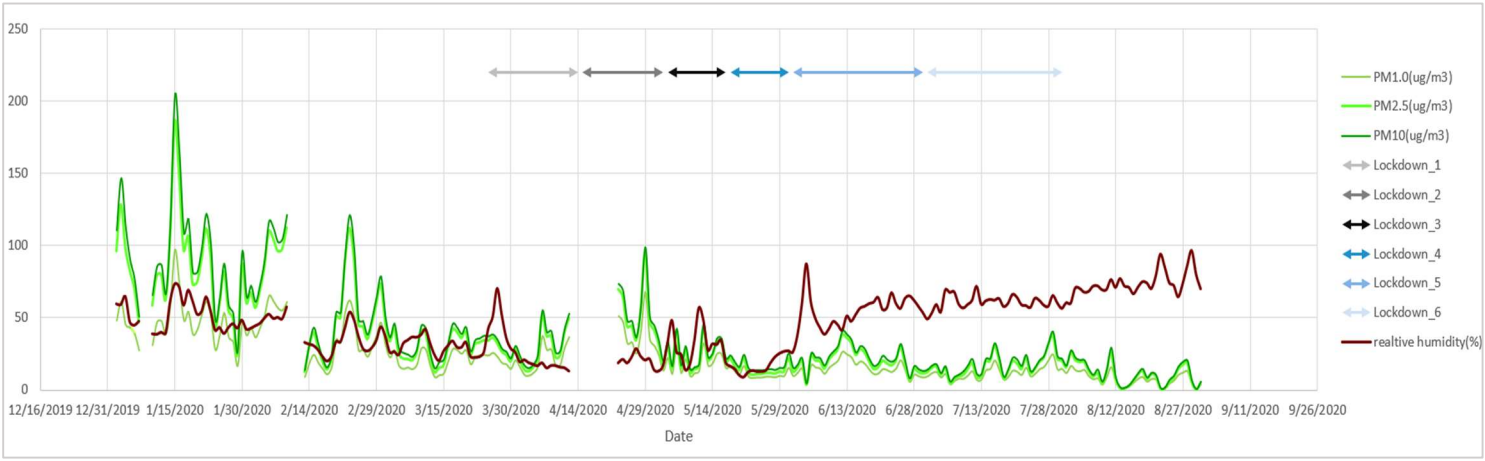
ES-209 HW_4

19d180006 Sanket Ambre
19d180019 Pranav Mahadkar
19d180008 Meghraj Chinchore

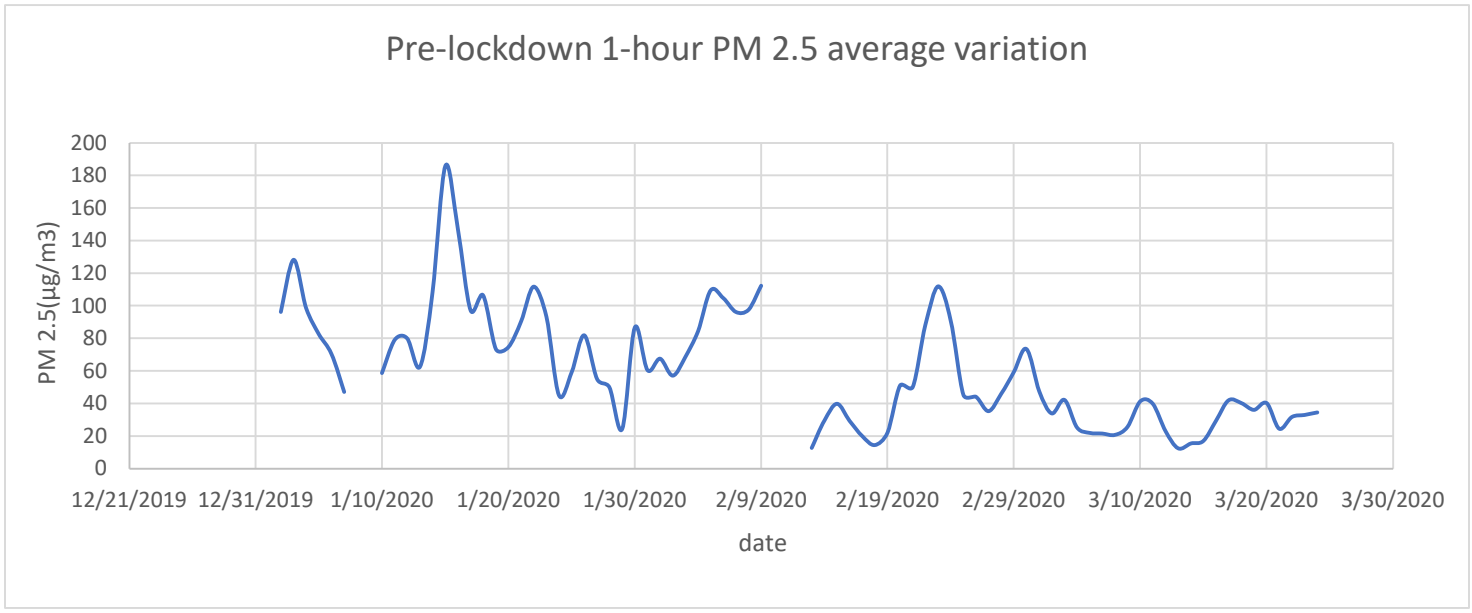
Q1.

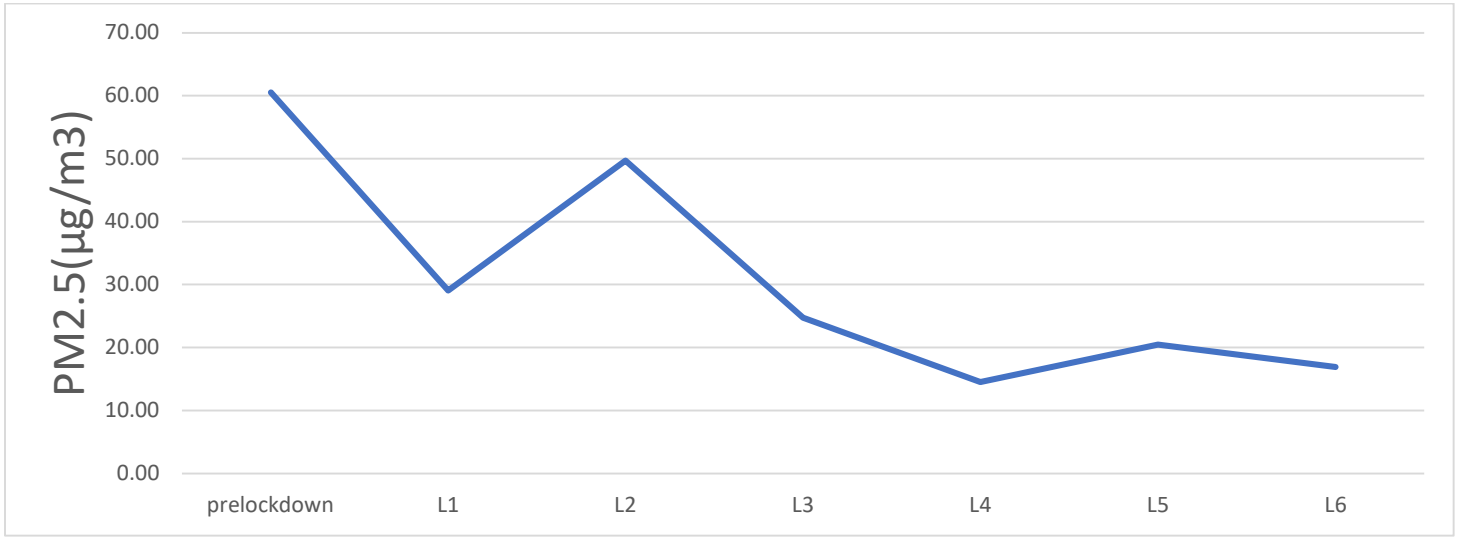


- The graph above shows the variation of PM(PM1,PM2.5,PM10) and Relative humidity with time (in days).
- We can observe that the PM values before the lockdown are quite high as compared to the general trend after the lockdown started.
- As the lockdown 1 started, relative humidity also started to decrease.
- We can observe a spike in PM concentration in Lockdown-2 which gradually lowered down as we were in lockdown-3. Contrast to it , the relative humidity increased comparatively in lockdown-3.
- The PM concentration as well as humidity achieved the lowest values in lockdown 4. Relative Humidity started to increase end of lockdown 4. The reason behind the increase may be the unlock 1 which lead to increase in pollution.
- The highest humidity point was noted in lockdown 5 where the highest points of PM1.0, PM2.5, PM10 were noted in lockdown 2.
- As we approach the end of the lockdown we can see that the PM concentrations are tending to a slightly fluctuating but low values and relative humidity is seen to be increasing slightly after achieving a minima in lockdown-4.

