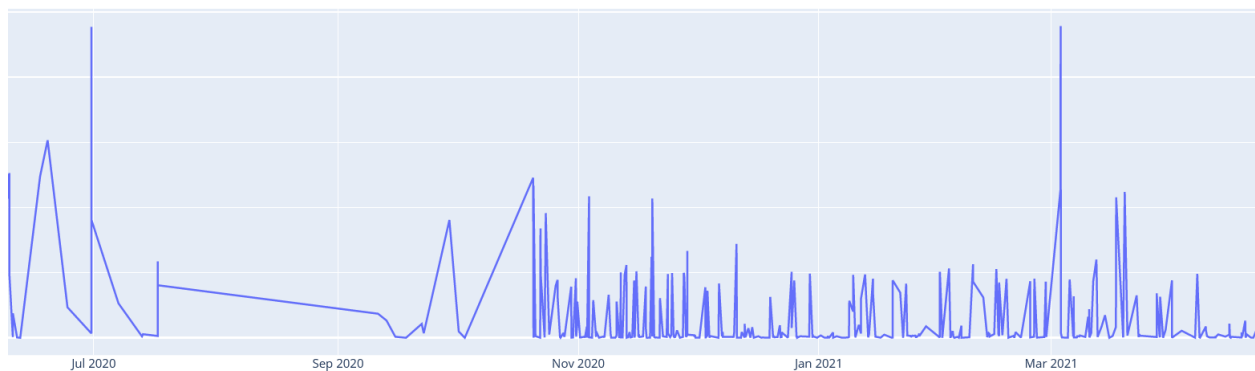
```
Timedelta('0 days 19:14:58.339130434')
```

```
timeDifDf['time'].std()
```

```
Timedelta('1 days 13:54:59.178749627')
```

This implies that this account is probably managed by a human. This is evident by the fact that there is a lot of variation in the response time as per the standard deviation. The max response time is huge and thus this can be reduced with the use of robots which can reply to certain types of tweets which don't require manual intervention.

b.



As per this graph, we can see that there are certain bursts of activity from the account implying that there are big dormant periods and then active periods. There are huge spikes seen in July 2020 and March 2021. Within a day, there is more activity from 8 am to 6 pm as compared to the rest of the day.

3. a.

	Tweet	Code
1	@DCPWestDelhi Jail Road from Tilak Nagar to Hari Nagar has been taken over by shopkeepers and parking. 75-80 pct o... https://t.co/92JLDBD7Sm	report
2	@Dranandking @anubhavkhanduri @PMOIndia @News18UP @ChaubeyMeena	spam
3	@DCPWestDelhi for necessary action and to inform the complainant.	accountability
4	@varungulati84 @DCPWestDelhi @DCPWestDelhi for necessary action.	accountability
5	Please help me for justice.I have a family have 2 little daughters we have nothing to survive. We can't survive th... https://t.co/9dsmRshdSK	requests
6	Death of SI Ramlal is a tragic loss both to his family as well as Delhi Police. He was a sincere and dedicated off... https://t.co/A0dk5qJE8u	appreciation
7	DJ playing around A3 block and Asalatpur Village, Janakpuri (a residential area) till 12:40am. Wondering what the... https://t.co/6x7VZWZTi9	report
8	@DCPWestDelhi @ArvindKejriwal this is how encroachment in govt land is being done brazenly... Rama Road industrial... https://t.co/Kpk19tKlpK	accountability report
9	Mr. SHO Inderpuri I request you to please visit E Block Bada Park Near Gas Filling Shop, every night group of bo... https://t.co/UbTmeRv366	report
10	@DelhiPolice @vshankarsingh @ArvindKejriwal @lgdelhi @DCPWestDelhi please consider it.	requests
11	TNobody wearing masks, next Corona Hotspot in Kirti Nagar G Block, Kirti Nagar, @DCPWestDelhi @ArvindKejriwal... https://t.co/CPiz4WSqwH	report
12	@DCPWestDelhi crackers still burning in the locality, Vikaspuri H-1 block	report
13	@DPSKumarJi @CMODelhi @aajtak @sudhirchaudhary Your tweet is forwarded to DCP/West @DCPWestDelhi	accountability
14	Thank you for cont... https://t.co/qlmKCWGcGf	appreciation
15	@LtGovDelhi @CPCB_OFFICIAL @DCPWestDelhi People start bursting crackers in Vishnu Garden but no action by any authorities	report

