

Title:

Using statistical methodology, give an estimate of the duration in which this novel coronavirus (COVID-19) outbreak may last in India in Data Analytics/ Discovery Science Challenges.

Authors:

Jayesh Khandelwal (B19EE040)

Shubh Doshi (B19EE080)

Guided By:

Nil Kamal Hazra

Introduction:

The COVID-19 has caught all of us on nerves. People all over the world are suffering from this deadly virus since last few months. This pandemic has caused severe economic damage to our country. Everybody is eagerly waiting for the end of the pandemic so that everything can become normal as it was earlier. So, we have finally made a model to estimate when the pandemic will end.

Model:

To understand this model, one should know some basic facts:

1. *This pandemic will end when there are zero new cases on a particular day i.e. total no. of cases on i^{th} day = total no. of cases on $(i+1)^{\text{th}}$ day. So by this, we can say that to end this pandemic we will have to make total number of cases to be constant.*

To make our calculations easy, we took the ratio of total no. of cases on i^{th} day ($x(t)$) / total no. of cases on $(i-1)^{\text{th}}$ day ($x(t-1)$).

$$\text{Ratio } X = x(t) / x(t-1)$$

When total no. of cases will become constant, ratio $X = 1$.

But we cannot only consider the factor of total no. of cases to be constant to calculate the time required for this pandemic to end. It will also depend on other factors.

To overcome this problem, we considered another basic fact.

2. *This pandemic will end when there are zero new deaths on a particular day i.e. total no. of deaths on i^{th} day = total no. of deaths on $(i+1)^{\text{th}}$ day. So by this, we can say that to end this pandemic we will have to make total no. of deaths to be constant.*

To make our calculations easy, we took the ratio of total no. of deaths on i^{th} day ($y(t)$) / total no. of deaths on $(i-1)^{\text{th}}$ day ($y(t-1)$).

$$\text{Ratio } Y = y(t) / y(t-1)$$

When total no. of deaths will become constant, ratio $Y = 1$.

So here, we have considered two parameters X and Y to calculate when this pandemic will end. But there are also other factors which affect this calculation and cannot be calculated like number of people following lockdown rules, immunity of people (it varies from person to person), facilities for curing the infected people (it varies in different cities) etc.

These mentioned factors are very complex to calculate which are out of our scope. So, we will include only two factors i.e. ratio X and Y.

Data:

We started our data from 16th April and collected till 15th July (91 days).

We collected data of total no. of cases and total no. of deaths on each day during the mentioned period.

We calculated ratio X and ratio Y from the collected data.

All the data is mentioned in our data set.

Data set:

Date	Time (t)	X(t) for cases	X(t-1)	X = X(t) / X(t-1)	Y(t) for death	Y(t-1)	Y = Y(t) / Y(t-1)
16-Apr	1	12675	11849	1.069710524	418	390	1.071794872
17-Apr	2	13751	12675	1.084891519	450	418	1.076555024
18-Apr	3	14708	13751	1.069594939	486	450	1.08
19-Apr	4	16032	14708	1.090019037	517	486	1.063786008
20-Apr	5	17572	16032	1.096057884	557	517	1.077369439
21-Apr	6	18901	17572	1.075631687	601	557	1.078994614
22-Apr	7	20387	18901	1.078620179	650	601	1.081530782
23-Apr	8	21616	20387	1.060283514	684	650	1.052307692
24-Apr	9	23368	21616	1.081051073	721	684	1.054093567
25-Apr	10	24858	23368	1.06376241	777	721	1.077669903

