IBM Applied Data Science capstone project

1. Introduction

The final module of the IBM Applied Data Science course on Coursera is a capstone project. For this project I have chosen to investigate the problem of where to set-up a food & beverage (F&B) outlet in Singapore.

Singapore is one of the most densely populated countries in the world. It is also one of the most expensive countries to buy a car. Therefore, over 5.5m Singaporeans rely on the excellent public transport system to get around. The railway network comprises of around 225 km of tracks and transports around 3.2 million people daily according to the Singapore Land Transport Authority.

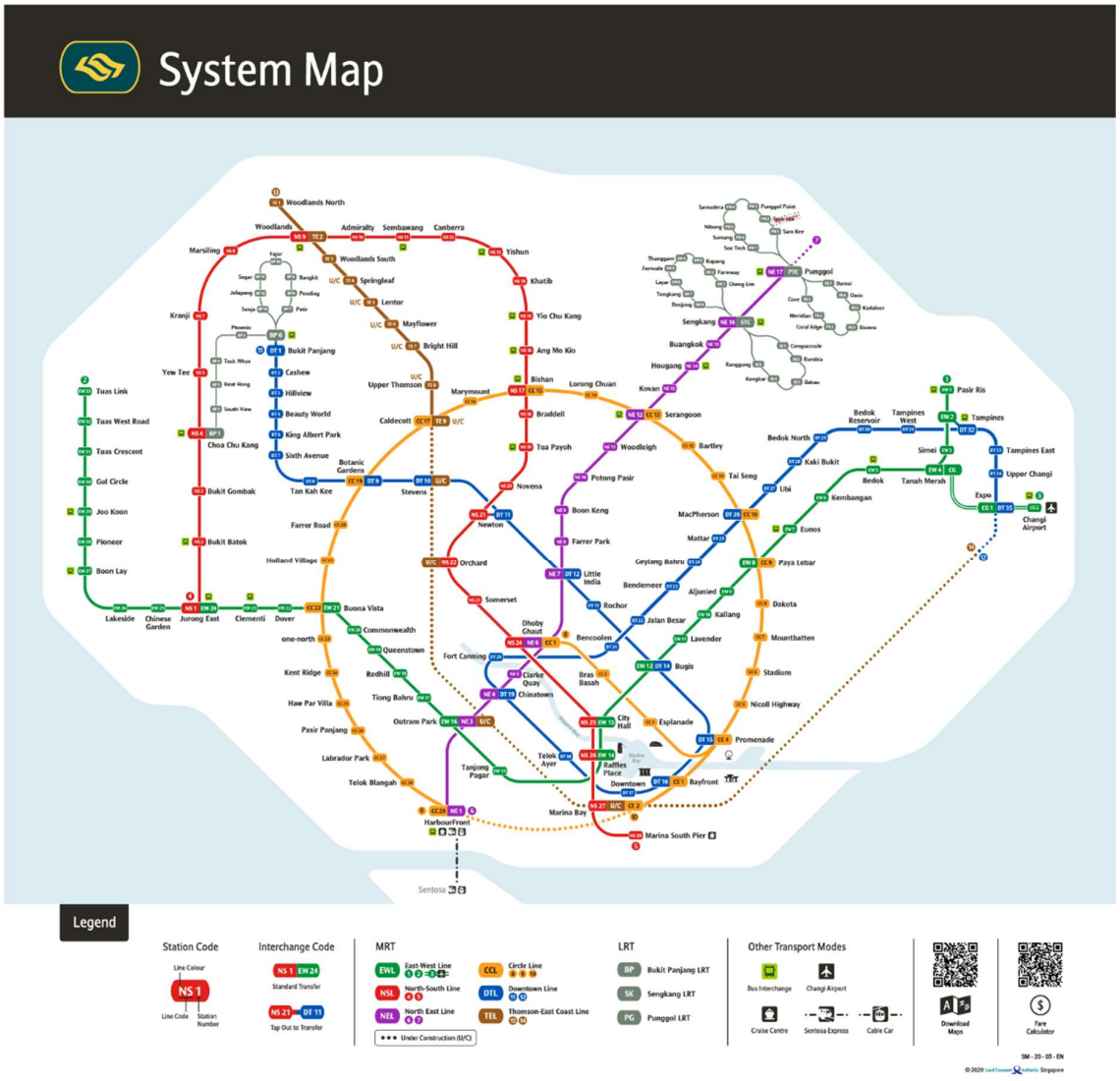

Figure 1. Singapore's rail network system.

