JOURNAL OF DATA SCIENCE
eISSN:2805-5160
Vol.2021:003

Data Analysis of Covid-19 Pandemic in Malaysia and Singapore

[1]Nelson Tang Kwong Kin, [1]Chong Jiet Vun, [1]Baek Dae Hui, [1]Wong Jeng Yang, [1]Dini Handayani

Email: dinioktarina.dwihandayani@taylors.edu.my

[1]School of Computer Science &. Engineering, Taylor's University. Selangor, Malaysia.

Abstract

COVID-19, an infectious disease that was caused by a newly discovered coronavirus. According to the World Health Organization (WHO), it will affect the human lung and airways hard to be moderated. The COVID-19 virus can spread through saliva drops or be released from the nose when an infected person coughs or sneezes. By washing hands or rubbing alcohol-based often and avoiding touching the face, one can prevent infection or slow down the spread rate. Currently, there are no specific vaccines or treatments for COVID-19. Some datasets are generated from COVID-19 epidemic in Malaysia and Singapore and the datasets can be used for analysis to improve understanding of the COVID-19. In this project, six datasets are collected and conducted with data selection, data pre-processing, data transformation, and data mining. To conduct visualization and analysis of the datasets, Microsoft Azure Machine Learning Studio is used during the methodology part while Microsoft Excel and R Studio are used to generate five different graphs based on the datasets. The objective of this project is to compare, investigate and analyse the COVID-19 epidemic and find out the impact of Movement Control Order (MCO) in Malaysia and Circuit Breaker (CB) in Singapore to overcome the COVID-19 epidemic in both countries based on the data visualization and analysis in this project.

Keywords: **Data Analysis, Covid19 Pandemic, Movement Control Order, Circuit Breaker**

1. Introduction

In Malaysia, the first patient was diagnosed with COVID-19 on 24th Jan 2020 (The Star, 2020), who was a Chinese national aged 61 who traveled in Malaysia on 18 Jan 2020 and started having mild fever on 23 Jan 2020. Three tourists from China were diagnosed with COVID-19 on January 25, 2020, in Malaysia on a trip. All entries to the country were installed with thermal scanners while 26 hospitals were identified by the government to carry out investigations and medical treatments to COVID-19 patients on 27th January 2020. The reported cases of COVID-19 have grown quickly and there are a total of 6002 COVID-19 cases which were updated on 30th April 2020. The second cluster of cases is related to a religious gathering in Petaling Jaya in early March.

To minimize the spread of the COVID-19, Malaysia had imposed a two-week period of Movement Control Order (MCO) starting on 18th March 2020. During the MCO, international borders were closed while traveling and public gatherings were prohibited, and non-essential business was shut down. Nonetheless, the number of reported cases surged to exceed 2000 during the month of March and approximately half of the cases were related to the second cluster. Therefore, the MCO was extended to avoid being exposed to the virus. According to Health & Pharmaceuticals, the daily number of reported cases has been

Submission: 16 August 2021
Acceptance: 23 August 2021

decreasing from three-digit numbers to two-digit numbers. Social distancing is one of the main methods to minimize the transmission of the disease.

In Singapore, there were a total of 16,169 confirmed cases updated on 30 April 2020. An article that was regarding this serious situation had been published by TodayOnline. The first imported case of COVID-19 in Singapore was confirmed on 23 January 2020. The number of reported cases had grown critically by late-March and April. Multiple dormitories for foreign workers were considered as COVID-19 clusters and it contributed to surged numbers of new cases in the country. Singapore announced a package of strict precautions, called "circuit breakers," from April 7, 2020 to May 4, 2020 to counter the rising number of reported cases. An extension of this "circuit breaker" has been announced on 21 April 2020 in which the "circuit breaker" will be extended to 1 June 2020.

