

Case Studies in Data Science (COSC2669)

Final Report WIL Project

Group 52

RMIT University

Somya Singhal (S3813520)

Chih-Ying, Ho (S3816723)

Milind Parvatia (S3806853)

Hill Modi (S3827516)

Yung-Chen, Niu (S3817129)

19 October 2020

TABLE OF CONTENT

1. INTRODUCTION

1.1 ABOUT

1.2 MOTIVATION

1.3 PROBLEM STATEMENT

2. SOLUTION

2.1 RESULTS

2.2 PLOTS/ VISUALIZATIONS

3. METHODOLOGY

3.1 TOOLS AND RESOURCES

3.2 ANALYTICS METHODS

4. IMPACT AND SIGNIFICANCE OF RESULTS

5. PROJECT MANAGEMENT

5.1 TEAM CAPABILITY

5.2 PLANNING

6. CONCLUSION

7. REFERENCES

STUDENT ID	NAME	CONTRIBUTION
S3813520	Somya Singhal	20%
S3816723	Chih-Ying, Ho	20%
S3806853	Milind Parvatia	20%
S3827516	Hill Modi	20%
S3817129	Yung-Chen, Niu	20%

INTRODUCTION

1.1 ABOUT

The spread of COVID-19 pandemic has a negative impact on our health and everyday life. Nowadays, the cases of testing positive are significantly increasing, there are total 39.9 million cases worldwide. Besides, more and more people died of the virus, up to 1.1 million deaths. Sadly, as the pandemic affects people for the first time, which means that the vaccine is not found and is not available for humans. What is even worse is that the virus mutates and becomes more powerful, which causes the second wave in some countries. The rise in people getting COVID-19 is a serious problem. The purpose of this project is to deal with this problem and suggest ways to different people to deal with COVID-19.

1.2 MOTIVATION

The outbreak of coronavirus severely causes people to get sick, including sore throat, fever and dry cough, and even pass away. The problem is the confirmed cases grow and it seems that the epidemic is difficult to control. The COVID-19 is more prevalent during the winter, which means that there are always places that are suitable for the virus to grow.

It is impossible to totally eradicate the virus in a short time. All we can do now is to remind ourselves not to go to the epidemic areas. To achieve this, the virus information becomes very essential to everyone. This analysis provides the updated information to business people, and help them to ensure the safety of the people and deal in their own field.

Therefore, using the data collected by the World in Data, this report identifies regions and people more likely to be affected by the virus. The analysis report helps to focus on the group that tends to get COVID-19, which achieves the goal of reducing the risk of being infected with the virus.

Having access to accurate information about the number of infected people and the infection degree for the countries can prevent virus from spreading uncontrollably.

1.3 PROBLEM STATEMENT

We proposed 7 questions on which we wished to work upon and find answers to them in order to provide information and benefit. These are related to different business areas and will profit the whole world in altogether in this situation of pandemic.

- 1) To doctor we want to suggest what all facilities are available in the most cases areas so that they can suggest government to make more arrangements if required.
- 2) To government we want to tell, how GDP and human index are affecting total covid-19 deaths and figure out some other measures to improve the conditions.
- 3) Giving insights to international trader of New Zealand, for better visibility of how his business is affecting over the years so that he can deal in some other way if needed.
- 4) Giving idea to an Economist on which events had highest effect on volatility in 2020.
- 5) Tell a doctor how harmful Covid-19 in comparison to other major viruses ever, so he can understand which steps I take to prevent spreading.
- 6) For government to know how Deaths are affecting GDP of countries?
- 7) For economist to know how much affect covid-19 has done to stock market?

SOLUTION

2.1 RESULTS

GRAPH 1:

We can find out a rough pattern that the higher the GDP the country has, the more beds it has for the citizens and hence less are the deaths due to more facility and vice versa.

Case 1: India, with GDP of 6426, only 0.53 beds are available for per thousand people and with 77427 deaths.

Case 2: South Korea, with GDP of 35938, up to 12.27 beds are available for per thousand people and with only 355 deaths. Hence as GDP increases, the country can afford more medical services and thus it is resulting in saving people lives.

However, there is an exception that although the USA has GDP 54225, only 2.77 beds are available for per thousand people, which causes 193016 deaths.

Therefore, I suggest to the government to put more financial resources on the medical services to help deal with the COVID-19.

GRAPH 2:

There is a positive correlation between the GDP and HDI, i.e. the higher the GDP is, the higher the human development index is, which means a better standard of living.

Case 1: Niger is with only 926 GDP and 0.354 HDI, so we can say Niger is a poor country and the number of COVID-19 deaths is only 2.85 per million.

Case 2: USA is with extremely high 54225 GDP and 0.924 human development index. However, the total deaths number is up to 583 per million people.

As a government, I can conclude that the higher the HDI is, the higher the death number will be. It is because the higher HDI means that citizens have the capability of traveling around the world, which can cause the prevalence of the COVID-19 whether internationally or domestically.

To government, I want to suggest those high-developed countries to increase their lockdown level.

GRAPH 3:

From 2015-2016 the import-export business earned almost the same value. And it had an increasing trend in the money value the following years until the year 2020, the pandemic year, where export business was slightly down, but on the other hand import business had a downfall as the country had put on high lockdown restrictions, travel ban, and restrictions on business which impacted import business as well.

GRAPH 4:

Plot 1:

We can deduce from analysis that CPI, CRUDE OIL INVENTORIES, GDP, INITIAL JOBLESS CLAIMS, and MANUFACTURING PMI are the most affected areas which influence the volatility rate the most. So, because of these events, variation in volatility is highest which is affecting the global economy as well as has an effect on currency value in the international market. These are the area most severely hit during the pandemic.

Plot 2:

Based on events how different countries are being affected. The graph depicts that, USA has been hit the most with highest volatility rate in crude oil inventories and jobless claims. UK comes on second number where variation in manufacturing PMI, GDP and CPI is highest. Australia has least amount of affect in CPI and GDP when compared to other countries.

I suggest to economists that they make relevant recommendations from these insights to improve the efficiency and economy of countries during the pandemic.

GRAPH 5:

Plot 1:

This plot shows the type of disease in logarithmic number of cases as ordinate. COVID-19 climbs the highest spot (at 17 cases) in terms of number of cases as compared to that of other deadly infections like H1N1, Malaria and SARS. Malaria comes second (at 12 cases) and H1N1 being third (at 9 cases). SARS has the least number (at 8 cases) of infectious cases falling just a little shy than H1N1.

Plot 2:

COVID-19 here as well (at 14 deaths) tops the chart by a huge margin. SARS comes second (at 7 deaths) while Malaria grabs third position by a small margin (at 6 deaths). H1N1 is the least deadly diseases amongst all 4 (at 4 deaths).

Plot 3:

This plot shows the type of disease in number of fatality rate as ordinate. Here, SARS gains top position at 9 fatality rates per 100 cases. COVID-19 comes second at 3 fatality rates per 100 cases. H1N1 and Malaria, respectively, have the least fatality rates of at 1 and 0.5 fatality rates per 100 cases respectively.

GRAPH 6:**Plot 1:**

We analyzed effects of GDP on total number of tests per million with increasing GDP the total number of testes per million shows sharp increase. Size of circles shows total number of confirmed cases per million. Considering Australia, their GDP is 44648.71 with testing per million is 181419. While India's GDP per capita is 6426.674 with testing capabilities is 16035. Even though Australia has done testing, total number of confirmed cases are still low. There is cluster between (10000 – 20000 GDP) and (0 – 100000) tests per million. (Peru, South Africa, Colombia and Brazil). This shows total number of tests per million and total cases per million. With increasing number of tests total number of cases have increased as well. This trend is also similar conserving total number of deaths. Some countries (New Zealand, Australia, Canada and Germany) show different trend because these countries have witnessed low number of cases and deaths due to COVID-19. Ergo, they eased the restriction of lockdown. But, for the safe side, they continued the practicing of tests, but noticed low number of confirmed COVID-19 cases.

Plot 3:

It shows total number of cased per million and total number of deaths per million. Total number of deaths per million shows a steep increase with increasing total number of cases per million. This is obvious, as if confirmed cases are huge, then confirmed deaths are also huge.

Plot 4:

Shrinking GDP and Total number of deaths per million are inversely proposal to each other. As death per million have been increased, GDP of countries declined. Comparing Peru with Deaths per million is 924, noticed highest drop in GDP of -30%. While Taiwan lost only 0.294, that's why drop of GDP is only -0.58%.