16	कृपया यथाशीघ्र आवश्यक कार्यवाही करें। उपरोक्त paisha और पासपोर्ट उन्हें विदेशों में रोजगार देने का... https://t.co/84lOuslmWk	
17	#Delhi Police After giving written Complaint on 5th Dec. No response from #Kirtinagar PS No Response of Email to... https://t.co/vd29Tfum1d	accountability report
18	@_pankajmarwah for necessary action and to inform the complainant @DCPWestDelhi	accountability
19	@DCPWestDelhi @CPDelhi @LtGovDelhi @HMOIndia Sirs, there has been a spate of burglaries in DDA MIG Flats/VATIKA Apa... https://t.co/3zi2b23U4h	report
20	Need Heavy pruning of tress on Major Sudesh Kumar Marg in front of DDA Green MIG Flats and replace of Bus stop by P... https://t.co/vczfloQOga	report requests
21	@DCPWestDelhi sir help regarding efir no 20460/2020 of police station paschim vihar east .the verification of ba... https://t.co/2TroD5oV2n	requests accountability report
22	बिंदापुर डीडीए फ्लैट A 184 से 6 लड़के मोबाइल और 5000 रुपए लेकर गाड़ी नंबर DI-11C-7351 से फरार हो गए हैं अकेली 70 सा... https://t.co/VOJmfkOyfR	report
23	@adv_aankit @ArvindKejriwal @narendramodi @CPDelhi @DCPWestDelhi for necessary action	accountability
24	@DCPWestDelhi @HMOIndia @LtGovDelhi @CPDelhi @DelhiPolice Mr DCP The real test will be enforcing the ban on Diwaan... https://t.co/AAEBDeQfb6	accountability
25	I hope @DelhiPolice and @DCPWestDelhi increases petrolling on this route as a lot such incidents are increasing... a... https://t.co/TA0MljN0EA	report requests
26	@DCPWestDelhi Sir please look it to https://t.co/dgqNM7PJyk	requests
27	@DCPWestDelhi @CPDelhi @ANI @DelhiPolice Kindly do some encroachment removal on outer ring road . Keshopur mandi to... https://t.co/h5q5VjQuuP	report requests
28	@CPDelhi @DCP_CCC_Delhi @DelhiPolice @DCPWestDelhi bank fraud took place from my mothers account through netbanking, please help 🙏	report requests

29	Please public ke upr juthe case dalne wale afsar ko suspend kia jaye na ki promotion di jaye sho anil sharma police... https://t.co/ieswaSpS8q	report requests accountability
30	क्या इस तरह की तलाशी हो रही है या एक धर्म विशेष के लोगओं को भ्रमित, भडकाया जा रहा है. कृपया स्पष्ट कीजिए ... https://t.co/C3xWXwvmpW	accountability

Different categories of tweets can be modelled by using the following parameters:

- In case there is a please in a tweet, it probably implies a **request** to a police department.
 - @DelhiPolice @vshankarsingh @ArvindKejriwal @lgdelhi @DCPWestDelhi please consider it.
- In case of mention of money, car or property, we can categorise **theft**.
 - बिंदापुर डीडीए फ्लैट A 184 से 6 लड़के मोबाइल और 5000 रुपए लेकर गाड़ी नंबर DI-11C-7351 से फरार हो गए हैं अकेली 70 सा... <https://t.co/VOJmfkOyfR>
- In case email accounts, net, hacking or blackmailing or such words are used, they can be characterized as **cyber crimes**.
 - @CPDelhi @DCP_CCC_Delhi @DelhiPolice @DCPWestDelhi bank fraud took place from my mothers account through netbanking, please help 🙏
- In case someone mentions some amount, we can probably categorise that as some **financial theft**.
 - बिंदापुर डीडीए फ्लैट A 184 से 6 लड़के मोबाइल और 5000 रुपए लेकर गाड़ी नंबर DI-11C-7351 से फरार हो गए हैं अकेली 70 सा... <https://t.co/VOJmfkOyfR>
- In case someone mentions women or ladies, we can probably categorise it as a **crime against women**.
- In case there is mention of traffic or road or blocks, it probably implies that it is something related to **traffic** and thus can be categorized as that.
- In case of a lot of positive words like good, super, amazing, thank you etc. it probably implies **appreciation**.
 - Thank you for cont... <https://t.co/qlmKCWGcGf>

