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

By: Choy Siew Fong

22 March 2020

## Introduction

Singapore’s rail network system consists of the Mass Rapid Transit (MRT), which is a heavy rail rapid transit system and the subsidiary light rail transit (LRT) system which serves the non-mature estates of Sengkang, Punggol and Bukit Panjang and function as a feeder network to the MRT. Figure 1 shows an example of an MRT train while Figure 2 shows a typical LRT train.



Figure 1: Photograph of an MRT train at Mandai Depot (Source: Mass Rapid Transit (Singapore). 18 March 2020. In *Wikipedia.* Retrieved March 19, 2020, https://en.wikipedia.org/wiki/Mass\_Rapid\_Transit\_(Singapore)



Figure 2: An LRT train (Source: Light Rail Transit (Singapore). 18 March 2020. In *Wikipedia.* Retrieved March 19, 2020, <https://en.wikipedia.org/wiki/Light_Rail_Transit_(Singapore)>

Due to the astronomical prices of cars in Singapore, most of the island’s population gets around by public transport. In 2019, the daily ridership of the MRT network was 3.4 million while the annual ridership was 1.2 billion. Since the MRT and LRT is such an integral part of people’s daily lives in Singapore, I was interested to find out what kind of venues (such as restaurants, cafes, or supermarkets et cetera) existed around the MRT and LRT stations and if they have any trends or commonalities which may be analysed using clustering methods such as KMeans. This idea came out of a personal experience where I needed to get something on the way home but wasn’t sure which MRT station I should alight at to buy it. By combining a dataset of the stations’ names, latitude and longitude coordinates with Foursquare location data, we should be able to gain some insight into the venues around each station.

### Potential Stakeholders

This information about the type and number of venues available around each MRT/LRT station would be of interest to various stakeholders. Firstly, passengers or commuters taking the train would be keen to know what amenities are around each MRT or LRT station so that they can make informed decisions during their commute, such as which station to drop off if they wish to buy a loaf of bread from the supermarket for tomorrow’s breakfast or which station would have an ATM from their bank.

Application developers may also have an interest in this information as it will enable them to develop mobile applications which can locate a certain type of venue (e.g. ATMs or cafes) which would be very useful to potential users.

Potential business owners who are contemplating starting a business around MRT stations due to their high footfall would also be interest in this information since it would allow them to know the number and type of competitors already existing in the area. For instance, a businessman interested in opening a Japanese restaurant in the Jurong East MRT area can use this information to find out how many *other* Japanese restaurants already exist in the area, and what are their specialities (from users tips in the Foursquare date), if any. This information would allow prospective business owners to make informed decisions on which niche area they wish to target or to decide to move to a different location altogether.

Similarly, current business owners would be interested in knowing which venues are trending in the locale of their business. This would highlight potential competitors or trends which would enable them to adjust their business model and menu accordingly.