Homework Week 10

FIT3179

David Tanuwijaya – 32156553

Dataset:

Data combination of Kaggle and government dataset in research of the Singaporean public transport from 2004-2014. The data is reasonably old, but no newer data can be found in the internet. Hence, the implication would be, the data would not be suitable for any future prediction due to its time obsolescence, but good for analyzing the past and observe how Singapore as a country has developed its way of being one of the country with the best public transport service across the world.

Source:

1. Kaggle’s Dataset

<https://www.kaggle.com/datasets/yxlee245/singapore-train-station-coordinates>

1. Lim Chee Aun’s Dataset

<https://observablehq.com/@cheeaun/list-of-bus-stops-in-singapore>

1. Singapore’s Government Public Data Bank

<https://data.gov.sg/dataset>

Data Cleaning Process:

Most of the data are taken directly from the dataset found in the internet. With minor data analysis require in changing the layout from wide to long formatting using excel.

Visualizations:

* Map: to visualize the exact location of the bus stops and railways stations placed across the country. It allows user to know which part of the country has more or less stops or source of public transportation.
* Stacked Bar: To show the proportion of the different type of public transportation method that the people could use. Better in terms of showing contrast and contribution toward the whole number
* Line Chart: To show the year to year commute time for each different public transportation available for people to choose from. The line chart provide a good comparing tool to depict the different easier.
* Donut Chart: To contrast the different proportion of the MRT and LRT stations, which in this case are very different.

Link to visualization:

