# Classification of Neighbourhood surrounding MRT Station in Singapore

*Chong Yau Loong*

*4th September 2019*

## 1. Introduction

Singapore Mass Rapid Transit (MRT) is one of the essential modes of transportation for many residents and tourists. With 119 stations in operation as of August 2019 and more than 60 stations under construction, the Singapore MRT system network is spanning across the island nation, forming the backbone of the public transport system in Singapore.

Most of the MRT stations in Singapore are strategically built according to the nation's masterplan for transportation, which will eventually support the growth and development of Singapore as a world class city. Many of the stations are integrated with other public transport services, such as the bus interchanges/terminals or Light Rail Transit (LRT), residential areas, business districts, shopping malls etc. As such, each MRT station can be regarded as a small neighbourhood, and existence of the MRT stations will gradually shape the uniqueness of the community around it. For this project, **we will be looking at the neighbourhoods surrounding the MRT stations and classify them accordingly using the data obtaining via Foursquare**. This analysis will be useful for city planner to get an insight of each of the neighbourhood shaped by the MRT stations (uniqueness quality that can be used to promote tourism), identify gaps or areas of improvement (lack of certain amenities compares to other neighbourhoods), and future expansion possibilities (line extension or new bus interchange etc.).

Apart from serving as a train station, Singapore MRT stations generally provide leasing space within the stations for commercial purposes (such as retails, eateries, services etc.). In the context of this project, XYZ Corporation, a food and retail company, who are interested in opening a few of its chained grocery stores and snack bars within the MRT stations. We will be using the results we obtain to advised XYZ Corporation on **which cluster of stations are more suitable for grocery stores or snack bars** accordingly.

## 2. Data

Several data sources will be used for this project but not limited to:

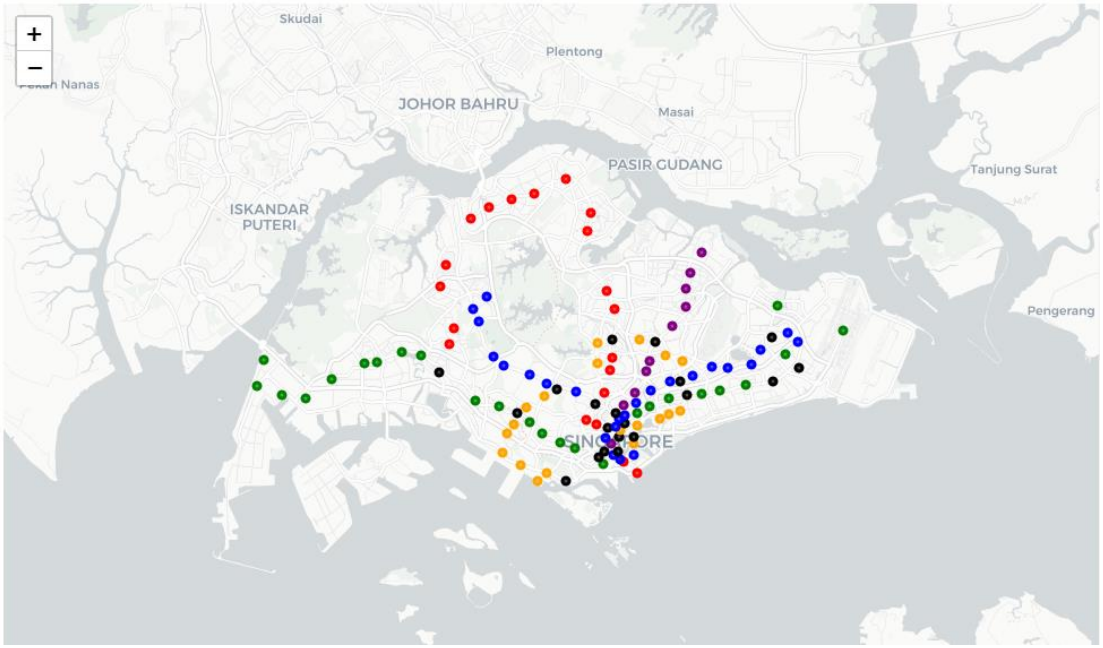
1. List of MRT stations (119 stations currently in-operation as of 2019) - Wikipedia
2. Explore venue types around each station - Foursquare

