UNIVERSITI KEBANGSAAN MALAYSIA

The National University of Malaysia

COURSE

STQD 6014 DATA SCIENCE

SEMESTER 1 2024/2025

TITLE

PROJECT 2

COVID-19: ANALYZING TRENDS AND GOVERNMENT INTERVENTIONS IN MALAYSIA

LECTURER

DR. NURUL AFIQAH BURHANUDDIN

NAME

AZRUL ZULHILMI BIN AHMAD ROSLI (P153478)

DECLARATION

I hereby declare that the work in this assignment is my own except for quotations and summaries which have been duly acknowledged.

20 January 2025

AZRUL ZULHILMI BIN
AHMAD ROSLI
P153478

1.1 INTRODUCTION

The data used is about Covid-19 pandemic in Malaysia. This is to find out about the government response and Covid-19 statistics in Malaysia. Analysing this data will helps the understanding of effectiveness and impact of the public health policies and interventions to curb the virus's spread on the population, as well as the progression of the pandemic within the country. This is can be done as descriptive analysis.

Data source: COVID-19 Open Data — Google Health

1.1.1 Dataset Overview

The datasets used in this analysis were sourced from Google's COVID-19 Open Data repository, which provides comprehensive, publicly available data related to the COVID-19 pandemic. This analysis specifically utilizes dataset COVID-19 Statistics Dataset (Malaysia) which are included together with government respond like policy and others.

a. COVID-19 Statistics Dataset (Malaysia)

The datasets provide aggregated data on Covid-19 cases and vaccination progress in Malaysia. The daily records include:

i. New confirmed case
ii. New deceased
iii. New recovered
iv. New tested
v. New persons fully vaccinated

Then, for government policy, the information on various government policy during the Covid-19 pandemic are selected which are:

i. School closing
ii. Workplace closing
iii. Cancel public events
iv. Restrictions on gatherings
v. Public transport closing

vi. Stay at home requirements
vii. Restrictions on internal movements
viii. International travel control
ix. Income support
x. Debt relief
xi. Public information campaigns
xii. Testing policy
xiii. Contract tracing
xiv. Facial coverings
xv. Vaccinations policy
xvi. Stringency Index

The datasets were used to gain some trends to predict the incoming case trends. For cases and vaccination, only six crucial criteria were chosen in this analysis to find the progression of the pandemic. For government policy, datasets were used evaluate the stringency of government actions to combat the pandemic. This also help the government to decide the crucial policy and be prepare for the next pandemic.

1.1.2 Objectives

i. To examine the trends in daily COVID-19 statistics, including new confirmed cases, recoveries, and deaths, to gain insights into the pandemic's progression over time.

ii. To analyse how government interventions, such as school closures, travel restrictions, and vaccination policies, correlate with changes in COVID-19 statistics.

1.2 DATA CLEANING

The datasets were imported and loaded into the Pandas DataFrames. The attributes were selected during the data preparation. Change some data structure to categorical variable. Handling the missing values and removing the duplicates and ensure the consistent formats.

1.2.1 Importing the Data

Load the raw datasets for Covid-19 Statistic in Malaysia into Pandas DataFrames.

```python
# Introduction
# Data source: https://health.google.com/covid-19/open-data/raw-data?loc=MY
# Data: Aggregated table at specific region (choose Malaysia)

# 1. Import the data
statistic = pd.read_csv('C:/Users/USER/3D Objects/Data Science/Project 2/MY.csv',header=0)
```

Figure 1.1 Importing data

1.2.2 Inspecting the Data

Inspecting the data is to understand the data structure and identify potential issues. This be can perform by using method like *.info ( )*, *.describe( )*,and *.head( )*. Detecting the data structure is crucial especially for categorical variables to ensure the correct data format and avoid error during the plotting. Government response is a dummy variable or in an integer form so changes for data structure are needed. The data structure in government response will be converted to a categorical format.

```python
statistic.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 991 entries, 0 to 990
Columns: 165 entries, location_key to relative_humidity
dtypes: float64(123), int64(23), object(19)
memory usage: 1.2+ MB
```

```python
# Checking attributes along with data types
for column in statistic.columns:
    print(f"Column: {column}, Data Type: {statistic[column].dtype}")
```

```
Column: location_key, Data Type: object
Column: date, Data Type: object
Column: place_id, Data Type: object
Column: wikidata_id, Data Type: object
Column: datacommons_id, Data Type: object
Column: country_code, Data Type: object
Column: country_name, Data Type: object
Column: iso_3166_1_alpha_2, Data Type: object
Column: iso_3166_1_alpha_3, Data Type: object
Column: aggregation_level, Data Type: int64
Column: new_confirmed, Data Type: float64
Column: new_deceased, Data Type: float64
Column: new_recovered, Data Type: float64
Column: new_tested, Data Type: float64
Column: cumulative_confirmed, Data Type: float64
Column: cumulative_deceased, Data Type: float64
Column: cumulative_recovered, Data Type: float64
Column: cumulative_tested, Data Type: float64
```

Figure 1.2 Inspecting the data

1.2.3 Selecting the Attributes

Selecting attributes is one of the important steps to produce a valuable insight from the raw data. There are 165 variables at Covid-19 statistic in Malaysia datasets so only some variables from Covid-19 statistic in Malaysia are selected to avoid redundancy in the information.

```
# 2. Choose suitable attributes only
statistics = statistic[['date', 'new_confirmed', 'new_deceased', 'new_recovered',
                        'new_tested', 'new_vaccine_doses_administered',
                        'new_persons_fully_vaccinated','school_closing','workplace_closing',
                        'cancel_public_events','restrictions_on_gatherings',
                        'public_transport_closing', 'stay_at_home_requirements',
                        'restrictions_on_internal_movement', 'international_travel_controls',
                        'income_support','debt_relief', 'public_information_campaigns',
                        'testing_policy','contact_tracing', 'facial_coverings',
                        'vaccination_policy', 'stringency_index']]
```