26-Apr	11	26833	24858	1.079451283	824	777	1.06048906
27-Apr	12	28296	26833	1.054522416	884	824	1.072815534
28-Apr	13	29890	28296	1.056333051	935	884	1.057692308
29-Apr	14	31703	29890	1.060655738	1006	935	1.075935829
30-Apr	15	33526	31703	1.057502445	1073	1006	1.066600398
01-May	16	35281	33526	1.052347432	1150	1073	1.071761417
02-May	17	37692	35281	1.068337065	1221	1150	1.06173913
03-May	18	40179	37692	1.065982171	1304	1221	1.067977068
04-May	19	42752	40179	1.064038428	1387	1304	1.063650307
05-May	20	46627	42752	1.090639034	1581	1387	1.139870224
06-May	21	49307	46627	1.057477427	1692	1581	1.070208729
07-May	22	52868	49307	1.072220983	1781	1692	1.052600473
08-May	23	56258	52868	1.064121964	1884	1781	1.057832678
09-May	24	59578	56258	1.059013829	1979	1884	1.050424628
10-May	25	62855	59578	1.055003525	2107	1979	1.064679131
11-May	26	67068	62855	1.067027285	2204	2107	1.046037019
12-May	27	70672	67068	1.053736506	2291	2204	1.039473684
13-May	28	74197	70672	1.049878311	2413	2291	1.053251855
14-May	29	77919	74197	1.050163753	2547	2413	1.055532532
15-May	30	81886	77919	1.050911844	2647	2547	1.039261877
16-May	31	85856	81886	1.048482036	2750	2647	1.038911976
17-May	32	90843	85856	1.058085632	2870	2750	1.043636364
18-May	33	96085	90843	1.057703951	3027	2870	1.054703833
19-May	34	101055	96085	1.051725035	3161	3027	1.044268252
20-May	35	106666	101055	1.055524219	3301	3161	1.044289782
21-May	36	112275	106666	1.052584704	3433	3301	1.039987882
22-May	37	118363	112275	1.054224004	3581	3433	1.043110982
23-May	38	125017	118363	1.056216892	3718	3581	1.03825747
24-May	39	131784	125017	1.054128639	3865	3718	1.039537386
25-May	40	138761	131784	1.052942694	4019	3865	1.039844761
26-May	41	145296	138761	1.047095365	4165	4019	1.036327445
27-May	42	151683	145296	1.04395854	4335	4165	1.040816327
28-May	43	158249	151683	1.043287646	4529	4335	1.044752018
29-May	44	165715	158249	1.047178813	4704	4529	1.038639876
30-May	45	173679	165715	1.048058414	4969	4704	1.056335034
31-May	46	182059	173679	1.048249932	5162	4969	1.038840813
01-Jun	47	190451	182059	1.046094947	5392	5162	1.044556373
02-Jun	48	198622	190451	1.042903424	5596	5392	1.037833828
03-Jun	49	207531	198622	1.044854044	5813	5596	1.038777698
04-Jun	50	216835	207531	1.044831856	6073	5813	1.044727335
05-Jun	51	226686	216835	1.045430858	6346	6073	1.044953071
06-Jun	52	236573	226686	1.043615398	6640	6346	1.046328396
07-Jun	53	246544	236573	1.042147667	6927	6640	1.043222892
08-Jun	54	256527	246544	1.040491758	7198	6927	1.039122275
09-Jun	55	266514	256527	1.038931574	7469	7198	1.037649347

10-Jun	56	276499	266514	1.037465199	7743	7469	1.036684965
11-Jun	57	286495	276499	1.036152029	8100	7743	1.04610616
12-Jun	58	297451	286495	1.038241505	8496	8100	1.048888889
13-Jun	59	308909	297451	1.03852063	8882	8496	1.045433145
14-Jun	60	320838	308909	1.038616551	9193	8882	1.035014636
15-Jun	61	332340	320838	1.035849868	9518	9193	1.035352986
16-Jun	62	343007	332340	1.032096648	9898	9518	1.039924354
17-Jun	63	353981	343007	1.031993516	11901	9898	1.202364114
18-Jun	64	366862	353981	1.036388959	12235	11901	1.028064868
19-Jun	65	380448	366862	1.037032999	12571	12235	1.027462199
20-Jun	66	394964	380448	1.038155017	12946	12571	1.029830562
21-Jun	67	410377	394964	1.03902381	13252	12946	1.023636645
22-Jun	68	425198	410377	1.036115572	13697	13252	1.033579837
23-Jun	69	440131	425198	1.035120109	14009	13697	1.022778711
24-Jun	70	456099	440131	1.036280108	14474	14009	1.033192947
25-Jun	71	473021	456099	1.037101594	14892	14474	1.02887937
26-Jun	72	490317	473021	1.036564973	15299	14892	1.02733011
27-Jun	73	508869	490317	1.037836746	15683	15299	1.02509968
28-Jun	74	528775	508869	1.039118123	16093	15683	1.026142957
29-Jun	75	548234	528775	1.036800151	16473	16093	1.023612751
30-Jun	76	566756	548234	1.033784844	16891	16473	1.025374856
01-Jul	77	585409	566756	1.03291187	17398	16891	1.030015985
02-Jul	78	604557	585409	1.032708756	17832	17398	1.024945396
03-Jul	79	625460	604557	1.034575731	18211	17832	1.021253926
04-Jul	80	648231	625460	1.036406805	18653	18211	1.024271045
05-Jul	81	673081	648231	1.0383351	19266	18653	1.032863346
06-Jul	82	697329	673081	1.036025382	19691	19266	1.022059587
07-Jul	83	719581	697329	1.031910332	20158	19691	1.023716419
08-Jul	84	742333	719581	1.0316184	20640	20158	1.023911102
09-Jul	85	767212	742333	1.033514609	21127	20640	1.023594961
10-Jul	86	793718	767212	1.034548469	21602	21127	1.022483079
11-Jul	87	820832	793718	1.034160747	22121	21602	1.024025553
12-Jul	88	849469	820832	1.034887772	22672	22121	1.024908458
13-Jul	89	878170	849469	1.033786989	23172	22672	1.022053634
14-Jul	90	906668	878170	1.032451575	23725	23172	1.023865009
15-Jul	91	936097	906668	1.032458408	24307	23725	1.024531085

