horizontal line

EGT204: DATA PREPARATION & VISUALIZATION

**Dengue Analysis**

**& Prevention**

**Report**

Group Name: Dengue Ops

**Report By:**

Stephanie Neo Sze Ying 222646Z

Tan Hao Wen 234733T

Sai Keerthan Satini 232594T

Tiah Wei Xuan 232287D

**Submitted To:**

Ms Zheng Xin Yu

Lecturer

Nanyang Polytechnic

Date of Submission: 12/08/2024

**TABLE OF CONTENTS**

Introduction

- Abstract page 3

- Purpose of report page 3

- Background Information page 3

Findings and Discussion

- Sai page 4 - 8

- Haowen page 8 - 9

- Wei Xuan pages 10 - 11

- Stephanie pages 11 - 15

Recommendations page 15 - 17

Conclusion page 17

References page 17

Contributions page 18

**ABSTRACT**

This paper delves into the in-depth analysis of the rising number of dengue cases within Singapore. The first step of the team’s analysis includes sourcing for datasets regarding the whereabouts of mosquitoes hotspots, top contenders for dengue outbreaks, and the measures other countries take, primarily Cuba.

The datasets are then cleaned using Google Colab, while Analysis & Visualisation is done on Power B.I. These analysed datasets will be proposed to the government for improvement strategies.

**Purpose of this Report**

The aim of this report is to analyse the growing number of dengue cases in Singapore and propose the government suggestions and guidance regarding potential dengue outbreak prevention. The purpose of this research project is to assist Singapore's government in developing objectives that effectively contain and prevent dengue outbreaks, hence maintaining the health and welfare of the general people.

# 

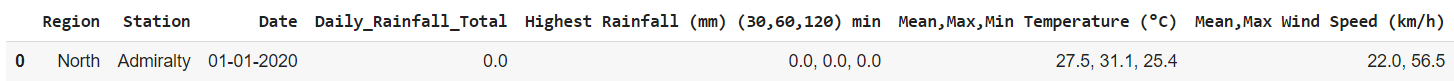
# Background/Current Situation

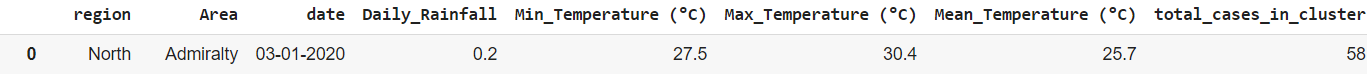
Dengue cases in Singapore are significantly increasing as of August 2024. By late July 2024, there will have been over 10,000 cases reported, surpassing the amount for 2023. This is a significant rise over the prior year. Regrettably, 13 fatalities linked to dengue this year also demonstrate the severity of the pandemic.

Numerous variables, such as a decline in population immunity to the different dengue serotypes and conducive climatic conditions for mosquito reproduction, have been linked to the increase in incidence. The virus has continued to spread quickly in spite of the National Environment Agency's (NEA) best efforts to stop it, including more stringent mosquito control measures.

**DATA PREPARATION:**

**RAW DATASET VISUAL:**

**CLEANED & MERGED DATASET VISUAL:**



The datasets gathered did not meet the expectations and requirements to begin analysis. To ensure that the dataset is usable, cleaning of the data is a necessity with the use of “dropna” and “drop\_duplicates”.

As for the “hospital disease” dataset, it is required to extract all the years into individual datasets so that it can be concatenated to merge the data from 2017 to 2020. Data melt was also used to turn columns into key-value pairs, making comparison and evaluation better.

The main dengue dataset utilised now was first separated into 256 different csv files. “Pre-processing.py” was the python file that merged all the 256 items together into one big dataset that can be cleaned and explored and ready for analysis

The 4 new columns added are:

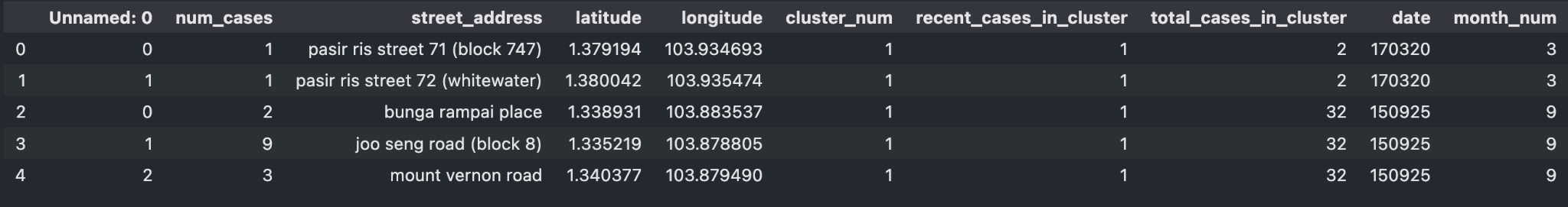
**Region:** The addition of the region group allowed for richer understanding of the spread of the dengue cases across the country, identifying potential hotspots across the country.

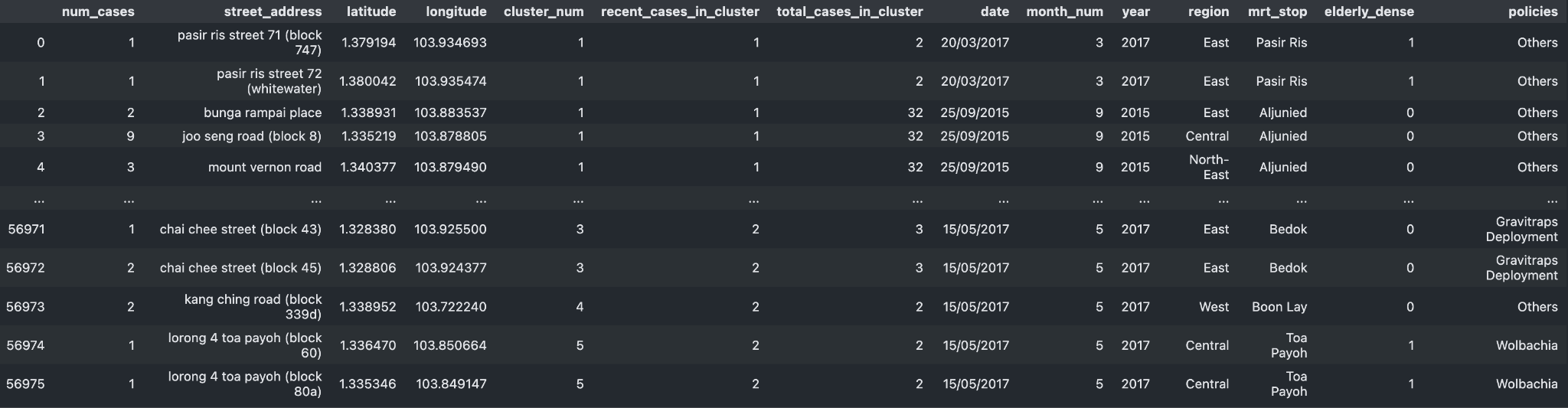
**Policy:** The Policy column was added after conducting a quick research on where “Wolbachia” and “Gravitrap Deployment” Policy was implemented. The street addresses from the original dataset were then associated with their respective policies. The addition of this column allows observation of the trend of dengue cases according to the places policies were implemented.

**MRT Stop:** The addition of the MRT stops identify places that have the most dengue cases, allowing for additional insights.

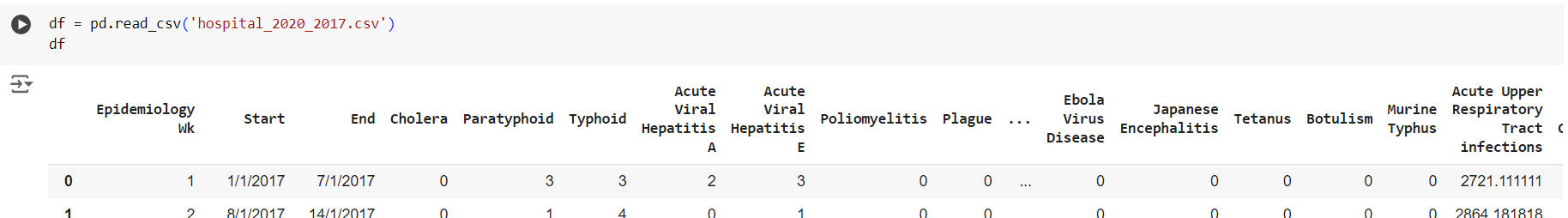
**Elderly Dense:** Quick and Brief research was conducted on the places which had a high elderly density and the respective street addresses were associated with the Elderly Dense column, with ‘1’ being the place dense with elderly population and ‘0’ being it isn't. This addition allowed for minor insights.