### 2.1 List of MRT Stations and Coordinates

The below table showing the data scrap from Wikipedia using BeautifulSoup4 and the coordinates of each MRT stations are determined using Geocoders.

| Station Abbr. | Station Name | Station Code | NSL | EWL | NEL | CGX | CCL | DTL | Overlapped | Lat | Long | Color |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HBF | HarbourFront | NE1, CC29 | 0 | 0 | 1 | 0 | 1 | 0 | 2 | 1.266055 | 103.819935 | black |
| LBD | Labrador Park | CC27 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1.266151 | 103.802741 | orange |
| TLB | Telok Blangah | CC28 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1.270839 | 103.808725 | orange |
| MSP | Marina South Pier | NS28 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1.270903 | 103.863195 | red |
| PPJ | Pasir Panjang | CC26 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1.275851 | 103.792596 | orange |
| TPG | Tanjong Pagar | EW15 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1.276419 | 103.842929 | green |
| MRB | Marina Bay | NS27, CE2 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1.277731 | 103.855259 | red |
| DTN | Downtown | DT17 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1.279274 | 103.852808 | blue |
| OTP | Outram Park | EW16, NE3 | 0 | 1 | 1 | 0 | 0 | 0 | 2 | 1.280276 | 103.840300 | black |
| BFT | Bayfront | CE1, DT16 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1.281808 | 103.861016 | blue |

The MRT stations are the plot onto the map of Singapore and the correctness of the stations locations are verifed.




### 2.2 Venues types information from Foursquare

There are 10 top level venue categories used by Foursquare:

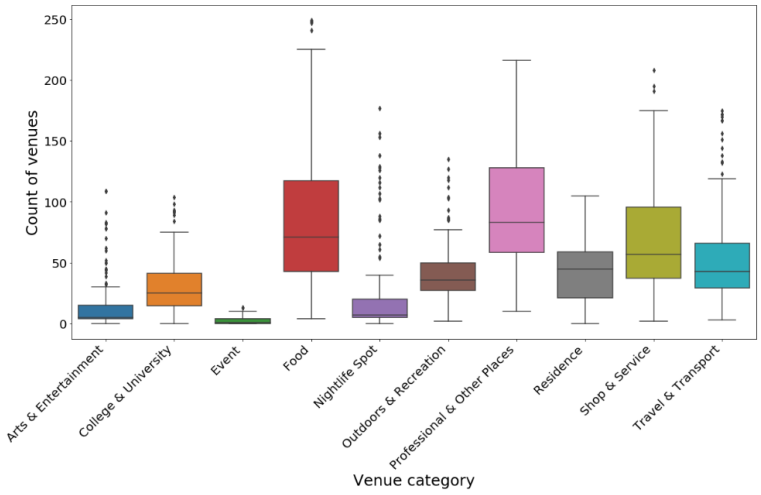
1. Arts & Entertainment
2. College & University
3. Event
4. Food
5. Nightlife Spot
6. Outdoors & Recreation
7. Professional & Other Places
8. Residence
9. Shop & Service
10. Travel & Transport

We will be exploring the number of venues of each category around the radius of 1km for each MRT station. The table below depicted the same dataset joined with the Foursquare venue's category counts.

