Applied Capstone Project - Analyzing the effect of Covid-19

1.Introduction:

Background -

Coronavirus disease 2019 (COVID-19) is defined as an illness caused by the novel coronavirus. It was initially reported to the WHO on December 31, 2019. On January 30, 2020, the WHO declared the COVID-19 outbreak as a global health emergency. On March 11, 2020, it was declared a global pandemic. With the worldwide cases nearing 9 million and 450 thousand people already succumbing to the virus, it has become a big concern. As countries strive to deal with the economic impacts caused by the COVID-19 pandemic, there is an urgent need for new evidence and research to improve our understanding of the economic crisis in affected countries and the appropriate policy responses.

Data Science has played an important role in the world's fight against Covid-19. Various dashboards of Covid-19 cases have been created and referred to by people around the world. This has generated tremendous amount of data, which has proven to be very useful. Further analysis of this data to give more meaningful conclusions may be invaluable and may give us patterns which have not been previously observed.

Description –

Data from various countries regarding Covid-19 cases will be explored and visual plot renditions will be produced for many variables. The data would include :

- Confirmed Cases
- Number of deaths
- Number of new cases
- Number of new deaths

This would be used to create plots for various countries all around the world. The cases in particular locations (e.g. New York) would then be

observed and it would be inferred which neighbourhoods have reported more cases. Data from Covid free countries (E.g. New Zealand) would also be analysed to see how their trend was and how the number of daily cases changed in these places.

Target Audience –

The data analysis performed would give results which would be useful to:

- The Government - To make strategies according to the measures which have successfully reduced Covid cases
- Health Representatives - It can be seen which places are in need of greater medical assistance
- Common People - To remain informed about the spread of the coronavirus in their locality and take measures accordingly
- Epidemiologists - To understand how the Covid 19 virus spread and how to prevent such pandemics/ be better prepared in the future
- And many more..

Making the analysis incredibly useful to a large target audience.

2.Data:

Data 1 :

The primary data which is to be used for this analysis would be the number of Covid cases in different countries on different dates. This will help us to see, for e.g. How Covid cases in different parts of the world have changed overtime.

This data is available online for free at
: https://covid.ourworldindata.org/data/ecdc/total_cases.csv

Data 2 :

We would also use the data for the total number of deaths which were caused by the coronavirus in different countries overtime to find trends and perform an analysis.

This data is available online for free at
: https://covid.ourworldindata.org/data/ecdc/total_deaths.csv

Data 3:

We would also use the data for new Covid cases as this will help us to for e.g. observe how the daily growth in Covid cases has changed overtime.

This data is available online for free at
: https://covid.ourworldindata.org/data/ecdc/new_cases.csv

Data 4:

We would also use the data for new Covid deaths as this will help us to for e.g. observe how the daily growth in Covid deaths have changed overtime.

This data is available online for free at
: https://covid.ourworldindata.org/data/ecdc/new_deaths.csv

Out[20]:

	date	World	Afghanistan	Albania	Algeria	Andorra	Angola	Anguilla	Antigua and Barbuda	Argentina	...	United States Virgin Islands	Uruguay	Uzbekistan	Vatican	Venezuela
0	2019-12-31	0	0.0	NaN	0.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
1	2020-01-01	0	0.0	NaN	0.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
2	2020-01-02	0	0.0	NaN	0.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
3	2020-01-03	0	0.0	NaN	0.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
4	2020-01-04	0	0.0	NaN	0.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN

5 rows × 212 columns

Data 5:

Lastly, data from New York was also used in my analysis in order to explore a particular area and compare how Covid has impacted different neighbourhoods in the city of New York.