**Main Dataset (Before and After the addition of the columns):**

****

****

**Raw data:**



**Cleaned data:**

****

**FINDINGS AND DISCUSSION**

**Sai’s PowerBi & Analysis:**

**Bar Charts** were used in this case as it breaks down the total number of dengue cases by different regions and places in Singapore and it is simple to evaluate from the bar charts on which places/regions are dengue-ridden in Singapore. Bar Charts are effective in comparing data from different categories and streamlines the process to derive insights.

**Line Carts** were used as it illustrates the trend of dengue cases over the desired period of time, which is either in months or years in this scenario. The Line charts are controlled by a slicer at the left corner where it adjusts based on the time period desired by the user. For showing the number of dengue cases over a period of years or months, line charts work well.Additionally, it makes it easier to comprehend and spot any notable variations in the quantity of dengue cases recorded over time.

**Map Visualisation** is used as it provides a spatial distribution of the dengue cases in Singapore. Using Maps allows the user to understand the spatial spread of dengue across the country and also streamlines the process to identify dengue hotspots across the country.

**Pie Chart** is used for the analysis on the elderly density as it allows the user to see the correlation. The Pie chart changes in accordance to the region and the place, showing the distribution of how dense the elderly population is in percentage.

**Slicers**: The three slicers, “Policy”, “Region” and “Year” affect all the visualisation tools and changes in accordance to the policy implemented, the region chosen and the desired period of time. Slicers enhance data interactivity and provide the user with dynamic filtering which allows the user to derive extensive insights as the different visualisations adapt to the slicer.

**Analysis:**

There is a noticeable spike in dengue cases after the year 2017. Following a brief research, the sudden increase in cases is caused by the shifts in the serotypes of dengue virus ([Link](https://www.ncid.sg/Health-Professionals/Articles/Pages/Epidemic-Dengue-in-Singapore-During-COVID-19-Pandemic.aspx#:~:text=The%20upward%20trend%20in%20dengue,found%2023%2C400%20mosquitoes%20breeding%20habitats.&In%202021%2C%20the%20number%20of,the%20year%20(Figure%201).&Singapore%20is%20an%20ideal%20environment,the%20increase%20in%20dengue%20cases.)). Hence, an upward trend from 2017 onwards.

When the filter is set to “Wolbachia” Policy, it is noticeable that most of the cases are being focused in the North-East and Central. This is where the Wolbachia policy was implemented in the early stages of the policy.

Looking at the Line Charts, the trend of the cases in the areas where the Wolbachia Policy was implemented based on the year and over the months.

When the filter is set to “Gravitrap” Policy, most of the cases seem to be focused on the West, Central and East. These regions are where the policy was first implemented.

Overall, both the line graphs show that cases have continued to surge after 2018. However, the Gravitrap Deployment Policy seems to be reducing cases between 2018 to 2020, with dengue cases taking a hit from 2019 to 2020. Whereas for Wolbachia, cases have continued to increase from 2018 to 2020. This shows that as a whole Gravitrap Deployment is a much more effective policy to be deployed throughout the year compared to Wolbachia and the decrease in cases from 2019 to 2020 shows that the Gravitrap Deployment Policy is able to contain the new Dengue Virus Serotype as compared to Wolbachia.

However,based on the line charts according to months, it is observable that cases have started to increase from the 4th month to the 8th month of the year. The increase in cases can be justified as it is mating season for mosquitoes in Singapore. During the months of the mating season, Wolbachia seems to contain the cases better as it shows peaks and troughs as compared to Gravitrap Deployment Policy, where the dengue cases are rising exponentially. This shows that during the mating season, The Wolbachia Policy seems to be more effective at containing the dengue cases compared to the Gravitrap Deployment.

**HaoWen’s Dataset**

**PowerBi & Analysis:**

**For the line graph:**

The purpose of the usage line graph is to give a visual of the trend of the years , so that identifying the patterns of the trends would be easier.