Figure 1.3 Select suitable attributes

1.2.4 Handling Missing Values

Handling missing values is one of the important steps in data cleaning. It might cause error or unreliable insight during the analysis. There are several imputations can be used such as remove the missing values or replace it with mean, median, or others value and with placeholder like “unknown”. In this case, remove the missing values and replacing missing numerical values was used.

```python
# Find columns with NA values
na_columns = statistics.columns[statistics.isna().any()].tolist()

# Print the columns
print("Columns with NA values:", na_columns)

# Optional: Count the number of NA values in each column
na_counts = statistics.isna().sum()
print(na_counts[na_counts > 0])  # Only show columns with NA values
```

```
Columns with NA values: ['new_confirmed', 'new_deceased', 'new_recovered', 'new_tested', 'new_vaccine_do
ses_administered', 'new_persons_fully_vaccinated', 'school_closing', 'workplace_closing', 'cancel_public
_events', 'restrictions_on_gatherings', 'public_transport_closing', 'stay_at_home_requirements', 'restri
ctions_on_internal_movement', 'international_travel_controls', 'income_support', 'debt_relief', 'public_
information_campaigns', 'testing_policy', 'contact_tracing', 'facial_coverings', 'vaccination_policy',
'stringency_index']
new_confirmed                        2
new_deceased                         2
new_recovered                       26
new_tested                          24
new_vaccine_doses_administered     423
new_persons_fully_vaccinated       423
school_closing                      54
workplace_closing                   54
cancel_public_events                54
restrictions_on_gatherings          54
public_transport_closing            54
stay_at_home_requirements           54
restrictions_on_internal_movement   54
international_travel_controls       54
income_support                      54
debt_relief                         54
public_information_campaigns        54
testing_policy                      54
contact_tracing                     54
facial_coverings                    54
vaccination_policy                  54
stringency_index                    54
dtype: int64
```

Figure 1.4 Missing values in the datasets

First, remove the missing values for government policy because on 17th September 2022 there are still some covid policy are still applicable. That was hard to find the accurate information due to frequent changing of information due to politic issues and others. Then, removing the missing values for government policy is a good choice. Information from missing value was used to check where is the starting point of missing value for government response. I can use *.tail(55)* and remove the entire rows.

```
statistics.tail(55)
```

e_doses_administered	new_persons_fully_vaccinated	school_closing	workplace_closing	cancel_public_events	restric
15114.0	3309.0	1.0	1.0	1.0	
16880.0	4599.0	NaN	NaN	NaN	
14263.0	2517.0	NaN	NaN	NaN	
17052.0	4045.0	NaN	NaN	NaN	
19911.0	6524.0	NaN	NaN	NaN	
11063.0	3458.0	NaN	NaN	NaN	

```
# 3. Ensure data consistency
statistics = statistics.drop(index=range(937, 991))  # Drop rows from index 937 to 990
```

Figure 1.5 Remove the missing values based on government policy

Second, replace the missing value with 0 is used to indicate the events or infection is not started yet. Then, remove the event which is not started yet.

```
statistics = statistics.drop(index=range(0, 23))
```

```
# Replace NA with zero values: indicates even not started yet
statistics.fillna(0, inplace=True)
```

Figure 1.6 Replace missing values with 0 values

1.2.5 Removing Duplicates

Removing duplicates was done to avoid the redundancy of the data like have the same date twice. If this happen, the insight gain is not well explained the events.

```
# Check for duplicates
statistics_duplicates = statistics.duplicated().sum()
```

Figure 1.7 Removing duplicates

1.2.6 Fixing Structural Issues

Dates are often stored in inconsistent formats because it is essential for sorting, filtering, and time-series analysis.

```
# 6. Factorization
# A. Factor the date
statistics['date'] = pd.to_datetime(statistics['date'])
```

Figure 1.8 Date formatting

Next, convert or categorize government policy variables into descriptive categorical values. This step ensures that these attributes are not misinterpreted as continuous variables during analysis. Additionally, renaming these variables with clear and meaningful labels improves readability, making it easier for non-technical users to understand the data visualizations without relying on binary or numerical codes.

1.3 ANALYSIS AND INSIGHT

Analysis and insight are key objectives in data science to uncover valuable patterns and trends. This is to drive informed decision-making, enabling organizations to optimize strategies and gain a competitive edge. There are 914 days was analyse from the Covid-19 statistics in Malaysia.

To provide a clearer overview, two new columns or attributes were added: positive test percentage and case fatality percentage. The positive test percentage was calculated by dividing the number of new confirmed cases by the number of new tests conducted, while the case fatality percentage was calculated by dividing the number of new deaths by the number of new confirmed cases. Data cleaning was carried out by replacing missing values with zeros to address cases where 0/0 resulted in NaN values.

First, government policy is plotted based on the date to see the overview government interventions influence the Covid-19 cases in Malaysia. Figure 1.9, Figure 1.10, Figure 1.11, and Figure 1.12 shows, the progression of government interventions during a specific period which is from 2020 to 2022.