This data was available for free on GitHub at :
https://raw.githubusercontent.com/nychealth/coronavirus-data/master/data-by-modzcta.csv

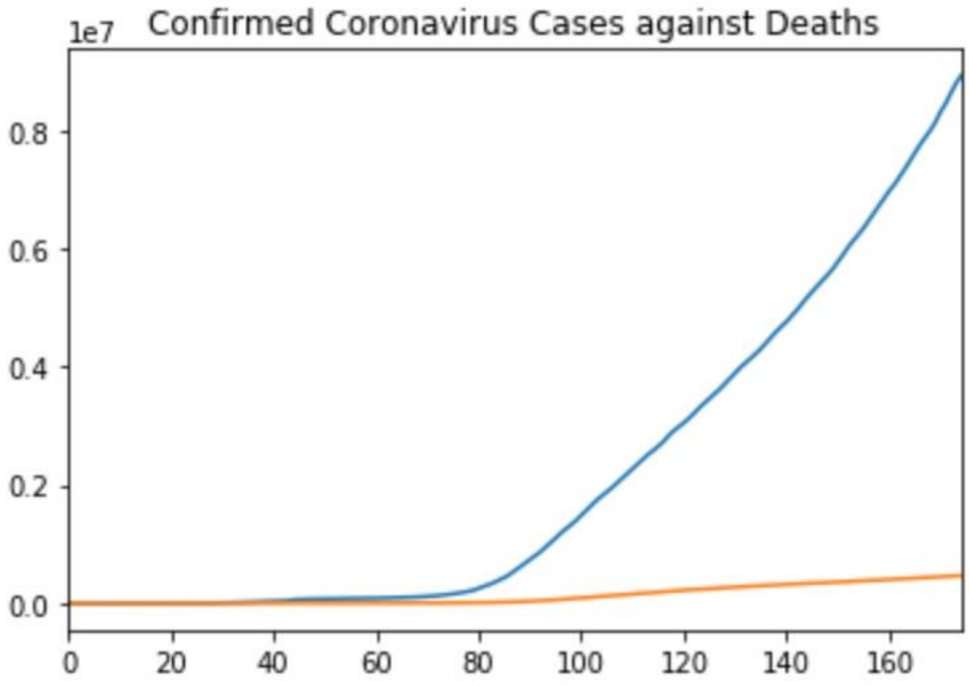
	MODIFIED_ZCTA	NEIGHBORHOOD_NAME	BOROUGH_GROUP	COVID_CASE_COUNT	COVID_CASE_RATE	POP_DENOMINATOR	COVID_DEATH_COUNT	COVID_DEATH_RATE	PERCENT_POSITIVE
78	10473	Castle Hill/Clason Point/Soundview	Bronx	1958	3161.88	61925.28	129	208.32	23.72
89	11201	Brooklyn Heights/DUMBO/Downtown Brooklyn	Brooklyn	726	1191.96	60908.15	89	146.12	9.02
1	10002	Chinatown/Lower East Side	Manhattan	1080	1407.07	76755.41	152	198.03	16.13
146	11377	Woodside	Queens	2316	2678.43	86468.55	185	213.95	20.14
46	10303	Graniteville/Mariner's Harbor/Port Ivory	Staten Island	895	3411.91	26231.66	44	167.74	24.68

3.Methodology:

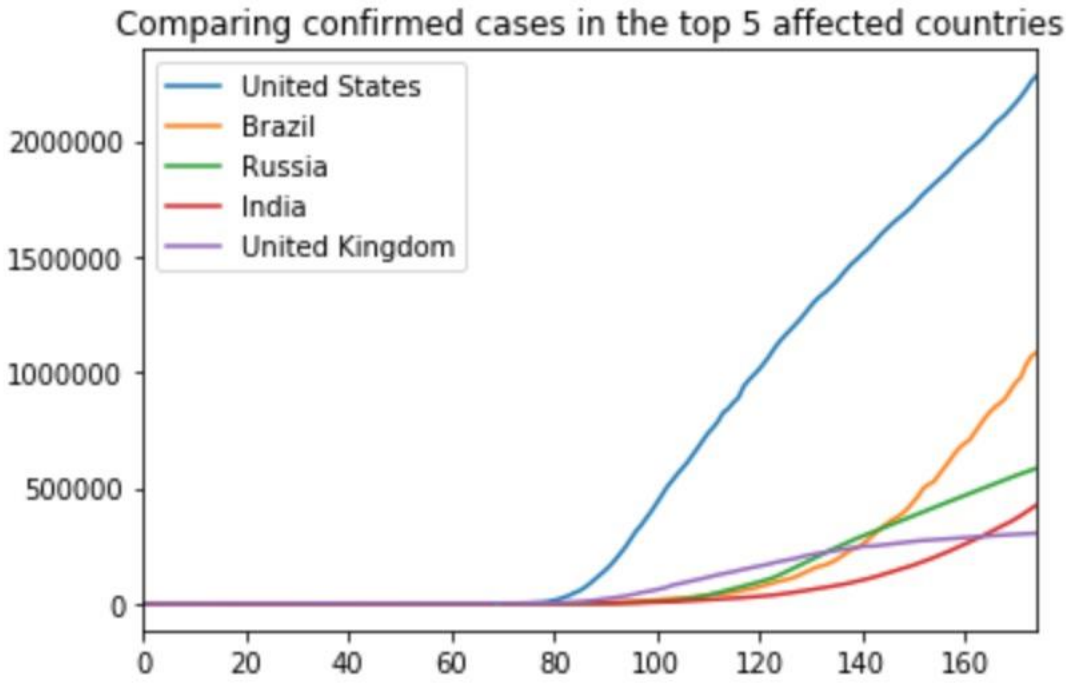
The main goal of this analysis was to analyse the Covid data available and use it in order to generate trends and gather meaningful information about the impact of the Coronavirus on various countries.

