THE BATTLE OF NEIGHBORHOODS

APPLIED DATA SCIENCE CAPSTONE – ASSIGNMENT

EXPLORING BUSINESS OPPORTUNITY AROUND TRAIN STATIONS'
NEIGHBORHOOD IN SINGAPORE

BACKGROUND INFO

- Singapore railway system is made up of five Mass Rapid Transit (MRT) lines and two Light Rapid Transit (LRT) lines. The MRT & LRT 230km system has over three million daily ridership.
- There are a total of 187 Mass Rapid Transit (MRT)/Light Rapid Transit (LRT) train stations as of Jan 2019. In general, the neighborhoods around MRT/LRT stations are highly populated and many shopping centers, shops, business offices are found near these stations.

BUSINESS PROBLEM

- To find out the existing common business or shops in the neighborhood around MRT/LRT stations and to explore business opportunity based on the analysis of the MRT/LRT trains' neighborhood data.

TARGET AUDIENCE

- People who are interested in opening a shop for certain business of their interest at location near any of the MRT/LRT stations in Singapore

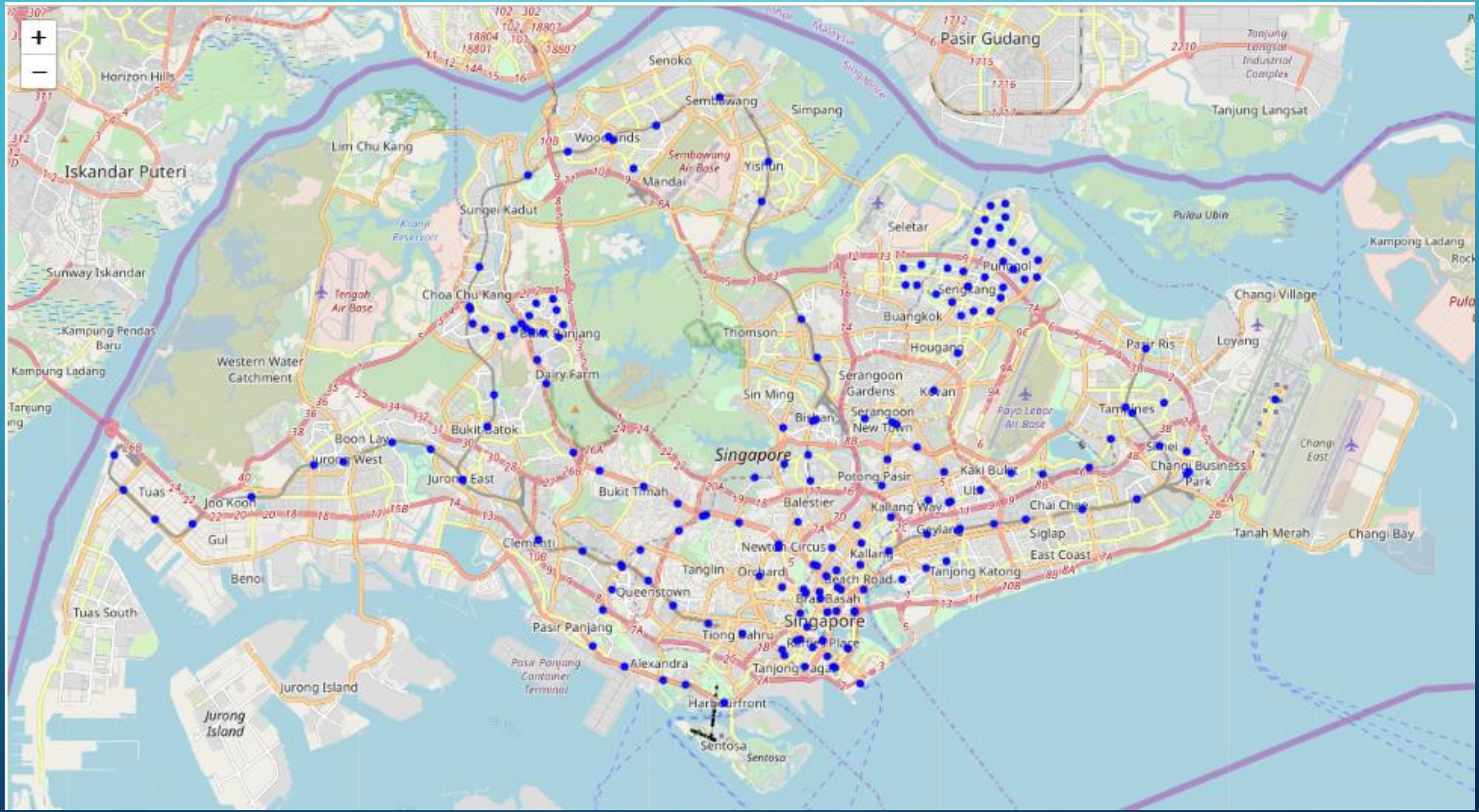
DATA

The following data are used:

- Location data (latitude & longitude coordinates) of all MRT & LRT stations in Singapore
  - Source: https://data.world/hxchua/train-stations-in-singapore/workspace/file? filename=mrtsg.csv
- Latitude and longitude data of Singapore
  - Obtained by Python's geopy library (Nominatin class)
- Foursquare's Places data about the neighborhood venues around MRT/LRT stations in Singapore
  - Foursquare Paces Data API
  - Foursquare company's URL: www.foursquare.com
  - Foursquare API documentation: https://developer.foursquare.com/docs