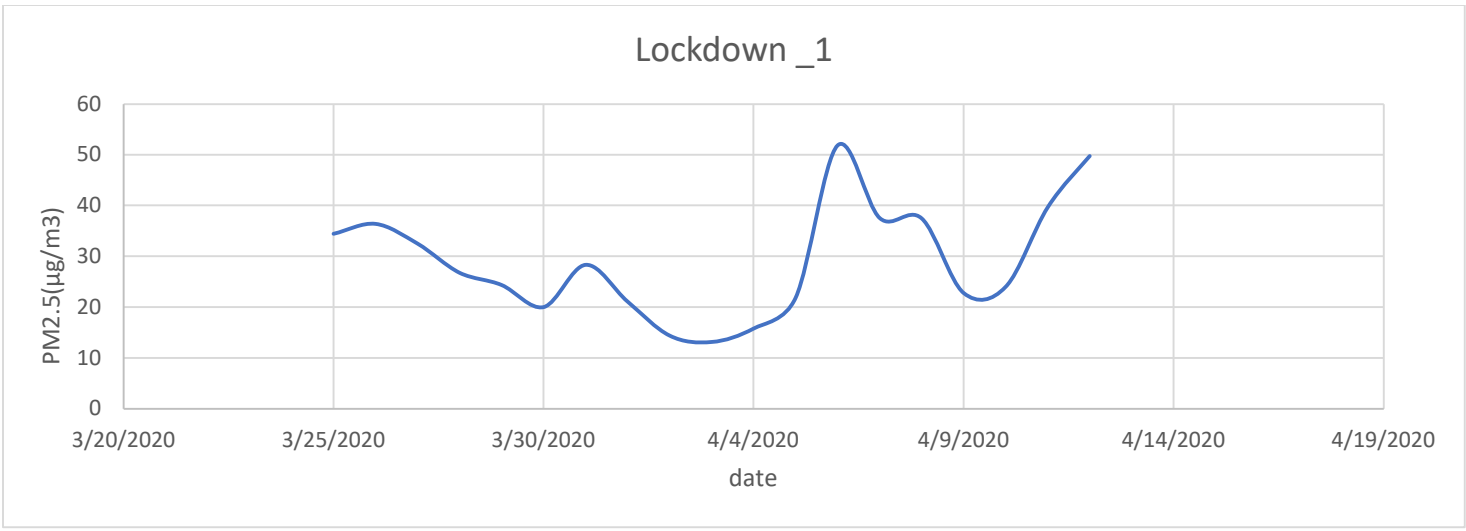
Q2.

- The graphs below show 1-hour PM2.5 average data over different days starting from the pre-lockdown period till lockdown-6.

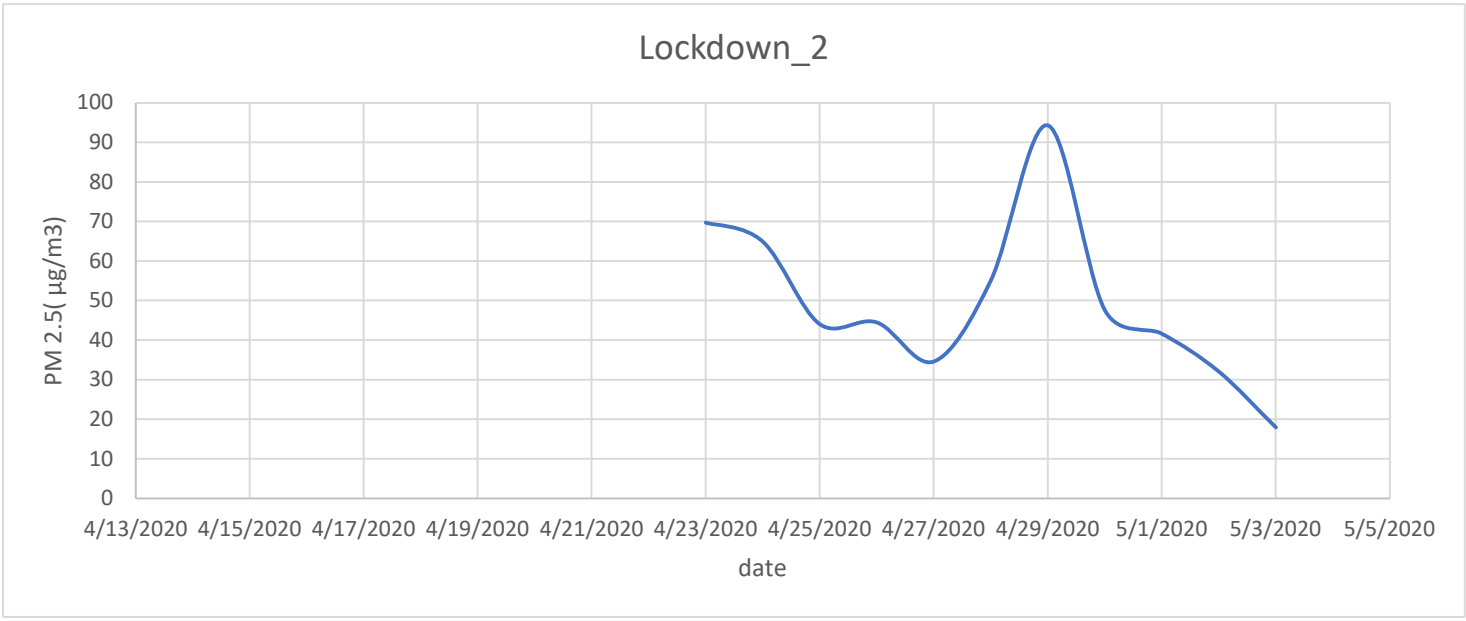




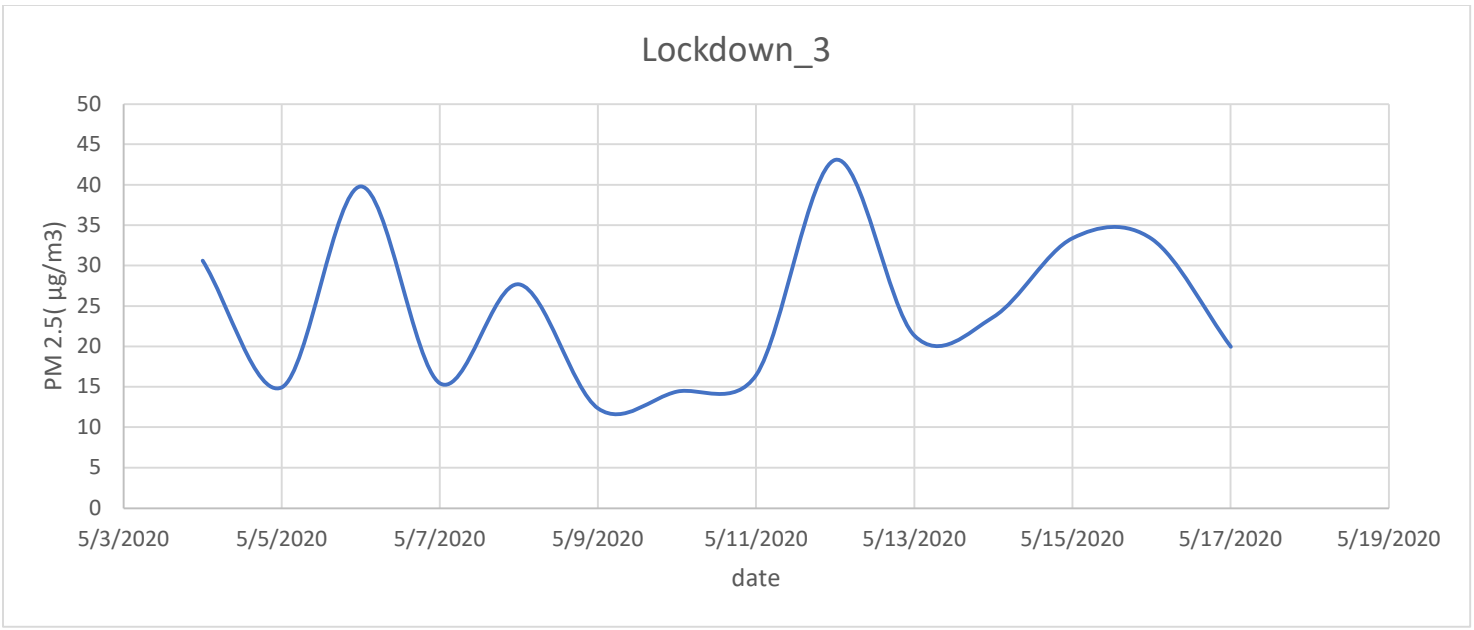
- In the pre-lockdown period, we can see that the PM values are quite high with maxima around 185.9 ug/m3 ani minimum of around 12.49 ug/m3 with an average of 60.5 ug/m3.



