COVID-19 Analyser

Team 23

Chow Sow Ying#1, Png Bao Huan\*2, Seah Su Qin#3 , Yong Javen#4 , Gian Meng Rong#5

#Singapore Institute Of Technology  
172 Ang Mo Kio Ave 8, Singapore 567739

[1](mailto:1first.author@first-third.edu)[2102211@sit.singaporetech.edu.sg](mailto:2102211@sit.singaporetech.edu.sg); [2](mailto:2second.author@second.edu)[2102788@sit.singaporetech.edu.sg](mailto:2000000@sit.singaporetech.edu.sg); [3](mailto:3third.author@first-third.edu)[2101842@sit.singaporetech.edu.sg](mailto:2000000@sit.singaporetech.edu.sg); 4[2100992@sit.singaporetech.edu.sg](mailto:2000000@sit.singaporetech.edu.sg); 5[2100657@sit.singaporetech.edu.sg](mailto:2000000@sit.singaporetech.edu.sg)

***Abstract*—This paper describes our group's first experience in Python Programming to develop a program using open-source libraries such as Pandas, and MatLibPlot etc. The project focuses on implementing a program for users to better understand the situation of Covid-19 in Singapore. Our approach used to develop the program was the waterfall methodology. Our major findings would be that there is a lot of data available on the internet but there isn't one platform where users are able to view Covid-19 data trends and statistics.**

***Keywords—*Covid-19, Datasets, Pandas, Plotly, Trends, Statistics**

# Introduction

The objective of this project is to create a program where users can better understand the current situation of Covid-19 through the different datasets loaded into the program. Users are also able to see the changes in data through the statistical graph and have the option to export those data.

## Purpose of Work

* Can users understand the data found in online reports?
* Can users search and filter for specific data within the dataset?
* Can users see statistical information obtained from the dataset?
* Can users export the data they identified as useful to them?

## Identified Problems

* No proper platform to manipulate data provided for Covid-19 on MOH’s website
* Users can see past data, but only in the form of reports and not a visual summary
* Users unable to filter relevant information online

Our project uses a waterfall methodology approach as we are mostly new to python programming. The Waterfall methodology uses a clear structure where it is easy for us to follow and implement. Through this, we can follow a structured approach and plan our time efficiently.

The intended outcome of our project is to create a program where users are able to analyse the data provided by the Singapore Government on Covid-19 and are able to view statistical data of their choice and export relevant information.

# Related Works

## Related Works: Gov.sg

Gov.sg digital channels for covid-19 updates. As stated in [1] through WhatsApp users will be able to receive daily updates on the current Covid-19 situation (Fig .1). It addresses the problem as it filters out data and sends only the important information to customers.

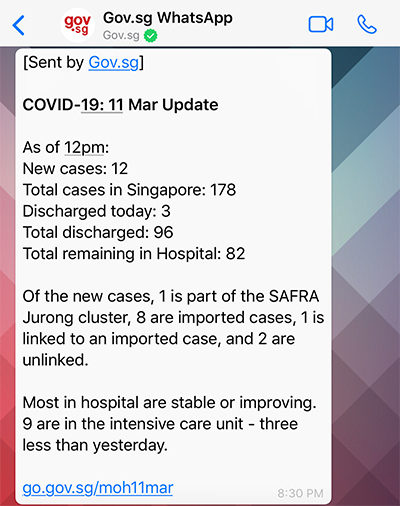


Fig. 1 Sample message of Gov.sg WhatsApp

Our application differs in a way where we are able to view all the different data available and get an overview of how the situation is. Whereas for the gov.sg we can only view daily information of the situation.

## Related Works: TraceTogether

TraceTogether is an application design for contact tracing. Users are able to view the number of people they came in contact with through anonymized proximity information using Bluetooth [2]. They are also able to see if they have any possible exposure with Covid-19 through the application (Fig. 2).