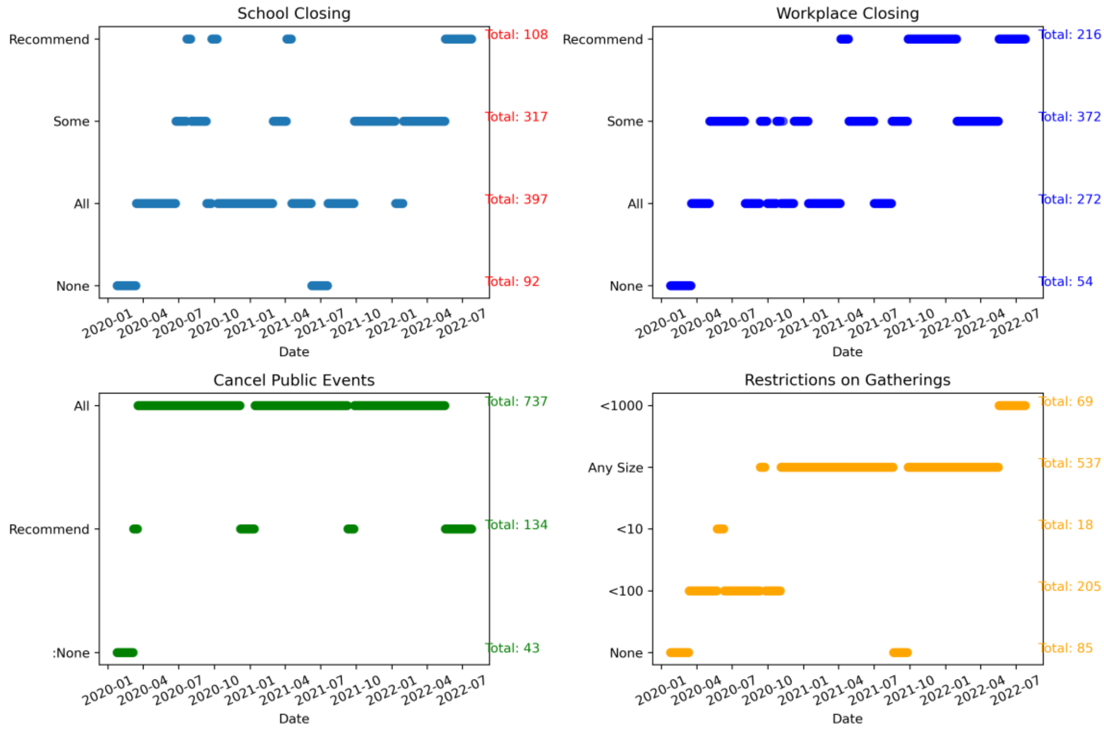

Figure 1.9 School closing, workplace closing, cancel public events and restrictions on gatherings policy over times

From Figure 1.9, for school closing, at first two months of 2020 there are not any government policy regarding school closures due to lower rate of infections. However, as awareness increased, the implementation of school closures was mandated during periods of higher transmission. During some periods when infections gradually slowed, schools were reopened for certain students, such as postgraduate students, to continue their research. Similarly, workplace closures followed a pattern similar to school closures, but not all workplaces were affected. Only certain industries were permitted to operate, while others transitioned to remote work. After the first vaccination program, most companies were recommended to have employees work from home, and businesses gradually began to reopen physically as the situation improved.

From 2020 to 2022, most public events were cancelled around 737 days, reflecting a proactive approach to limiting large gatherings. As for restrictions on gatherings, the policy initially allowed any size of gatherings, but it gradually tightened, first to gatherings of fewer than 1,000 people, then to fewer than 100. During periods of stricter measures, even more stringent restrictions were applied, with gatherings limited to fewer than 10 people.

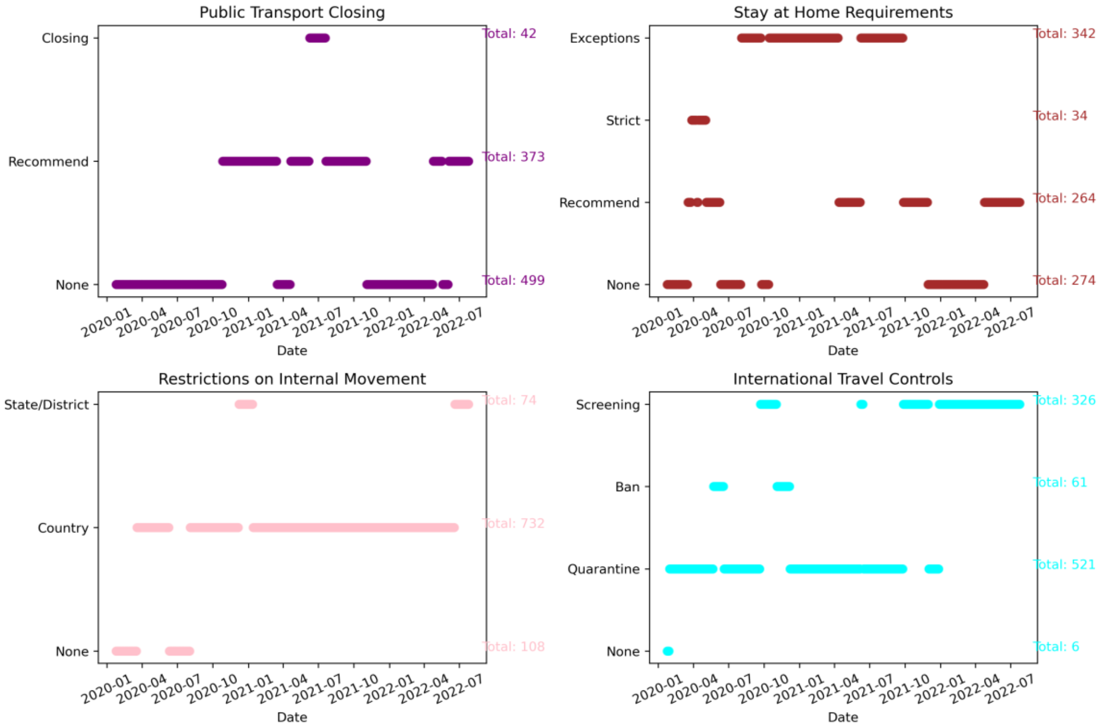

Figure 1.10 Public transport closing, stay at home requirements, restrictions on internal movements and international travel controls policy over times

Figure 1.10 highlights four key policy areas which are public transport closures, stay-at-home requirements, restrictions on internal movement, and international travel controls. Public transport closures were the dominant policy for most of the observed period, with recommendations and exceptions closures implemented during periods of higher transmission. Full closures were only enforced for shorter durations, likely during stricter lockdowns. Stay-at-home requirements, including recommendations and exceptions, were implemented for extended periods, influenced by varying working conditions. Stricter stay-at-home mandates were enforced during times of higher transmission to limit movement and reduce infection rates.

Restrictions on internal movements reveal that there were only 74 days during which the government prohibited crossing state borders, while crossing the country's borders was restricted for 732 days. For international travel controls, during the early phase of COVID-19, all traveller entering Malaysia were required to undergo quarantine to mitigate the spread of infections. After the vaccination rollout, screening measures transitioned to primarily swab tests for a significant period. During stricter lockdowns, international travel bans were enforced, completely restricting cross-border movement.

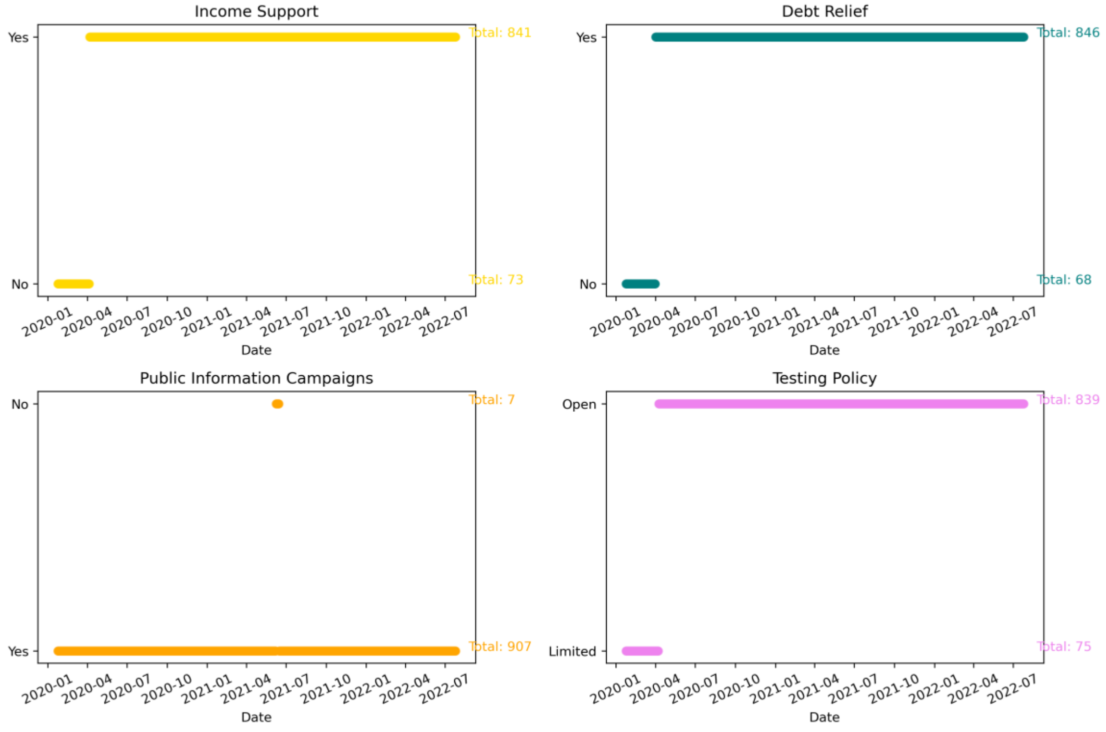

Figure 1.11 Income support, debt relief, public information campaigns and testing policy over times

From Figure 1.11, most of the time, the government provided various initiatives to the public. These initiatives were categorized based on their target audience, such as companies or individuals. Similarly, debt relief measures were introduced, including increased subsidies and discounted bills, such as internet services. Furthermore, the government frequently conducted public information campaigns to raise awareness among citizens. Testing policies became a routine practice, encouraging individuals to get tested if they suspected exposure to the COVID-19 virus. To meet the rising demand, testing kits were mass-produced to manage the growing number of infections effectively.

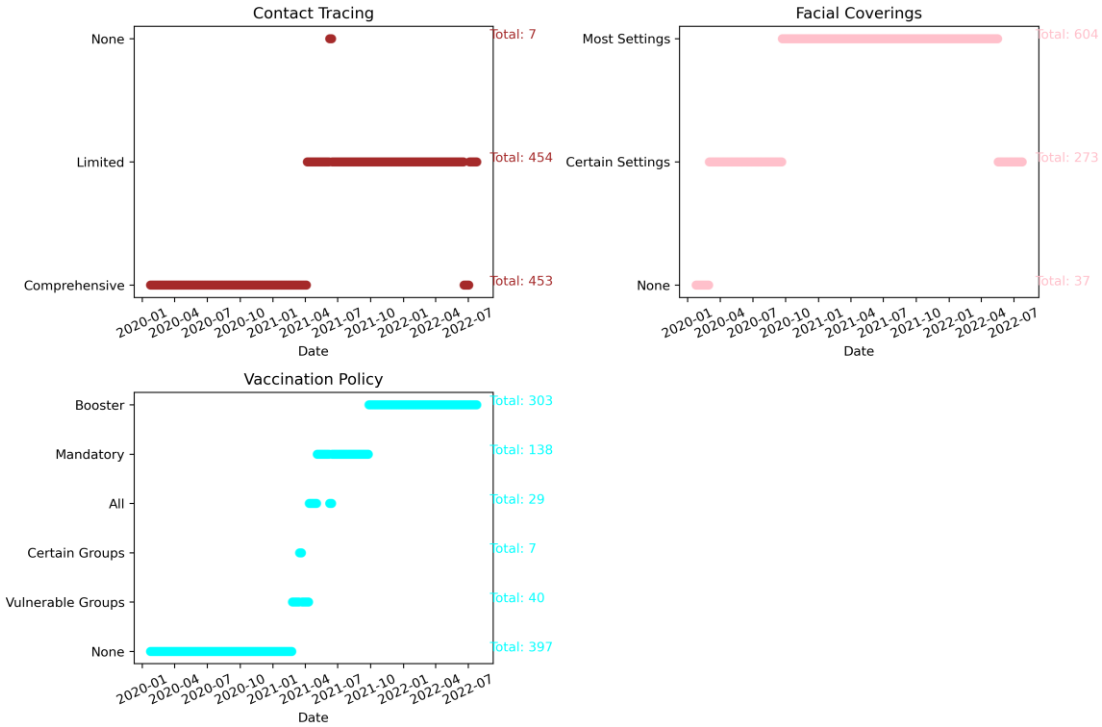

Figure 1.12 Contact tracing, facial coverings, and vaccination policy policy over times

From Figure 1.12, during the early stages of Covid-19, comprehensive contact tracing was conducted for approximately 453 days to curb the spread of infections. Following this, limited contact tracing, focusing only on specific clusters, was carried out for 454 days. In terms of facial coverings, initial policies required their use only in specific settings, such as healthcare facilities, reflecting targeted measures. Over time, most policies mandated the public to wear face masks when going outside, emphasizing widespread compliance.

