Data Section

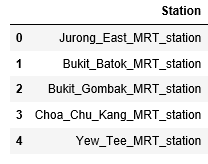
To solve this problem, we need to understand and collect the data we need. Below are the data we need to used :

1. **Longitude and Latitude of the old and new office**

These data are obtained from Wikipedia, which are International business park (1.329583, 103.7475) , Changi Business Park (1.334972, 103.965167).

1. **Singapore MRT station list**

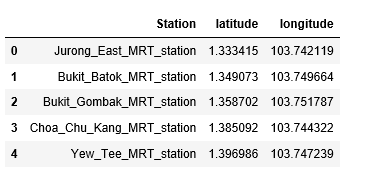
This data is exacted from Wikipedia link <https://en.wikipedia.org/wiki/List_of_Singapore_MRT_stations>. After extraction, the data is in the form of data-frame as shown in Figure 1.



**Figure 1: Singapore MRT table head**

1. **Longitude and Latitude of all the Singapore MRT station**

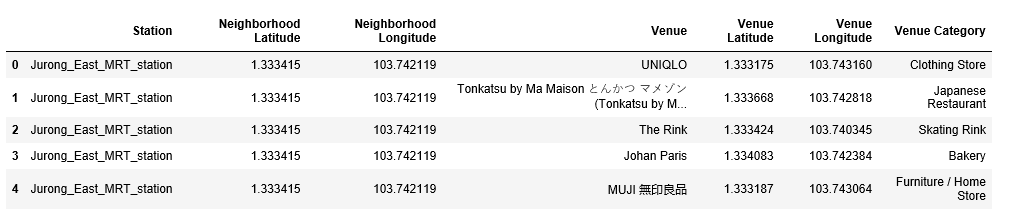
Since no available free library can be used to extract the Longitude and Latitude, it was decided to use a simple piece of code to automatically exact the Longitude and Latitude by searching the Wikipedia web page. The obtained data are shown in Figure 2.



**Figure 2: Singapore MRT with Longitude and Latitude table head**

1. **Use the Foursquare API to extract the neighborhoods in SG MRT Stations**

For each MRT station, the searching range is set as the top 100 venues within a radius of 500 meters. Then a data-frame with venues of each MRT station can be seen in Figure 3.



**Figure 3: venues of each MRT station**

With this data set, we can cluster all these MRT into 5 clusters and sort out the most similar neighborhood that near to John’s new office.