Overall story is, Countries with higher GDP have afford the more testing, and because they conduct more testing, their total number of 'Confirmed Cases' are also big. As Confirmed Cases have been increases, all countries run towards health of their civilians and imposed the lockdown. As 'Confirmed Cases' were huge, 'Confirmed deaths' number is also whooping. Due to lockdown, whole nation or whole countries were freeze. After certain period, when some countries noticed slowdown of pandemic, they eased the restriction and their industries were back to track, like Taiwan. But countries where death are continuously increasing, they stuck to the lockdown which cost then whopping amount of declination in GDP.

GRAPH 7:

Plot 1 & 3:

Plot 1 shows performance of Bombay Stock Exchange during COVID-19 pandemic. The trading value dropped sharply in late March due to the imposed lockdown in the country. However, upon reducing lockdown, the trading increased sharply and shows a steady increase thereafter. Similar is the trend for NASDAQ in Plot 3.

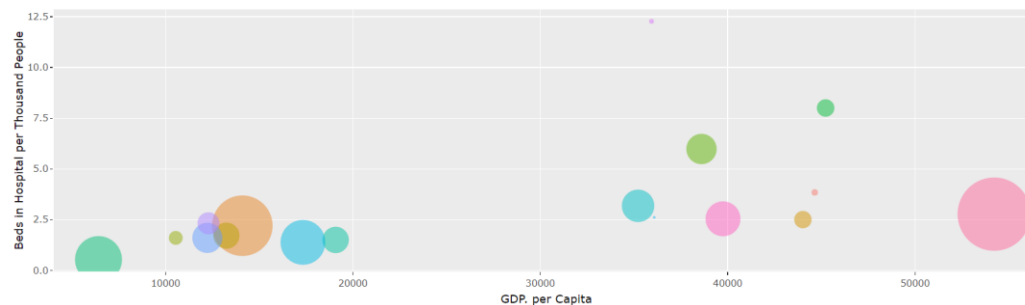
Plot 2 & 4:

Analysis shows Hong Kong and New York Stock Exchange performance during COVID-19. During late March, the trading showed steep decrease in value from nearly \$14000 to \$2000 due to reduced business activities during lockdown. However, upon releasing the lockdown, both shows a near-steady increase in performance with some abrupt stops due to absence of trading.

2.2 PLOTS/ VISUALIZATIONS

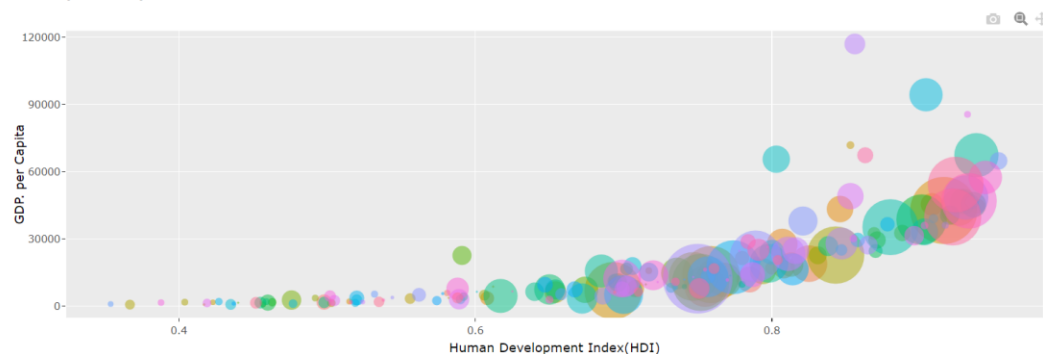
Graph 1:

GDP vs. Hospital beds per Deaths



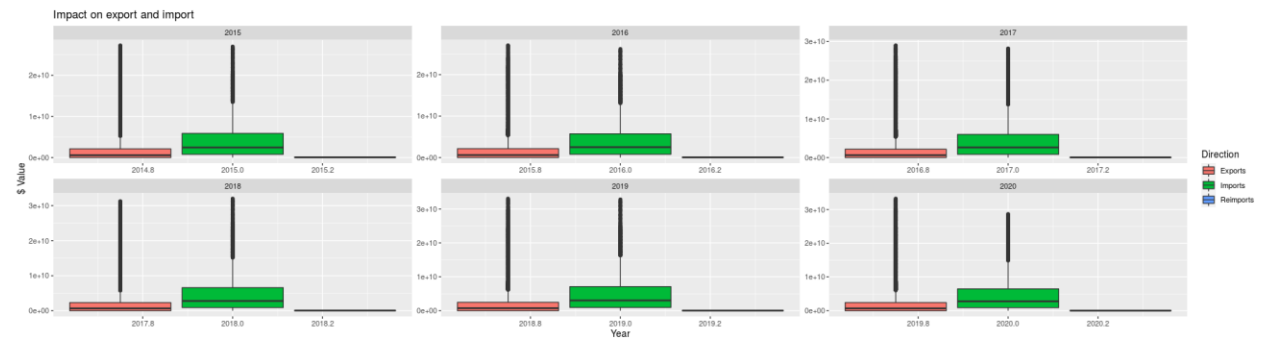
Graph 2:

GDP per capita vs. Human Index



Graph 3:

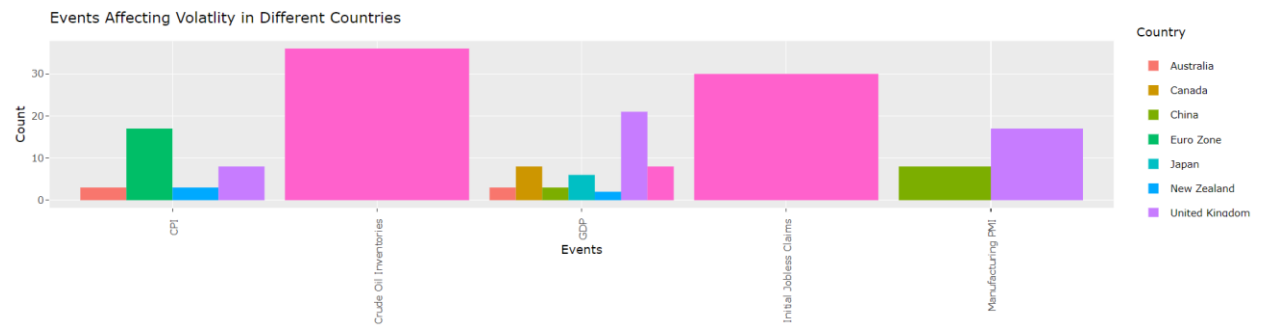
Effect on Import/Export Business



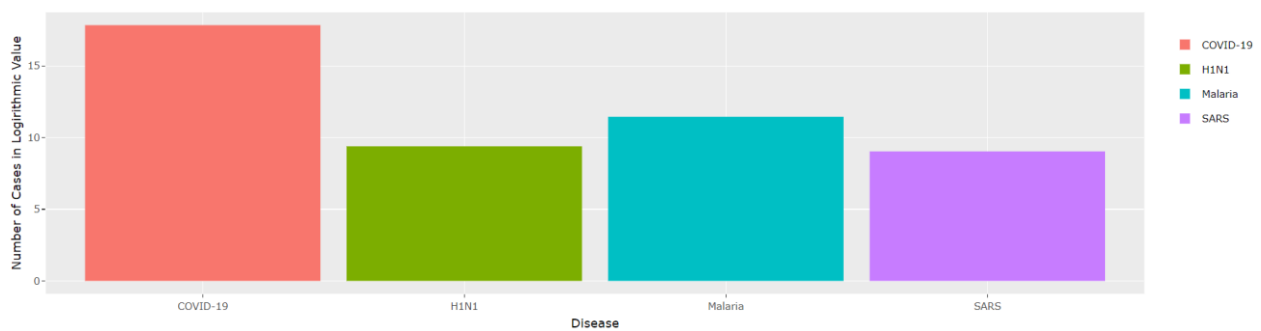
Graph 4: Plot 1:



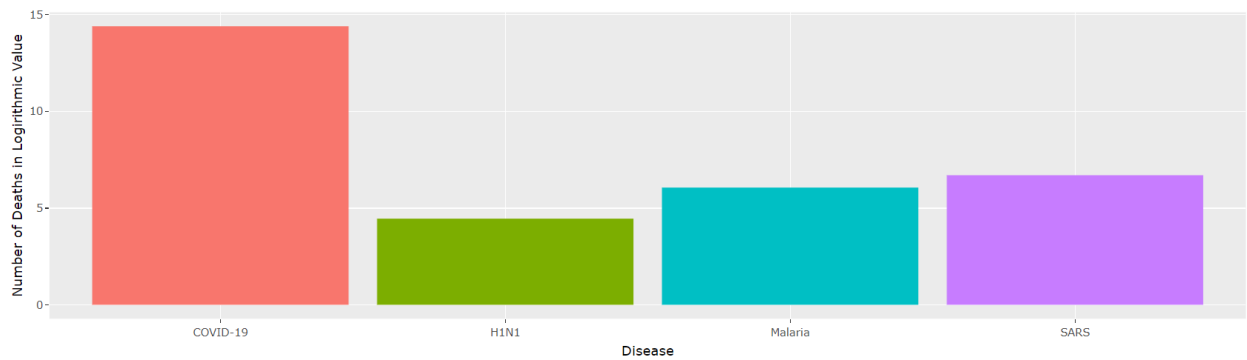
Plot 2:



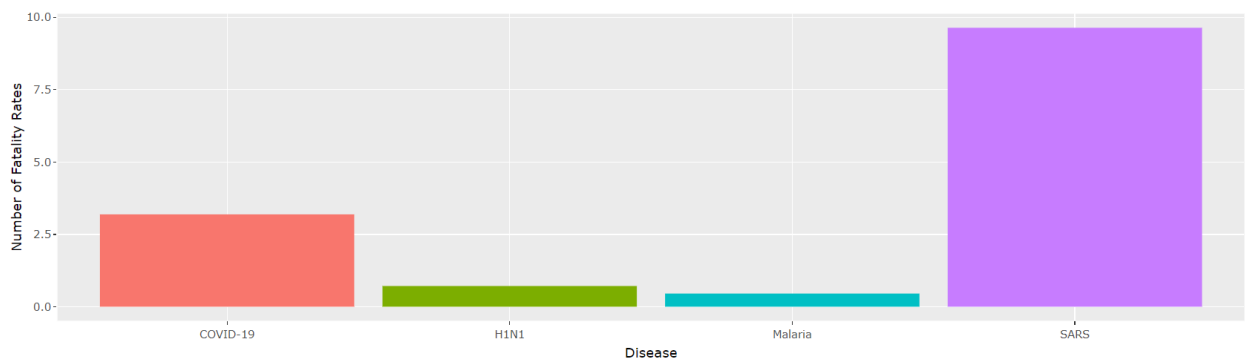
Graph 5: plot 1:



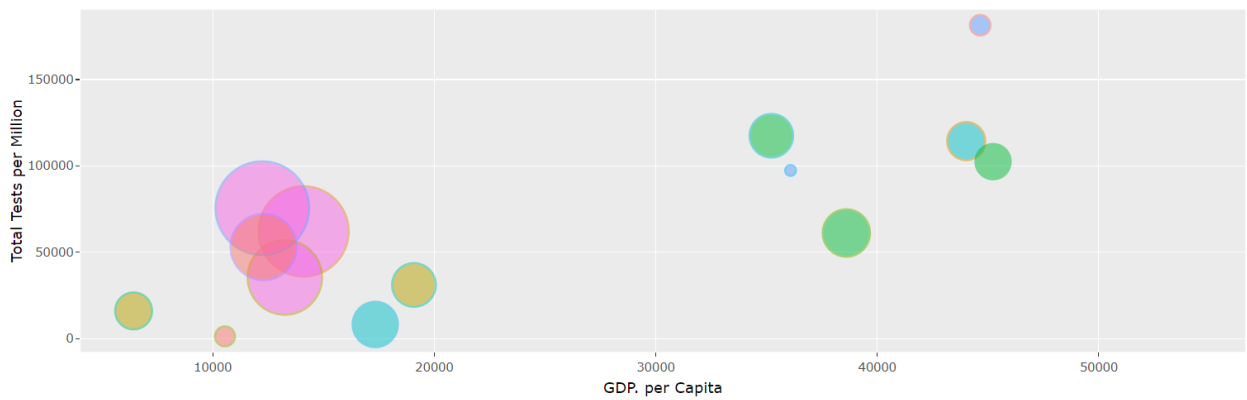
plot 2:



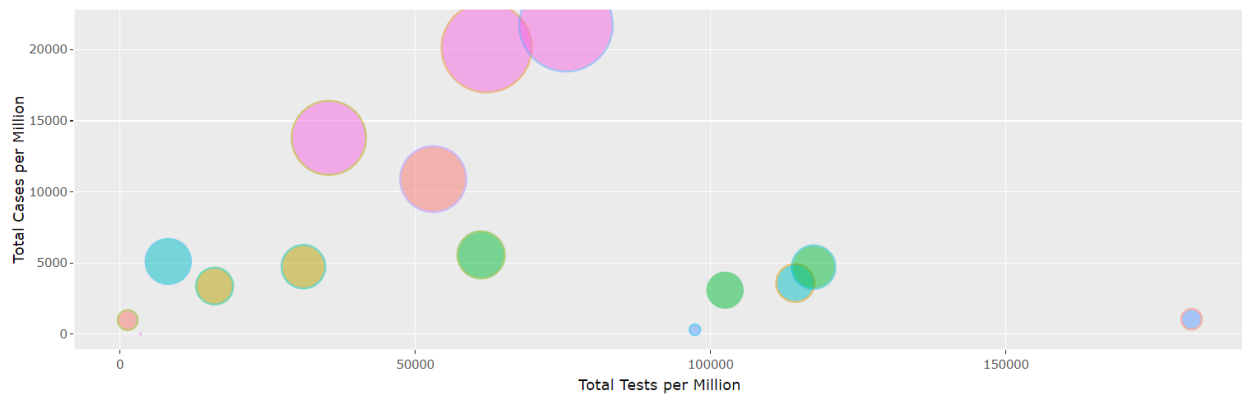
plot 3:



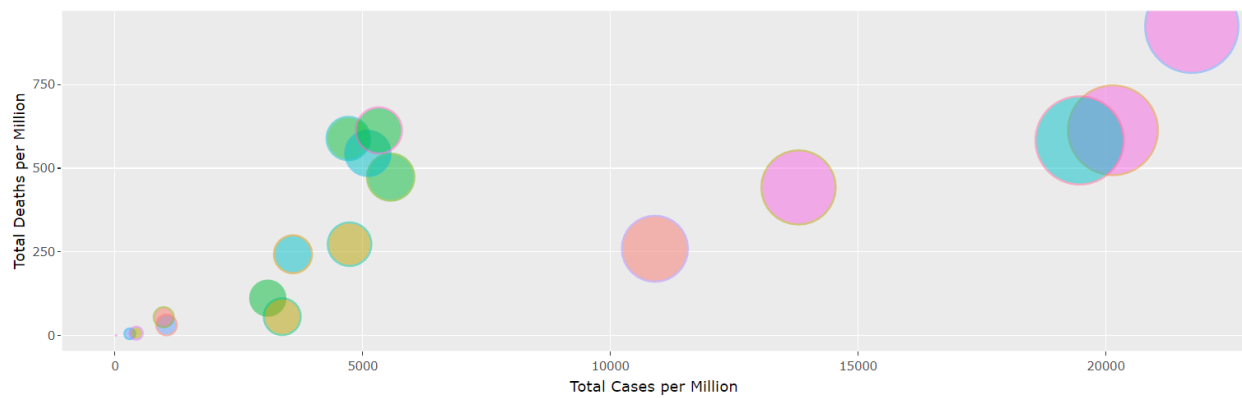
Graph 6: plot 1:

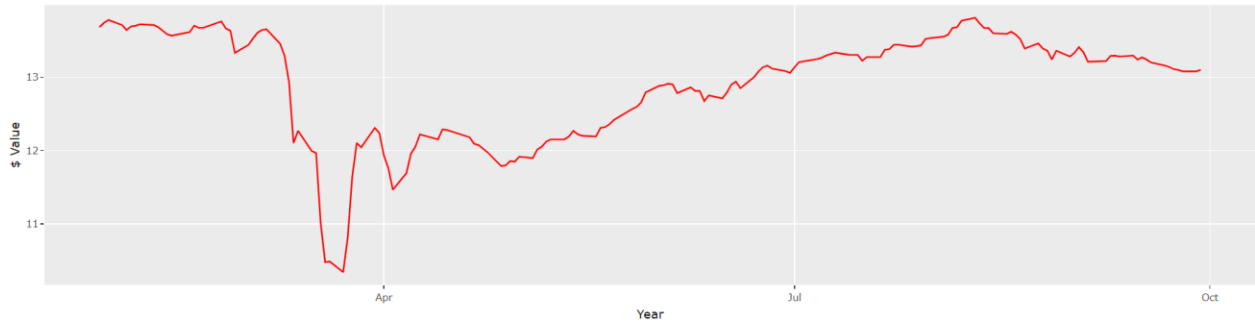


plot 2:

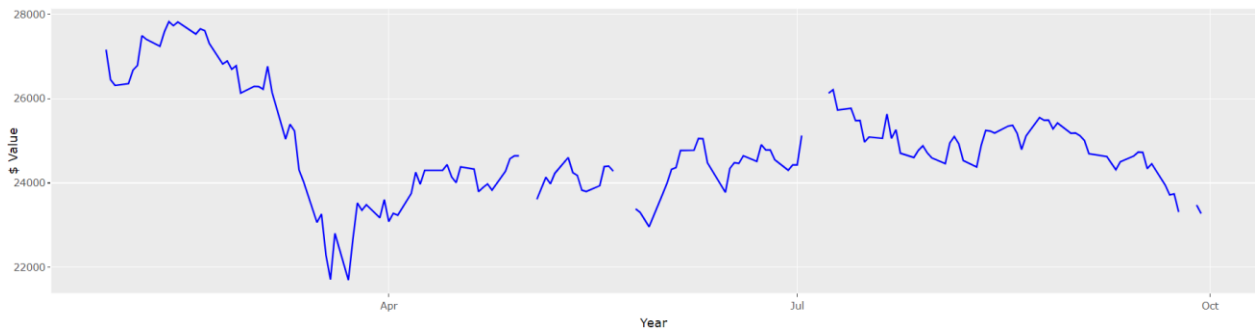


plot 3:





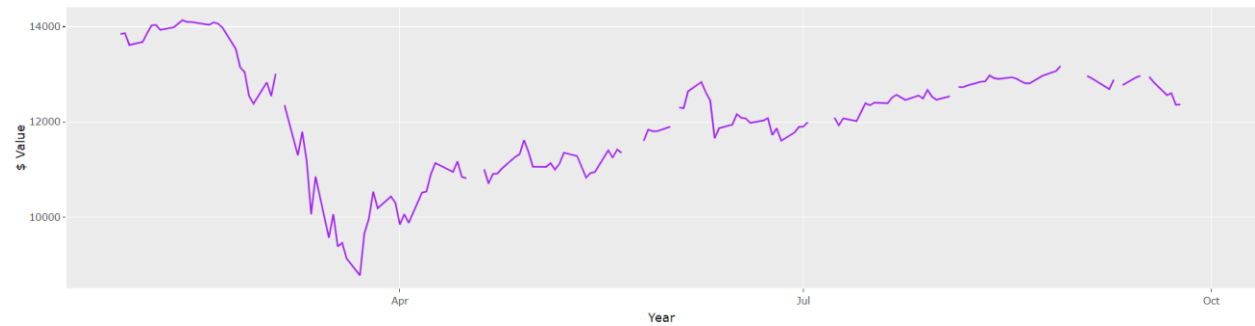
plot 2:



plot 3:



plot 4:



3. METHODOLOGY

3.1 Tools and Resources

The report analyzed covid-19 using the dataset from Kaggle. The data includes the number of positive cases, death cases and testing cases. The other data is related to economic information, including the degree of volatility in countries around the world. The other data is about the impact on trade due to the virus.

R is a useful tool for statistical and data analysis. In this report, we use R package, including tidyr, dplyr, magrittr, ggplot2 and lubricate library, to analyze data and build interactive dashboard using Shiny.

3.2 Analytics Methods

To understand how the covid-19 affects the economics of countries and how do we reduce the negative influence of covid-19, we considered GDP, the number of hospital beds, human develop index and analyze the impact on profit of import and export business, volatility rate and stock market. In addition, we compare covid-19 with other diseases to understand how harmful it is. We use ggplot library to present the relationship between variables through scatter plot, bar chart and line graph. Then we apply shining and shiny dashboard library to create powerfully interactive web applications. Through displaying the visualizations, we got some meaningful insights.

4. IMPACT AND SIGNIFICANCE OF RESULTS

From the analysis, we suggest to the government to put more financial resources on the medical services to help deal with the COVID-19. And also, the government should increase their lockdown level in those high-developed countries. From plot 6 analysis, we strongly suggest that deaths are directly affecting GDP, and we can see that from Peru and Korea status hence they should deal accordingly.