Vaccination policies were introduced later due to the time needed for vaccine development. Around 397 days after the onset of Covid-19, citizens began receiving their first dose of the vaccine. However, distribution was based on specific criteria, prioritizing certain groups. Initially, certain group such as healthcare workers and vulnerable populations, were prioritized for vaccination. Subsequently, the vaccination program expanded to include all individuals. At one point, mandatory vaccination policies were implemented, reflecting a stricter approach to enhance vaccination rates. Booster policies were also introduced, encouraging or mandating booster shots to sustain immunity levels over time.

Secondly, scatter plot between date with new confirmed case, new deceased, new recovered and tested are done at Figure 1.13. This is to visualize the trends over time of several key Covid-19 related metrics. All the attributes are representing the daily number of newly confirmed COVID-19 cases, number of people recovering from COVID-19, number of new tests conducted, and number of deaths due to COVID-19. This is done to helps track the spread of the virus over time and identifies how quickly new infections are being reported. Then, it provides insight into how well people are recovering and how healthcare systems are managing recoveries. Testing plays a crucial role in detecting new infections and understanding the overall situation. It shows the severity of the pandemic and provides critical data for healthcare and policy planning.

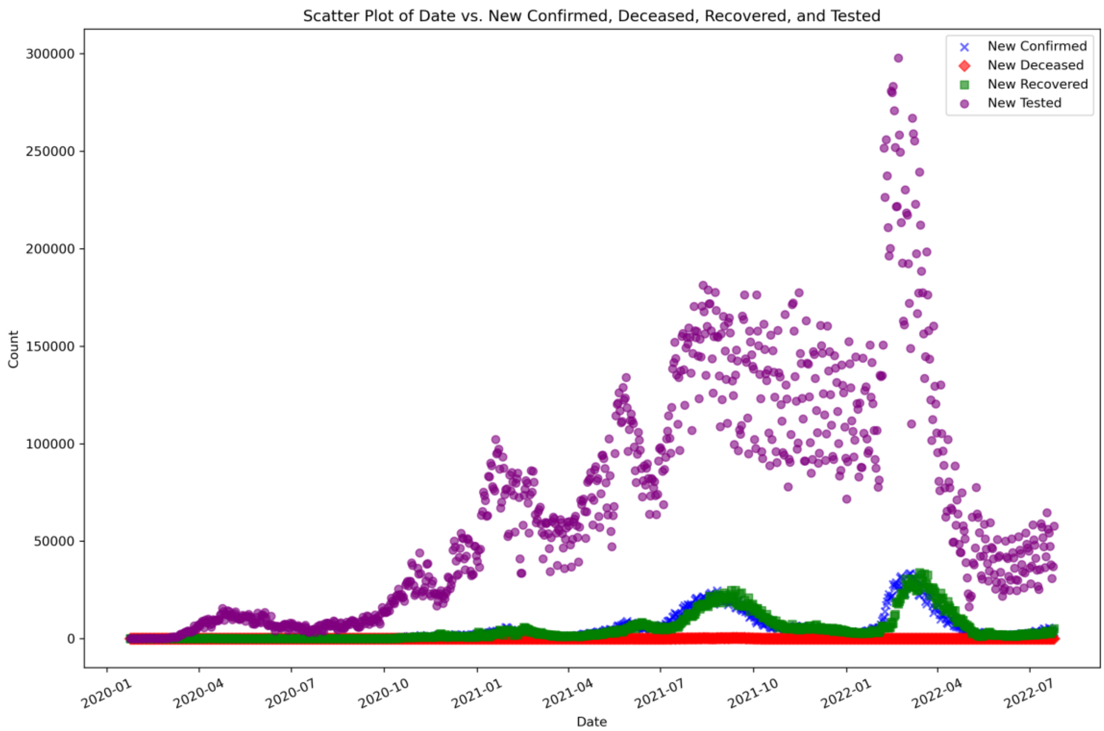

Figure 1.13 Scatter Plot of Date vs. New Confirmed, Deceased, Recovered, and Tested

From the plotting, it is observed that higher numbers of new tests for COVID-19 generally lead to an increase in new confirmed cases, which is followed by a rise in new recoveries approximately a week later. The higher peaks in new confirmed cases indicate periods of increased transmission. The values for new confirmed cases and new recoveries are often similar but occur on different dates, with recoveries lagging behind infections.

The rise in the number of tests likely contributed to the identification of more cases. However, this trend is also influenced by factors such as expanded testing capacity and changes in testing strategies. Two major spikes in infection rates are evident: one from July to October 2021 and another from February to May 2022. These surges align with two significant events.

The first peak, from July to October 2021, corresponds to the emergence of the Delta variant and the creation of 42 new COVID-19 clusters, 23 of which were linked to workplaces, as reported by FMT News (19 August 2021). The second peak, from February to May 2022, aligns with the Omicron variant, known for its high transmissibility. These events highlight the impact of new variants and workplace-related clusters on infection rates during these periods

