ANALYSIS OF VENUES AROUND MRT STATIONS IN SINGAPORE USING FOURSQUARE LOCATION DATA

Choy Siew Fong

Introduction

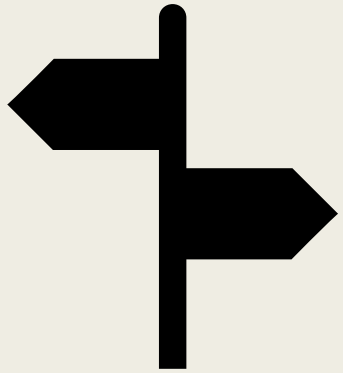
- Singapore's rail transport network consists of the Mass Rapid Transit (MRT), which is a heavy rail rapid transit system; and the subsidiary light rail transit (LRT) system
- LRT serves as feeder system to MRT
- In 2019, the daily ridership of the MRT network was 3.4 million while the annual ridership was 1.2 billion
- Since MRT and LRT are such an integral part of people's lives, what venues exist around them?
- Do these venues have any commonalities or trends which may be analyzed by clustering methods such as KMeans?

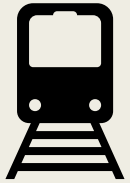
MRT train[1]

LRT train

Source: Mass Rapid Transit (Singapore). 18 March 2020. In Wikipedia. Retrieved March 19, 2020, https://en.wikipedia.org/wiki/Mass_Rapid_Transit_(Singapore)

Potential Stakeholders





- Passengers or commuters taking the train

Where should I drop off to get bread for tomorrow?

- Application developers

What mobile apps can I develop to locate a certain venue?

- Potential business owners

Who is already here?

- Current business owners

Which venues are trending in the locale? Any new competitors?

Information about the venues around MRT stations would help stakeholders make informed decisions

About the Dataset

- 'Singapore Train Station Coordinates' dataset from Kaggle (https://www.kaggle.com/yxlee245/singapore-train-station-coordinates)
- Comma separated values (.csv) format
- Four columns listing the station name, type (i.e. whether it is an MRT or an LRT station) and positional coordinates (latitude and longitude) of each MRT and LRT station in Singapore

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

Screenshot from Excel showing the top 5 rows of the dataset

- Coordinates for Admiralty Station were found to be identical to Woodlands Station during Foilum map plotting
- Correct coordinates for Admiralty obtained from here: https://www.findlatitudeandlongitude.com/l/Admiralty+mrt/1214628/
- The Foursquare information about each station would be obtained by defining a search URL and sending the ***GET*** request to Foursquare API.

Methodology section –exploratory data analysis

- Dataset was read into Pandas dataframe with pd.read_csv

```
In [2]: # read data file
        df = pd.read_csv('mrt_lrt_data2.csv')
        df.head()
```

Out[2]:

	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

Functions used and data obtained

- **Dtypes:** station_name' and 'type' fields were found to be of type *Object* while the 'lat' and 'lng' fields were of type *Float*.
- **Groupby.count:** there are 119 MRT stations and 38 LRT stations in the dataset

Define scope

- Decision made to confine analysis to MRT stations only as LRT stations are in suburbs with few interesting venues if any and limited calls on free Foursquare developer sandbox account may be an issue if too many stations are analyzed
- New dataframe *df_MRT* containing only the MRT stations was then defined

```
#Get MRT data into new dataframe - confine analysis to MRT stations
df_MRT = df.loc[df['type'] == 'MRT']
df_MRT.head()
```