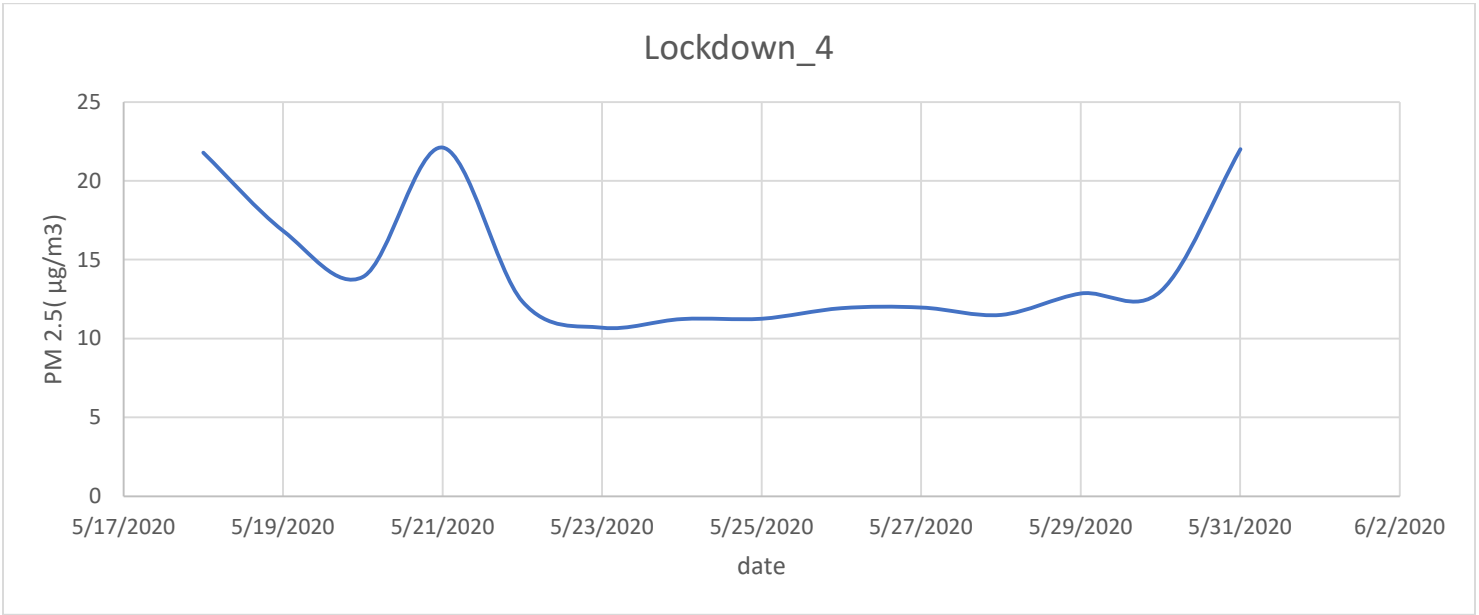
- In the lockdown-1 , we can see that the PM values are lowered due to a sudden stop to human activities. We can observe a Maxima around 52 ug/m3 ani minimum of around 13 ug/m3 with an average of 29 ug/m3.



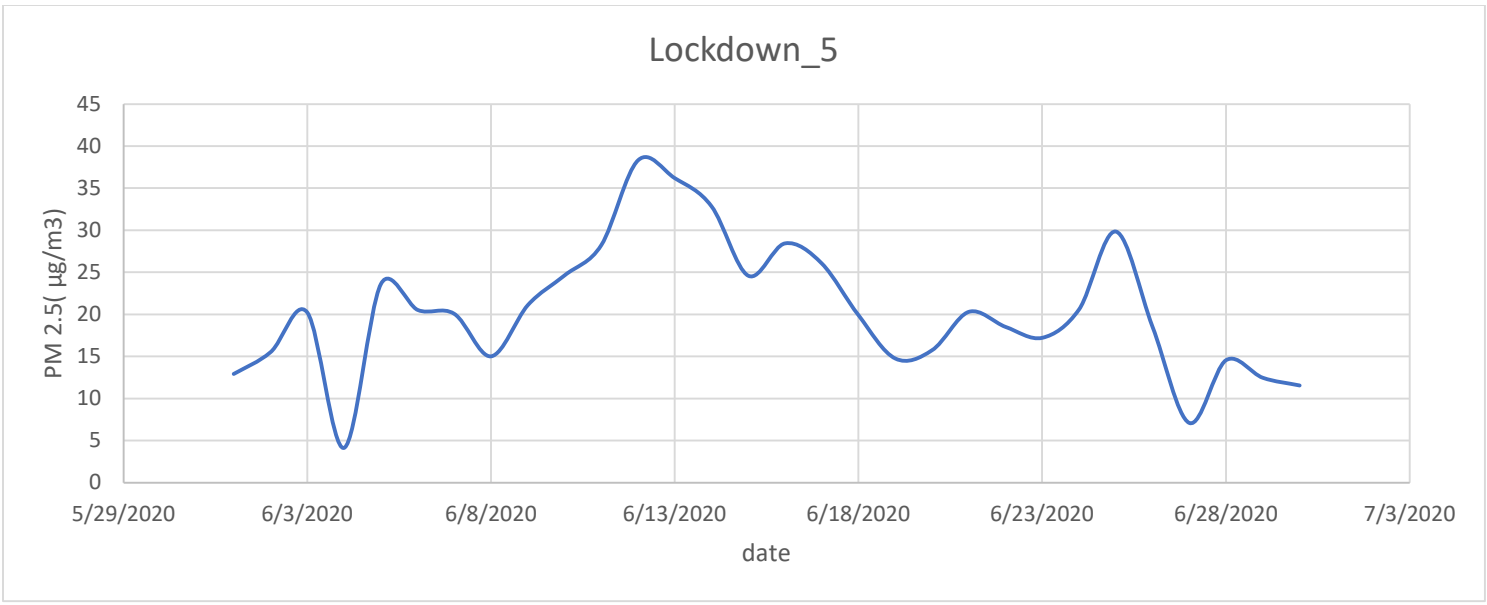
- In the lockdown-2, we can see that the PM values have a slight spike. We can observe a Maxima around 94 ug/m3 ani minimum of around 18 ug/m3 with an average of 49 ug/m3.