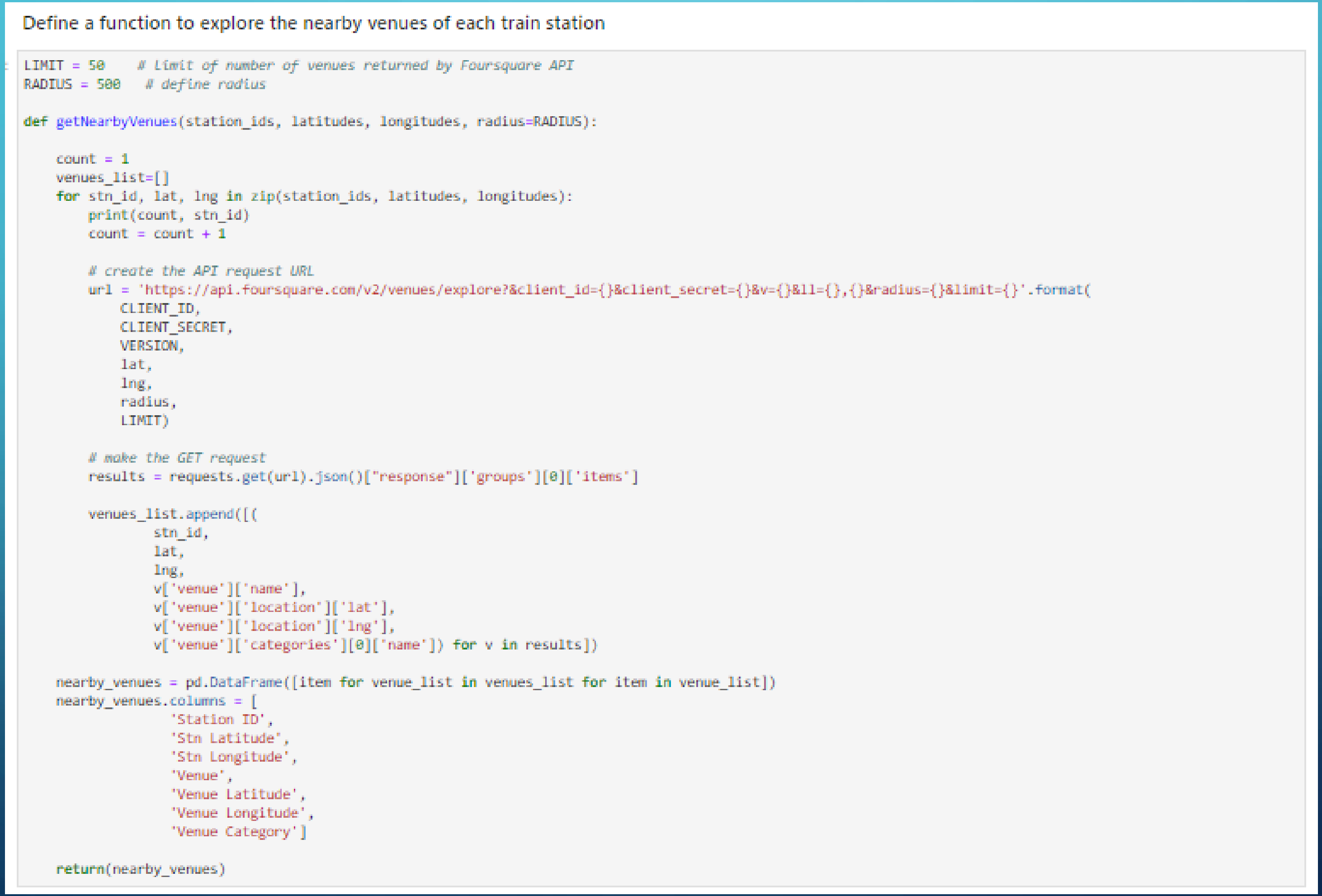
DATA – SINGAPORE MAP & TRAIN STATION LOCATIONS

- Generating map of Singapore with MRT/LRT station markers display



DATA – NEARBY VENUES OF TRAIN STATIONS

- Using Foursquare Places Data APIs to identify the popular venues around each MRT/LRT station and the associated venue category of each venue