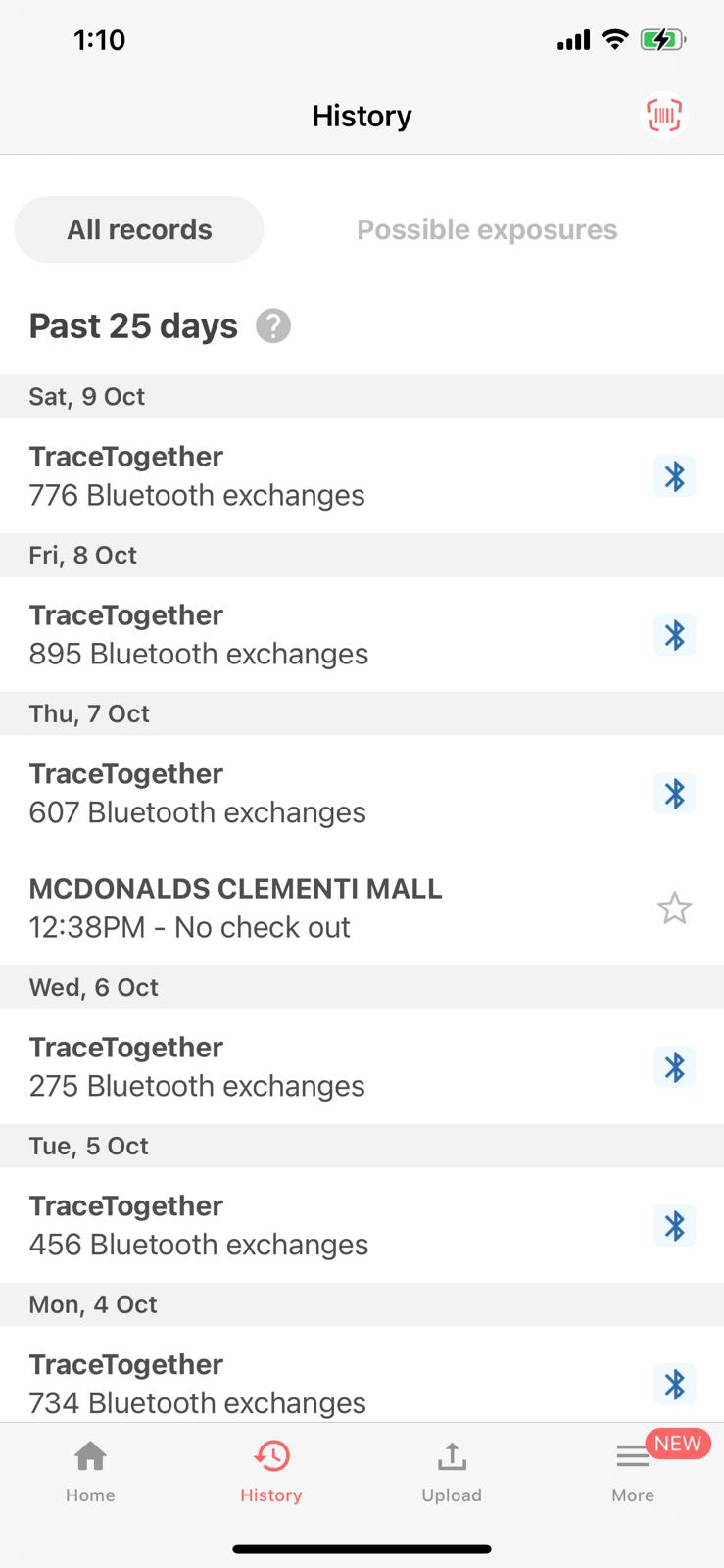


Fig. 2 Image of TraceTogether Application

TraceTogether differs in a way where it's more directed at a certain user providing information relevant to the user while our app focuses on the general covid-19 situation providing the different relevant data gathered in Singapore.

## Related Works: MOH Covid Sitrep

The government has a Situation Report website that also tracks key indexes that are of relevance to the public such as the number of Total Cases and Active Cases currently and they do not show the previous day's figures on the site. Thus, the project’s application would be to allow users to see the trend of these numbers using the help of graphs, and to compare the difference with the data gathered in the past days.

