IBM Applied Data Science Capstone Project

Introduction

▶ Singapore is densely populated and one of the most expensive countries to buy a car. Therefor Singaporeans rely on using the excellent public transport system to get around.

▶ The rail network consists of 225km of tracks and transports 3.2m people daily according to the Singapore Land Transport Authority

▶ Singaporeans are foodies and love to eat out.

▶ For those interested in opening up a food & beverage (F&B) outlet to cater for these travelers, a location close to a train station makes sense. But which train stations to consider ?
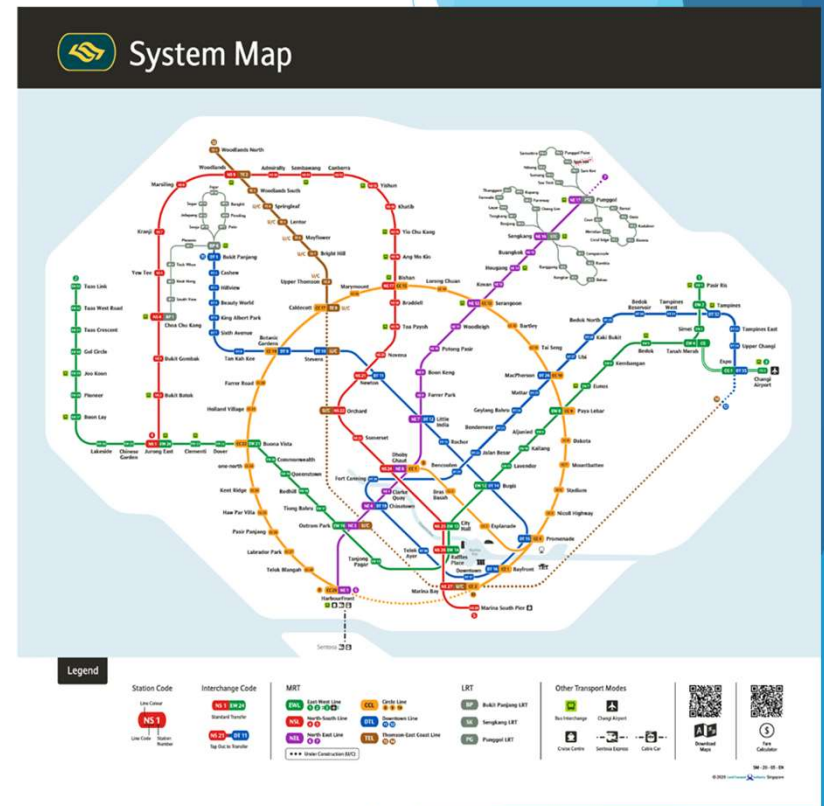
Data

▶ **Train station coordinates:** available in a csv file obtained from Kaggle. Original data was obtained by manually entering the train stations name in Google maps and copying the longitude and latitude.

▶ **Nearby venues:** obtained form Foursquare (www.foursquare.com) through an API

Methodology

1. CSV file was uploaded into a pandas dataframe and cleaned
2. Using Folium the train stations were plotted onto a Singapore map

Methodology

3. Venues within a range of 500m where obtained from Foursquare through an API.

4. These venues were ranked in order of most common ones for each train station as shown 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

Methodology

5. One-hot encoding was used to replace the categorical values in the previous table with numerical ones.

6. Using the Silhouette score the optimal number of clusters was determined for K-means clustering. The optimal number of clusters is 5.
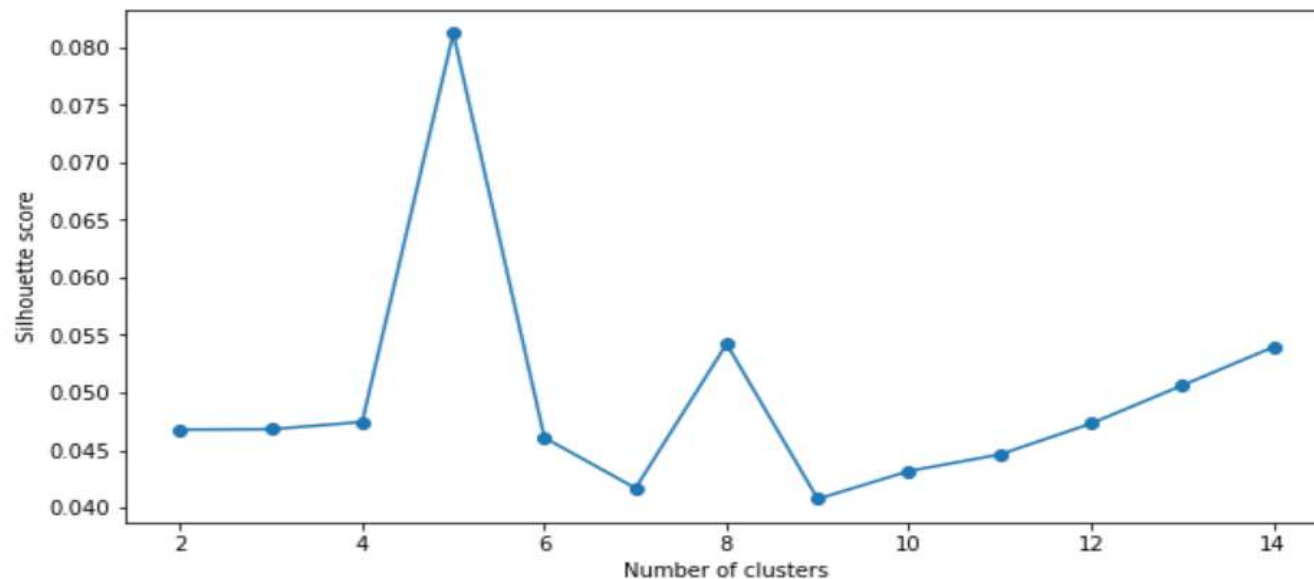
Methodology