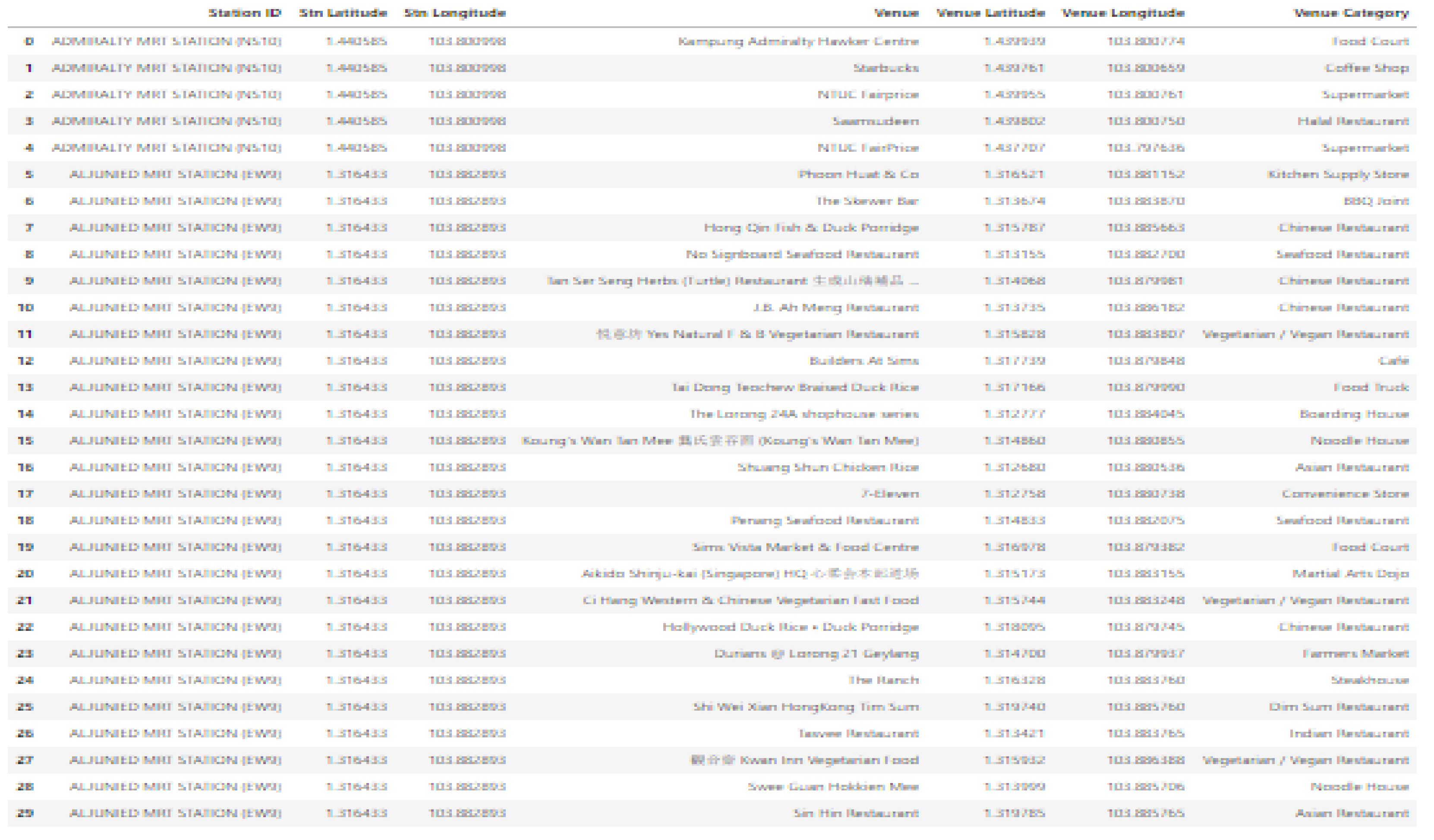


Define a function to explore the nearby venues of each train station

```
LIMIT = 50    # Limit of number of venues returned by Foursquare API
RADIUS = 500   # define radius

def getNearbyVenues(station_ids, latitudes, longitudes, radius=RADIUS):

    count = 1
    venues_list=[]
    for stn_id, lat, lng in zip(station_ids, latitudes, longitudes):
        print(count, stn_id)
        count = count + 1

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        venues_list.append([(
                stn_id,
                lat,
                lng,
                v['venue']['name'],
                v['venue']['location']['lat'],
                v['venue']['location']['lng'],
                v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = [
                  'Station ID',
                  'Stn Latitude',
                  'Stn Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

DATA – VENUE & VENUE CATEGORY DATA

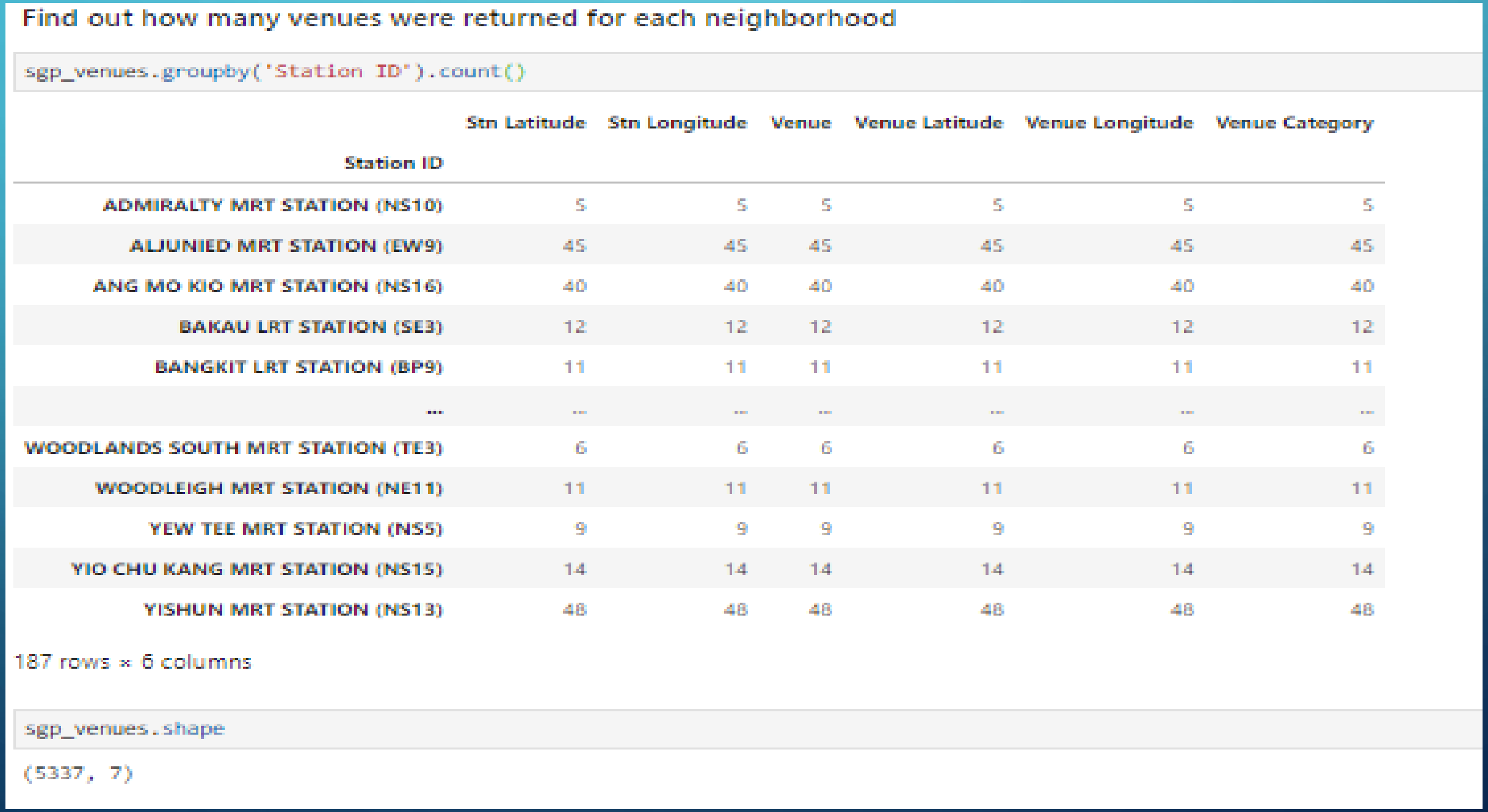
- The venue, venue latitude & longitude as well as venue category obtained via making Foursquare Places Data API calls are tabulated



	Station ID	Stn Latitude	Stn Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ADMIRALTY MRT STATION (NS10)	1.440585	103.800998	Kampung Admiralty Hawker Centre	1.439939	103.800774	Food Court
1	ADMIRALTY MRT STATION (NS10)	1.440585	103.800998	Starbucks	1.439761	103.800659	Coffee Shop
2	ADMIRALTY MRT STATION (NS10)	1.440585	103.800998	NTUC Fairprice	1.439955	103.800761	Supermarket
3	ADMIRALTY MRT STATION (NS10)	1.440585	103.800998	Saamudeen	1.439802	103.800750	Halal Restaurant
4	ADMIRALTY MRT STATION (NS10)	1.440585	103.800998	NTUC FairPrice	1.437707	103.797636	Supermarket
5	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Phoon Huat & Co	1.316521	103.881152	Kitchen Supply Store
6	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	The Skewer Bar	1.313674	103.883870	BBQ Joint
7	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Hong Qin Fish & Duck Porridge	1.315787	103.885663	Chinese Restaurant
8	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	No Signboard Seafood Restaurant	1.313155	103.882700	Seafood Restaurant
9	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Tan Ser Seng Herbs (Turtle) Restaurant 陈成山龟補品 ...	1.314068	103.879981	Chinese Restaurant
10	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	J.B. Ah Meng Restaurant	1.313735	103.886182	Chinese Restaurant