These keyword search approaches can help to categorize the tweets into the relevant divisions and thus be approached in a more systematic way.

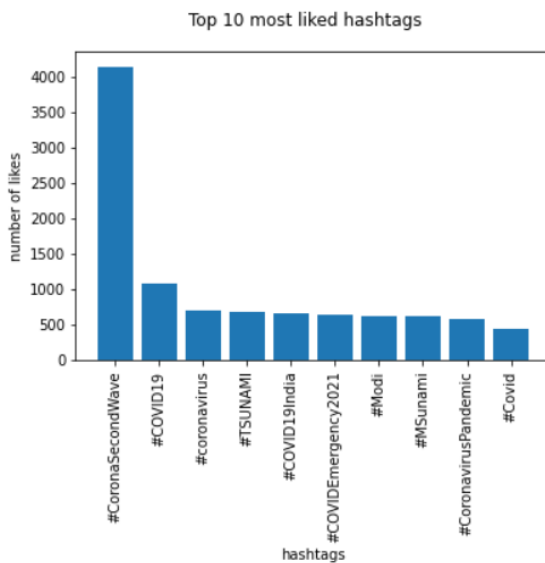
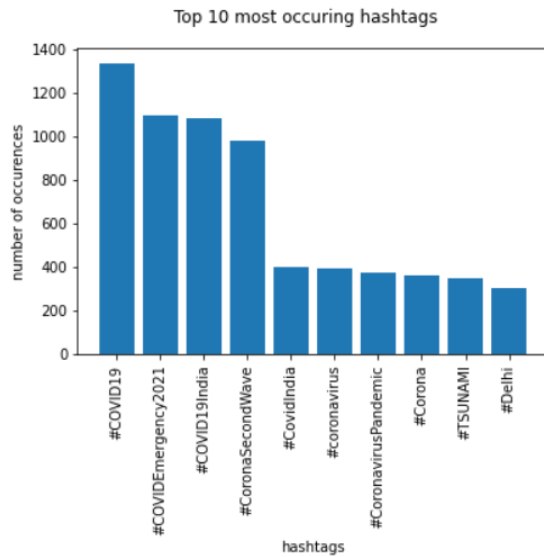
b. This can bridge the communication gap between citizens and police and help improve their accountability in the following ways:

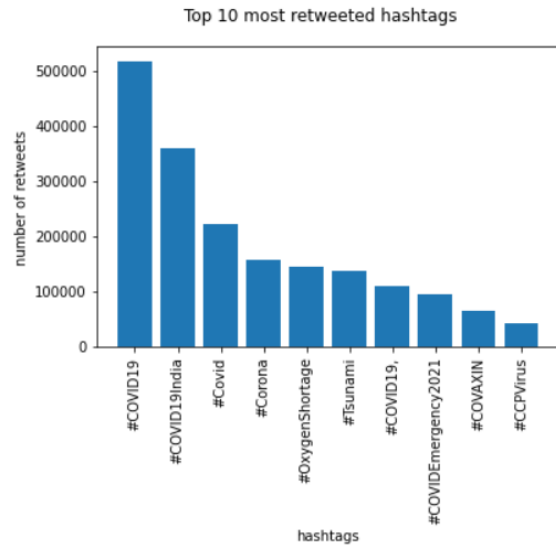
- The tweets can be categorised and forwarded to the correct departments within the police and thus the best possible responses can be given to each tweet.
- Twitter can also be used as a public forum to hold police responsible because in case police doesn't follow up, it'll reflect badly.
- This can also be used as a means to amplify certain important cases and bring them to the correct authority's jurisdiction.
- The data can be collected and stored, using it for future benchmarks and making sure that all problems are solved.
- The data can be annotated to make the coding process automated and better.
- The statistics from police activity can be used to facilitate further research.