Each problem of these two countries mentioned above can contribute to different data and results. Malaysia's Movement Control Command (MCO) and Singapore's Circuit Breaker (CB) are considered as measures based on different cultures, conditions and environments of each country. And with these two measures, effective prediction based on data can be modelled to minimize the spread of COVID-19 and improve public awareness. It can also be linked to our project of extracting information from multiple data to help us understand and act on the threat of viruses. In addition, the main problem with this virus is lack of information, but the test was slow, even with a diagnostic kit. As a result, the lack of visibility into virus proliferation made it difficult to respond globally (Svět Lustig Vijay, 2020). As COVID-19 progresses, the effects of national efforts against virus eradication may depend on the ability of governments in the country to accurately measure proliferation and target public health efforts using that information and data.

The data about this pandemic collected from different nations plays a significant role to combat against the threat. The primary objective of this project is to compare the data of COVID-19 in both Malaysia and Singapore in order to interpret the data and extract information from it. Comparing data across different nations can help to learn the progression of the threat in other countries, especially the first infections and where leaders have enforced community restrictions (News Literacy Project, 2020). Robust data collection includes data segmented by ethnicity and race is important to gain thorough understanding of the virus's impact on every community (American Medical Association, 2020). Data comparison is essential to learn working countermeasures against the threat and to respond appropriately to the pandemic (Hasell et al., 2020).

Comparisons of data across countries can be a great help to learn lessons from other countries and to help relevant associations to draw conclusions (News Literacy Project, 2020). In addition, the objects of our main project will be described and predicted as below.

- Number of New Cases (find out effectiveness of MCO(Movement Control Order) and CB(Circuit Breaker) in 2 countries)
- Death (calculate the death rate in 2 countries and the fatality rate of the virus can be expected.)
- Recovered and confirmed (find out the effectiveness of medical treatment within the 2 countries)
- Age with number of deceased (We can predict where we are vulnerable)
- Compare the cases related to clusters and other cases not related to clusters in order to know the role of clusters in spreading the virus.

Submission: 16 August 2021
Acceptance: 23 August 2021

2. Methods

The methods consist of four stages: 1) data selection, 2) data pre-processing, 3) data transformation, and 4) data mining. We had collected the dataset sources through the Internet and create some new datasets based on the following four datasets.

The first dataset was made by the Outbreak.my and it has shown us about the statistics of Malaysia's COVID-19.From the statistics, it had recorded the detail about the infected patient such as case number, age, nationality, status, confirmed date, recovered date and so on. The record begins from the January of 2020 until the May of 2020.

The second dataset was found on the Kaggle and it contains data on COVID-19 confirmed cases in Singapore which was published by the Singapore's Ministry of Health. Based on this dataset, we had created a new dataset about the age and the infected patient whom are related to cluster (COVID-19) in Singapore.

The third dataset is about the data on COVID-19 confirmed cases in Malaysia which was made by Erhan Azrai and posted on Data World. This dataset contains the information about the COVID-19 including the state, hospitals, and patients and so on. He had created this dataset by monitoring the COVID-19 daily status which was updated by Ministry of Health, Malaysia.

The fourth dataset was created by Hui Xiang Chua and it can be found on Data World too. The dataset will be updated daily to improve the accuracy of COVID-19's information in Singapore. This dataset consists of daily confirmed, daily discharged, daily deaths and so on.

We perform data transformation for COVID-19 data in the two countries, Malaysia and Singapore by dividing it based on government restrictions and only selecting important columns since huge data volume were involved in data transformation.

In data mining stage, we train data using AzureML. First, each data of Malaysia's Covid-19 and Singapore's Covid-19 are split into two datasets. Second values in the second dataset can be predicted by a two-class logistic regression model, which is used by the trained first dataset. The accuracy of the predictions is able to be obtained. A two-class logistic regression model is imported, and the data is split. The ratio used to split the data is 70:30. Therefore, 70% of the data is trained in the first dataset and 30% of the data is used in the second dataset in order to obtain the accuracy of the predictions. The outputs of the two-class regression model and split data are connected to the trained model.