7. The 5 resulting clusters were visualised on a Singapore map using Folium.
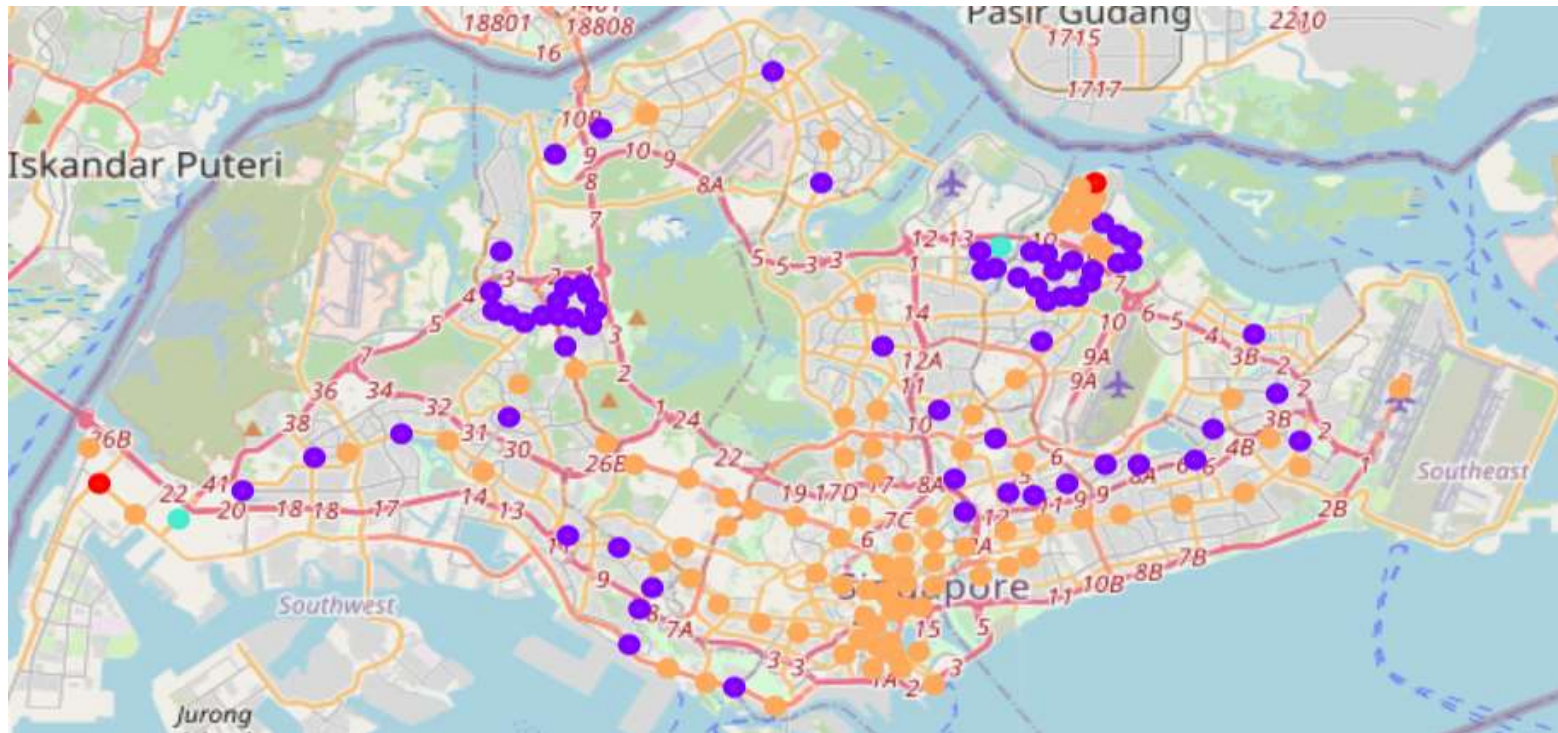
Results

The clusters of interest are cluster 2 and 4

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

▶ 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.

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

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

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

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.

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.

References

▶ The Jupyter notebook can be found on Github: [IBM applied data science capstone project.ipynb (github.com)](github.com)