11	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	悦意坊 Yes Natural F & B Vegetarian Restaurant	1.315828	103.883807	Vegetarian / Vegan Restaurant
12	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Builders At Sims	1.317739	103.879848	Café
13	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Tai Dong Teochew Braised Duck Rice	1.317166	103.879990	Food Truck
14	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	The Lorong 24A shophouse series	1.312777	103.884045	Boarding House
15	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Koung's Wan Tan Mee 龔氏雲吞麵 (Koung's Wan Tan Mee)	1.314860	103.880855	Noodle House
16	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Shuang Shun Chicken Rice	1.312680	103.880536	Asian Restaurant
17	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	7-Eleven	1.312758	103.880738	Convenience Store
18	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Penang Seafood Restaurant	1.314833	103.882075	Seafood Restaurant
19	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Sims Vista Market & Food Centre	1.316978	103.879382	Food Court
20	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Aikido Shinju-kai (Singapore) HQ 心寿会本部道场	1.315173	103.883155	Martial Arts Dojo
21	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Ci Hang Western & Chinese Vegetarian Fast Food	1.315744	103.883248	Vegetarian / Vegan Restaurant
22	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Hollywood Duck Rice • Duck Porridge	1.318095	103.879745	Chinese Restaurant
23	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Durians @ Lorong 21 Geylang	1.314700	103.879937	Farmers Market
24	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	The Ranch	1.316328	103.883760	Steakhouse
25	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Shi Wei Xian HongKong Tim Sum	1.319740	103.885760	Dim Sum Restaurant
26	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Taswee Restaurant	1.313421	103.883765	Indian Restaurant
27	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	觀音齋 Kwan Inn Vegetarian Food	1.315932	103.886388	Vegetarian / Vegan Restaurant
28	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Swee Guan Hokkien Mee	1.313999	103.885706	Noodle House
29	ALJUNIED MRT STATION (EW9)	1.316433	103.882893	Sin Hin Restaurant	1.319785	103.885765	Asian Restaurant

METHODOLOGY – EXPLORATORY ANALYSIS

- A total of 5337 venues retrieved and the number of venues for each station is tabulated
- A total of 308 unique venue categories.



Find out how many venues were returned for each neighborhood