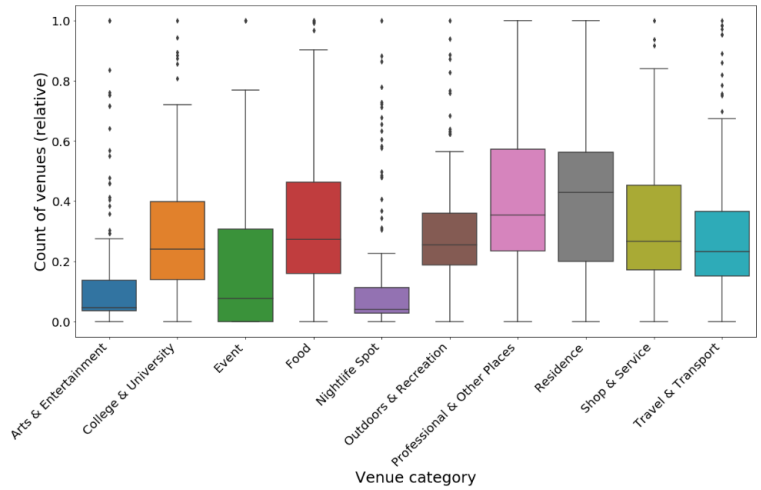
| | Station Abbr. | Station Name | Station Code | NSL | EWL | NEL | CGX | CCL | DTL | Overlapped | ... | Arts & Entertainment | College & University | Event | Food | Nightlife Spot | Outdoors & Recreation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | HBF | HarbourFront | NE1, CC29 | 0 | 0 | 1 | 0 | 1 | 0 | 2 | ... | 16 | 10 | 1 | 119 | 27 | 35 |
| 1 | LBD | Labrador Park | CC27 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | ... | 6 | 3 | 0 | 43 | 5 | 20 |
| 2 | TLB | Telok Blangah | CC28 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | ... | 15 | 10 | 0 | 38 | 8 | 38 |
| 3 | MSP | Marina South Pier | NS28 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | ... | 3 | 1 | 1 | 7 | 5 | 2 |
| 4 | PPJ | Pasir Panjang | CC26 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | ... | 0 | 7 | 0 | 52 | 11 | 15 |

## 3. Descriptive Analysis

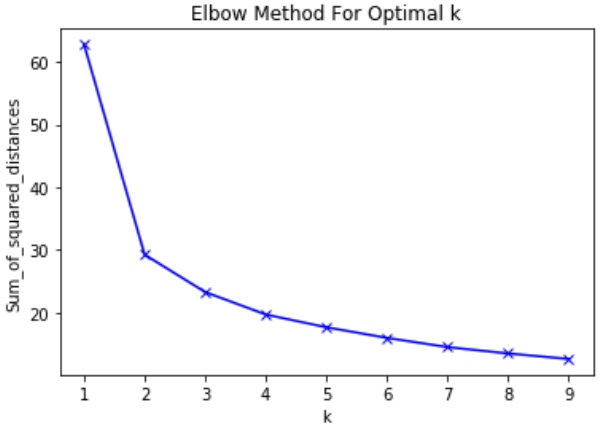
In the section we will be checking the general distribution of the number of venues using boxplots (showing the average count, spread and outliers).




We can see that the most frequent venue categories are Professional & Other Places, Food, and Shop & Service. As the total number of counts for each category are vary from each other significantly (e.g. the Event category count is about 25 times smaller than Food), we shall then normalize the data using min-max scaling. This both normalizes the data for the K-mean clustering algorithm we will be using in the later part.