A score model is connected by the output of the trained data. The results of the predictions are able to be generated by using the score model. The data can be visualized since the output of the score model is connected to an evaluated model. For our data of the number of total cases in Malaysia, an accuracy of 0.909 is achieved while an accuracy of 0.955 is achieved for our data of the number of total cases in Singapore. The accuracy of the results show that the results provided by the regression model are considered as very accurate to predict the future number of total cases in both Malaysia and Singapore.

Besides, a two-class decision forest is used to compare the accuracy of the model in different algorithms. The result shows the accuracy of 0.909 for our data of the number of total cases in Malaysia and the accuracy of 0.955 for our data of the number of total cases in Singapore. Both results are the same as the results using the two-class logistic regression and this indicates the high accuracy of the model.

Submission: 16 August 2021
Acceptance: 23 August 2021

3. Result and Discussion

This section discusses the results that we obtained after data mining stage using AzureML.

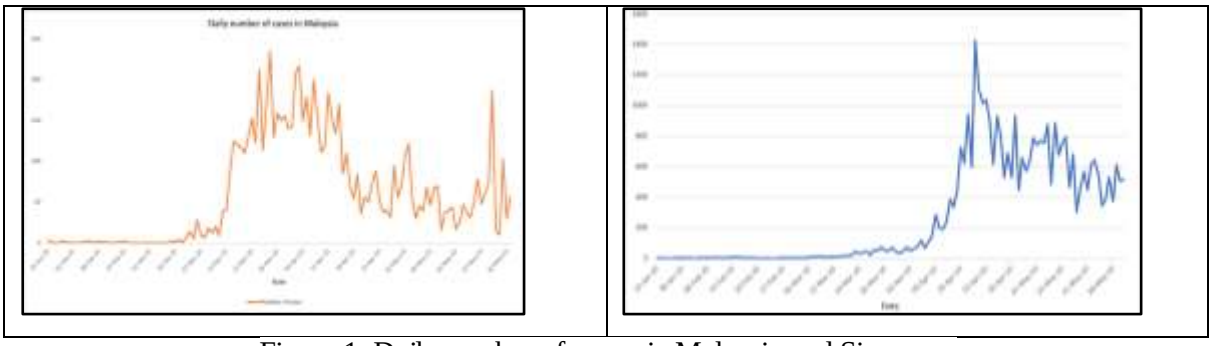

Figure 1. Daily number of cases in Malaysia and Singapore

By investigating the result of graphs in figure 1, we find out the effectiveness of Movement Control Order in Malaysia and Circuit Breaker in Singapore. The first graph shows the daily number of cases in Malaysia from 25 January 2020 until 30 May 2020.The peak had growth continuously for one month which was started from 29 February 2020 to 28 March 2020 and it stopped increasing after the day of 28 March 2020.We had noticed that the most highest peak of cases is on 28 March 2020 which is around 230 confirmed cases. Malaysia had imposed a two week period of Movement Control Order (MCO) starting on 18 March 2020.The peak is still increasing even though there were a two week period of Movement Control Order (MCO), so Malaysia's Prime Minister had announced that first extension of MCO to last until 14 April 2020.The peak is significantly decreases within the extension of MCO, it had shown the effectiveness of Movement Control Order (MCO).

From the second graph in figure1, it shows the daily number of cases in Singapore from 23 January 2020 until 30 May 2020.The peak had growth constantly from 23 January 2020 until 19 March 2020 and it had growth rapidly started from the day of 19 March 2020. From our observations, the highest peak is on 23 April 2020 which the total daily number of cases is more than 1400. Circuit Breaker (CB) had been imposed on 7 April 2020 to 4 May 2020 by Singapore. We had noticed that the number cases are started to decrease after two week duration of Circuit Breaker which is similar to Malaysia's Movement Control Order, this had shown the effectiveness of Circuit Breaker (CB).