Code:

We have used machine learning to find the equation depending on two variables X and Y and the graph of the equation. Code for the same can be found from the link given below:

https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/5df18bbf-8533-418d-8744-4ed461152342/view?access_token=74e020f8dbd92faa34c39ae89075d1555e81d7e2b260835af84903cof9654433

Explanation of code:

We performed the coding and calculation part in Jupyter Notebook, the link is provided above.

First we upload the data as a Data Frame in the Notebook. We add the data set in Db2 Cloud (which is an API). Then we connect our Notebook to Db2 Cloud database and get the data table as Data Frame.

Then we use Machine Learning Techniques to estimate when the new cases ratio and Total Death Ratio will be equal to 1.

For this we use various libraries like sklearn , Polynomial Features to transform the data X , Y ratios in a equation of n degree.

We find the most suitable model by calculating accuracy of models with different n to avoid over fitting or under fitting. The best degree which came n = 2.

So we found our equation on which the ratios and days are dependent.

Futher we use our Data Visualization techniques to plot the graphs for better understanding on how our model ensures that the pandemic will end.

Result:

$$Z = 33,473.689 - 43,256.25 (X) - 18,565.935(Y) + 16,486.82(X*Y) + 7,042.04(X^2) + 5,018.95(Y^2)$$

Here, X = total no. of cases on i^{th} day ($x(t)$) / total no. of cases on $(i-1)^{th}$ day ($x(t-1)$)

Y = total no. of deaths on i^{th} day ($y(t)$) / total no. of deaths on $(i-1)^{th}$ day ($y(t-1)$)

Z = no. of days

So to calculate when the pandemic will end, we have to put value of $X = 1$ and $Y = 1$.

