# Applied Data Science Capstone Project (as part of the IBM Professional Certificate in Data Science)

By: Choy Siew Fong

20 March 2020

## Introduction

## Description of the Dataset

The dataset ‘Singapore Train Station Coordinates’ was obtained from Kaggle (see <https://www.kaggle.com/yxlee245/singapore-train-station-coordinates>). It is in comma separated values (.csv) format and consists of four columns listing the station name, type (i.e. whether it is an MRT or an LRT station) and positional coordinates (latitude and longitude) of each MRT and LRT station in Singapore. The information about the dataset indicated that it was uploaded by Lee Yu Xuan around eight months ago. Figure 3 shows the screenshot from Kaggle with some details about this dataset.

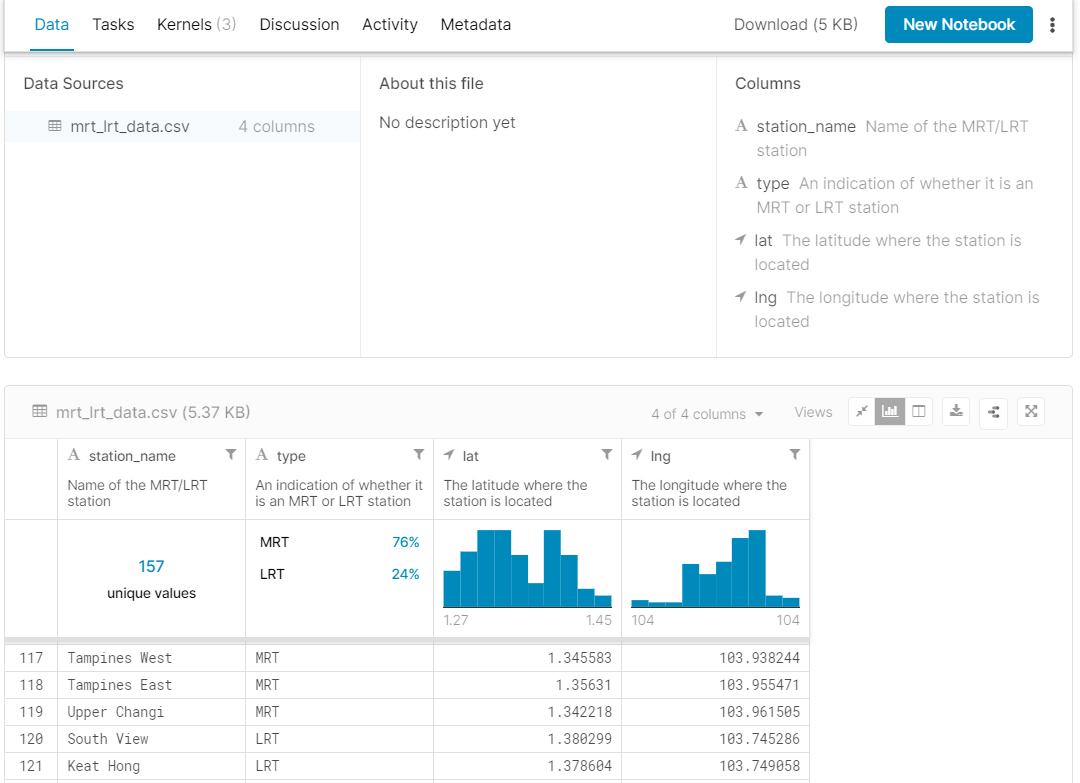


Figure 3: Screenshot from Kaggle showing some details of the dataset

Each row lists the station name, the type of station, and its latitude and longitude coordinates. For instance, Tampines East station is an MRT station located at latitude 1.345583˚ and longitude 103.938244˚ while South View is an LRT station located at 1.380299˚, 103.745286˚ respectively. Preliminary inspection of the dataset found no missing or null values. However, it was later discovered during Folium map plotting that the coordinates for Admiralty Station were identical to Woodlands Station, probably due to a typographical error by the author of the dataset. To resolve this issue, I searched online for the coordinates of Admiralty station and replaced the erroneous latitude and longitude values for Admiralty station in the dataset with those from this site\* (latitude = 1.44069˚, longitude = 103.8009˚).

\*(<https://www.findlatitudeandlongitude.com/l/Admiralty+mrt/1214628/>).

The Foursquare information about each station would be obtained by defining a search URL (limiting the number of venues returned for each station to 100 within a radius of 500 m from each station) and sending the ***GET*** request to Foursquare API. The output would be a **json** file containing information such as the venue id, name, location coordinates, tips, rating e.t.c.