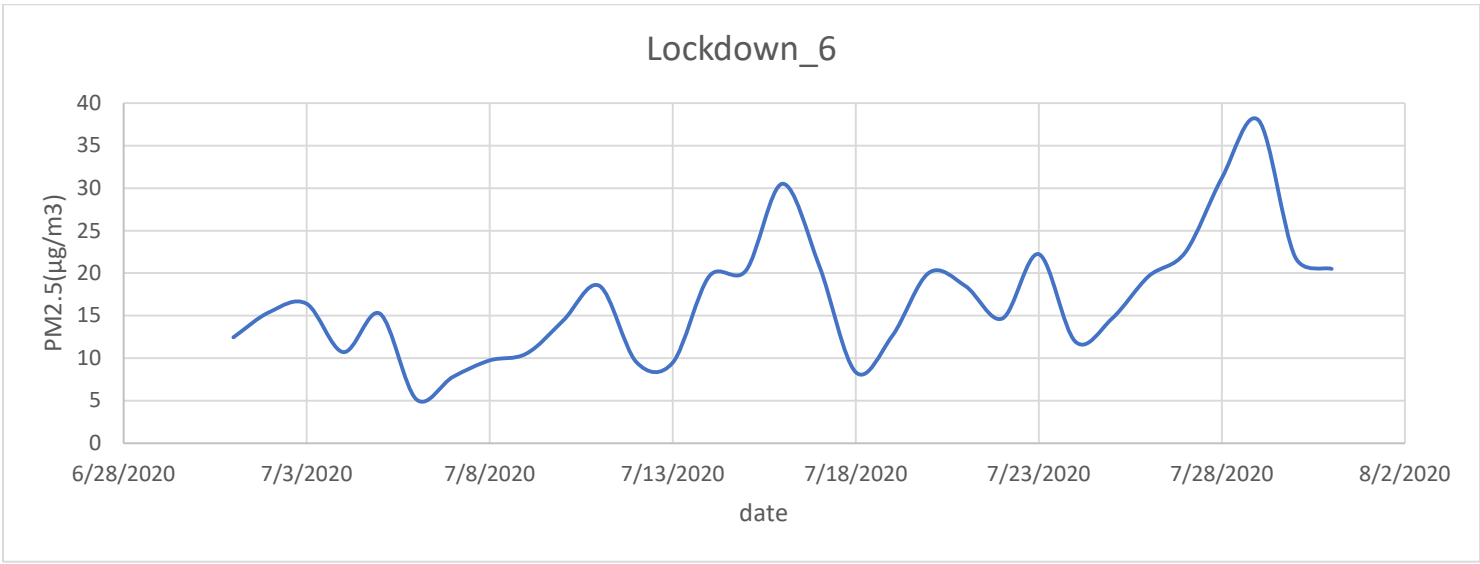
- In the lockdown-3, we can see that the PM values again are lowered may be due to stricter actions by government. We can observe a Maxima around 43 ug/m3 ani minimum of around 12.34 ug/m3 with an average of 25 ug/m3.



- In the lockdown-4, we can see that the PM values are lowered due to a sudden stop to human activities. We can observe a Maxima around 22 ug/m3 ani minimum of around 10.7 ug/m3 with an average of 14.54 ug/m3.



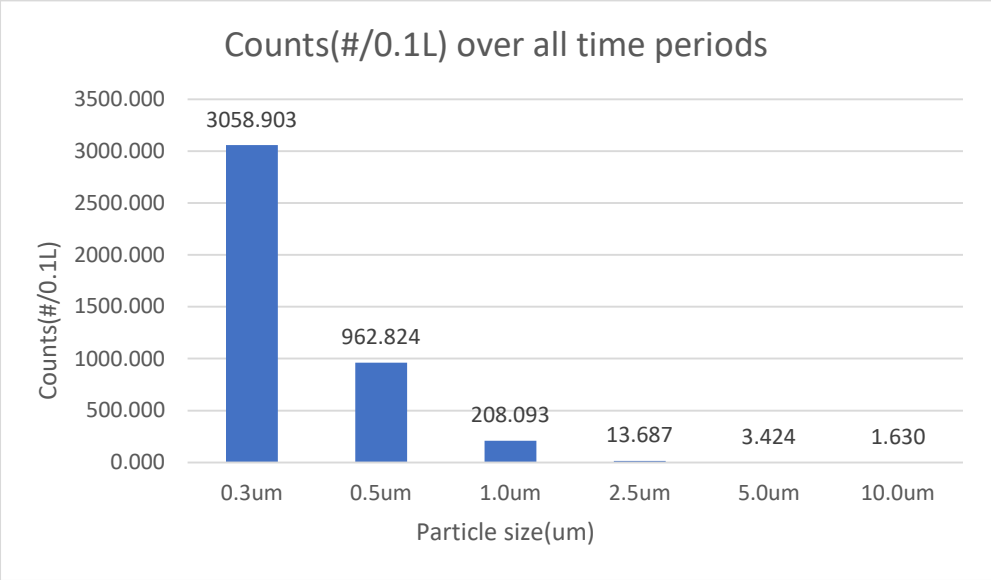
- In the lockdown-5, we can see that the PM values are fluctuating a bit around a lower value. We can observe a Maxima around 38.33 ug/m3 ani minimum of around 4.15 ug/m3 with an average of 20.46 ug/m3.

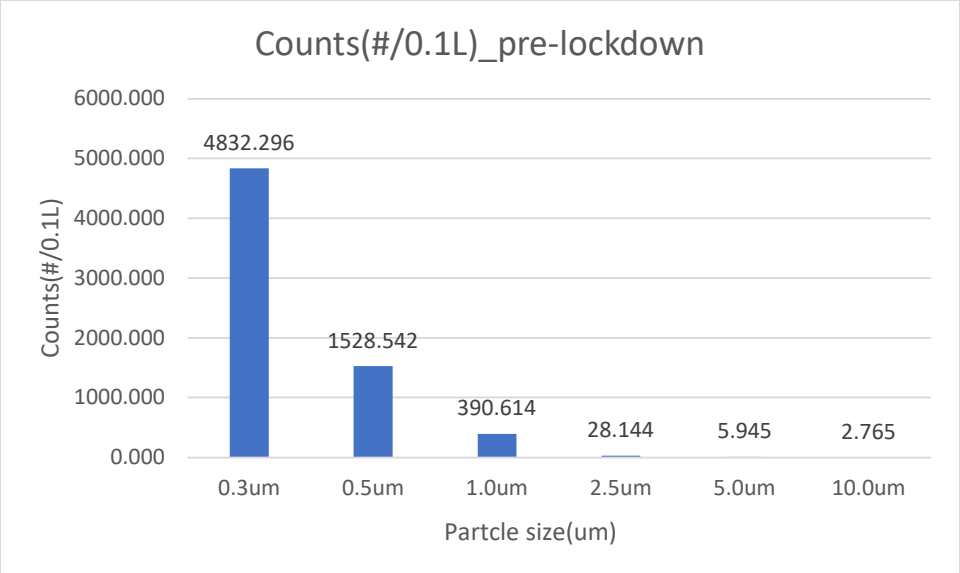


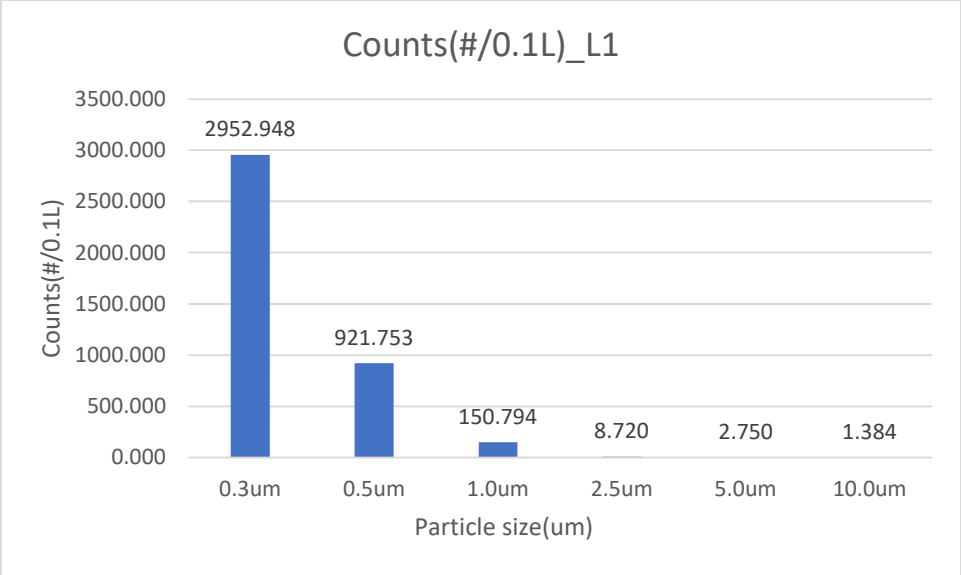
- In the lockdown-6, we can see that the PM concentration has a maxima around 34 ug/m3 ani minimum of around 5.18 ug/m3 with an average of 16.9 ug/m3.