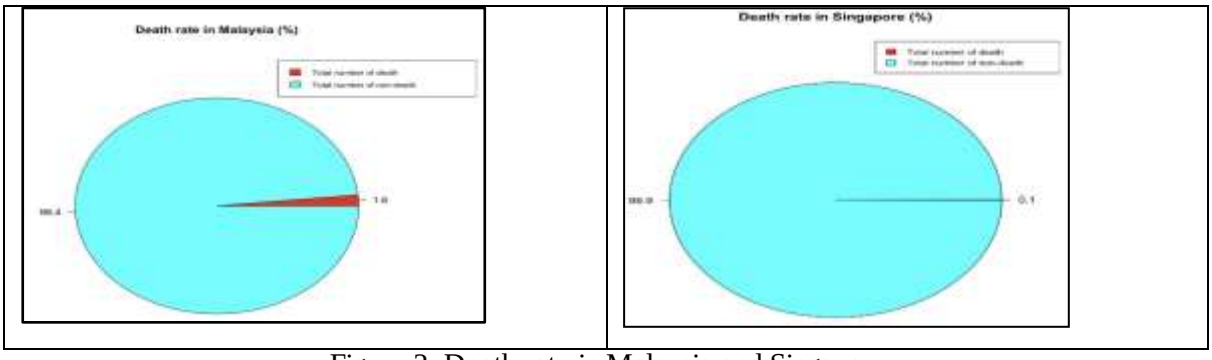

Figure 2. Death rate in Malaysia and Singapore

Submission: 16 August 2021
Acceptance: 23 August 2021

From the figures above, although the death rate in Malaysia is higher than the death rate in Singapore, there is a 1.5% difference only. In these figures, red color is representing the total number of death and blue color is representing the total number of non-death. The death rate in Malaysia is 1.6% while the death rate in Singapore is 0.1%.On the other hand, the non-death rate in Malaysia is 98.4% and the non-death rate in Singapore is 99.9%.To find out the fatality rate of COVID-19,we had compared the result of two pie charts. The death rate of COVID-19 patients in both countries is extremely low.

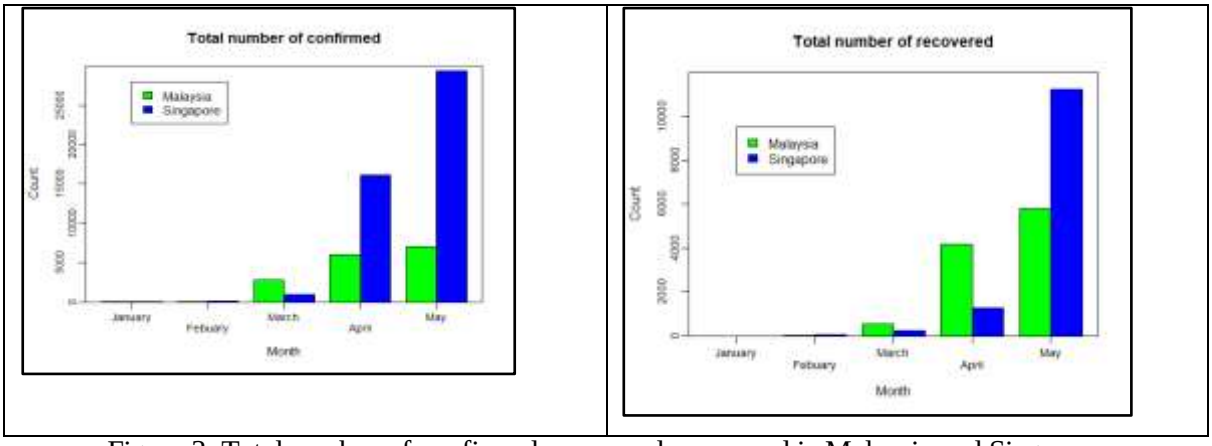

Figure 3. Total number of confirmed cases and recovered in Malaysia and Singapore

The bar-plots above illustrate the result of total number of confirmed cases and total number of recovered cases in both Malaysia and Singapore respectively. The first bar-plot shows the result of the total number of confirmed cases in Malaysia and Singapore from January until May. There were 7059 of total confirmed cases in Malaysia up until May while there were 29409 cases in total in Singapore, which is significantly higher than confirmed cases in Malaysia. The second bar-plot shows the result of the total number of recovered cases in Malaysia and Singapore. Up until May, the number of total recovered cases in Malaysia was 5796 while the total number of recovered cases in Singapore was 11227.