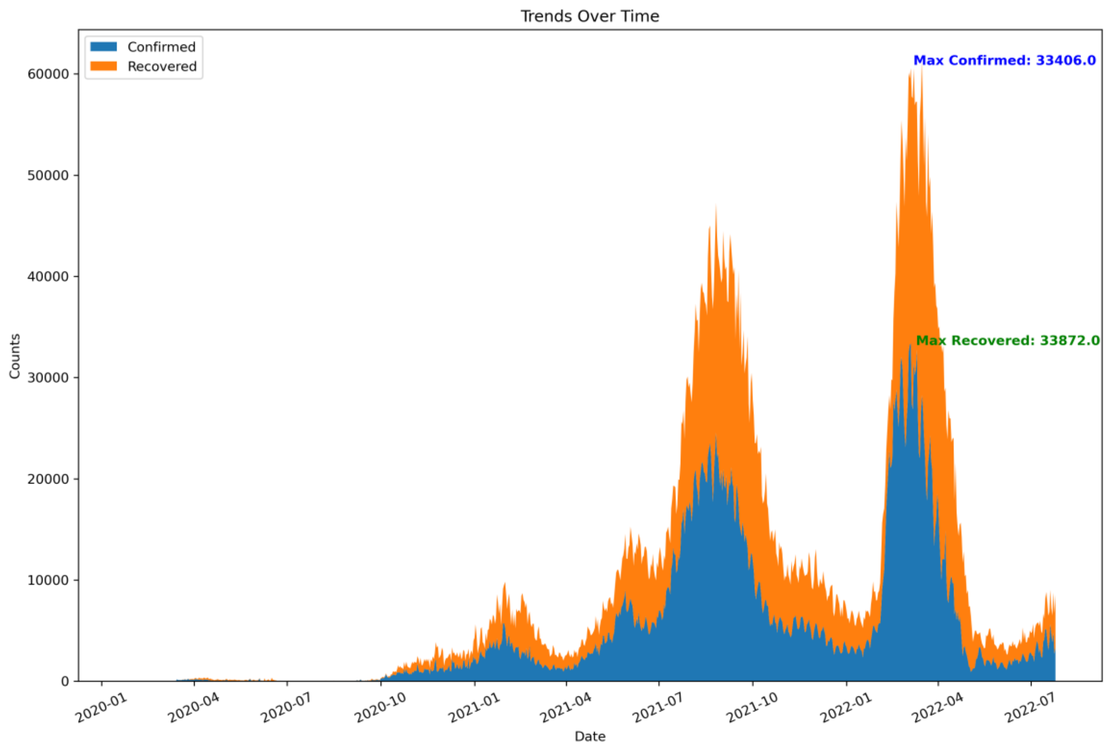

Figure 1.14 Trends over time between confirmed and recovered cases.

Figure 1.14 illustrates the trends over time for confirmed and recovered cases. The blue line, representing confirmed cases, shows an upward trend with peaks driven by events like the emergence of Delta and Omicron variants. The orange line, for recovered cases, follows a similar pattern, with peaks and valleys aligned to infection surges, indicating most infected individuals recover. The graph highlights multiple waves of infections, each with a sharp rise in cases followed by a gradual decline, mirroring global COVID-19 trends. The highest figures are seen in February to March 2022, with 33,406 confirmed cases and 33,872 recoveries.

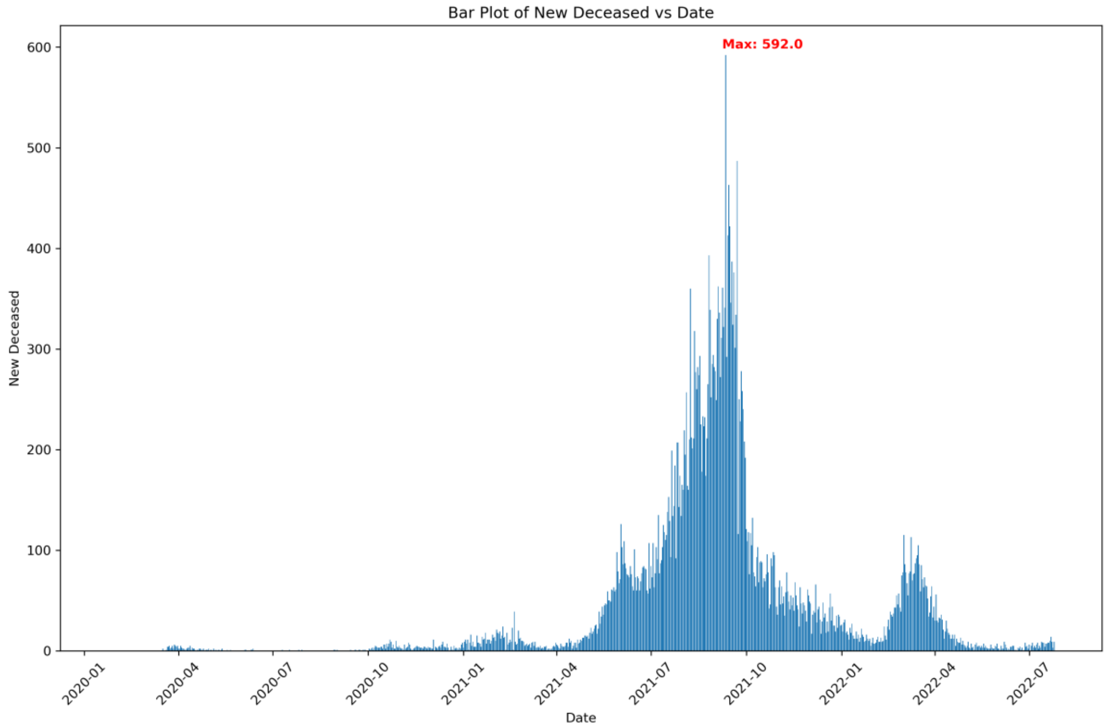

Figure 1.15 Bar Plot of New Deceased with Date

Figure 1.15 shows an upward trend in COVID-19-related deaths, with distinct peaks and valleys. The highest peak, at 592 deaths per day, occurred in late 2021 to early 2022, likely due to the Omicron variant's surge. Smaller peaks reflect fluctuations influenced by public health measures, testing rates, and new variants. Despite vaccination efforts, booster shots were not widely administered until September 2021, contributing to the rise in fatalities.