All these commuters need to eat, and it is a well-known fact that Singaporeans are foodies, who frequently eat out. Restaurants, coffee shops, bars and smaller eateries come and go across the island. To cater for these commuters, it makes sense to set-up a business close to one of these

train stations. This report aims to provide some insights on the locations to consider for a F&B outlet.

2. Data

The analysis is based on the two sets of data:

- A .csv file created by Lee Yu Xuan available on Kaggle containing a list of 157 Singapore train stations names, their type (MRT/Metro Rail Transit or LRT/Light Rail Transit) and their GPS coordinates. The coordinates were manually collected by searching for every train station on Google Maps and copying their latitude and longitude. A sample of the dataset can be seen from the table below.

	station_name	type	lat	lng
0	Jurong East	MRT	1.333207	103.742308
1	Bukit Batok	MRT	1.349069	103.749596
2	Bukit Gombak	MRT	1.359043	103.751863
3	Choa Chu Kang	MRT	1.385417	103.744316
4	Yew Tee	MRT	1.397383	103.747523

- Information on F&B outlets and other venues located around these train stations obtained from Foursquare (www.foursquare.com) through an API.

3. Methodology

First the .csv file was loaded into a pandas data frame and reduced to the needed information by taking out the 'type' column. Using Folium the train stations locations were then visualised onto a map of Singapore as shown in Figure 2.

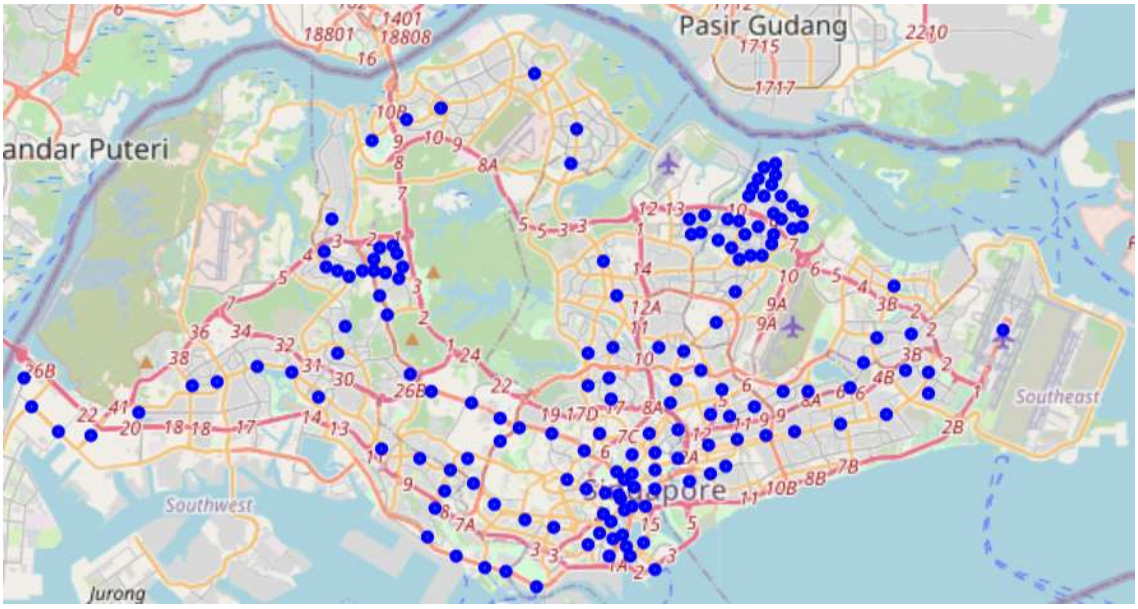

Figure 2. Singapore's MRT and LRT stations

In the next step venues within a 500m range of each train station were collected from Foursquare through an API.

These venues were then ranked in order of most common ones for each train station. An extract is shown in the table below.

	Station Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Admiralty	Japanese Restaurant	Shopping Mall	Café	Indian Restaurant	Electronics Store	Coffee Shop	Clothing Store	Fast Food Restaurant	Chinese Restaurant	Snack Place
1	Aljunied	Chinese Restaurant	Coffee Shop	Noodle House	Asian Restaurant	Vegetarian / Vegan Restaurant	Café	Dim Sum Restaurant	Seafood Restaurant	Food Court	Food Truck
2	Ang Mo Kio	Coffee Shop	Dessert Shop	Food Court	Bubble Tea Shop	Supermarket	Japanese Restaurant	Sandwich Place	Frozen Yogurt Shop	Fried Chicken Joint	Noodle House
3	Bakau	Bus Station	Shopping Mall	Indonesian Restaurant	Fast Food Restaurant	Breakfast Spot	Japanese Restaurant	Basketball Court	Supermarket	Trail	Sandwich Place
4	Bangkit	Food Court	Convenience Store	Noodle House	Fast Food Restaurant	Trail	Bike Trail	Miscellaneous Shop	Piano Bar	Coffee Shop	Supermarket

Subsequently one-hot encoding was used to convert the categorical venue values in the above table into numeric ones in order to be able to apply machine learning.

Using the unsupervised machine learning algorithm K-means, the train stations were then clustered into the optimal number of clusters. To determine this optimal number of clusters the Silhouette score was used. Figure 3 shows that the optimal number of clusters is five.

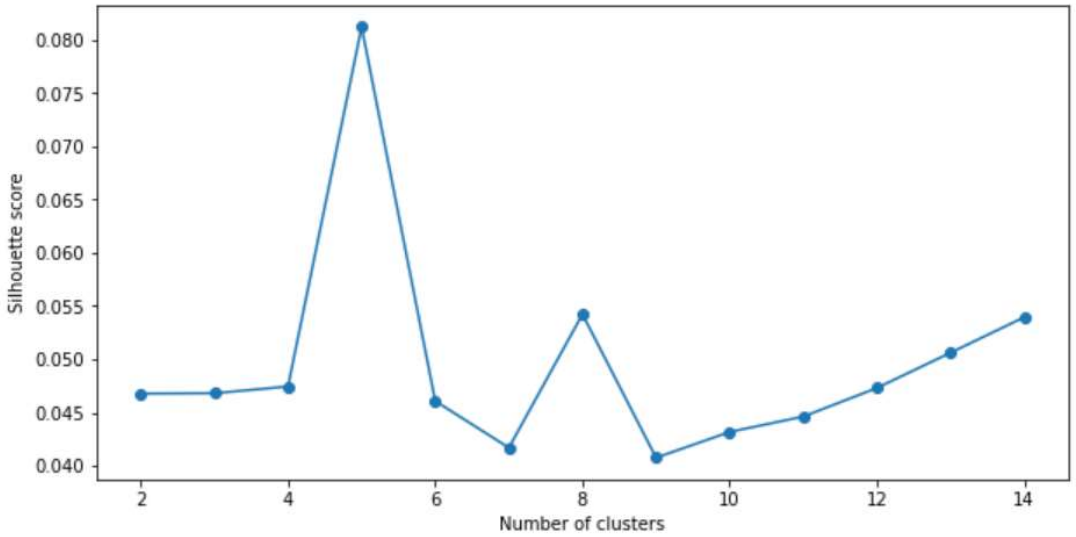

Figure 3. Silhouette score vs number of clusters

The five resulting clusters can be visualised using Folium as shown in figure 4.

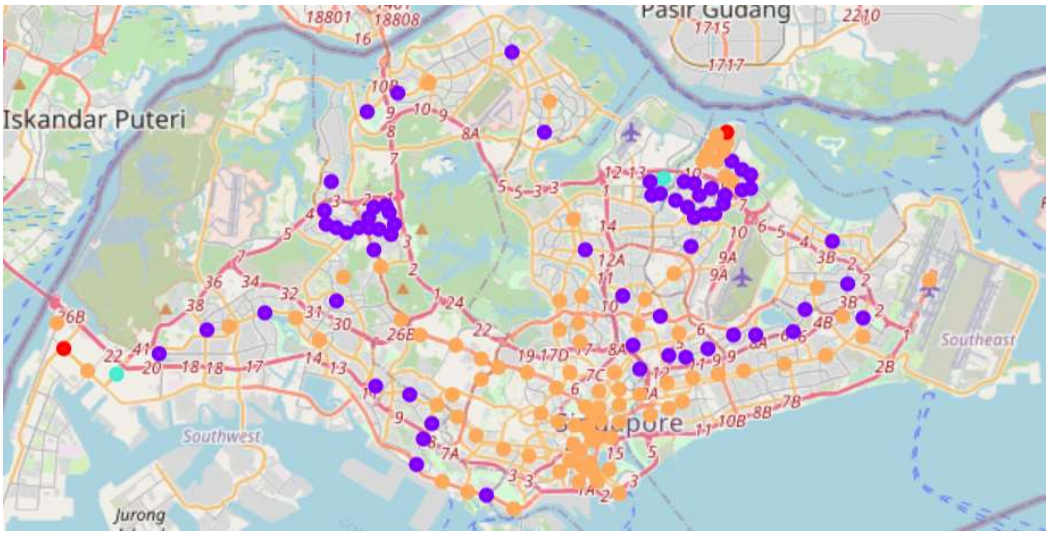

Figure 4. Singapore's train stations clusters

As Singapore is a rather small country, the analysis was repeated using a range of 200m instead of 500m to find nearby venues. The results proved to be the same.

4. Results

The resulting 5 clusters can be described as follows:

Cluster 1 (red – North) where remote venues are most common followed by some F&B outlets.

	Station Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
152	Punggol Point	Light Rail Station	Campground	Yunnan Restaurant	Fishing Spot	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	Fish & Chips Shop	Fish Market

Cluster 2 (dark blue) where "quick & easy" F&B outlets, like cafés, coffee shops, food courts and fast-food restaurants are most common. These are mainly located in the more residential areas.

	Station Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Bukit Batok	Coffee Shop	Fast Food Restaurant	Food Court	Malay Restaurant	Chinese Restaurant	Mobile Phone Shop	Sandwich Place	Shopping Mall	Bowling Alley	Café
3	Choa Chu Kang	Café	Fast Food Restaurant	Coffee Shop	Portuguese Restaurant	Bakery	Bus Line	Food & Drink Shop	Bubble Tea Shop	Supermarket	Noodle House
4	Yew Tee	Fast Food Restaurant	Cosmetics Shop	Pool	Café	Sandwich Place	Diner	Playground	Food Court	Japanese Restaurant	Shopping Mall
5	Kranji	Racetrack	Night Market	Noodle House	Jazz Club	Bus Line	Bus Station	Go Kart Track	Café	Flea Market	Field
6	Marsiling	Food Court	Bus Stop	Grocery Store	Market	Flower Shop	Fast Food Restaurant	Hainan Restaurant	Paintball Field	Halal Restaurant	Coffee Shop
...	...	...	...	...	...	...	...	...	...	...	...
145	Coral Edge	Park	Bus Stop	Soccer Field	Café	College Cafeteria	Track	Fast Food Restaurant	Scenic Lookout	Bookstore	Beer Garden
146	Riviera	Gastropub	Fast Food Restaurant	Fishing Spot	Golf Course	Bus Stop	Soccer Field	Café	Seafood Restaurant	Scenic Lookout	Beer Garden
147	Kadaloor	Golf Course	Music Venue	Fishing Spot	Park	Fast Food Restaurant	Gym	Track	Light Rail Station	Other Great Outdoors	Café
148	Oasis	Fast Food Restaurant	Supermarket	Café	Coffee Shop	Bridge	Steakhouse	Japanese Restaurant	Food Court	Plaza	Playground
149	Damai	Music Venue	High School	Basketball Court	Grocery Store	Food Truck	Food Stand	Fried Chicken Joint	Farm	Farmers Market	Fast Food Restaurant

Cluster 3 (light blue) where “quick & easy” F&B outlets in more “greener” areas (parks, fields and farms) are most common.

	Station Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
29	Gul Circle	Coffee Shop	Park	Electronics Store	Exhibit	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	Fish & Chips Shop	Fish Market
137	Kupang	Coffee Shop	Bridge	Park	Pizza Place	Farm	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	Fish & Chips Shop

Cluster 4 (orange) where (non-fast food) restaurants are most common. These are mainly located in the Central Business District and surrounding areas.

	Station Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Jurong East	Japanese Restaurant	Coffee Shop	Chinese Restaurant	Café	Shopping Mall	Food Court	Bistro	Sandwich Place	Bubble Tea Shop	Sushi Restaurant
2	Bukit Gombak	Food Court	Vegetarian / Vegan Restaurant	Stadium	Coffee Shop	Chinese Restaurant	Halal Restaurant	Train Station	Fast Food Restaurant	Steakhouse	Flea Market
7	Woodlands	Japanese Restaurant	Shopping Mall	Café	Indian Restaurant	Electronics Store	Coffee Shop	Clothing Store	Fast Food Restaurant	Chinese Restaurant	Snack Place
8	Admiralty	Japanese Restaurant	Shopping Mall	Café	Indian Restaurant	Electronics Store	Coffee Shop	Clothing Store	Fast Food Restaurant	Chinese Restaurant	Snack Place
10	Yishun	Food Court	Chinese Restaurant	Coffee Shop	Pharmacy	Arcade	Supermarket	Fried Chicken Joint	Italian Restaurant	Hainan Restaurant	Furniture / Home Store
...	...	...	...	...	...	...	...	...	...	...	...
151	Teck Lee	Light Rail Station	Thai Restaurant	Café	Bridge	Music Venue	Yoga Studio	Steakhouse	Pool	Fast Food Restaurant	Spa
153	Samudera	Soccer Field	Bar	Beer Garden	Beer Bar	Light Rail Station	Club House	Harbor / Marina	Surf Spot	Farm	Sporting Goods Shop
154	Nibong	Light Rail Station	Soccer Field	Bar	Beer Garden	BBQ Joint	Market	Scenic Lookout	Beer Bar	Club House	Coffee Shop
155	Sumang	Japanese Restaurant	Coffee Shop	Food Court	Chinese Restaurant	Fast Food Restaurant	Toy / Game Store	Clothing Store	Sandwich Place	Market	Seafood Restaurant
156	Soo Teck	Ice Cream Shop	Playground	Bridge	Japanese Restaurant	Food Court	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	Fish & Chips Shop

Cluster 5 (red - West) where restaurants in a more remote setting are most common.

	Station Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
27	Tuas West Road	Malay Restaurant	Yunnan Restaurant	Fishing Spot	Farm	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	Fish & Chips Shop	Fish Market

The results show two distinct F&B outlet clusters. In the dark-blue cluster the “quick & easy” F&B outlets like coffee shops and fast-food restaurants are more prevalent, whereas in the orange cluster restaurants are more common.

5. Discussion

Depending on the type of F&B business you are looking to set-up, these clusters can be further analysed to provide more detailed information. For example, the restaurant cluster could be further analysed based on what type of cuisine they offer.

Note that the Singapore MRT/LRT lines will be expanded in future and this analysis could be updated with future train stations.

6. Conclusion

This report can be used as an initial step in understanding how the different types of F&B businesses are clustered close to Singapore metro stations. This will help future F&B business owners make more informed decisions on where to set up their businesses.

7. Reference

The Jupyter notebook can be found on Github: guschreu75/Coursera_Capstone: IBM Data Science course 9 (capstone project) (github.com)