```
sgp_venues.groupby('Station ID').count()
```

	Stn Latitude	Stn Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Station ID						
ADMIRALTY MRT STATION (NS10)	5	5	5	5	5	5
ALJUNIED MRT STATION (EW9)	45	45	45	45	45	45
ANG MO KIO MRT STATION (NS16)	40	40	40	40	40	40
BAKAU LRT STATION (SE3)	12	12	12	12	12	12
BANGKIT LRT STATION (BP9)	11	11	11	11	11	11
...	...	...	...	...	...	...
WOODLANDS SOUTH MRT STATION (TE3)	6	6	6	6	6	6
WOODLEIGH MRT STATION (NE11)	11	11	11	11	11	11
YEW TEE MRT STATION (NS5)	9	9	9	9	9	9
YIO CHU KANG MRT STATION (NS15)	14	14	14	14	14	14
YISHUN MRT STATION (NS13)	48	48	48	48	48	48

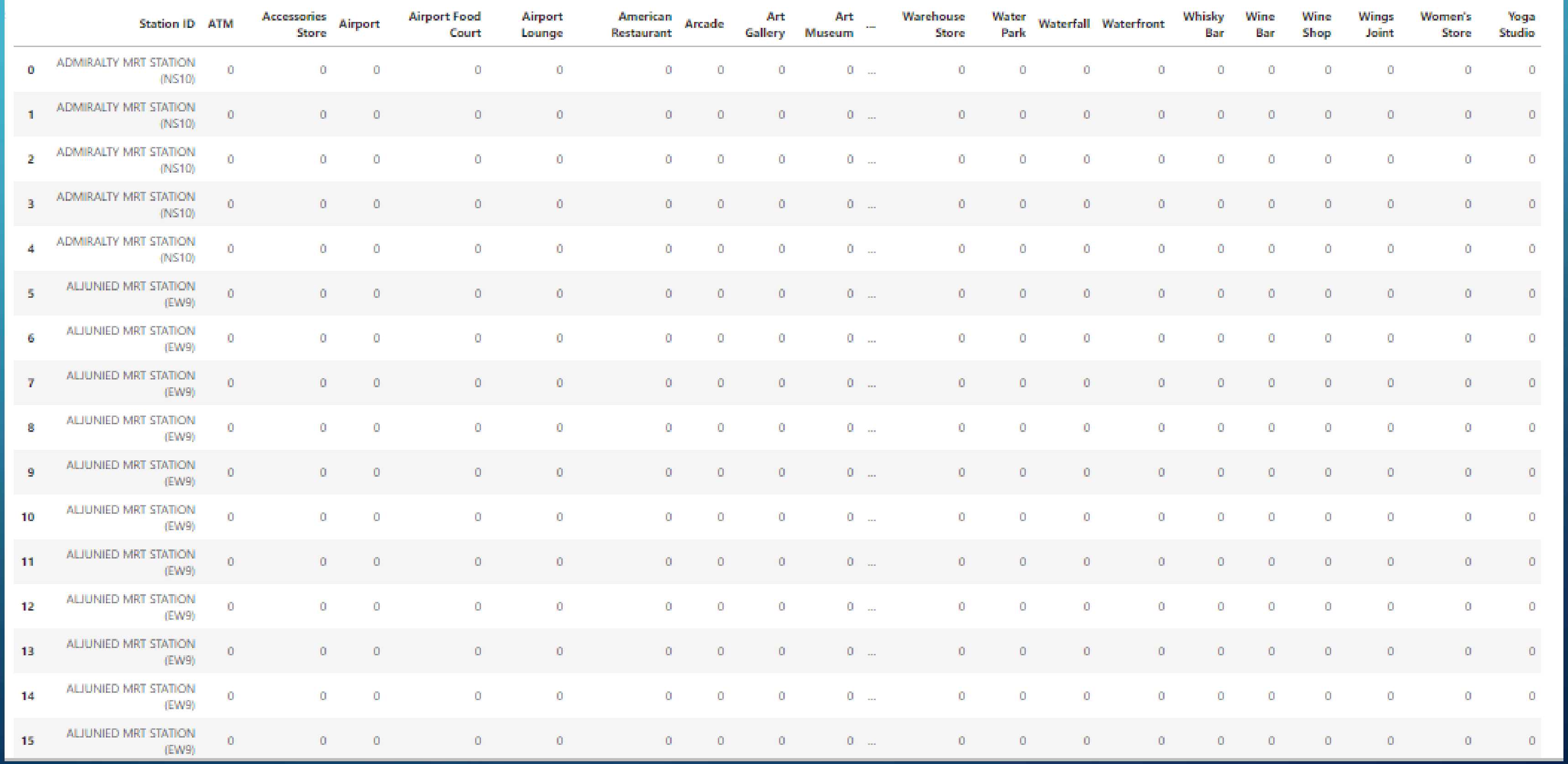
187 rows × 6 columns