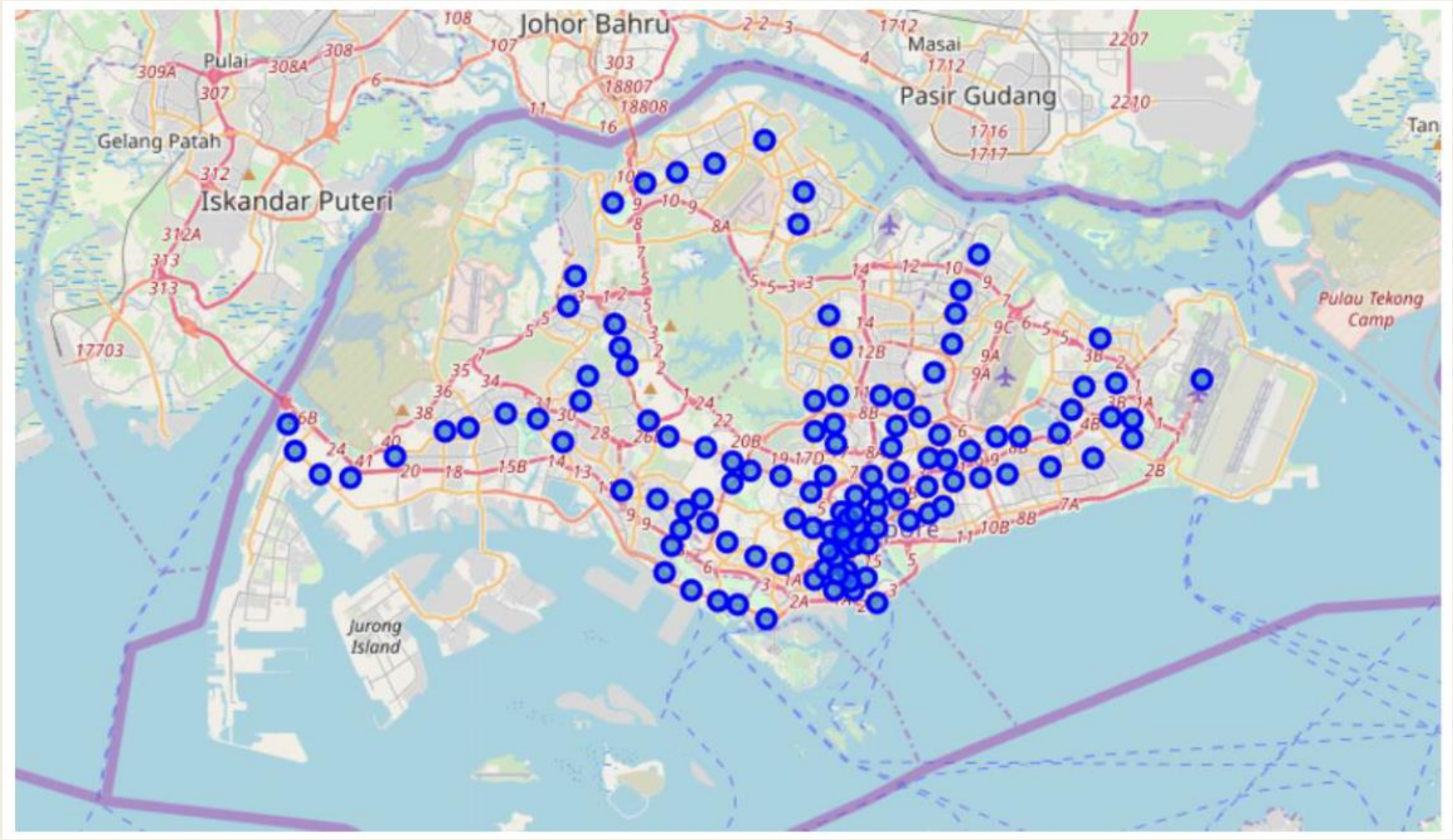
- Size of new dataframe *df_MRT* checked using ***shape*** attribute
- Output confirmed that *df_MRT* captured all MRT stations from the original dataset

```
In [6]: # check size of MRT data
        df_MRT.shape

Out[6]: (119, 4)
```

- Data was then visualised using **Folium.Map** function (see next slide)

Folium Map of MRT Stations



Exploring the first MRT station

- Foursquare credentials defined and first station identified using **.loc** method
- ***GET*** URL request was sent to Foursquare API to get information of 100 venues located within 500 m from the station.
- A *get category type* function was defined and the json output normalized to a new *nearby_venues* dataframe
- Output was then checked by calling the ***head*** function

Out[15]:

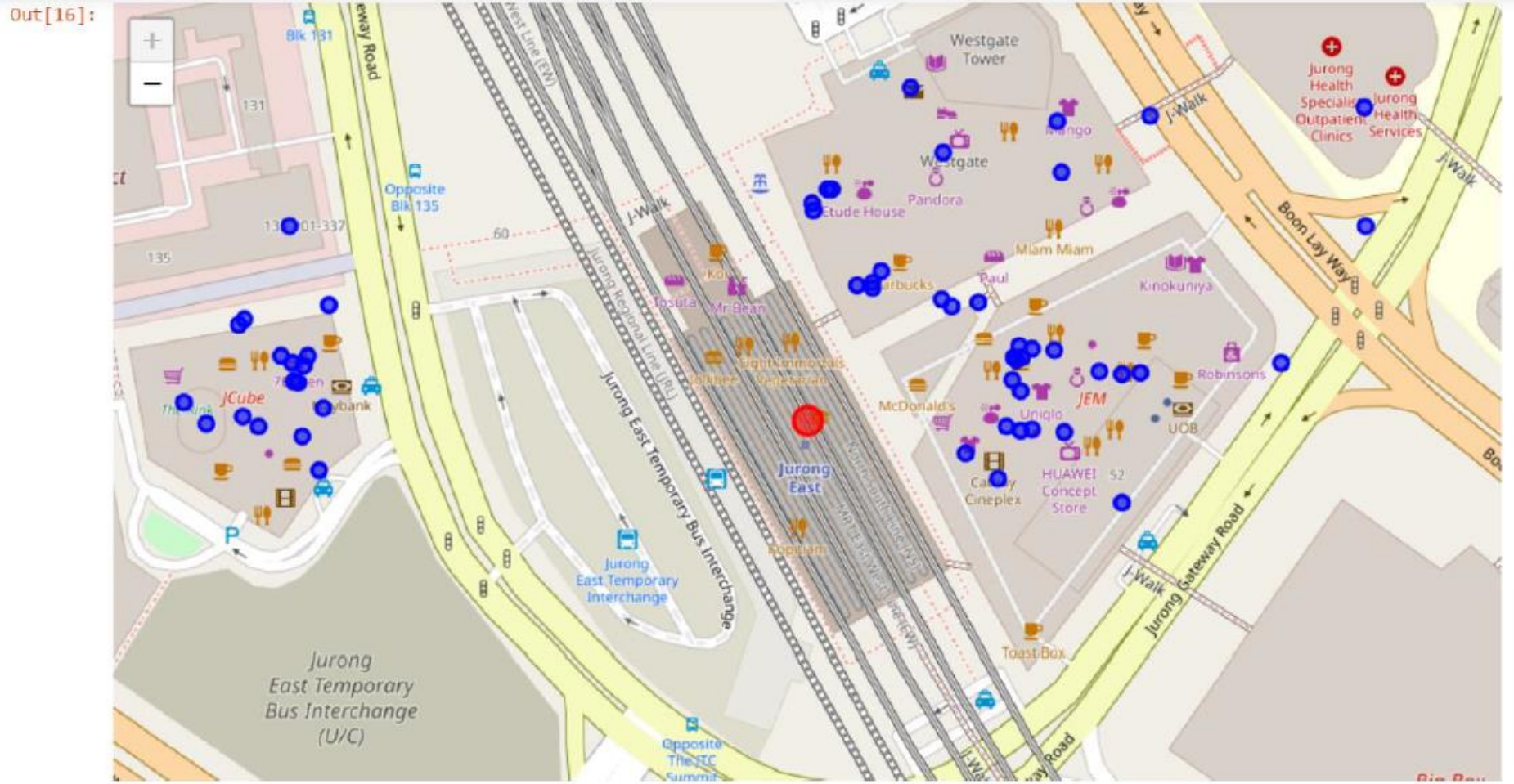
	name	categories	lat	lng
0	UNIQLO	Clothing Store	1.333175	103.743160
1	MUJI 無印良品	Furniture / Home Store	1.333187	103.743064
2	Song Fa Bak Kut Teh 松發肉骨茶	Chinese Restaurant	1.333394	103.743420
3	Johan Paris	Bakery	1.334083	103.742384
4	The Rink	Skating Rink	1.333424	103.740345

Output showing first 5 rows of nearby_venues dataframe

- Venues found around Jurong East station was visualized using ***Folium.Map*** function

Folium Map of Venues near Jurong East Station

Out[16]:



Using the ***shape*** function on nearby_venues dataframe, 73 venues were found around Jurong East Station by Foursquare

Expanding the analysis to the full dataset

- The same analysis was applied to the full dataset
- A *getNearbyVenues* function was defined. It was then applied to df_MRT and the results were output to a new dataframe called *station_venues*
- Using **shape** function, the resultant dataframe was found to consist of consisted of 4712 rows and 7 columns

```
In [20]: #check size of station_venues
         print(station_venues.shape)
         station_venues.sample(10)

         (4712, 7)
```

- The ***sample*** function was used to return a random sample of 10 stations' data

Sample of 10 Stations' Data

	Station	Station_Latitude	Station_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Category
3728	Telok Blangah	1.270769	103.809878	Telok Blangah Drive Blk 82 Market	1.273392	103.807595	Market
83	Bukit Batok	1.349069	103.749596	NTUC FairPrice	1.348814	103.749248	Grocery Store
3130	Esplanade	1.293995	103.855396	Narrative Coffee Stand	1.297118	103.854371	Coffee Shop
89	Bukit Batok	1.349069	103.749596	Bee Cheng Hiang	1.349794	103.748235	Asian Restaurant
1947	Eunos	1.319809	103.902888	Benji Pet Kennel	1.318012	103.906195	Pet Store
2343	HarbourFront	1.265453	103.820514	Play Court | Vivocity	1.263891	103.822152	Playground
4606	Geylang Bahru	1.321479	103.871457	Beng Soon Seafood 明顺海鲜	1.323005	103.868693	Seafood Restaurant
2070	Simei	1.343237	103.953343	Each-A-Cup	1.342812	103.953248	Juice Bar
1562	Tanjong Pagar	1.276385	103.846771	Quan Ji @ Amoy Street Food Market	1.279100	103.847392	Chinese Restaurant
3665	Haw Par Villa	1.283149	103.781991	Haw Par Villa	1.283855	103.781421	Sculpture Garden