Analytic Approach –
For this analysis, various countries around the world were used, but, in order to see clear trends and make it easier to understand and interpret data, extremes (for e.g. top 5 most affected countries) were used. The data from these places was interpreted and exploratory data analysis was conducted.

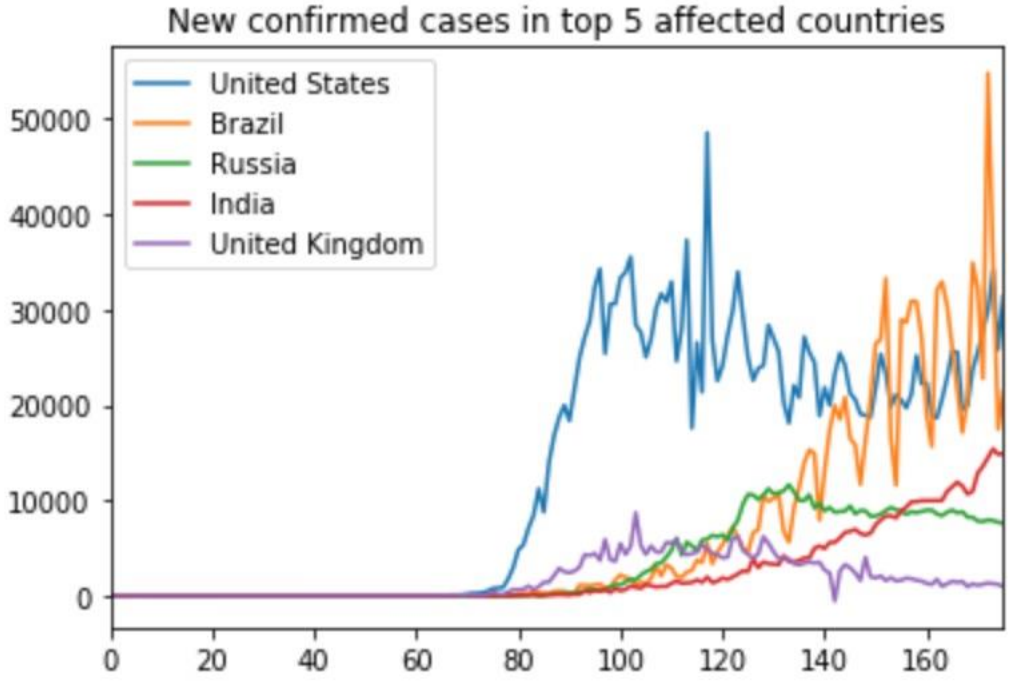
Exploratory Data Analysis –

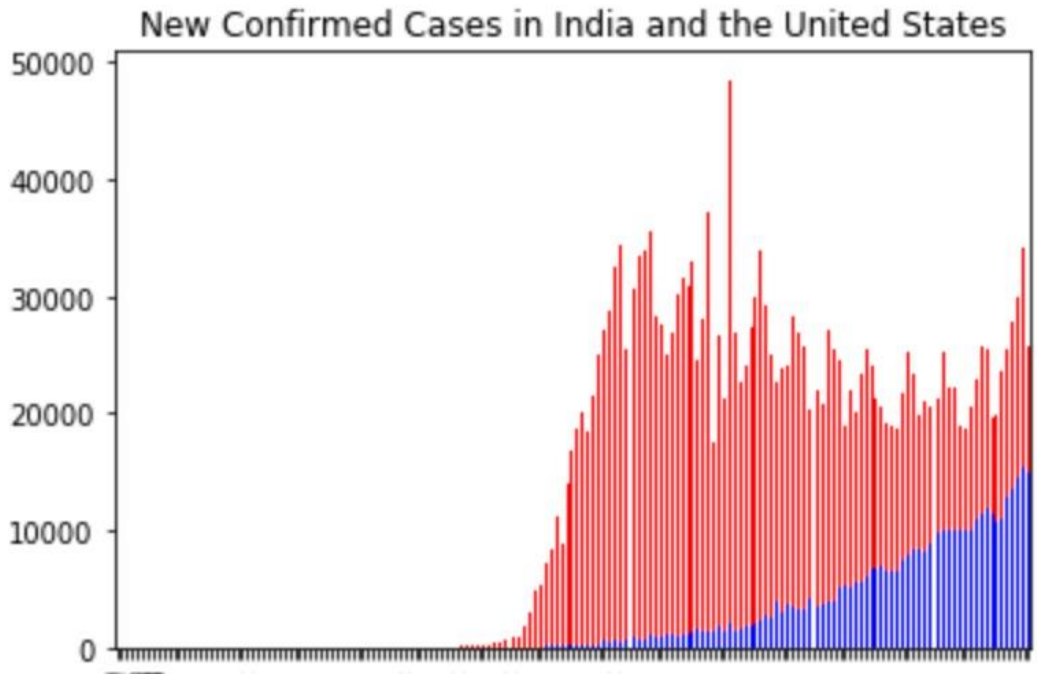


For the first analysis, the situation was taken on a macro level and confirmed coronavirus cases in the world were taken into consideration. Confirmed cases were compared to the deaths over a period in excess of 160 days. As the graph suggests, the deaths and the number of cases were similar in quantity until about 45 days, but then the number of cases started spreading much quickly, resulting in a large gradient increase and, thankfully, a decreased mortality rate.



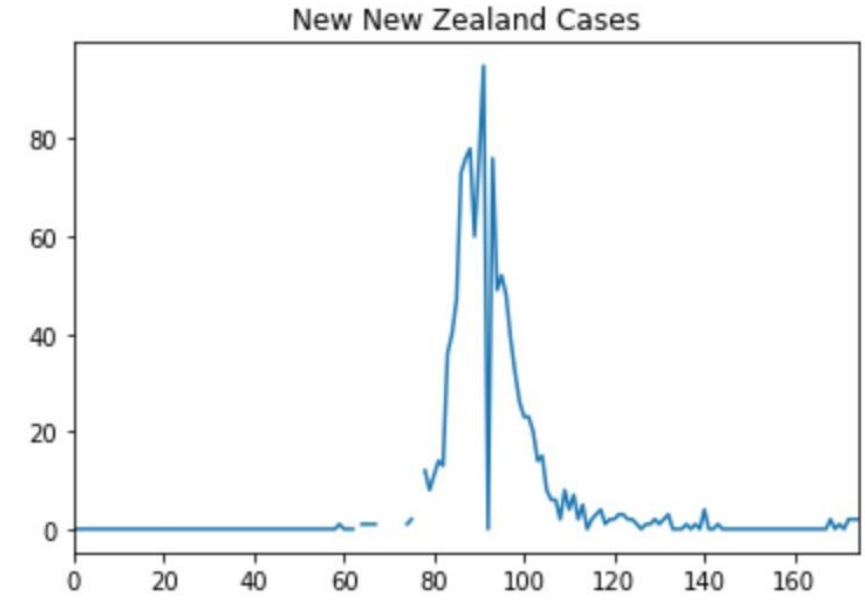
The next analysis was performed on the number of confirmed cases in the top 5 affected countries (i.e. United States, Brazil, Russia, India, and the United Kingdom). This analysis was highly insightful. As the graph suggests, the United States picked up speed very quickly and recorded the greatest number of cases. Brazil had less cases compared to Russia and India until around 130 days, but it soon picked up speed and gathered more cases than both of these countries. Steady rise in the cases in India also shows that the lockdown imposed had little or no effect in controlling the spread of the virus. Russia has shown a steady growth in the number of cases throughout whereas India has increasing cases (as it ideally should be), similar to a logarithmic graph. The United Kingdom, however, first grew quickly compared to other countries, being behind to only the United States, but soon “flattened the curve” as clearly visible in the graph, becoming the fifth most affected country.