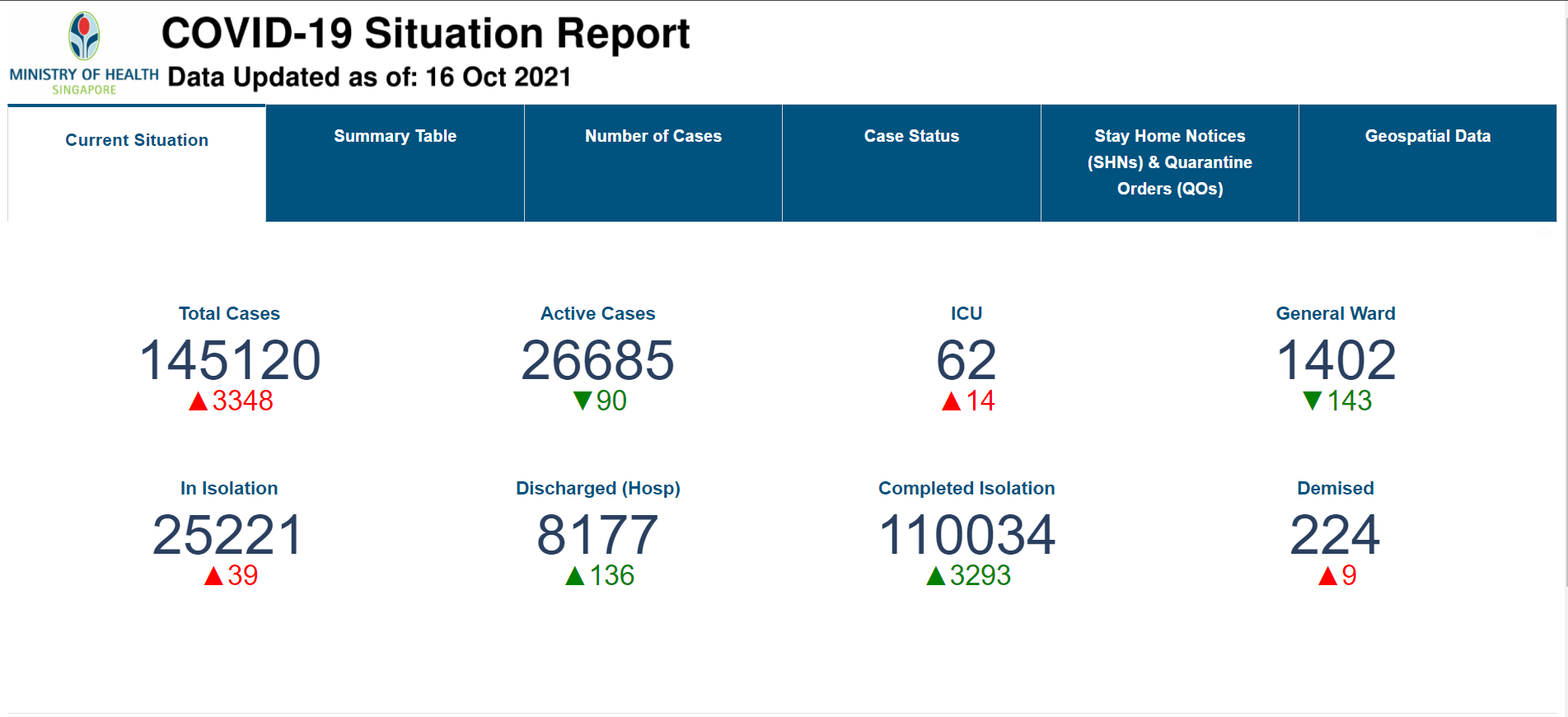


Fig. 3 MOH Covid-19 Situation Report Website

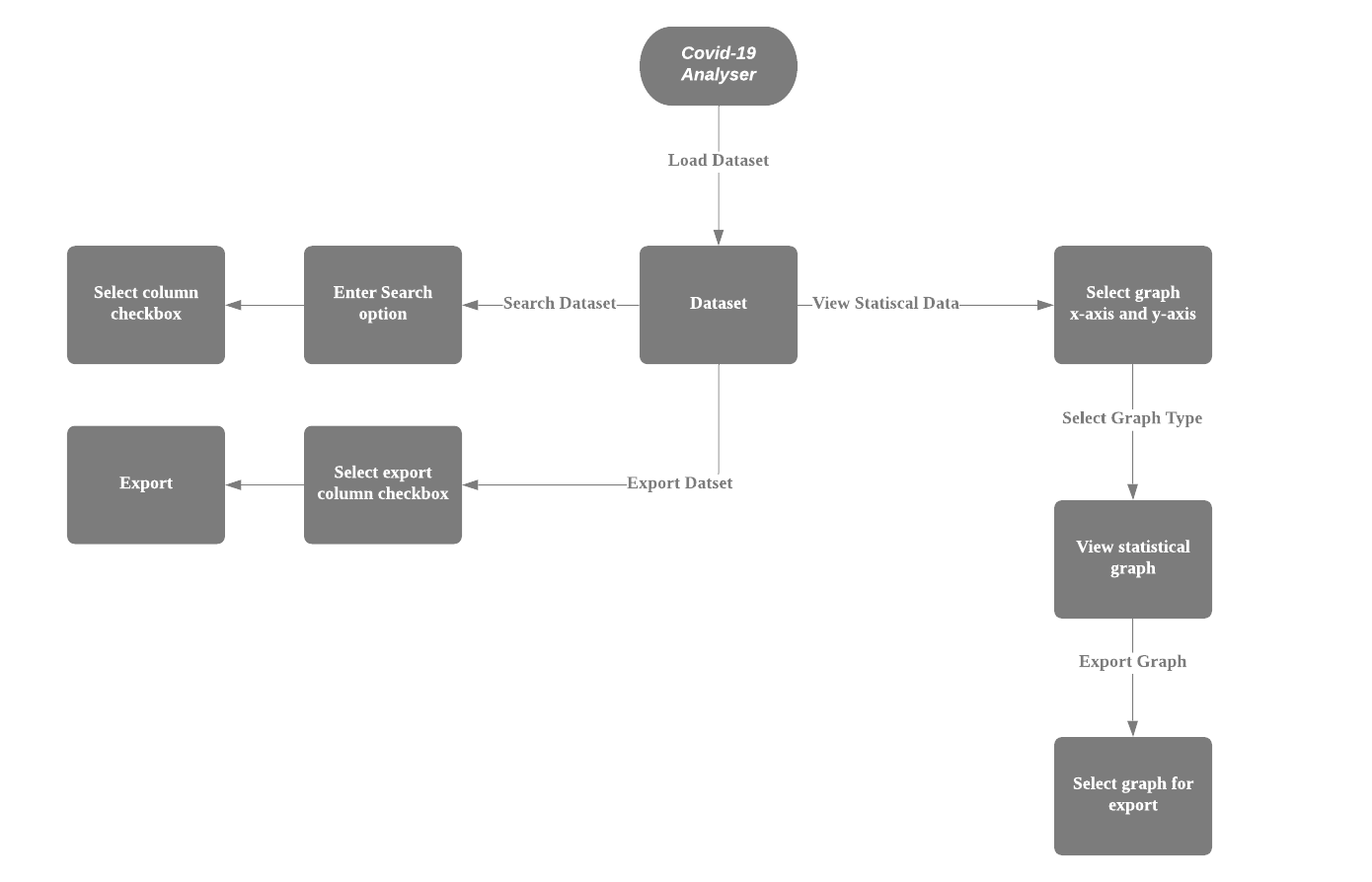
# Methods

## Dataset

The dataset used is Covid-19 Singapore excel sheets created by Hui Xiang Chua.