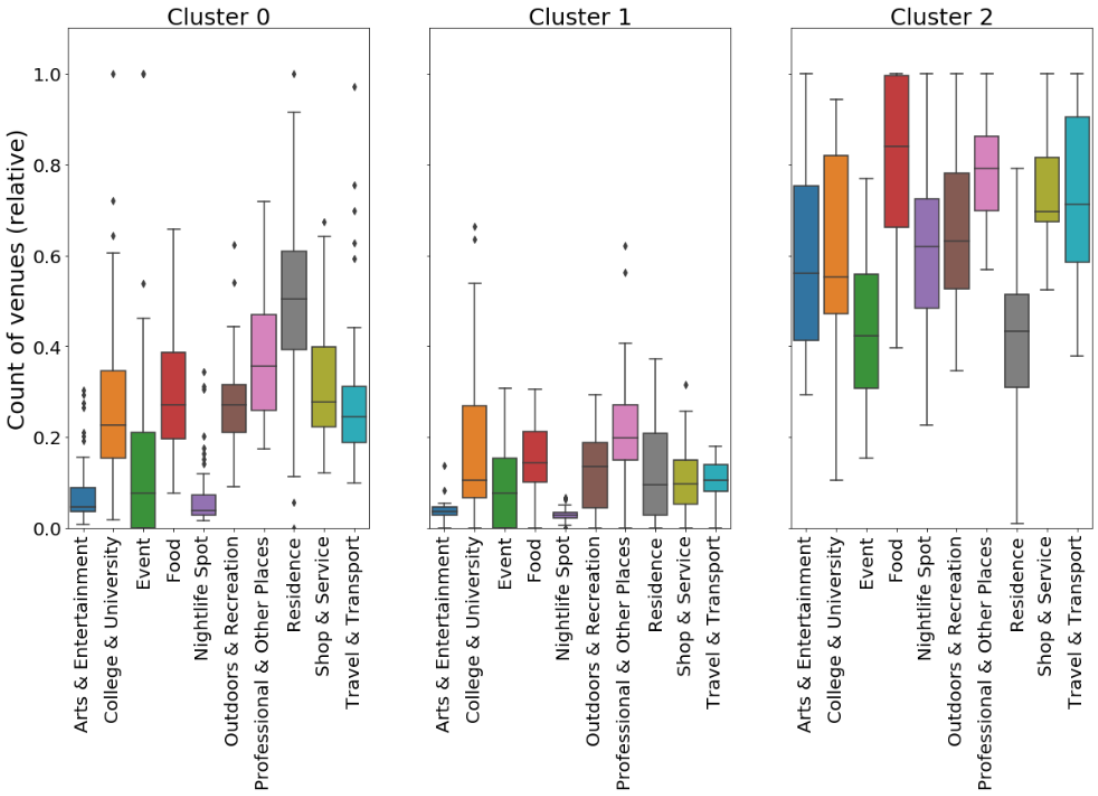



Next, we used the Elbow Method to determine the optimal number of clusters, where the input being the normalized dataset. From the below chart, the optimal number of clusters is found to be 3.



## 4. K-Mean Clustering

Using SKlearn library, K-Mean Clustering algorithm is used to classify 119 MRT stations into 3 groups. The distribution of each venue category in each cluster are represented by the boxplots below:

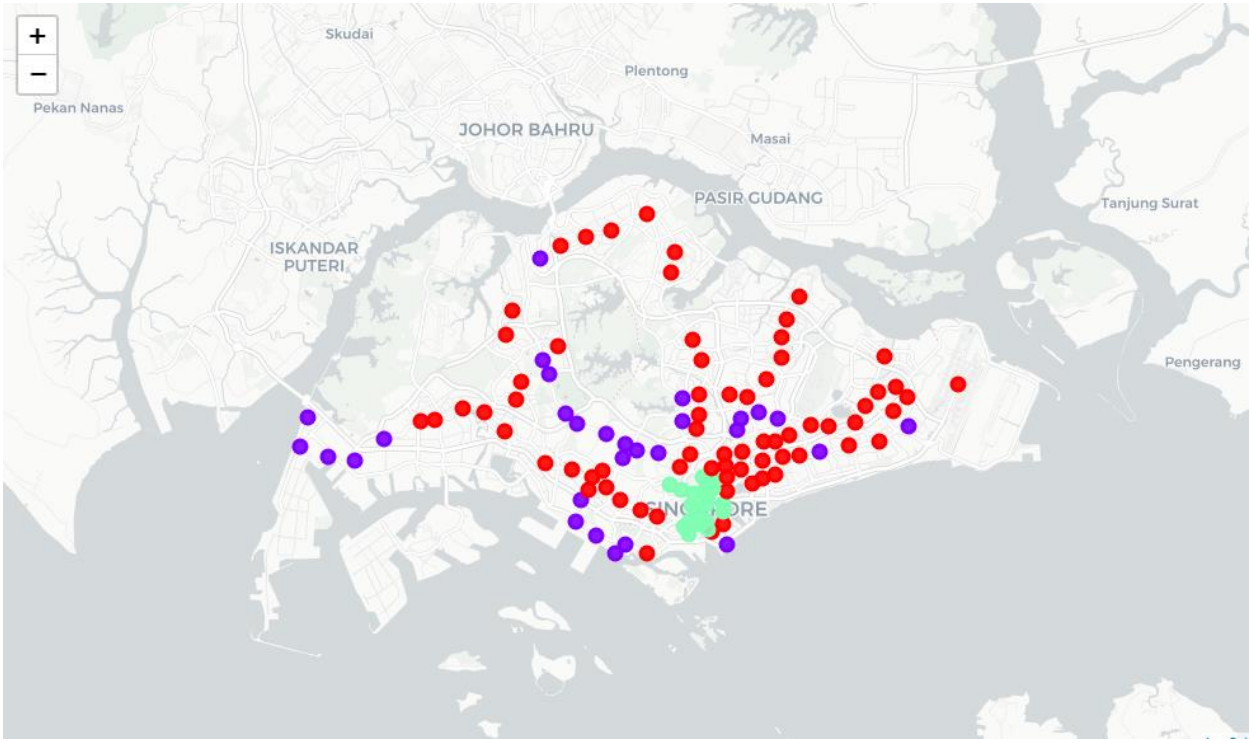



Respective colour coded markers are superimposed onto the Singapore map, where

**Red** = Cluster 0

**Purple** = Cluster 1

**Green** = Cluster 2



## 5. Result

From the previous section, we classified the MRT stations into 3 groups based on the venue's categories surrounding the stations provided by Foursquare API.

- Cluster 0 (Red): Having significant higher proportion of residential areas, and intermediate rating for rest of the categories.
- Cluster 1 (Purple): Appear to be a less developed area with limited number of places recorded in Foursquare.
- Cluster 2 (Green): Exhibit highest rating in almost all types of venues.

By visualizing the clusters on map, it's now making more sense that:

- Cluster 0: Belong to typical developed HDB estates (Singapore public housing), typically with one or more shopping malls.
- Cluster 1: Are mostly less populated areas and close to industrial parks or shipping ports.
- Cluster 2: Highly developed, fast-paced, Central Business District, with tourist attractions nearby.

## 6. Discussion

Despite Foursquare provide a comprehensive database to explore the surrounding of any point of interest, but the data may not be able to tell the whole insight about a community. Generally, most of the data are gathered from its users, and we are only able to rely on the sources from a specific group of people who uses Foursquare. Furthermore, due to the preceding reason, not all venues in the neighbourhood will be capture by Foursquare, as not everyone in Singapore uses the App, and a venue will only be recorded if a Foursquare user posted anything about the place.

## 7. Conclusion

In this project, we successfully clustered all the 119 MRT stations in Singapore into 3 groups:

1. Housing Estates
2. Sub-urban Industrial Zone
3. Central Business District (i.e. Downtown)

Back to the problem of XYZ Corporation, as most of the Singaporean work in the Central Business District (CBD) (Cluster 2) and live in the Housing Estates (Cluster 0). Therefore, one of the recommendations will be opening their snack bar business in the CBD targeting office workers in the morning peak, and the grocery stores at Housing Estate targeting the same group of people working in CBD who returning to their home at the evening.