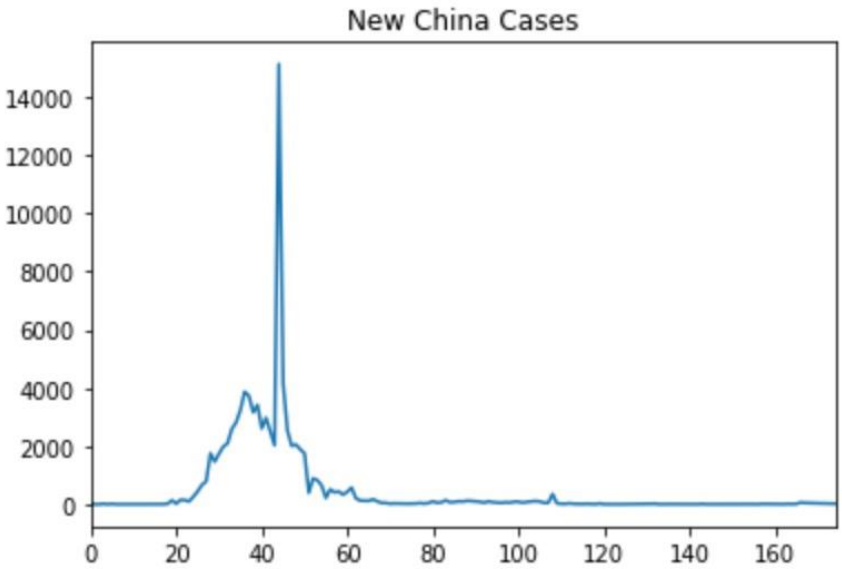




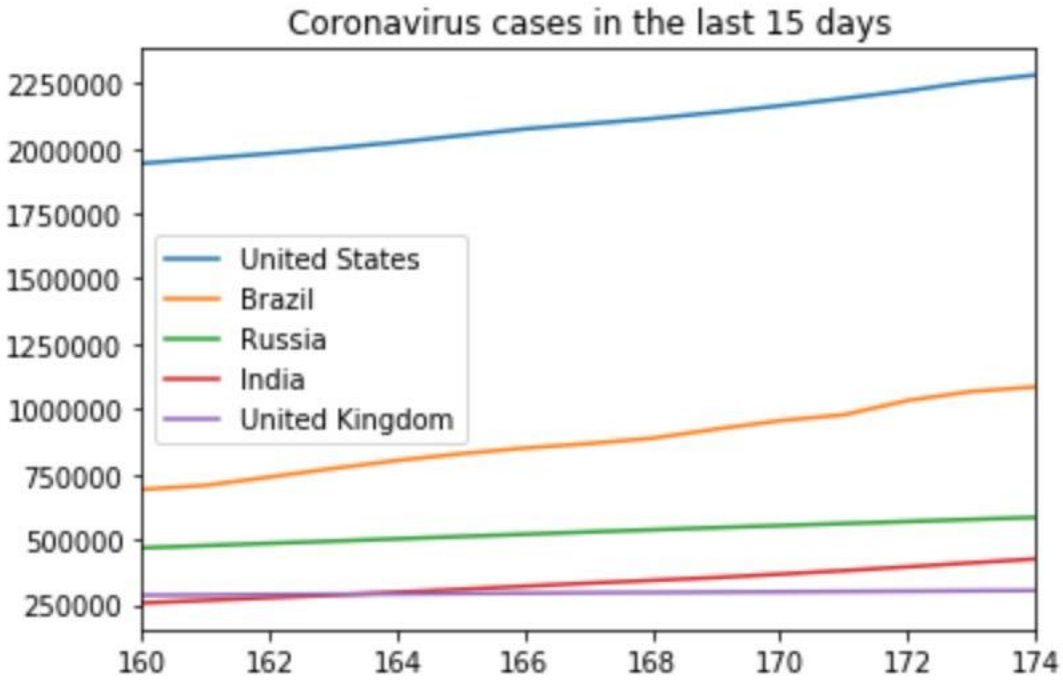
The next analysis is similar to the previous one, however, it focuses on the number of new cases instead of the number of total cases. It provides evidence for the patterns observed in the first graph. For example, the United Kingdom “flattened their curve” and we can see, in this graph, that the number of new cases recently has been close to zero, resulting in no new cases and therefore, a smaller number of total