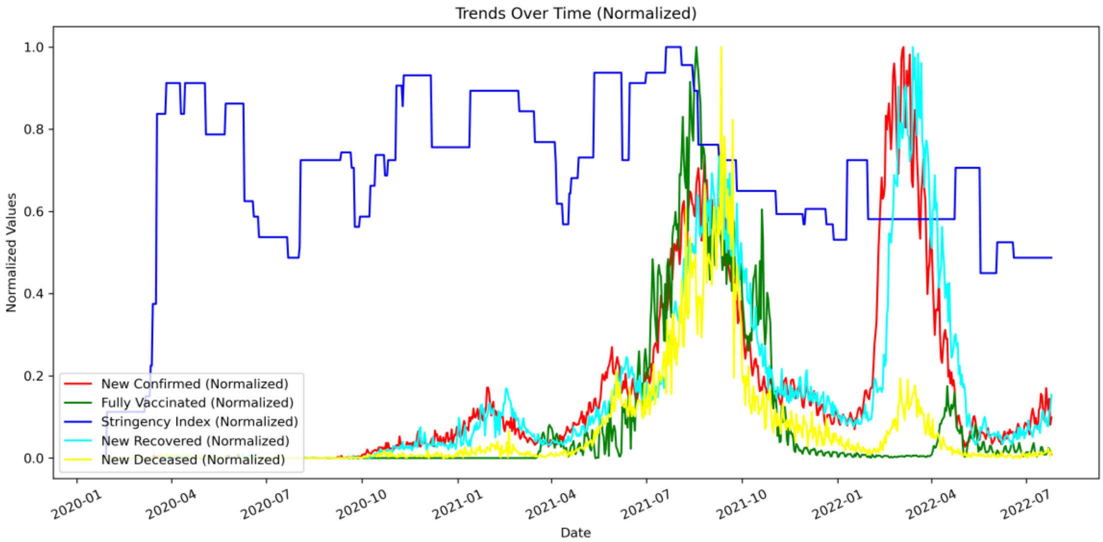

Figure 1.16 Trends Over Time (Normalized) for New Confirmed, Fully Vaccinated, Stringency Index, New Recovered, and New Deceased

Figure 1.16 shows normalized trends for new confirmed cases, fully vaccinated individuals, stringency index, new recoveries, and new deaths. Here min-max normalization was used to allows for easy comparison across metrics. The stringency index, shown as step-mid, reflects policies that remained in place over time and generally followed case trends. From July 2021 to January 2022, lower cases led to a lower stringency index. Then, the graph highlights the effectiveness of vaccination in reducing the severity of COVID-19 outcomes. During the early phase of vaccination, when only a small portion of the population had been vaccinated, the number of deaths remained high despite rising infections. This reflects the initial vulnerability of the population before widespread immunization took effect. However, as more people received their vaccines, a noticeable shift occurred even though the number of new confirmed cases remained high, the number of deaths decreased significantly. This indicates that vaccines were effective in preventing severe cases and fatalities, breaking the direct correlation between high infection rates and high mortality.

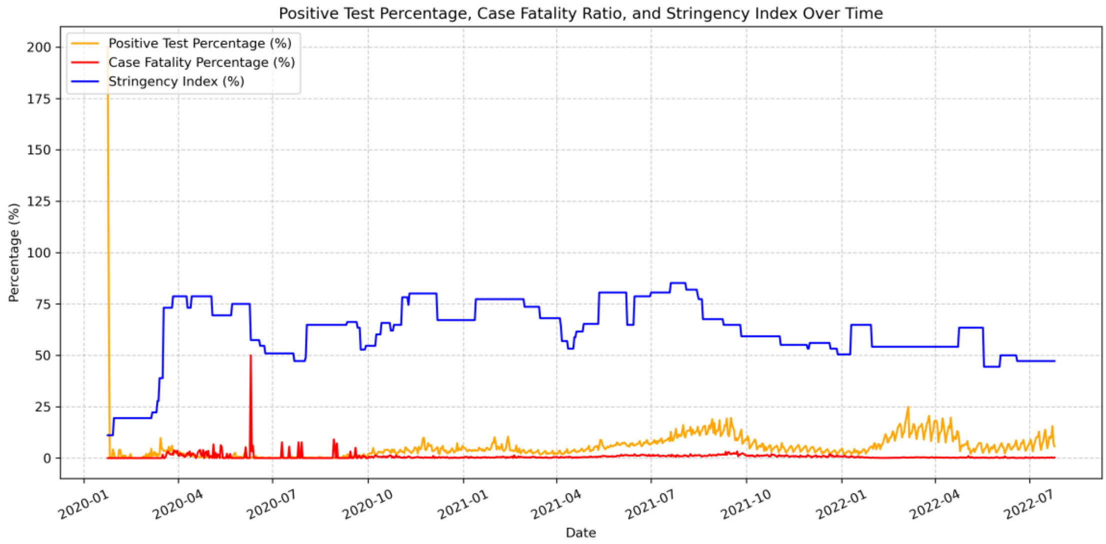

Figure 1.17 Positive Test Percentage, Case Fatality Ration and Stringency Index Over Time

Figure 1.17 shows that at the beginning of the pandemic, without any policies, the positive test percentage was higher. As government policies like lockdowns and social distancing were introduced, the positive test percentage went down. Similarly, the case fatality percentage was also higher early on, as more people got infected, and it took time for the infection to cause death. As stricter policies were put in place, the case fatality percentage decreased. This suggests that stronger measures helped reduce the spread of the virus and lower the number of deaths.

1.4 CHALLENGES

There are several challenges when doing a data cleaning and gaining the insight.

1.4.1 Selecting the Attributes

Selecting attributes from a dataset with 165 columns can be time-consuming and requires extensive reading and research. The system only provides an initial preview of the first 5 and last 5 column names when using *.columns( )* directly, making it difficult to quickly identify the relevant attributes. After careful selection, 22 columns were chosen, and two new attributes were created: positive test percentage and case fatality percentage. The data consists of two types of attributes: categorical data and continuous data. The figure below shows the selected attributes.

```
statistics.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 914 entries, 23 to 936
Data columns (total 24 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   date                                914 non-null    datetime64[ns]
 1   new_confirmed                       914 non-null    float64
 2   new_deceased                        914 non-null    float64
 3   new_recovered                       914 non-null    float64
 4   new_tested                          914 non-null    float64
 5   new_persons_fully_vaccinated        914 non-null    float64
 6   school_closing                      914 non-null    category
 7   workplace_closing                   914 non-null    category
 8   cancel_public_events                914 non-null    category
 9   restrictions_on_gatherings          914 non-null    category
 10  public_transport_closing            914 non-null    category
 11  stay_at_home_requirements           914 non-null    category
 12  restrictions_on_internal_movement   914 non-null    category
 13  international_travel_controls       914 non-null    category
 14  income_support                      914 non-null    category
 15  debt_relief                         914 non-null    category
 16  public_information_campaigns        914 non-null    category
 17  testing_policy                      914 non-null    category
 18  contact_tracing                     914 non-null    category
 19  facial_coverings                    914 non-null    category
 20  vaccination_policy                  914 non-null    category
 21  stringency_index                    914 non-null    float64
 22  positive_test_percentage            914 non-null    float64
 23  case_fatality_percentage            914 non-null    float64
dtypes: category(15), datetime64[ns](1), float64(8)
memory usage: 80.1 KB
```