```
sgp_venues.shape
```

```
(5337, 7)
```

METHODOLOGY – CLUSTER ANALYSIS

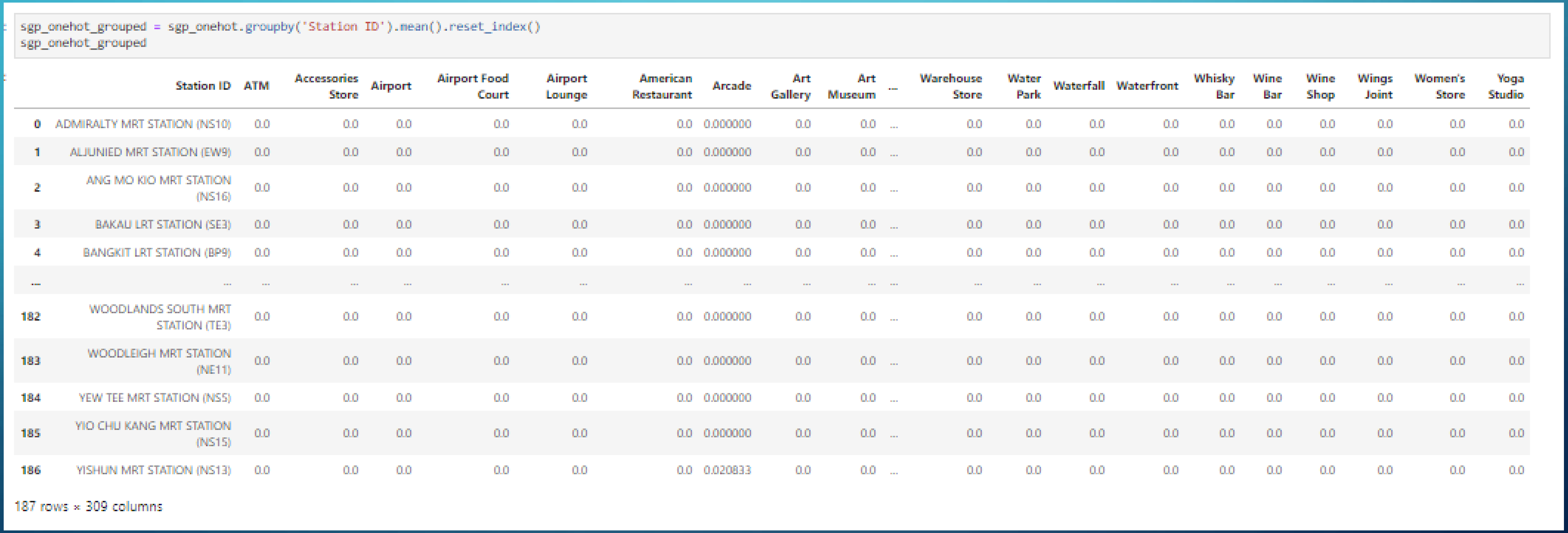
- The objective is to find out the top 5 most common venue categories of each MRT/LRT station's neighborhood.
- Firstly, applying the one hot encoding



	Station ID	ATM	Accessories Store	Airport	Airport Food Court	Airport Lounge	American Restaurant	Arcade	Art Gallery	Art Museum	...	Warehouse Store	Water Park	Waterfall	Waterfront	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
0	ADMIRALTY MRT STATION (NS10)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	ADMIRALTY MRT STATION (NS10)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	ADMIRALTY MRT STATION (NS10)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	ADMIRALTY MRT STATION (NS10)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	ADMIRALTY MRT STATION (NS10)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
5	ALJUNIED MRT STATION (EW9)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
6	ALJUNIED MRT STATION (EW9)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
7	ALJUNIED MRT STATION (EW9)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
8	ALJUNIED MRT STATION (EW9)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
9	ALJUNIED MRT STATION (EW9)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
10	ALJUNIED MRT STATION (EW9)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
11	ALJUNIED MRT STATION (EW9)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
12	ALJUNIED MRT STATION (EW9)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
13	ALJUNIED MRT STATION (EW9)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
14	ALJUNIED MRT STATION (EW9)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
15	ALJUNIED MRT STATION (EW9)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

METHODOLOGY – CLUSTER ANALYSIS

- Next, grouping each row of the one hot encoding table by Station ID and then taking mean of the frequency of occurrence for each venue category



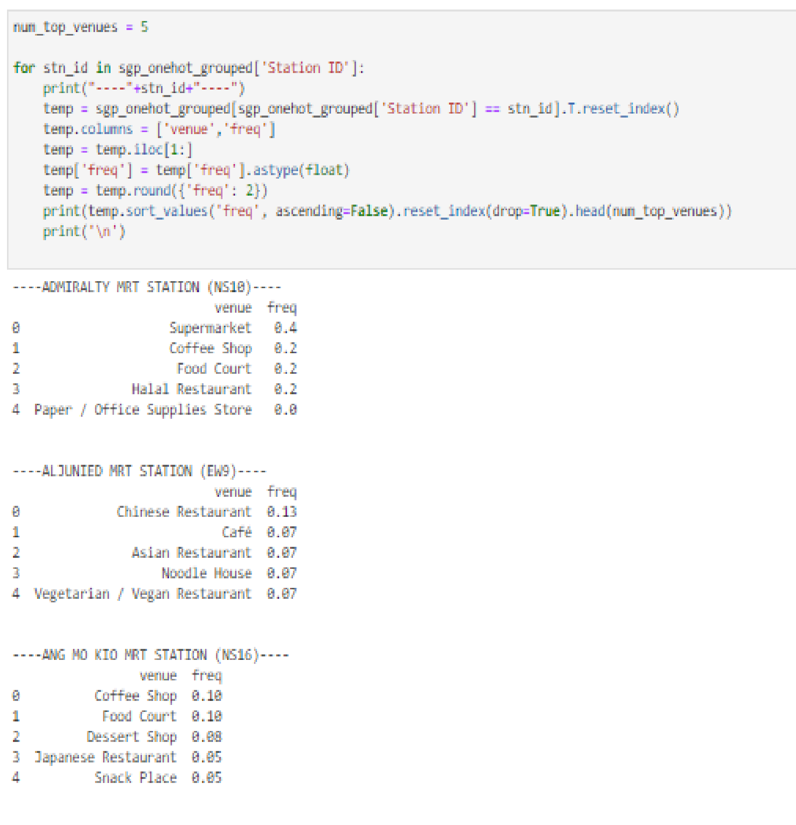
```
sgp_onehot_grouped = sgp_onehot.groupby('Station ID').mean().reset_index()
sgp_onehot_grouped
```

	Station ID	ATM	Accessories Store	Airport	Airport Food Court	Airport Lounge	American Restaurant	Arcade	Art Gallery	Art Museum	...	Warehouse Store	Water Park	Waterfall	Waterfront	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
0	ADMIRALTY MRT STATION (NS10)	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	ALJUNIED MRT STATION (EW9)	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	ANG MO KIO MRT STATION (NS16)	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	BAKAU LRT STATION (SE3)	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	BANGKIT LRT STATION (BP9)	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
182	WOODLANDS SOUTH MRT STATION (TE3)	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
183	WOODLEIGH MRT STATION (NE11)	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
184	YEW TEE MRT STATION (NS5)	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
185	YIO CHU KANG MRT STATION (NS15)	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
186	YISHUN MRT STATION (NS13)	0.0	0.0	0.0	0.0	0.0	0.0	0.020833	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

187 rows × 309 columns

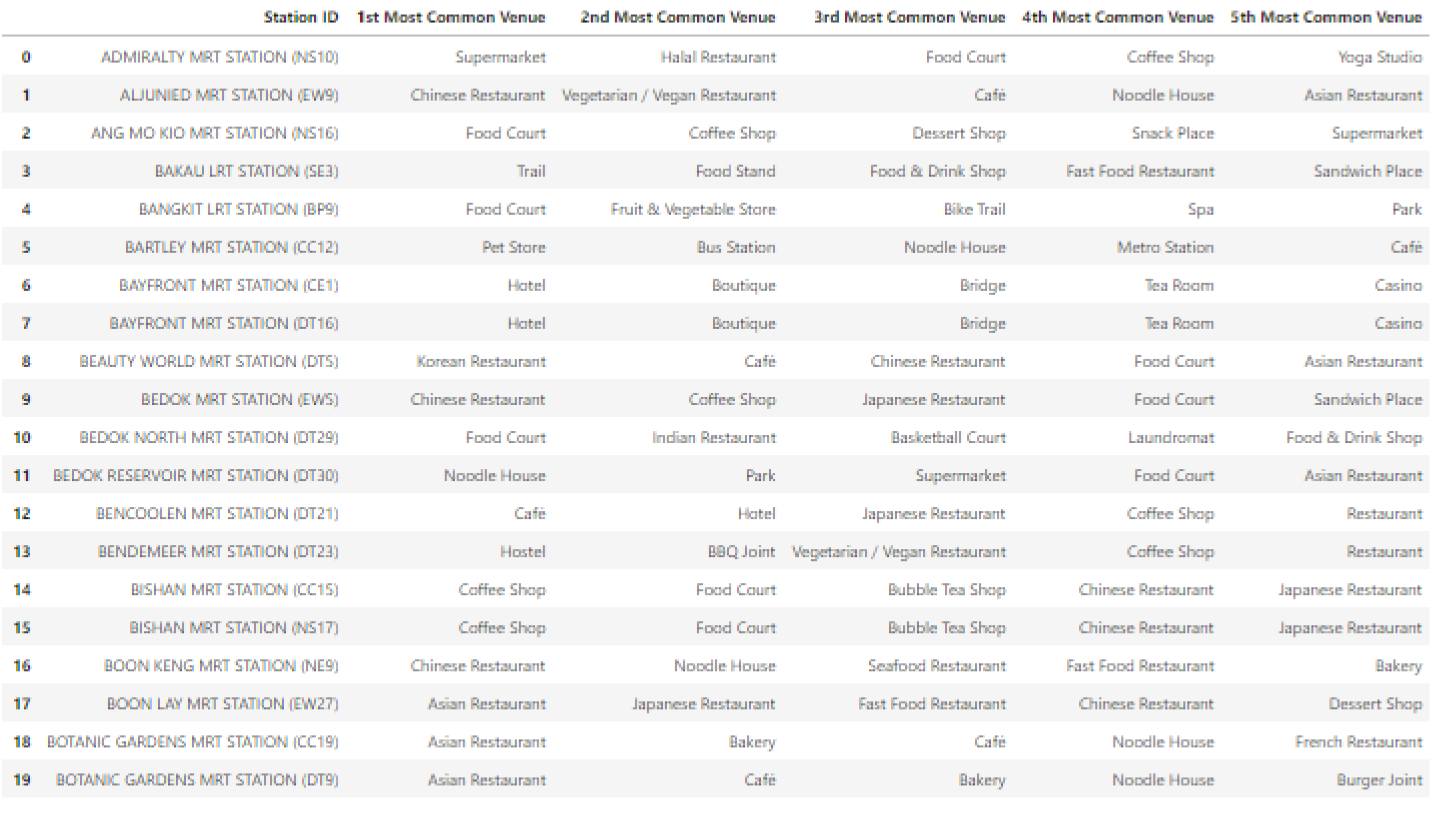
METHODOLOGY – CLUSTER ANALYSIS

- Print out the top 5 most common venue categories for each station and keep the result in a data frame