Covid-19 Singapore dataset is a timed series of data on Covid-19 cases in Singapore. It includes numeric counts on confirmed, discharged, hospitalized, deaths, imported cases.

## System Diagram



|  |  |
| --- | --- |
| **Features Name** | Features Description |
| **Load Function** | Users are able to load datasets that are excel files into the program through the load button. |
| **Search Function** | Users can search and filter through the dataset through the search function. Users first enter what they want to search into the input text box and select the column checkbox of the respective column they want to search in. |
| **View Statistic Graph Function** | Users are able to visualize the data through the different function buttons. Users can check the check box to set the x and y-axis. Then select the graph that he/she would like to view. |
| **Export Function** | Users are able to filter the dataset through the checkbox function. Users will select the column checkbox of the respective column they want to export into a CSV file. |

## Data-Preprocessing

First, raw data must be obtained from the dataset. Then the columns which are essential for our program are identified. The key columns are essential in highlighting new statistics by manipulating the data within different columns to get new sets of columns. The initial columns identified were, **'Date', 'Daily Confirmed', 'Cumulative Confirmed', ‘Daily Deaths’ 'Daily Imported', 'Daily Local transmission', 'Intensive Care Unit (ICU)', 'General Wards MOH report', 'In Isolation MOH report'.**

After identifying key columns that are to be used. The data in the columns need to be cleaned and be removed of ‘dirty’ data that is within it. Such ‘dirty’ data includes blank rows at the end of excel files or CSV files. Instead of being truncated, Python will read every line of the file which includes the blank data and include it as a row of entry into the DataFrame since it is not a null value.

To clean the data, it is not ideal to just drop all entries of rows that contain NA (blank value), as some additional columns of the dataset were only updated once the Government has issued a new set of official situation report, e.g. ‘Intensive Care Unit’ and ‘General Wards’ being used for Covid-19 at government hospitals. Thus, it is valid for these entries to contain ‘NA’ values for dates before their release. To overcome this, the python program removes the ‘NA’ values found in only the ‘Date’ column. As every date is unique, it will not be skipped and thus all the blank data after the current date of extraction will be removed from the DataFrame and marked as the end of the data table using Pandas module.

## Data Analysis

From the initial columns, we can identify 2 subcategories from all the columns, one being a daily Covid-19 situation report and the other being the situation of the current healthcare resources utilized and expended for patients of Covid-19. The following are the expected data analysis to be retrieved from individual DataFrames.

|  |  |
| --- | --- |
| **Table Name** | **New Data Analysis from DataFrame** |
| **Daily Covid-19 DataFrame** | 1. Mortality Rate in of infected for the year 2. Mortality Rate of infected for the month |
| **Gov Healthcare Resources DataFrame** | 1. “Current Active Cases” can be made by combining daily “Intensive Care Unit (ICU)”, “General Wards MOH report” and “In Isolation MOH report” 2. Able to plot a graph to show whether the overall active patients are increasing or decreasing with government efforts. 3. Daily Pie-Chart of the percentage of people that were hospitalized, in ICU or isolated recovery at home. To show the severity of health conditions, the majority of the patients underwent. |

## Results and Insights

**\*Dataset updated as of 16 October 2021**

|  |  |
| --- | --- |
| Total Confirmed Cases | 145200 |
| Total Confirmed Death | 225 |
| Total Confirmed Cases (Past 3 months) | 80139 |
| Total Confirmed Death (Past 3 months) | 188 |

Through analysing the data, we manage to identify the columns needed to calculate the mortality rate which is the following ‘Confirmed Cases’ and ‘Confirmed Death’. The formula used to do the calculations is

Confirmed Death Confirmed Cases

|  |  |
| --- | --- |
| Mortality rate (Total) | 0.15% |
| Mortality rate (Past 3 months) | 0.23% |

As shown the mortality rate for the past 3 months is higher by 0.08 percent as compared to the total mortality rate. Although the vaccination rate has gone up the surge in the mortality rate for the past 3 months could be that there are still a group of seniors who are not getting vaccinated as the number of confirmed deaths are mostly dependent on the seniors. Another reason would be the Delta Variant [3]. The variant which spreads at a rate of 2 to 4 times faster is now affecting people in the community. As this variant has a much higher viral load the vaccines are less effective against it.