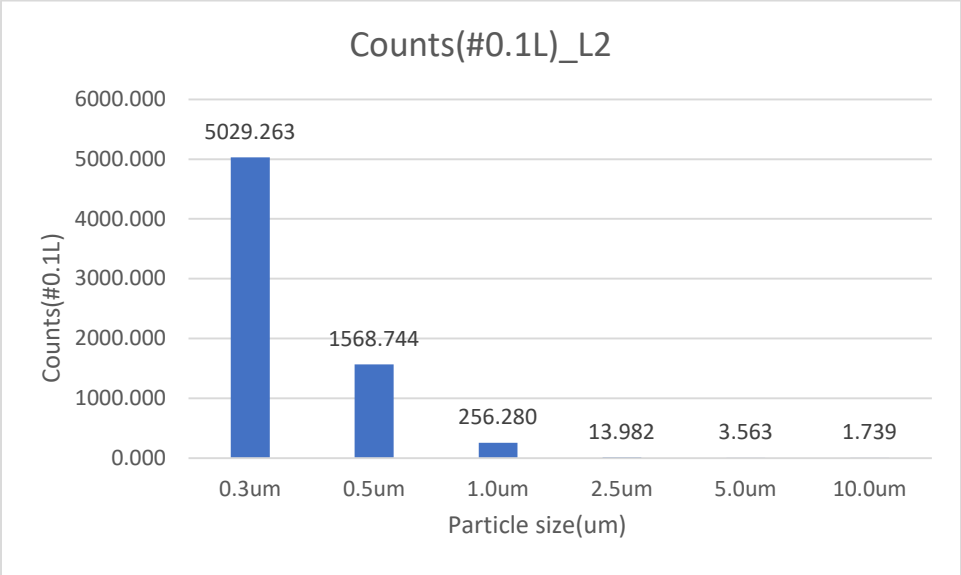
Q3.

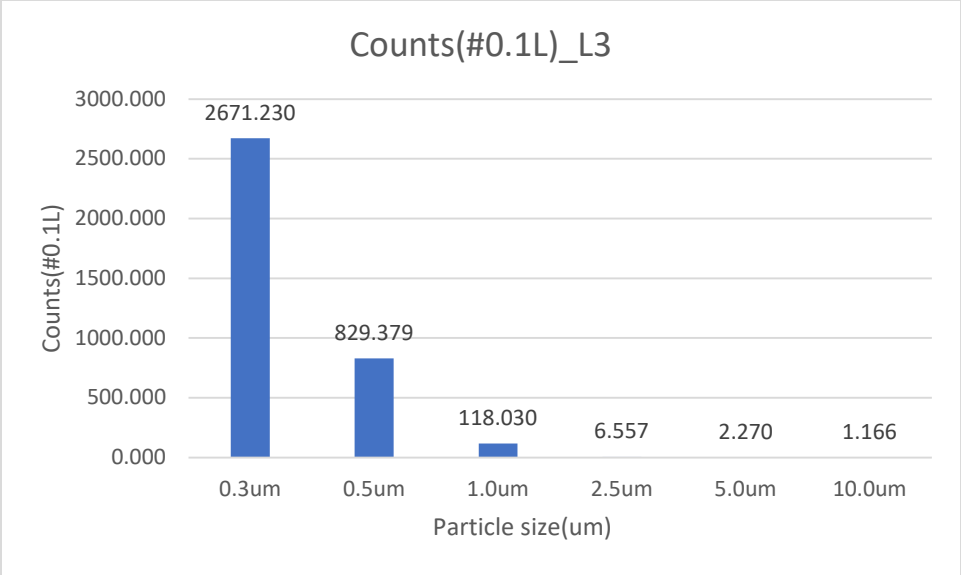
The graphs below show the Particle Size Distribution for the whole dataset and then separately for each lockdown. Each column represents the average of daily means of a specific size particulate matter.