We compare the result of both bar plots to find out the effectiveness of medical treatment in both countries. To find out the effectiveness of medical treatment, we divide the total number of recovered cases by the total number of confirmed cases. As a result of this method, Malaysia's medical effectiveness rate is 5796/7059, which is 0.82. In Singapore, the medical effect is 11227/20409, which is 0.38: 82% in Malaysia and 38% in Singapore. In simple terms, we can predict that Malaysia's medical care effects are more effective than Singapore's. Also, as you can see from the graph, the numbers continue to increase. The reason is that our datasets are aligned based on the cumulative values.

Submission: 16 August 2021
Acceptance: 23 August 2021

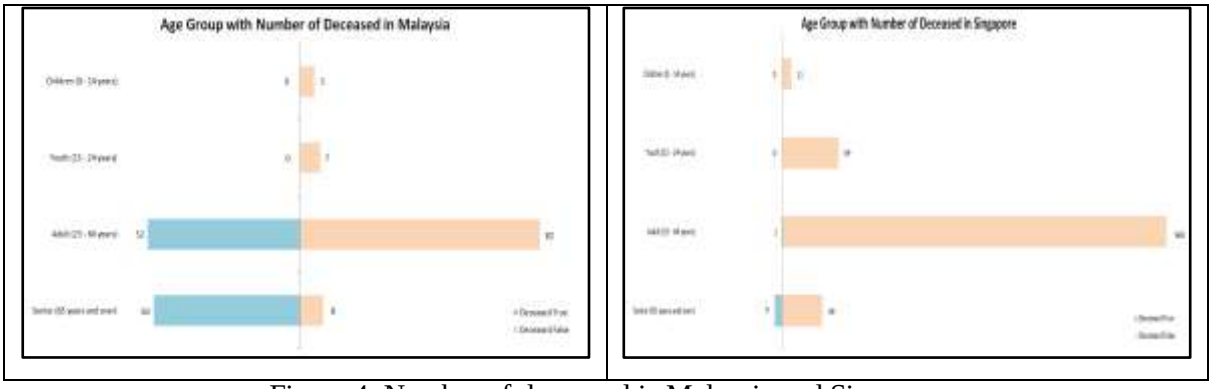

Figure 4. Number of deceased in Malaysia and Singapore

The population pyramids above illustrate the result of the number of total deaths grouped by age groups. The age was categorized into 4 categories in which they are Children (0-14) years old, Youth (15-24) years old, Adult (25-64) years old, and Senior (65 and above). The result of the age groups with the number of deceased in Malaysia shows that the age group "Adult" and "Senior" has the higher total number of deaths while other age groups have 0 deaths. Although Adult group has a slightly higher total number of deaths than Senior, however, it also has a significantly higher number of confirmed cases than Senior's. This means that the senior age group has a significantly higher death rate than Adult. To find out the death rate for a particular age group, we divide the total number of deaths by the total number of confirmed cases of the age group. The death rate of Adult in Malaysia is 52/134 which is 0.39. For the death rate of Senior, the death rate is 50/58 which is 0.86. So, the senior age group in Malaysia has the highest death rate than other age groups and is considered more vulnerable to the disease.

Similar situation happened in Singapore where the senior age group in Singapore also has the highest total number of deaths. Adult age group has a significantly low number of deaths while other age groups have 0 deaths. The death rate in Singapore is calculated using the same method, which is dividing the total number of deaths by the total number of confirmed cases of the particular age group. So the death rate of Adult is 1/469 and the death rate of Senior is 9/58 which is 0.16. The results show that the senior age group also has the highest death rate. In conclusion, based on the results in both Malaysia and Singapore, we can conclude that the disease is most lethal to senior citizens.