Figure 1.18 Selected Attributes and New Attributes

1.4.2 Missing Values in the Dataset

The dataset contained missing values across most attributes, requiring the use of appropriate methods to handle them. One option was to consult experts or further research the data. Several methods for imputing missing values include removing them or replacing them with the mean, median, or a placeholder like "unknown." In this case, missing values in categorical variables, such as government policies, were removed, along with rows where all values were zero, indicating that the event had not started. When creating new variables, some missing values appeared due to division by zero, which was manually replaced with zero. Other continuous variables with missing values were also replaced with zero, as removing them would not be ideal, especially for health-related data that cannot be controlled.

1.4.3 Structural Issues in the Data

In this dataset, categorical variables were in a dummy format (0s and 1s), requiring conversion for proper analysis. Functions were necessary to perform this transformation. It was important to first understand the factors behind the 0s and 1s and carefully analyse why these values appeared before renaming them for clarity. This process involved significant time and effort to read through the data and gain a deeper understanding of it. The figure below illustrates the functions used and the renaming process applied to the categorical variables, specifically for government policies.

```
statistics['testing_policy'].unique()
```

```
array([1., 3.])
```

```
# Mapping factors for categorical variables
def map_factors(df, column, mapping):
    df[column] = df[column].map(mapping).astype('category')
```

```
map_factors(statistics, 'school_closing', {0: 'None', 1: 'Recommend', 2: 'Some', 3: 'All'})
map_factors(statistics, 'workplace_closing', {0: 'None', 1: 'Recommend', 2: 'Some', 3: 'All'})
map_factors(statistics, 'cancel_public_events', {0: ':None', 1: 'Recommend', 2: 'All'})
map_factors(statistics, 'restrictions_on_gatherings', {0: 'None', 1: '<1000', 2: '<100',
                                                       3: '<10', 4: 'Any Size'})
map_factors(statistics, 'public_transport_closing', {0: 'None', 1: 'Recommend', 2: 'Closing'})
map_factors(statistics, 'stay_at_home_requirements', {0: 'None', 1: 'Recommend',
                                                      2: 'Exceptions', 3: 'Strict'})
map_factors(statistics, 'restrictions_on_internal_movement', {0: 'None', 1: 'State/District',
                                                              2: 'Country'})
map_factors(statistics, 'international_travel_controls', {0: 'None', 2: 'Screening',
                                                          3: 'Quarantine', 4: 'Ban'})
map_factors(statistics, 'income_support', {0: 'No', 1: 'Yes'})
map_factors(statistics, 'debt_relief', {0: 'No', 2: 'Yes'})
map_factors(statistics, 'public_information_campaigns', {0: 'No', 2: 'Yes'})
map_factors(statistics, 'testing_policy', {1: 'Limited', 3: 'Open'})
map_factors(statistics, 'contact_tracing', {0: 'None', 1: 'Limited', 2: 'Comprehensive'})
map_factors(statistics, 'facial_coverings', {0: 'None', 2: 'Certain Settings', 3: 'Most Settings'})
map_factors(statistics, 'vaccination_policy', {0: 'None', 1: 'Certain Groups',
                                               2: 'Vulnerable Groups', 3: 'All',
                                               4: 'Mandatory', 5: 'Booster'})
```

Figure 1.19 Mapping Factors for Categorical Variables

1.4.4 Visualization Complexity

Creating effective visualizations was challenging due to the large number of variables and the complexity of the relationships. To simplify, the data was grouped into meaningful categories, such as combining the type of policy with its respective dates and counting the number of days. This approach reduced the number of individual plots needed. By grouping continuous variables in a single plot, it became easier to observe the relationships between them and identify potential influences. Additionally, normalizing the data ensured that the visualizations were more accurate and allowed for clearer comparisons.

1.5 CONCLUSION

In conclusion, when preparing the data, careful consideration of imputation strategies for missing values is essential, as it can impact the insights gained. Domain knowledge is crucial to avoid spending excessive time understanding the data or producing inaccurate outcomes. Iterative refinement of visualizations is also necessary for clarity and effective communication. From the analysis, it can be concluded that the stringency index plays a role in influencing the rate of new confirmed cases and the positive test percentage. Additionally, vaccination policies demonstrate the effectiveness of vaccines, as reflected in the reduced case fatality percentage.

REFERENCES

Ceder, V.L. (2010). The Quick Python Book (2nd ed.). Manning.

FMT Reporters. (2021, August 19).42 new Covid-19 clusters, 23 involving workplaces. *Free Malaysia Today (FMT).* Retrieved from: https://www.freemalaysiatoday.com/category/nation/2021/08/19/record-42-new-covid-19-clusters-23-involving-workplaces (on 19 January 2025)

Grus, J. (2015). Data Science from Scratch: First Principles with Python (1st ed.). O'reilly.

K K, Ramachandran. (2024). Data Science in the 21st Century: Evolution, Challenges, and Future Directioons. International Journal of Business and Data Analytics (IJBDA). pp:1. 1-01.

Mansor, Z., Mokhtar, S.A. (2022).Covid-19 Clusters in Malaysia: A Descriptive Analysis. *Advances in Health Sciences Research.* Vol.44

Matthes, E. (2019). Python Crash Course: A Hands-on, Project-Based Introduction to Programming (2nd ed.). No Starch Press, Inc.

McKinnery, W. (2012). Python for Data Analysis: Agile Tools for Real-World Data (1st ed.). O'reilly.

Pricing & Tariffs. (2014) TNB Better. Brighter. Retrieved from https://www.tnb.com.my/residential/pricing-tariffs

VanderPlas, J. (2016). Python Data Science Handbook (1st ed.). O'reilly.