**For the pie chart:**

The pie chart is designed to show which top 5 diseases have the highest impact in Singapore. It effectively visualises the proportion each disease contributes to the total, allowing for easy comparison of the most significant contributors to health outcomes in the region.

**For the map:**

The map is utilised to show the geographical concentration of the death rate, providing a spatial perspective on where the impact is most significant. This allows for regional comparisons and can help identify areas that may require more focused health interventions.

**The card**

The purpose of this is to provide a total amount of data about dengue over the years, death rate , trends , and diseases.It helps to get the attention of the audience.

**The Slicer**

The purpose of slicer is to filter the specific year and countries.

1.Google Trends:

* There is an increasing trend during the period 2021 to 2022(during covid times) from 511 to 1544.

2.Hospital of disease:

* Dengue has been the top 5 diseases over 2017 to 2020 affecting people.

3. World death rate:

* There has been a constant above 4000 death rate between 2000 to 2010 , and an increasing trend after 2010 and hit 6300 death rate in 2020.

**WeiXuan’s Dataset**

Using the Line graph to visualise the dashboard is effective because it is useful for visualising data trends over time. It shows how the amount of dengue cases increases and decreases within years. This allows comparison between Singapore and Cuba to show that Singapore’s dengue cases increased drastically as compared to Cuba.

1. Dengue Cases Trends:
   * Singapore: There is an increase in dengue cases between 2015 and 2020. In 5 years, there were 147k reported cases.
   * Cuba: Dengue cases increased around the year 2000, with numbers reaching around 10,000, but then decreased and maintained its low number of reported cases. In 20 years, there were 15k reported cases.
2. Methods Used for Dengue Control:
   * Both Singapore and Cuba use similar methods such as using Bacillus Thuringiensis Israelensis (BTI), Wolbachia bacteria, chemical control, public education, reduction of stagnant water, surveillance & monitoring, and regulations & policies.
   * The only difference is that Cuba uses Gambusia (mosquitofish) and Copepods as control methods, while Singapore isn’t using this method.
3. Gambusia vs. Copepods:
   * Gambusia: Effective in various environments but can cause environmental harm and require high maintenance.
   * Copepods: Environmentally friendly, low maintenance, and effective in targeting mosquito larvae, despite the fact that both require careful management to prevent ecosystem disruption and their effectiveness can vary.

**Stephanie’s Dataset**

**Power B.I Analysis:**

1. **Pie Chart** - Dengue Cases by Region:

Pie Chart was used to showcase the highest amount of dengue cases by region as it brings ease to the eyes, especially since the Central region holds a large percentage.

1. **Line & Stacked Column Chart** - Sum of Rainfall & Correlation to dengue cases arranged by region:

This graph was utilised as it comes with a secondary y-axis and a visual that is not the same as to what the 1st y-axis provides. What is meant is that, the 1st visual provided by the 1st y-axis is in the form of a bar chart, while the 2nd visual is in the form of a line graph. Using a graph with a secondary y-axis is important.

The 2 selected columns getting compared have a huge numerical difference between them. The rainfall column’s peak value is at 4k, while the sum of dengue cases has a maximum value of 150k. When these are plotted in a bar chart, there is no secondary y-axis and both columns are plotted against the same scale, making the rainfall column’s bar chart barely visible.

Having different visuals when you’re planning to plot more than 1 visual is also important. It is easier to identify trends. Multiple of the same visuals can be overwhelming.

1. **Clustered Column Chart**- Correlation between temperature & Number of Dengue Cases:

This graph was used as it has a simple visual. Initially, the temperature column was meant to be plotted against the dengue cases column, with the x-axis as region. However, it is discovered that when any column is moved to the Y-axis, it is automatically summed up together. This would cause the temperature to go into the thousands, which is illogical. Hence, the temperature column is only plot against 1 other column, the dengue cases column at the y-axis. The summation of dengue cases is more helpful.

1. **Line Chart**- Correlation between Humidity & Dengue Cases:

Line chart was utilised to show the trends over time. As the bar chart was already used earlier, a repetitive visual was unwanted, and the next best option was a line chart.

1. **Decomposition Tree** - By region, The months with the highest counts of rainfall:

Allows for breaking down of data into hierarchies for easier exploration/visualisation. This was utilised, as it is an essential graph to see which months my recommendations should be implemented to. The ‘base’ analyser used is the count of rainfall columns. It is explained further by the region column, and showcases the sum of rainfall per region. It is then explained further by the months column, showcasing the highest rainfall per month.