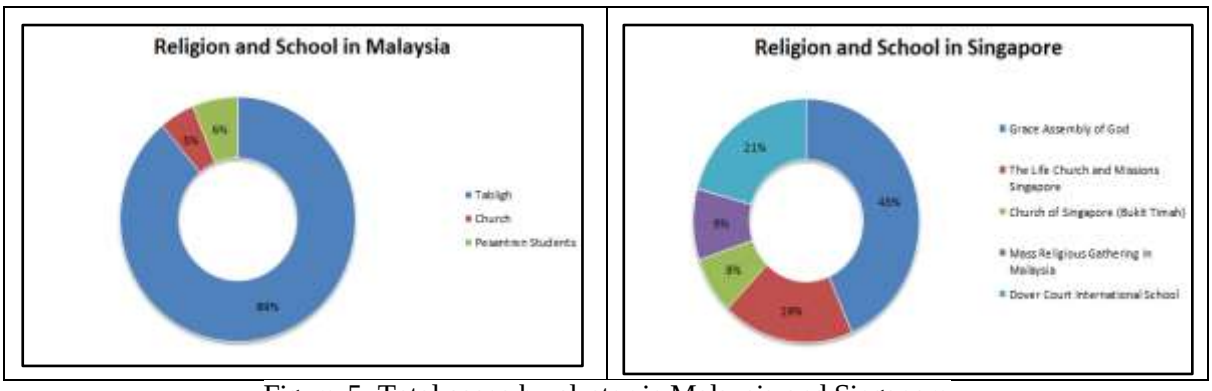

Figure 5. Total cases by cluster in Malaysia and Singapore

Submission: 16 August 2021
Acceptance: 23 August 2021

From the doughnut graphs above, the graphs had shown us about the percentage of total cases by each cluster in Malaysia and Singapore respectively. The clusters in these two countries had been categorized into religious and school. Each color had represented their own cluster in Malaysia and Singapore. For Malaysia, blue color is representing for Tabligh in which there are 89%of total cases were caused by this cluster. Green color is representing for Pesantren Students in which there are 6% of total cases were caused by this cluster. Red color is representing for Church in which there are 5% of total cases were caused by this cluster.

For Singapore, blue color is representing for Grace Assembly of God in which there 43% of total cases were caused by this cluster. Red color is representing for The Life Church and Missions Singapore in which there are 19% of total cases were caused by this cluster. Green color is representing for Church of Singapore (Bukit Timah) in which there are 8% of total cases were caused by this cluster. Purple color is representing for Mass Religious Gathering in Malaysia in which there are 9% of total cases were caused by this cluster. Light blue color is representing for Dover Court International School in which there are 21% of total cases were caused by this cluster.

We had compared these graphs and find out that the percentage of total cases for religious' cluster is higher than the percentage of total cases for school's cluster in both countries. From the comparison, we can conclude that religious' cluster is playing a main role in spreading the virus while school's cluster is playing the secondary role in spreading the virus.

4. Conclusions

As a summary of the results obtained above, the differences of the data of COVID-19 in both Malaysia and Singapore has given us a great deal of useful information. From the results gathered above, we found that the situation is way worse than in Malaysia by comparing the total number of cases and the total number of deaths cases in Singapore exceeded those in Malaysia. To prevent the spread of the disease, the governments of Malaysia enforced an order named Movement Control Order that restricts the activities and contact of people. The government of Singapore has also enforced a similar order called Circuit Breaker.

To find out which order is more effective, we compare the daily number of new cases. We found that Singapore has significantly higher number of daily new cases during the period of Circuit Breaker where the highest number it reached was over 1400 cases. Thus, theoretically speaking, MCO is more effective than Circuit Breaker.

Apart from that, we compare the death rate caused by the disease in both countries to find out how deadly the virus is. The result of death rate in both countries varies due to the huge difference in numbers of two countries. Besides that, we are to find out the effectiveness of medical treatment in both countries. From the results of the recovery rate we collected above, we found that Malaysia has more effective medical treatment than Singapore. Also, we compared the age group that has the highest number of death cases in both countries to find out whether we are vulnerable to the virus. The results of both countries indicate that senior citizens are most vulnerable to the disease and have the highest number of death cases. According to WHO (2020), due to physiological changes that come with aging and potential underlying health conditions, elderly people face significant risk of developing severe illness if they contract the disease. As for the result of clustering of the disease's spread, the figure we presented above shows that in Malaysia, a large portion of the spread came from the tabligh cluster. Similarly, in Singapore, religion gathering is also the biggest cluster of spreading.