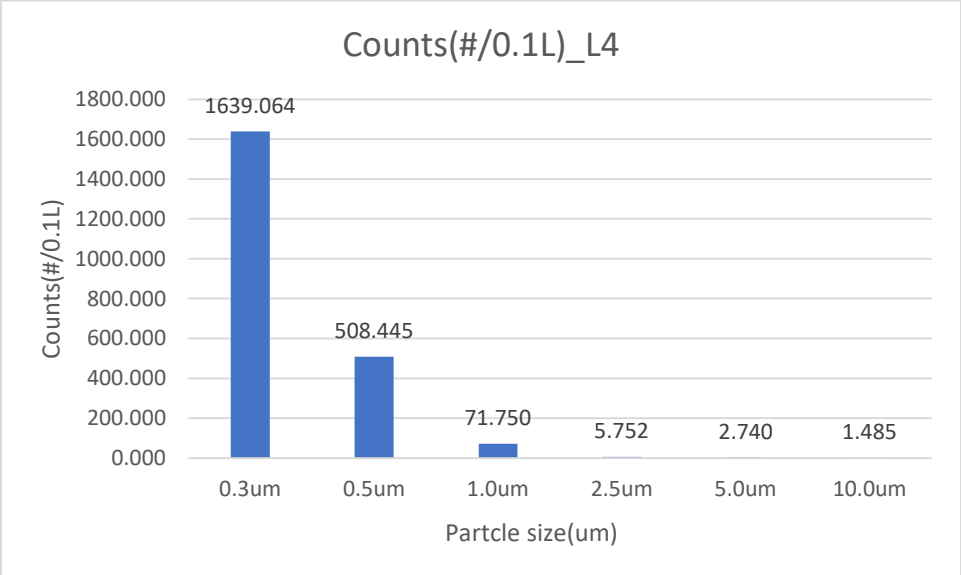


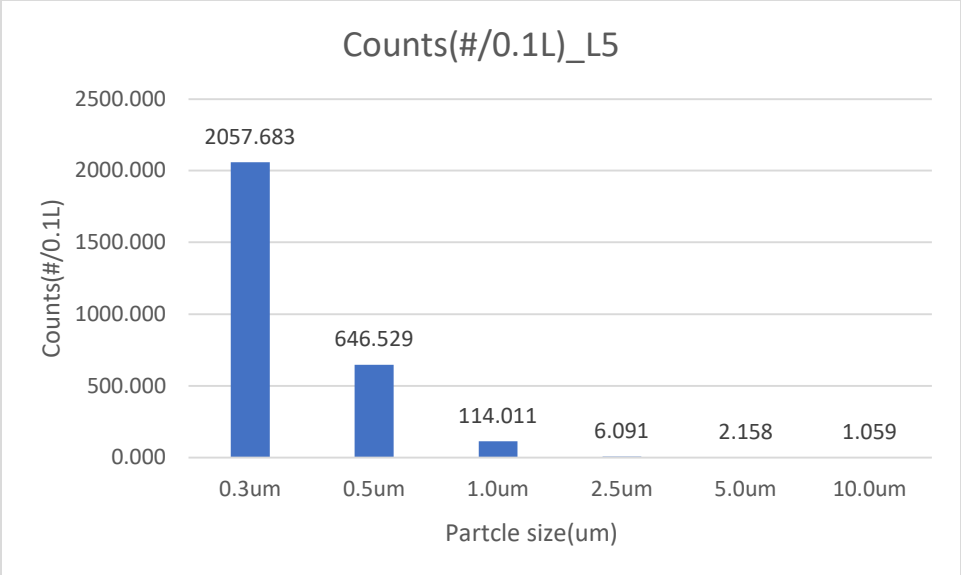


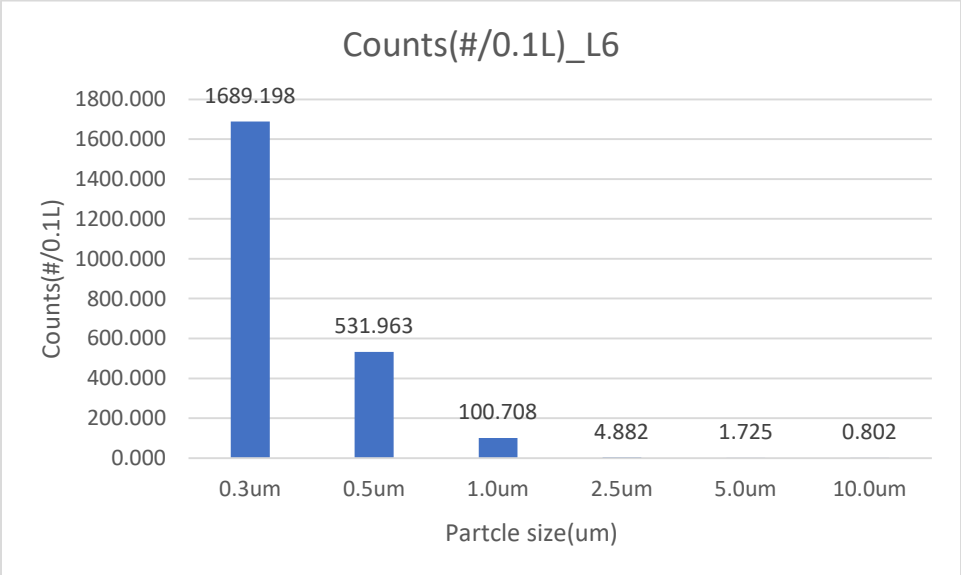

Counts(#/0.1L)_L1
3500.000
3000.000
2500.000
2000.000
1500.000
1000.000
500.000
0.000
Counts(#/0.1L)
2952.948
921.753
150.794
8.720
2.750
1.384
0.3um
0.5um
1.0um
2.5um
5.0um
10.0um
Particle size(um)

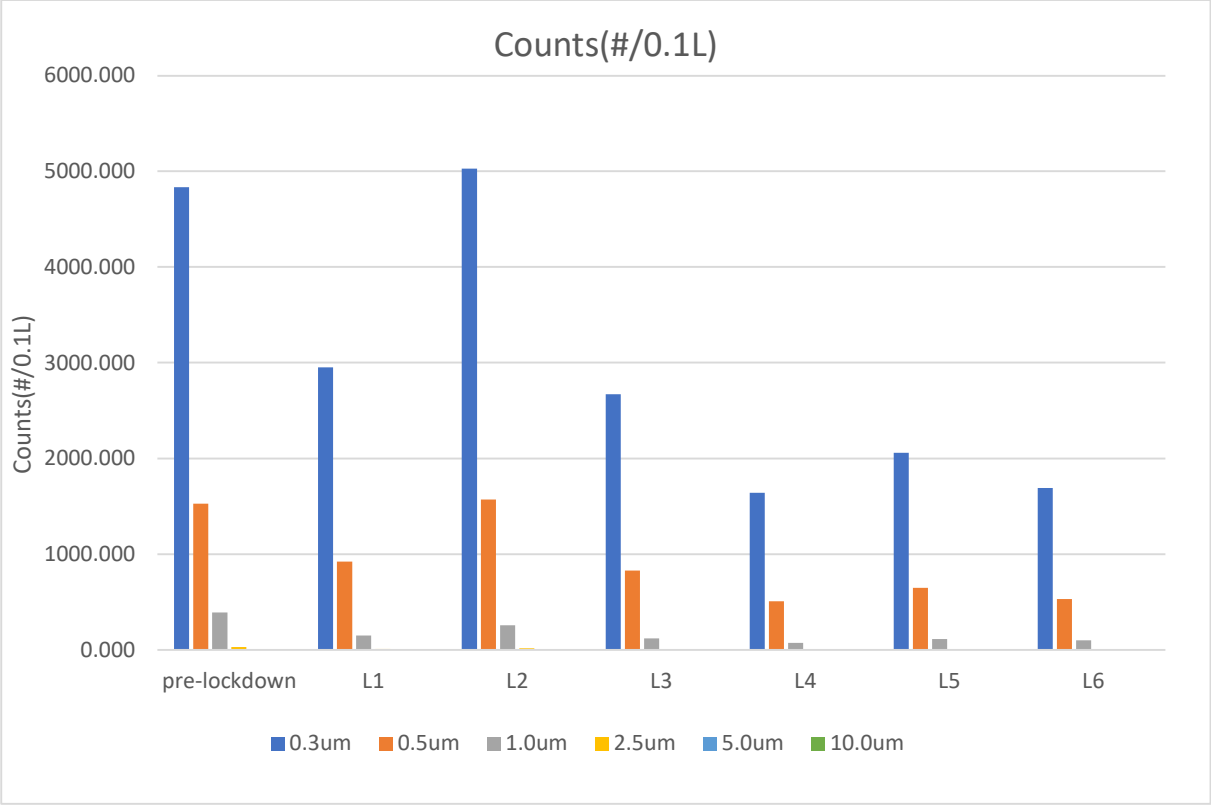

Counts(#0.1L)_L2
6000.000
5000.000
4000.000
3000.000
2000.000
1000.000
0.000
Counts(#0.1L)
5029.263
1568.744
256.280
13.982
3.563
1.739
0.3um
0.5um
1.0um
2.5um
5.0um
10.0um
Particle size(um)


Counts(#0.1L)_L3
3000.000
2500.000
2000.000
1500.000
1000.000
500.000
0.000
Counts(#0.1L)
2671.230
829.379
118.030
6.557
2.270
1.166
0.3um
0.5um
1.0um
2.5um
5.0um
10.0um
Particle size(um)


Counts(#/0.1L)_L4
1800.000
1600.000
1400.000
1200.000
1000.000
800.000
600.000
400.000
200.000
0.000
Counts(#/0.1L)
1639.064
508.445
71.750
5.752
2.740
1.485
0.3um
0.5um
1.0um
2.5um
5.0um
10.0um
Partcle size(um)


Counts(#/0.1L)_L5
2500.000
2000.000
1500.000
1000.000
500.000
0.000
Counts(#/0.1L)
2057.683
646.529
114.011
6.091
2.158
1.059
0.3um
0.5um
1.0um
2.5um
5.0um
10.0um
Partcle size(um)


Counts(#/0.1L)_L6
1800.000
1600.000
1400.000
1200.000
1000.000
800.000
600.000
400.000
200.000
0.000
Counts(#/0.1L)
1689.198
531.963
100.708
4.882
1.725
0.802
0.3um
0.5um
1.0um
2.5um
5.0um
10.0um
Partcle size(um)



- Considering the whole dataset, we can observe that count of 0.3um(#/0.1L) is considerably higher than any other size of particulate matter.
- Lockdown 1 followed a similar trend but the absolute values of counts got lowered due to lesser human activities because of the strict lockdown restrictions.
- In lockdown 2, the PM counts(0.3um and 0.5um) increased to larger values, in fact larger than the pre-lockdown period. Relative trends were same as lockdown 1.
- Lockdown 3 experienced an decrease in counts of all size particulate matter and hence significant decrease in pollution as well.
- In lockdown 4, more decrease in the pollutants can be seen. PM0.3um still being the dominating pollutant size.
- Lockdown 5 saw a small increase in counts of all the sizes of particulate matter, relatively the distribution remaining the same.
- In lockdown 6, the counts decreased as compared to Lockdown 5. As compared to pre-lockdown stage, the counts in lockdown were considerably low.

Q4.