confirmed cases. The trend India has shown in this analysis too, is consistent, suggesting its logarithmic function-like graph in the previous analysis. The number of new cases in Russia have been declining and that is why their growth in cases is rather steady. The number of new cases in the United States and Brazil are really high (and inconsistent depending on certain situations), causing many total confirmed cases.

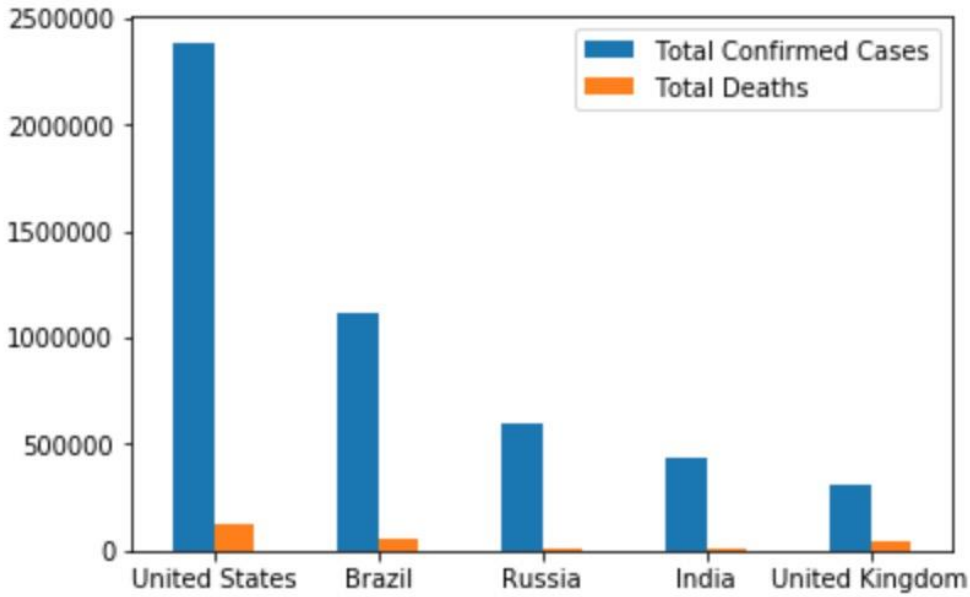




The next analysis was conducted on countries New Zealand and China, which have done well in controlling the spread of the Coronavirus. As observed, these two countries have nearly zero new cases every day, which is why the affected people are recovering and no new diagnosis is taking place, resulting in less cases.



This analysis explores the Coronavirus cases in the last 15 days. As it can be observed, all countries have steady growth other than the UK, which suggests that the cases are increasing. The United Kingdom has done well in terms of ceasing the spread of the virus as the cases are decreasing. The cases in Russia also remain, fairly steady, acting as an indicator of good performance.

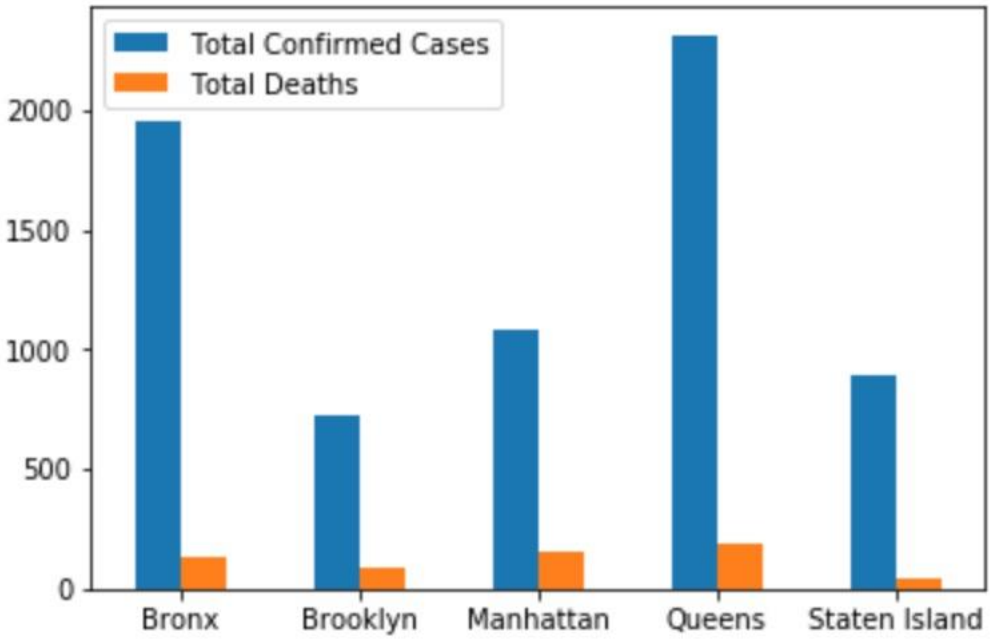


This analysis shows proves the low mortality rate of the coronavirus. All of the countries have reported very few deaths when compared to the number of cases and this can be used as an indicator for governments to remove the lockdown impositions and start the economy.

```
In [153]: from pandas.io.json import json_normalize
          url='https://api.foursquare.com/v2/venues/trending?&client_id=T2MI0S
          results = requests.get(url).json()
          results

Out[153]: {'meta': {'code': 200, 'requestId': '5ef1ed32e826ac00212958ed'},
           'response': {'venues': []}}
```

Using foursquare api to find trending locations in different places resulted in giving no values. The snippet above is from a request made for trending venues in New York. It can be clearly seen that “venues” is empty suggesting that no places are trending due to the coronavirus. This reflects the impact of the virus on the economy and on people visiting public places.



It can be observed that Queens has the greatest number of confirmed cases and Bronx has the second most confirmed cases. The number of deaths remain low in all places, again suggesting a low mortality rate.

4.Discussion:

1. Lockdowns have helped in some countries like the UK while they have not proven beneficial in countries like India and the United States.
2. The mortality rate of the coronavirus is very low as few people are dying, governments could open the economy for people who are at less risk.
3. The coronavirus has significantly decreased public activity. People aren't visiting public places and the direct effect of this is observed on the economy.
4. More healthcare resources should be devoted to Queens and Bronx as they have reported the greatest number of cases among places in New York City.

5.Conclusion:

The effects of Coronavirus can clearly be seen in my analysis. It can be discerned that:

1. The coronavirus has prevented people from visiting public places outside as the trending section of foursquare api suggests.
2. The United States has shown a large increase in the number of cases. It has recorded the most cases in the world and the lockdown hasn't helped control the situation.
3. Russia and the United Kingdom have done well in controlling the coronavirus. The number of daily cases has decreased in both countries and they are reporting less new cases every day and are "flattening the curve".
4. India and Brazil have a rapid growth of Covid cases. The imposition of the lockdown hasn't helped India as it still has a trend of increasing cases similar to a logarithmic function.
5. Countries like New Zealand and China have done very well in combating the coronavirus and can be ideals for other countries.
6. In New York, Bronx and Queens have reported the most Coronavirus cases.