Submission: 16 August 2021
Acceptance: 23 August 2021

5. BIBLIOGRAPHY

AMA underscores importance of science, data in COVID-19 fight. (2020, April 07). Retrieved June 9, 2020, from https://www.ama-assn.org/press-center/press-releases/ama-underscores-importance-science-data-covid-19-fight

Beng, H. (2020, April 15). Singapore's COVID-19 cases. Retrieved June 1, 2020, from https://www.kaggle.com/rhodiumbeng/singapores-covid19-cases

Cases by States | COVID-19 | Malaysia Outbreak Monitor ... (n.d.). Retrieved June 3, 2020, from https://www.outbreak.my/states

Covid-19 Malaysia - dataset by erhanazrai. (2020, April 13). Retrieved June 2, 2020, from https://data.world/erhanazrai/httpsdocsgooglecomspreadsheetsd15a43eb68lt7ggk9vavy/workspace/file?filename=Covid19-KKM.xlsx

COVID-19 Singapore - dataset by hxchua. (2020, June 29). Retrieved June 3, 2020, from https://data.world/hxchua/covid-19-singapore/workspace/file?filename=Covid-19+SG.xlsx

Current Issue - July 2020, Volume 14, Issue No. 2. (n.d.). Retrieved June 1, 2020, from https://www.morthoj.org/2020/v14n2/COVID-19-reference.php

Lab, K. (2020, June 27). Covid-19 in Malaysia. Retrieved June 2, 2020, from https://newslab.malaysiakini.com/covid-19/en

Pung, R., Chiew, C., Young, B., Chin, S., Chen, M., Clapham, H., . . . Ang, L. (1970, January 01). Investigation of three clusters of COVID-19 in Singapore: Implications for surveillance and response measures: Semantic Scholar. Retrieved June 2, 2020, from https://www.semanticscholar.org/paper/Investigation-of-three-clusters-of-COVID-19-in-for-Pung-Chiew/fffc87eb6a9ff5567e186ef0968859c2d854d3ea

Ritchie, R. (n.d.). Coronavirus (COVID-19) Testing - Statistics and Research. Retrieved June 6, 2020, from https://ourworldindata.org/covid-testing

UKK DOSM. (n.d.). Retrieved June 29, 2020, from https://ukkdosm.github.io/covid-19

Understanding COVID-19 data: Comparing data across countries. (2020, May 22). Retrieved June 11, 2020, from https://newslit.org/updates/understanding-covid-19-data-countries/

Understanding COVID-19 data: Comparing data across time. (2020, April 14). Retrieved June 11, 2020, from https://newslit.org/updates/understanding-covid-19-data-part-two-comparing-data-across-time/

Vijay, S. (2020, March 30). COVID-19 Testing Trends - Globally & Regionally. Retrieved June 1, 2020, from https://healthpolicy-watch.org/covid-19-testing-trends-globally-regionally/

WHO delivers advice and support for older people during COVID-19. (n.d.). Retrieved June 7, 2020, from https://www.who.int/news-room/feature-stories/detail/who-delivers-advice-and-support-for-older-people-during-covid-19

Wong, J., Leo, Y., & Tan, C. (1970, January 01). COVID-19 in Singapore-Current Experience: Critical Global Issues That Require Attention and Action.: Semantic Scholar. Retrieved June 4, 2020, from https://www.semanticscholar.org/paper/COVID-19-in-Singapore-Current-Experience:-Critical-Wong-Leo/c1d60df95f43b664e7d13b7847b4477d19b471c8

Zp, & Academy, U. (n.d.). COVID-19 Singapore Dashboard. Retrieved June 5, 2020, from https://co.vid19.sg/singapore/

Submission: 16 August 2021
Acceptance: 23 August 2021