- PM2.5(ug/m3) data for different time-periods:

	N	std dev	mean	difference(pre-lockdown-other lockdowns)
total	228.00	29.97	34.38	
Pre-lockdown	78.00	35.11	60.51	
L1	19.00	11.14	29.07	31.44
L2	11.00	20.80	49.70	10.81
L3	14.00	10.00	24.76	35.75
L4	14.00	4.31	14.54	45.97
L5	30.00	7.91	20.46	40.05
L6	31.00	7.31	16.90	43.61

Mean(x)	34.60
std-dev	12.786

a	0.05
t-critical	2.015

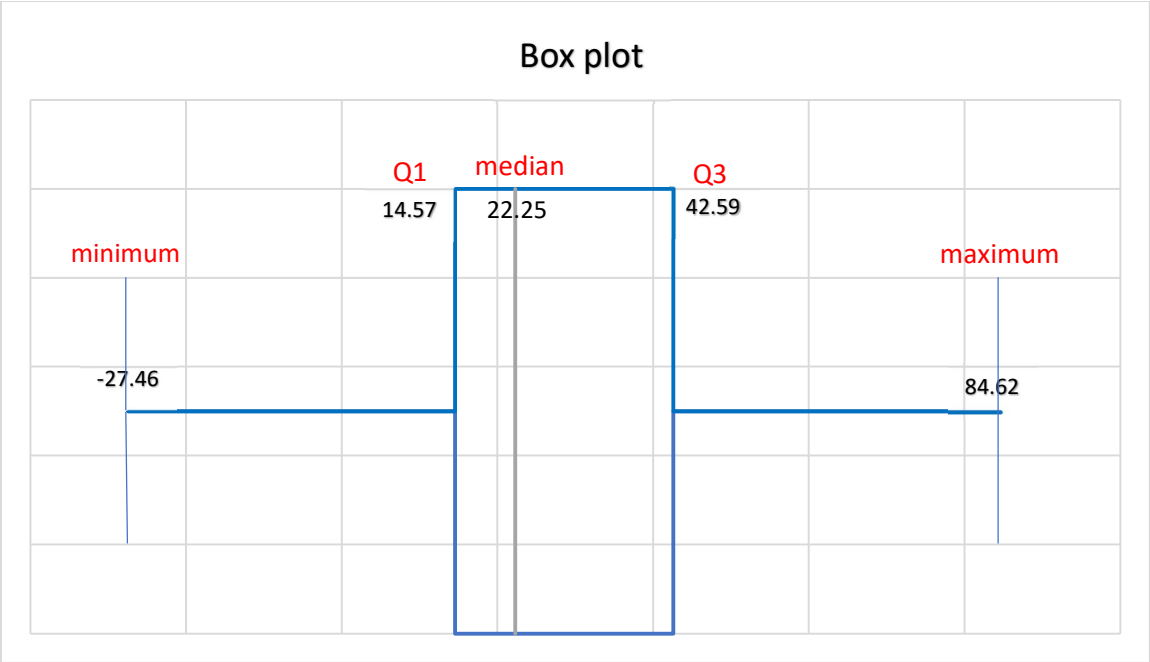
Considering a=0.05, that is confidence level of 95%, we get two tailed t-value=2.015
Assuming null Hypothesis ; Ho = difference b/w pre-lockdown and each lockdown is zero , u=0

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

X = 34.60
u = 0
s = 12.76
n = 6
On solving , t = 6.63
t-critical = 2.015
t > t-critical ==> We can conclude that our assumption is wrong and there is significant difference b/w PM2.5 concentration means of pre-lockdown period and other lockdowns

- BOX PLOT for whole data of PM2.5 avg concentration data (ug/m3) :

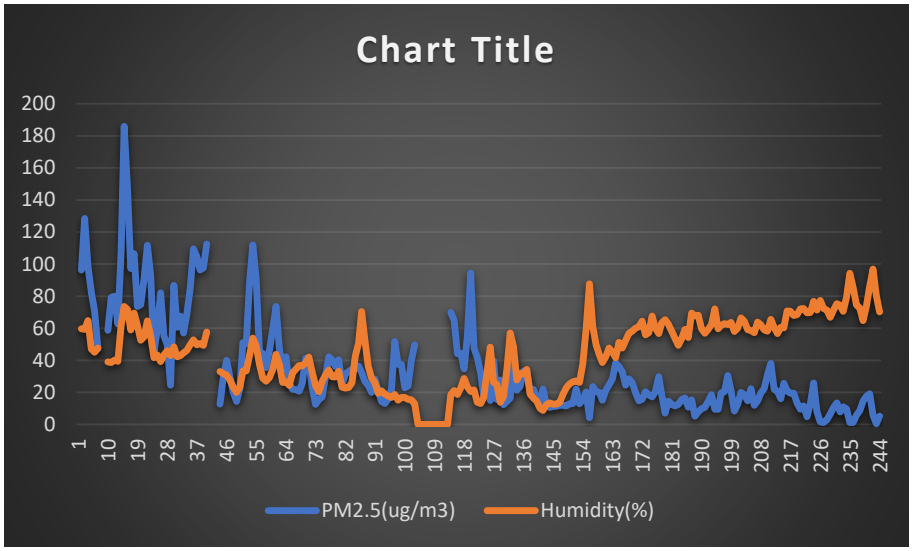
IQR		28.02
Q1-1.5*IQR		-27.46
Q1	57.25	14.57
Q3	171.75	42.59
Q3+1.5*IQR		84.62



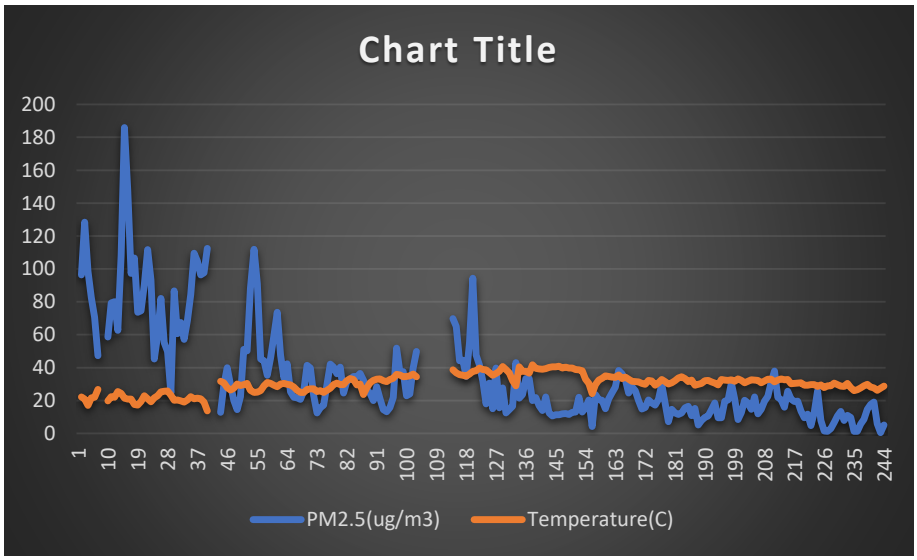
- We can infer from the box plot above that most of values lie between that the data values are concentrated in the region b/w 14.57 ug/m3 and 42.59 ug/m3 with a median value of 22.25 ug/m3.
- Values beyond 84.62 ug/m3 and values below -27.46 ug/m3 can be considered as outliers.

Q5.

Graph PM2.5 vs RH



Graph PM2.5 vs Temperature



For 70% data, we will take 5 months and 18 days i.e. from 01-01-2020 to 18-06-2020 (170 days).

Now, Applying linear regression,

Here, y=PM2.5 & x=RH,

So, our regression line would be y=mx+c.

Mean of x=5508.44/170=32.40

Mean of y=6781.33/170=39.89

Sx=(235867.6-(5508.44^2)/170)/169=339.525

Sy=(454020.8-(6781.33^2)/170)/169=1085.87

Sxy=(285400.1-(5508.44*6781.33)/170)/169=388.56

So, Co-relation coefficient (r)=388.56/(339.525*1085.87)=0.00105

So, m= r*Sy/Sx =0.00105*1085.87/339.525=0.00335 &

c=39.89-0.00335*32.40=39.781

So, linear regression equation is y=0.00335*x +39.781.