```
num_top_venues = 5

for stn_id in sgp_onehot_grouped['Station ID']:
    print("----"+stn_id+"----")
    temp = sgp_onehot_grouped[sgp_onehot_grouped['Station ID'] == stn_id].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

```
----ADMIRALTY MRT STATION (NS10)----
                         venue  freq
0                  Supermarket   0.4
1                  Coffee Shop   0.2
2                   Food Court   0.2
3              Halal Restaurant   0.2
4  Paper / Office Supplies Store   0.0


----ALJUNIED MRT STATION (EW9)----
                         venue  freq
0           Chinese Restaurant  0.13
1                         Café  0.07
2             Asian Restaurant  0.07
3                 Noodle House  0.07
4  Vegetarian / Vegan Restaurant  0.07


----ANG MO KIO MRT STATION (NS16)----
                venue  freq
0         Coffee Shop  0.10
1          Food Court  0.10
2        Dessert Shop  0.08
3  Japanese Restaurant  0.05
4         Snack Place  0.05
```



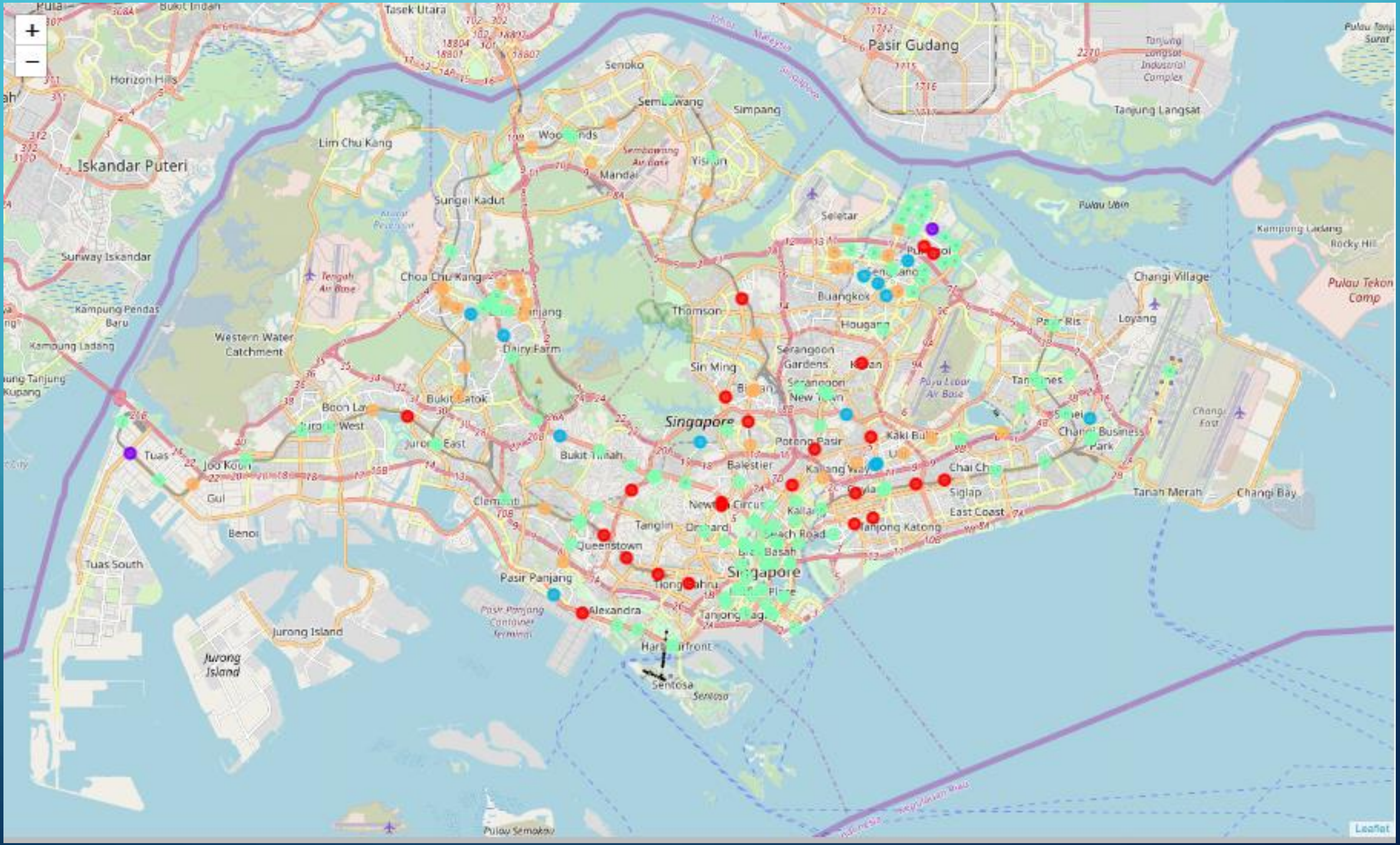
	Station ID	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	ADMIRALTY MRT STATION (NS10)	Supermarket	Halal Restaurant	Food Court	Coffee Shop	Yoga Studio
1	ALJUNIED MRT STATION (EW9)	Chinese Restaurant	Vegetarian / Vegan Restaurant	Café	Noodle House	Asian Restaurant
2	ANG MO KIO MRT STATION (NS16)	Food Court	Coffee Shop	Dessert Shop	Snack Place	Supermarket
3	BAKAU LRT STATION (SE3)	Trail	Food Stand	Food & Drink Shop	Fast Food Restaurant	Sandwich Place
4	BANGKIT LRT STATION (BP9)	Food Court	Fruit & Vegetable Store	Bike Trail	Spa	Park
5	BARTLEY MRT STATION (CC12)	Pet Store	Bus Station	Noodle House	Metro Station	Café
6	BAYFRONT MRT STATION (CE1)	Hotel	Boutique	Bridge	Tea Room	Casino
7	BAYFRONT MRT STATION (DT16)	Hotel	Boutique	Bridge	Tea Room	Casino
8	BEAUTY WORLD MRT STATION (DT5)	Korean Restaurant	Café	Chinese Restaurant	Food Court	Asian Restaurant
9	BEDOK MRT STATION (EW5)	Chinese Restaurant	Coffee Shop	Japanese Restaurant	Food Court	Sandwich Place
10	BEDOK NORTH MRT STATION (DT29)	Food Court	Indian Restaurant	Basketball Court	Laundromat	Food & Drink Shop
11	BEDOK RESERVOIR MRT STATION (DT30)	Noodle House	Park	Supermarket	Food Court	Asian Restaurant
12	BENCOOLEN MRT STATION (DT21)	Café	Hotel	Japanese Restaurant	Coffee Shop	Restaurant
13	BENDEMEER MRT STATION (DT23)	Hostel	BBQ Joint	Vegetarian / Vegan Restaurant	Coffee Shop	Restaurant
14	BISHAN MRT STATION (CC15)	Coffee Shop	Food Court	Bubble Tea Shop	Chinese Restaurant	Japanese Restaurant
15	BISHAN MRT STATION (NS17)	Coffee Shop	Food Court	Bubble Tea Shop	Chinese Restaurant	Japanese Restaurant
16	BOON KENG MRT STATION (NE9)	Chinese Restaurant	Noodle House	Seafood Restaurant	Fast Food Restaurant	Bakery
17	BOON LAY MRT STATION (EW27)	Asian Restaurant	Japanese Restaurant	Fast Food Restaurant	Chinese Restaurant	Dessert Shop
18	BOTANIC GARDENS MRT STATION (CC19)	Asian Restaurant	Bakery	Café	Noodle House	French Restaurant
19	BOTANIC GARDENS MRT STATION (DT9)	Asian Restaurant	Café	Bakery	Noodle House	Burger Joint

METHODOLOGY – CLUSTER ANALYSIS

- With the MRT/LRT station's 5 top most common venue categories data obtained in previous steps, the following two clustering algorithms are applied to the data:
    1. k-Means Clustering
    2. Hierarchical Clustering
- For the clusters generated, each cluster will be examined to determine the discriminating venue categories that distinguish each cluster.