We would suggest international businessman, to do domestic trading or trade in other areas until restrictions are lifted so that their business doesn't get affected and they can continue doing business.

To doctor we suggest that even though the cases are very high, the fatality rate is quite low compared to SARS, so it's not that dangerous, but still its Higher than malaria and H1N1.

Analysis will give insights to economists so that they make relevant recommendations on volatility to improve the efficiency and economy of countries during the pandemic. Also, when corona started all stock market crashed, and it started with recovering as lockdown starting to open up, hence economist should take steps accordingly.

5. PROJECT MANAGEMENT

5.1 TEAM CAPABILITY

Our team was combination of people with different skills: management, visualizations, front end, python, R, report writing, speaking and collaborating, understanding and keeping everyone on same page.

First, everyone's skills were understood and work was divided in the same way. High level planning was carried out and hence tasks were carried out.

5.2 PLANNING

We took Milestones as our targets for the month and did group meeting on WhatsApp. Since we were treating milestones as the target, we sort of did scrums one every week to check on how much work we have done and how much remains. During scrums, everyone takes part and one of us takes the role of scrum master. This we everyone got an equal chance to participate and take the leadership role.

From the early stage, we have used tasks priority as a standard condition to estimate how much time and work should we allocate to a particular task, we worked on tasks with priorities early on and then later proceeded with low priority tasks.

We decided to take covid-19 as our main problem and planned to create a dashboard for the Government to better understand the situation of their country in comparison to other countries and proceed to make more improvements for next pandemic. To create this dashboard, we first tried to find as much data related to Covid-19 as possible, once we found few datasets, we started processing it and making a plot to see if we can find any problems.

For logging, we use spreadsheets throughout the project to write work progress and use google docs for writing documentation. Once we started working codebase, we used MS Teams to share code and its working demo with each other.

Once the project was completed, we worked on creating presentations and videos for demonstrating our dashboard. To create the video, we used Zoom screen recorder to record screen with our voice inputs, and to edit video we used iMovie.

6. CONCLUSION

On 5 October 2020, our team demonstrated a solution in the form of a browser-based interactive working prototype for the subject in our Semester 2 2020 Case Studies in Data Science at RMIT. It aims to provide insight to different business users about the effects of COVID-19 pandemic in different fields and how they can deal with it. We used various datasets for analysis and build our solution. With the help of our analysis and results people could benefit and be able to apply our insights onto their business or field and do better for other people.

7. REFERENCES

- Our World in Data: <https://ourworldindata.org/coronavirus>
- BBC 2020, Coronavirus: What is a second wave and is one coming?
<https://www.bbc.com/news/health-53113785>
- GitHub 2020, Data on COVID-19 (coronavirus) by Our World in Data, viewed 22 Sep,
<<https://github.com/owid/covid-19-data/tree/master/public/data>>
- Kaggle 2020, Economic calendar Invest Forex, Kaggle, viewed 28 Sep,
<<https://www.kaggle.com/devorvant/economic-calendar?select=D2019-20.xls>>
- Kaggle 2020, Effects of COVID19 on Trade, Kaggle, viewed 28 Sep,
< <https://www.kaggle.com/kkhandekar/effects-of-covid19-on-trade?select=Effects-of-COVID19-on-trade-1Feb-12Aug2020.csv>>
- <https://ourworldindata.org/coronavirus-testing>
- World Health Organization 2020, WHO Coronavirus Disease (COVID-19) Dashboard,
<https://covid19.who.int/>
- Healthline:[https://www.healthline.com/health-news/how-deadly-is-the-coronavirus-compared-to-past-outbreaks#20022004-severe-acute-respiratory-syndrome-\(SARS\)](https://www.healthline.com/health-news/how-deadly-is-the-coronavirus-compared-to-past-outbreaks#20022004-severe-acute-respiratory-syndrome-(SARS))
- Our World in Data: <https://ourworldindata.org/blog>
- Our World in Data: <https://ourworldindata.org/coronavirus-testing>
- Shiny from R studio: <http://shiny.rstudio.com/>
- Shiny from R studio: <https://shiny.rstudio.com/tutorial/>
- Shiny from R studio: <https://shiny.rstudio.com/reference/shiny/>