Now, to verify remaining 30% data,

y1 = m*x1 + c

Date	PM2.5(y)	Humidity(x)	y1=0.00335*x+39.781	z=(y1 - y)/y1	z^2
19-06-2020	14.757123	60.96594858	39.98523593	0.630935703	0.398079861
20-06-2020	15.74209246	64.33055266	39.99650735	0.606413322	0.367737117
21-06-2020	20.31931897	55.85684503	39.96812043	0.491611846	0.241682207
22-06-2020	18.52918694	57.176164	39.97254015	0.536452103	0.287780859
23-06-2020	17.22747826	67.49530435	40.00710927	0.569389577	0.32420449
24-06-2020	20.61870003	60.09176225	39.9823074	0.484304399	0.234550751
25-06-2020	29.8402223	56.65022577	39.97077826	0.253449055	0.064236423
26-06-2020	18.55995829	63.51129649	39.99376284	0.535928681	0.287219551
27-06-2020	7.111652174	65.3426087	39.99989774	0.822208241	0.676026392
28-06-2020	14.56636553	62.41556637	39.99009215	0.635750638	0.404178874
29-06-2020	12.4923558	58.09867964	39.97563058	0.687500719	0.472657239
30-06-2020	11.55323591	53.69798191	39.96088824	0.710886409	0.505359486
01-07-2020	12.46089677	49.23218631	39.94592782	0.688055893	0.473420912
02-07-2020	15.48957609	53.38082001	39.95982575	0.61237128	0.374998585
03-07-2020	16.41113043	59.27269565	39.97956353	0.589512016	0.347524417
04-07-2020	10.72446143	54.16018068	39.96243661	0.731636448	0.535291892
05-07-2020	15.24965229	69.45514604	40.01367474	0.618888982	0.383023573
06-07-2020	5.183460737	67.74774149	40.00795493	0.870439248	0.757664484
07-07-2020	7.84358707	68.13868613	40.0092646	0.80395573	0.646344816
08-07-2020	9.766504517	59.6066713	39.98068235	0.755719414	0.571111833
09-07-2020	10.54308548	56.91452397	39.97166366	0.73623601	0.542043463
10-07-2020	14.40841154	59.36670143	39.97987845	0.639608421	0.409098932
11-07-2020	18.5266087	62.35408696	39.98988619	0.536717644	0.288065829
12-07-2020	9.557678944	71.97567755	40.02211852	0.761190079	0.579410337
13-07-2020	9.504517026	59.6542738	39.98084182	0.762273214	0.581060453
14-07-2020	19.66620306	61.81571627	39.98808265	0.508198399	0.258265613
15-07-2020	20.32036136	62.90826963	39.9917427	0.491886075	0.241951911
16-07-2020	30.53284672	62.3274244	39.98979687	0.236484076	0.055924718
17-07-2020	20.8515299	63.56467316	39.99394166	0.478632787	0.229089345
18-07-2020	8.379909628	57.98574904	39.97525226	0.790372564	0.62468879
19-07-2020	12.70629128	60.07542579	39.98225268	0.682201716	0.465399182
20-07-2020	20.08133472	66.39346542	40.00341811	0.498009528	0.24801349
21-07-2020	18.49965254	64.36865879	39.99663501	0.537469776	0.28887376
22-07-2020	14.67292318	59.18526243	39.97927063	0.632986722	0.40067219
23-07-2020	22.23662265	58.59277276	39.97728579	0.443768575	0.196930548
24-07-2020	11.93884642	57.28005559	39.97288819	0.7013264	0.491858719
25-07-2020	14.68682656	63.72019465	39.99446265	0.6327785	0.40040863
26-07-2020	19.63955509	62.17309698	39.98927987	0.508879501	0.258958346
27-07-2020	22.4923558	58.8533704	39.97815879	0.4373839	0.191304676
28-07-2020	31.20542231	58.11261731	39.97567727	0.219389778	0.048131875
29-07-2020	37.98957609	65.54586518	40.00057865	0.050274337	0.002527509
30-07-2020	21.90852174	60.48869565	39.98363713	0.452062811	0.204360785
31-07-2020	20.52	56.6306087	39.97071254	0.486624113	0.236803027
01-08-2020	15.99131341	60.56045865	39.98387754	0.600055963	0.360067158
02-08-2020	25.65380605	60.26590198	39.98289077	0.358380408	0.128436517
03-08-2020	20.91657977	70.81786583	40.01823985	0.477323844	0.227838052
04-08-2020	19.44614315	70.56219597	40.01738336	0.514057604	0.26425522
05-08-2020	19.50729672	67.91973593	40.00853112	0.512421572	0.262575867

06-08-2020	13.46226087	68.45217391	40.01031478	0.663530244	0.440272384
07-08-2020	9.569725421	72.1438999	40.02268206	0.760892451	0.578957322
08-08-2020	11.61737318	72.31688673	40.02326157	0.709734471	0.50372302
09-08-2020	4.764186371	69.57753825	40.01408475	0.880937265	0.776050465
10-08-2020	12.78484005	69.80980529	40.01486285	0.680497716	0.463077142
11-08-2020	25.86714184	76.68671766	40.0379005	0.35393361	0.125269
12-08-2020	7.849235042	71.2701669	40.01975506	0.80386599	0.64620053
13-08-2020	1.498054203	77.38985407	40.04025601	0.962586298	0.926572381
14-08-2020	1.279999998	72.18434783	40.02281757	0.968018244	0.93705932
15-08-2020	2.893326381	71.31803962	40.01991543	0.927702836	0.860632552
16-08-2020	6.555076495	66.79346314	40.0047581	0.836142579	0.699134412
17-08-2020	10.58353023	71.33912439	40.01998607	0.73554388	0.5410248
18-08-2020	13.42444367	75.17002782	40.03281959	0.664664048	0.441778297
19-08-2020	7.816052821	74.36101459	40.0301094	0.804745654	0.647615568
20-08-2020	11.04035454	70.53632256	40.01729668	0.724110436	0.524335923
21-08-2020	9.668682649	79.27250608	40.0465629	0.758563983	0.575419316
22-08-2020	1.202712101	94.26738526	40.09679574	0.970004783	0.940909278
23-08-2020	1.249391726	85.46159194	40.06729633	0.968817668	0.938607674
24-08-2020	5.74659486	74.51737318	40.0306332	0.856445067	0.733498153
25-08-2020	8.564650683	72.5950643	40.02419347	0.78601316	0.617816688
26-08-2020	14.57940257	64.72282042	39.99782145	0.635495083	0.403854001
27-08-2020	17.53531457	73.83350713	40.02834225	0.561927535	0.315762554
28-08-2020	18.98217889	86.20361991	40.06978213	0.526271972	0.276962188
29-08-2020	5.869318962	96.86970118	40.1055135	0.853653065	0.728723556
30-08-2020	0.449878261	80.25565217	40.04985643	0.988767044	0.977660268
31-08-2020	5.081584431	70.27275886	40.01641374	0.873012498	0.762150821
				Sum of z^2 =	33.22317632

normalize error = sq. root of (33.22317632)

= 5.7639

Q6.

Let x1=RH, x2=Temperature, y=PM2.5 & a,b,c be constants.

So, our linear regression equation would be

y = a*x1 + b*x2 + c

Now,

Mean of x1 = 10413.11/244 = 42.67

Mean of x2 = 6909.128/244 = 28.31

Mean of y = 7838.098/244 = 32.12

Sx1 = (566863.9 - (10413.11^2)/244)/244 = 501.91

Sx2 = (216368.7 - (6909.128^2)/244)/244 = 84.95

Sx1y = (353268.2 – (10413.11*7838.098)/244)/244 = 76.90

Sx2y = (216126.1 – (6909.128*7838.098)/244)/244 = -23.84

Sx1x2 = (303807.7 – (10413.11*6909.128)/244)/244 = 36.67

We know that,

a= (Sx2*Sx1y – Sx1x2*Sx2y)/(Sx1*Sx2-(Sx1x2^2))

= (84.95*76.90-36.67*(-23.84))/(501.91*84.95-(36.67^2))

= 0.179

b= (Sx1*Sx2y – Sx1x2*Sx1y)/(Sx1*Sx2-(Sx1x2^2))

= (501.91*(-23.84)-36.67*76.9)/(501.91*84.95-(36.67^2))

= -0.358

c= 32.12 – 0.179*42.67 + 0.358*28.31

= 34.617

Therefore, our linear regression equation is

y = 0.179*x1 – 0.358*x2 + 34.617 .

Q7.

Hypothesis Test :

Ho : No relation between x and y (in question 5 x is RH and in question 6 x1 is RH and x2 is temperature and y is same in both i.e. PM2.5)

Ha : there exist relation between x and y .

In Q.6 P value is very close to zero and r value is high which signifies that Ho is false and Ha is true hence there is significant relation between PM 2.5 and RH , Temperature.

In Q.5 P value is small but r value is less which means relation between PM2.5 and RH is weak , hence Ho holds true.

Hence Q.6 regression is is very suitable for determining PM2.5 from temperature and RH as compared to Q5.