# Conclusion

To conclude our project, we designed a program where users can load raw datasets and through our program view the cleaned information and data.

Some possible future upgrades would be:

1. Compatible to add more than 1 data file and plot graphs based on new columns added
2. Data model and prediction of the following day and following month estimate of ‘New Cases’ (not accounting people who recovered) or ‘Active Cases’ (accounting people who recovered)

## Task Allocation: Chow Sow Ying

Chow Sow Ying was tasked to develop the base skeleton of the program which includes loading and viewing of data. She was also tasked to develop and implement the search function of the program to filter out data when searching. She also implemented the GUI for the program.

|  |  |
| --- | --- |
| **Task Name** | **Task Description** |
| **Base Skeleton and GUI** | The program layout and design were created with Tkinter widgets. |
| **Loading and Viewing Data** | Pandas libraries were used to load in the dataset and NumPy array was used to sort the data. To view the data a treeview was created with Tkinter. |
| **Search Function** | Search Function: The search function was implemented with the Tkinter checkbox, treeview and input box. |

## Task Allocation: Gian Meng Rong

Gian Meng Rong was tasked to build upon the base skeleton GUI that Sow Ying has laid upon. Raw data was provided, however before isolating and working on the data, the raw data contained unclean data within them, and the team needed to first clean through before being able to make use of it.

|  |  |
| --- | --- |
| **Task Name** | **Task Description** |
| **Plotting different graphs using DataFrame and stacking to get a better visualisation.** | Data retrieved after cleaning the raw dataset are parsed into the X and Y axis of the graph to create a visualization. |
| **New Analysis Figures based on pre-existing data** | By using the data given in the dataset, figure out useful and new statistics that can be helpful in observing trends between different statistics.  E.g., Mortality Rate in of infected for the year |

## Task Allocation: Png Bao Huan

Png Bao Huan was tasked to build upon the GUI functionality that Sow Ying has implemented. With the cleaned data set provided, the next step is to be able to provide a function to allow the user to view the data in a graphical manner.

Useful data visualization libraries such as matplotlib and seaborn are imported to implement different kinds of visualization. Data visualization gives a clear idea and information delivered. The data is easier to comprehend by humans, compared to displaying a huge chunk of data in front of them. This makes it relatively easy to identify patterns, trends, outliers and others within the given dataset.

|  |  |
| --- | --- |
| **Task Name** | **Task Description** |
| **Create function for graph** | Matplotlib and Seaborn libraries are used to create graphs for visualization purposes. To view the graph, a treeview is created with Tkinter  Different buttons for displaying the various graph |
| **Create checkbox for graph** | Checkbox for the user to select the respective x and y-axis to create a graph |

## Task Allocation: Seah Su Qin

Seah Su Qin was tasked to develop a module export data to allow users to export requested information and store it in a new CSV file. She worked together with Javen.

|  |  |
| --- | --- |
| **Task Name** | **Task Description** |
| **Export function** | Retrieving requested data columns for user by storing the data into a list before exporting it to a new CSV file |

## Task Allocation: Yong Javen

Yong Javen was tasked to develop a module export data that allows users to choose specific information and export them. He worked together with Su Qin.

|  |  |
| --- | --- |
| **Task Name** | **Task Description** |
| **Create checkbox for export** | To create multiple checkboxes for users to select on a specific dataset. |
| **Export function** | Getting the specific data columns that the user requested. Storing the data into a list before exporting it as a CSV file. |

# 

# References

## Text Reference

[1] “Gov.sg launches new channels to keep the public informed about COVID-19,” *Gov.sg*, 02-Apr-2020. [Online]Available: https://www.mci.gov.sg/pressroom/news-and-stories/pressroom/2020/4/gov-sg-launches-new-channels-to-keep-the-public-informed-about-covid-19. [Accessed: 08-Oct-2021].