RESULTS

- Clusters obtained via k-Means Clustering algorithm, with k=5



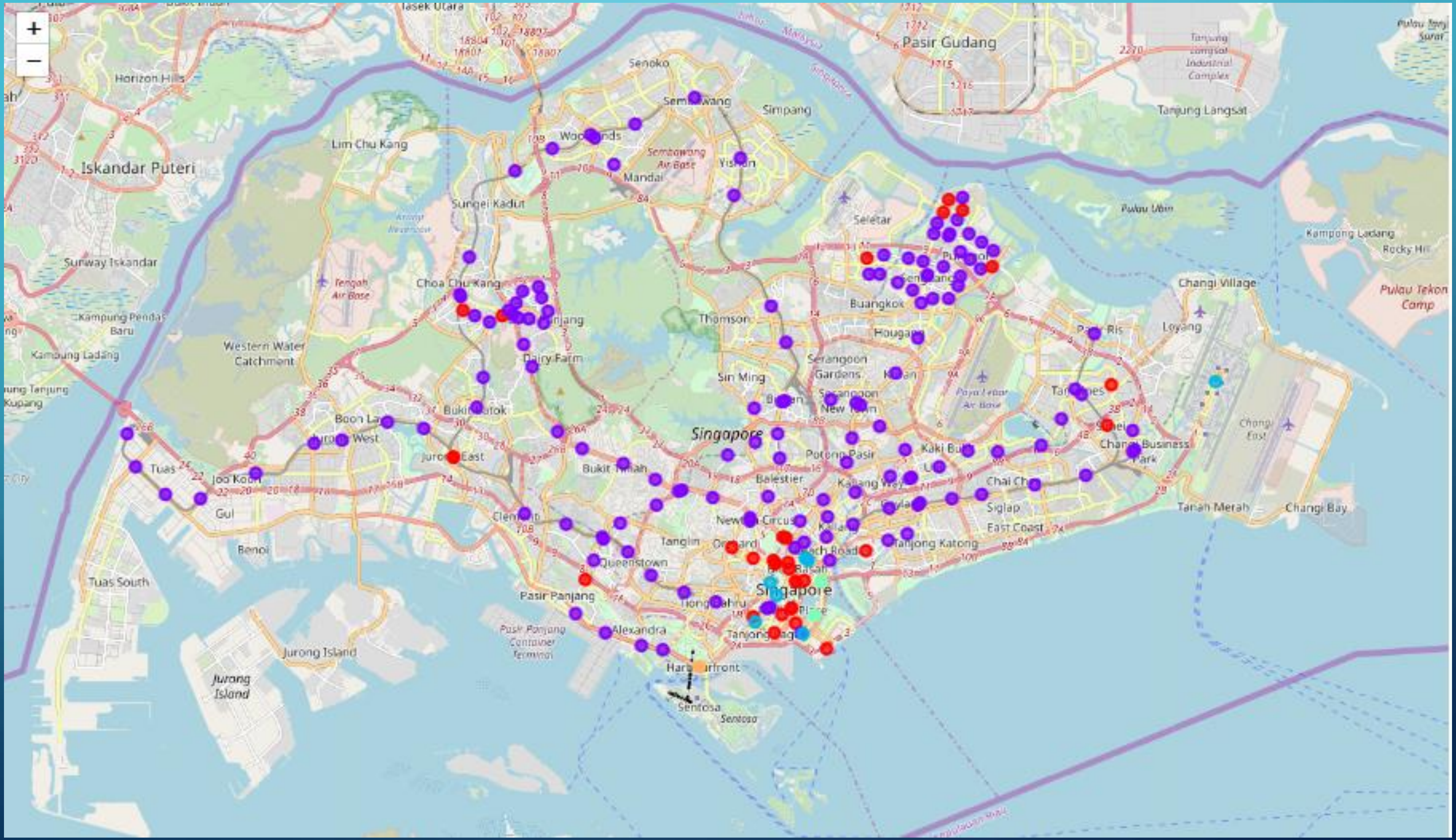
RESULTS

- Clusters obtained via k-Means Clustering algorithm, with k=5

1. **Cluster 1** – marked with red circle in the above map and consists of 23 train stations and characterized by having plenty of restaurants, particularly Chinese & Indian restaurants.
2. **Cluster 2** – marked with purple circle in the above map and consists of only 2 train stations and does not exhibit any obvious characteristic other than both are having yoga studio.
3. **Cluster 3** – marked with light blue circle in the above map and consists of 13 train stations and mainly characterized by having bus station/stop or gym
4. **Cluster 4** – marked with light green in the above map. It is the biggest cluster consisting 115 train stations and somewhat characterized by having café, hotel, Japanese/Korean restaurant in their top 5 most common venue categories.
5. **Cluster 5** – marked with light orange circle in the above map and consists of 34 train stations, mainly characterized by having food courts and coffee shops.

RESULTS

- Clusters obtained via Hierarchical Clustering with agglomerative approach, up to 5 flat clusters formed



RESULTS

- Clusters obtained via Hierarchical Clustering with agglomerative approach, up to 5 flat clusters formed
  1. **Cluster 1** – marked with purple circles in the above map and it is the biggest cluster comprising of 142 stations, primarily characterized by having eating places e.g. food courts, Chinese restaurants, fast food restaurants, coffee shops.
  2. **Cluster 2** – marked with light blue circles and consists of 7 stations, primarily characterized by having Japanese restaurant, hotel or bakery.
  3. **Cluster 3** – marked with light green circles and consists of 4 train stations, characterized by having hotel as the 1st or 2nd most common venue category.
  4. **Cluster 4** – marked with light orange circles and consists of 2 train stations located very near to each other, characterized by having "clothing store" as the 1st most common venue category
  5. **Cluster 5** – marked with red circles in the above map and consists of 32 train stations, characterized by most of them having café, coffee shop or hotel as their top 5 most common venue categories

DISCUSSION

- Observations from comparing the two clustering algorithms' outcomes:
  1. Even though both the clustering algorithms are set up to produce 5 clusters, the clustering outcomes are quite different.
  2. There is a dominant cluster which covers a big majority of the train station, for instance the Cluster 4 of k-means clustering (includes 114 out of 187 stations) and Cluster 1 of hierarchical clustering (includes 143 out of 187 stations). However, the main characteristics of these two dominant clusters are quite different.
  3. Both algoritms also produced a insignificant cluster which contains only two train stations, i.e Cluster 3 of k-means clustering and Cluster 4 of hierarchical clustering. Again, these two insignificant clusters are pretty different.
- It is unclear to the author of this report at this stage what are the main factors contributing to the significantly different clustering outcome of these two algorithms.

CONCLUSION

- The analysis of the venue category data allows us to find out the top 5 most common venue categories in the neighborhood of each train station.
- The clustering of the train stations either by k-means clustering algorithm or hierarchical clustering algorithm provides another guiding means for user to identify a group of train stations with similar characteristic (in terms of their most common venue categories) so as to narrow down the selection of train station of their interest.