- The communication is fully transparent and public on twitter and thus it lessens the scope for corruption.

Section II

I analysed the data of the trending hashtag #CoronaSecondWave for this section. This data has been saved in dataHW2Section2.txt. Then I basically found the top 10 most occurring hashtags in this data and then I compiled the full data in dataHW2Section2Full.txt.





	Hashtag	R/10	F	U	C
0	#COVID19	6.484655	4.503786	1.199904	12.188346
1	#COVIDEmergency2021	5.967886	4.906334	1.268100	12.142319
2	#COVID19India	5.529197	4.014599	1.215078	10.758873
3	#CoronaSecondWave	6.090819	3.356367	1.246002	10.693189
4	#CovidIndia	3.891403	6.334842	1.270115	11.496359
5	#coronavirus	5.525672	8.557457	1.231928	15.315057
6	#CoronavirusPandemic	3.582474	11.855670	1.315254	16.753399
7	#Corona	6.135726	5.384184	1.310066	12.829976
8	#TSUNAMI	8.850575	4.885057	1.689320	15.424953
9	#Delhi	8.000000	8.000000	1.198083	17.198083

Top 10 occurring hashtags

	Hashtag	R/10	F	U	C
0	#CoronaSecondWave	6.090819	3.356367	1.246002	10.693189
1	#COVID19	6.484655	4.503786	1.199904	12.188346
2	#coronavirus	5.525672	8.557457	1.231928	15.315057
3	#TSUNAMI	8.850575	4.885057	1.689320	15.424953
4	#COVID19India	5.529197	4.014599	1.215078	10.758873
5	#COVIDEmergency2021	5.967886	4.906334	1.268100	12.142319
6	#Modi	7.248460	6.570842	1.453731	15.273033
7	#MSunami	9.858156	4.255319	2.238095	16.351570
8	#CoronavirusPandemic	3.582474	11.855670	1.315254	16.753399
9	#Covid	5.449102	6.159110	1.316441	12.924654

Top 10 liked hashtags

	Hashtag	R/10	F	U	C
0	#COVID19	6.484655	4.503786	1.199904	12.188346
1	#COVID19India	5.529197	4.014599	1.215078	10.758873
2	#Covid	5.449102	6.159110	1.316441	12.924654
3	#Corona	6.135726	5.384184	1.310066	12.829976
4	#OxygenShortage	4.750000	5.000000	1.071429	10.821429
5	#Tsunami	4.412955	51.012146	2.627660	58.052761
6	#COVID19,	9.493671	1.265823	1.000000	11.759494
7	#COVIDEmergency2021	5.967886	4.906334	1.268100	12.142319
8	#COVAXIN	10.000000	0.000000	1.000000	11.000000
9	#CCPVirus	9.861111	4.166667	1.531915	15.559693

Top 10 retweeted hashtags

CTM captures inorganic behaviour. This is characterised by people propagating the same opinions, some people driving the whole narrative. Looking at the CTM values, there do not seem to be any inorganic activity except for the #Tsunami. But this hashtag can be justified because a lot of it could have to do with influencers getting involved with this hashtag as a means to get help and their sentiments getting echoed. However, it can also be contributed to inorganic activity because the F value and U value is extremely high.

There were hashtags out of context including: #TSUNAMI, #Modi, #MSunami, #Tsunami. But on deeper reflection and research, I saw that even these tweets were also related to COVID. Tweets related to tsunami refers to the sudden spike of COVID cases and #Modi is trending because of the way his administration handled the pandemic and is thus facing scrutiny.