[2] “How does tracetogether work? – tracetogether faqs,” *TraceTogether*, Apr-2021. [Online]Available:https://support.tracetogether.gov.sg/hc/en-sg/articles/360043543473-How-does-TraceTogether-work-.[Accessed: 08-Oct-2021].

[3] KHALIK, S., 2021. *Why the surge in Covid-19 deaths in S'pore, and what it means for the future*. [Online]The Straits Times. Available at: <https://www.straitstimes.com/singapore/health/why-the-surge-in-covid-19-deaths-and-what-it-means-for-the-future> [Accessed 17 October 2021].

## Image Reference

* https://www.gov.sg/-/media/gov/covid-19/govsg-covid-update.jpg

## Code Reference

* https://stackoverflow.com/a/60949800/4352930
* https://www.youtube.com/watch?v=PgLjwl6Br0k&list=LL&index=1&t=1425s&ab\_channel=RamonWilliams
* https://www.youtube.com/watch?v=WdhNkabUAVU&ab\_channel=softwareManiac
* https://www.semicolonworld.com/question/42826/switch-between-two-frames-in-tkinter#comment-21

## Dataset Reference

Hui Xiang Chua, “Covid-19 in Singapore: Latest Detailed Data.” Kaggle, 2021, doi: 10.34740/KAGGLE/DSV/2699267.

# Appendix

## Program Screenshots

Startup:

Graphical user interface

Description automatically generated

Search:

A screenshot of a computer

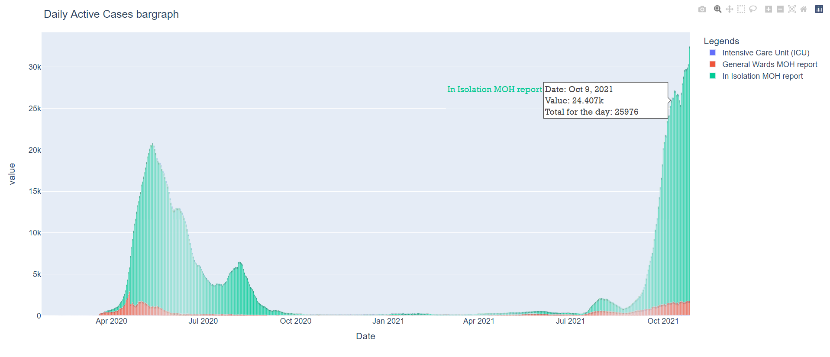
Description automatically generated with medium confidence

View Statistics:

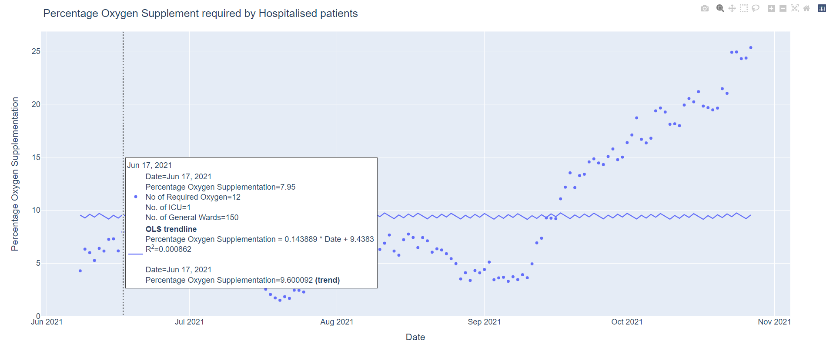
A screenshot of a computer

Description automatically generated with medium confidence

Daily Active Cases Bargraph:



Percentage Required Oxygen Supplementation Scatterplot:



Total ICU based on Age Group PieChart:

Chart, pie chart

Description automatically generated

Total ICU Based on Age Group:

Chart, pie chart

Description automatically generated