1. **Slicer**: *Allows for sorting by month/region.*

**Recommendations**

**Sai’s Recommendations:**

Although it has limits, the Wolbachia policy is an excellent mosquito control tool. It is most effective from June to August, when mosquito activity is at its highest due to it being the mating season. Outside of that time of year, its effectiveness wanes. It has been demonstrated that gravitraps are more reliable in year-round mosquito population control.

A coordinated strategy is advised to control mosquito populations and stop dengue transmission. Better overall control can be achieved by using Wolbachia during mosquito peak season and augmenting it with techniques like Gravitraps and targeted spraying during off-peak seasons.

Moreover, to combat the newer virus strains, the government should identify regions where the new serotype are predominant and implement more of the Gravitrap policy as it is shown to be the superior policy in combating the new virus strains as compared to Wolbachia

**WeiXuan’s Recommendations:**

The Singaporean government should make the most of copepods' potential in order to combat the rising number of dengue cases in the nation. Copepod efficacy has been well established, especially when contrasting the positive results attained in Cuba. Singapore should think about using this tactic as part of its all-encompassing dengue strategy, given Cuba's shown success in lowering mosquito populations and, consequently, dengue incidence. Through the utilisation of copepods, the government can implement a proactive and scientifically supported strategy to mitigate the dengue epidemic, safeguarding public health and lessening the strain on medical resources.

**Stephanie’s Recommendations:**

Rainwater often ends up accumulating in drains or canals, which may cause the stagnation of water to occur. As stagnant water is a breeding ground for mosquitoes, it is recommended that the government conduct scheduled maintenance checks on months with the highest rainfall in the region with the highest amount of dengue cases.

The government can increase the use of larvicides in water bodies that cannot be drained and adulticides during peak mosquito activity periods.

Next, another solution is that investment can be made to improve drainage systems to reduce stagnant water from high rainfall.

**CONCLUSION**

1. In conclusion, the main suggestion to combat our dengue cases is to make use of copepods. It is a solution that the Singapore Government has never implemented before, and Cuba has proven its effectiveness in stopping the reproduction of mosquitoes.
2. The Government should alternate between Gravitrap and Wolbachia policy, deploying Wolbachia Policy during the mating season, specifically on the 4th and the 8th month of the year, and Gravitrap during the remaining months of the year.
3. With inference from the pie chart and decomposition tree respectively, more focus should be on the Central region as it has the highest reported amount of dengue cases. With that information in mind, after setting up the decomposition tree, it is needed to narrow down the months within the Central region that has the highest rainfall. April is the peak rainfall season within the central region. Hence, such measures should be imposed during this period to make the solutions more effective. Since NEA conducts frequent search & destroy operations of mosquito breeding grounds within Singapore, they can also fumigate these areas more during these periods as well, aside from the recommendations provided above.

**References**

<https://www.moh.gov.sg/resources-statistics/infectious-disease-statistics/2020/weekly-infectious-diseases-bulletin>

<https://trends.google.com/trends/explore?date=2020-07-01%202024-08-03&geo=SG&q=dengue&hl=en-SG>

<https://data.worldbank.org/indicator/SP.DYN.CDRT.IN>

Cuba: <https://opendengue.org>

[Main Dengue Dataset](https://outbreak.sgcharts.com/data)

Climate:

<https://www.kaggle.com/datasets/whitecl0ud/weather-data-of-singapore?select=combined_output.csv>

<https://beta.data.gov.sg/datasets?q=&ext_type=dataset&groups=&organization=&query=humidity&resultId=d_31dbf162e94f76065e1e4c60a80a4264>

**Individual Contribution**

**Hao Wen:**

* Assist Group members with data visualisation on how to use PowerBI.
* Data Cleaning
* Slides.
* Report.

**Wei Xuan:**

* Personal Dashboard on the comparison between Singapore and Cuba
* Power B.I Analysis
* Data Cleaning
* Slides
* Report

**Sai:**

* Assist in search of Datasets
* Data Cleaning
* Power B.I Analysis
* Slides

**Stephanie:**

* Power B.I analysis
* Slides
* Report
* Data Cleaning