By this, we get value of $Z = 199.314$ days

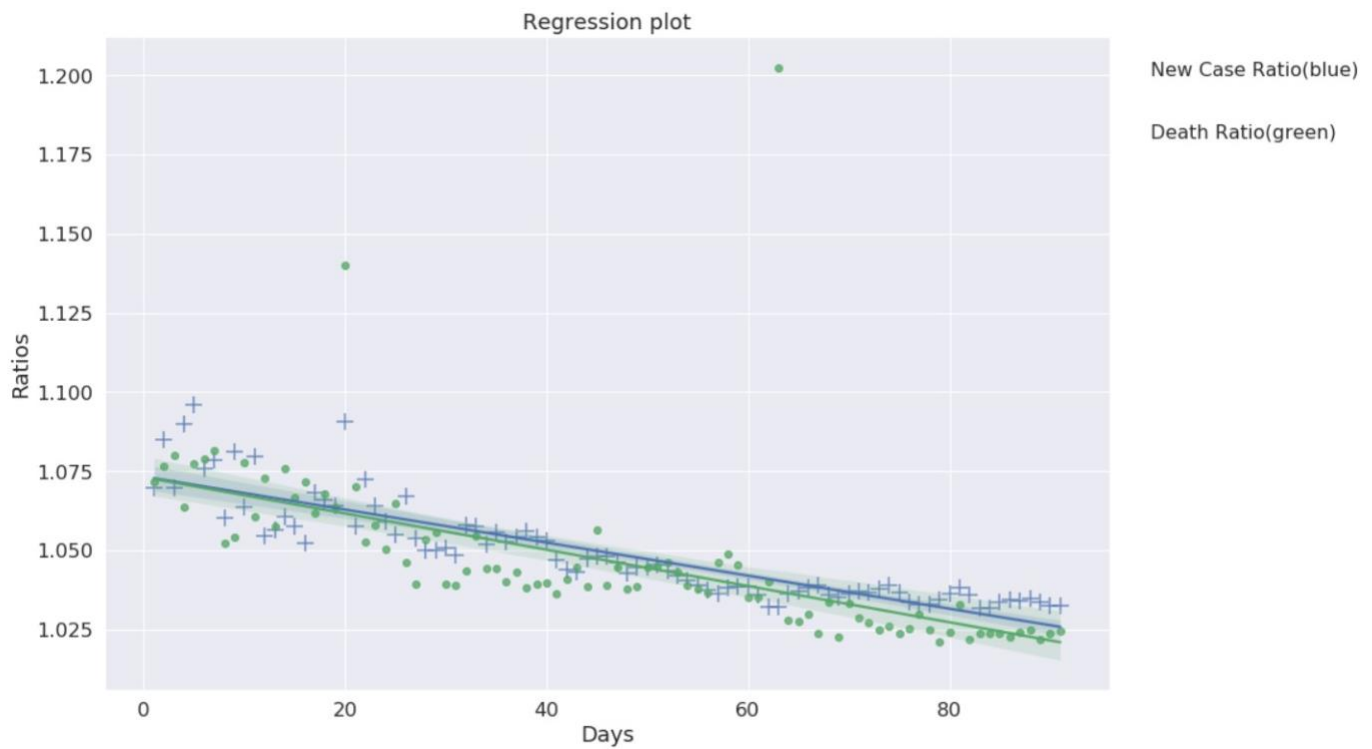
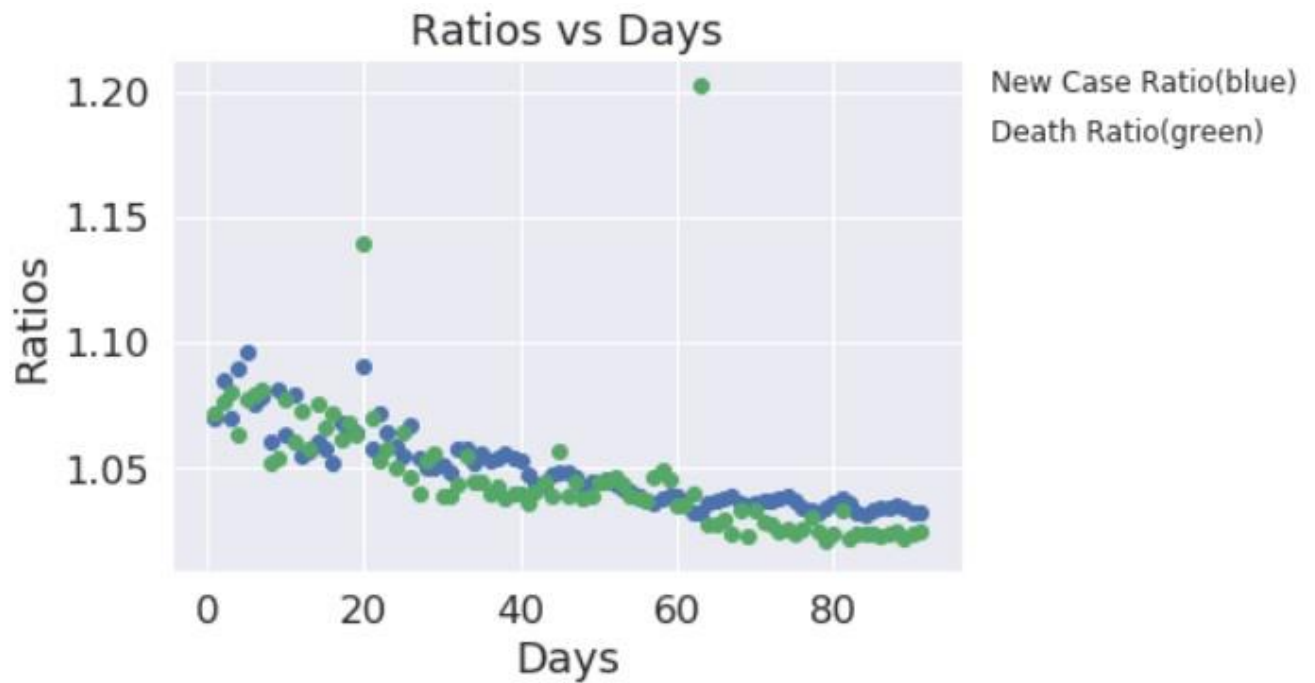
$Z = 1$ was on 15th April, 2020

So, $Z = 200$ will be on 1st November, 2020

Also, the accuray of our calculation came out to be 0.97938 which can be maximum 1. So, our calculation is quite accurate. This accuracy was also calculated from the code.

So according to our model, this pandemic will be over on 1st November, 2020.

Visualization:



Bibliography:

Data : <https://www.mygov.in/covid-19>

Machine Learning : Coursera, Udemy and YouTube

Other Sources : Wikipedia , WHO