- Using the **unique** function, 327 unique venue categories and 3226 unique venues were found around the 119 MRT stations in the dataset
- The **groupby** and **count** functions to find the total number of venues for each MRT station

Output of groupby and count functions showing total number of venues around each station

Out[23]:

Station	Station_Latitude	Station_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Category
Admiralty	8	8	8	8	8	8
Aljunied	47	47	47	47	47	47
Ang Mo Kio	41	41	41	41	41	41
Bartley	12	12	12	12	12	12
Bayfront	50	50	50	50	50	50
Beauty World	76	76	76	76	76	76
Bedok	56	56	56	56	56	56
Bedok North	17	17	17	17	17	17
Bedok Reservoir	5	5	5	5	5	5
Bencoolen	100	100	100	100	100	100

Converting categorical values and getting top 5 venues around each station

- The categorical values for Venue_Category to numerical values using **get_dummies** function
- The new dataframe contained 4712 rows and 328 columns

```
In [25]: #size of new one-hot dataframe
         station_onehot.shape

Out[25]: (4712, 328)
```

- The dataframe was then grouped by stations and taking the mean of occurrence of each venue category
- Top 5 venue category around each station were printed to get an idea of which venues were common

```
----Admiralty----
                venue  freq
0         Supermarket  0.25
1        Optical Shop  0.12
2  Sushi Restaurant    0.12
3          Food Court  0.12
4         Snack Place  0.12


----Aljunied----
                           venue  freq
0             Chinese Restaurant  0.11
1               Asian Restaurant  0.09
2                   Noodle House  0.09
3                    Coffee Shop  0.06
4  Vegetarian / Vegan Restaurant  0.06


----Ang Mo Kio----
               venue  freq
0        Coffee Shop  0.12
1       Dessert Shop  0.07
2         Food Court  0.07
3  Bubble Tea Shop    0.05
4        Supermarket  0.05
```

Top 5 Venues output for the first 3 stations

Top 10 venues for each MRT Station

- The venue categories were then sorted in descending order and a new dataframe, *station_venues_sorted* containing the top ten venue categories for each MRT station created

Out[30]:

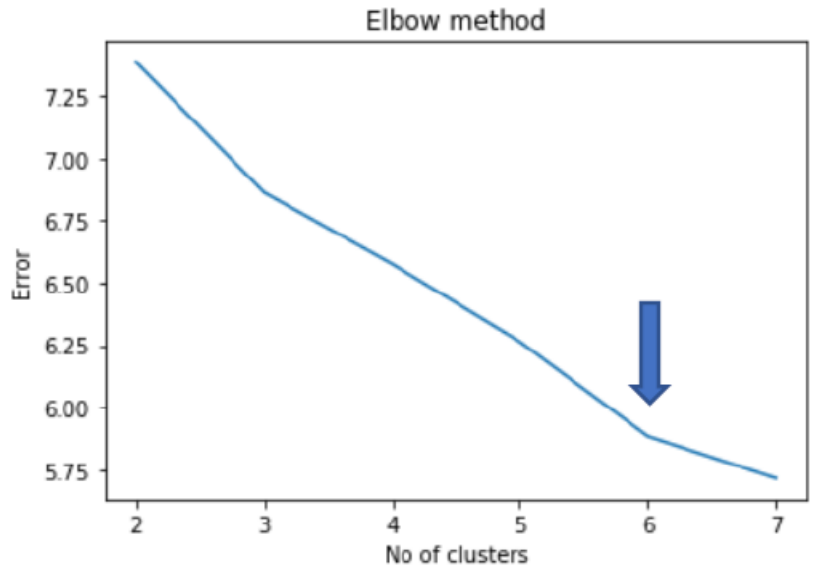
	Station	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Admiralty	Supermarket	Coffee Shop	Park	Sushi Restaurant	Optical Shop	Snack Place	Food Court	Food & Drink Shop	Filipino Restaurant	Fish & Chips Shop
1	Aljunied	Chinese Restaurant	Noodle House	Asian Restaurant	Coffee Shop	Vegetarian / Vegan Restaurant	Bus Station	Indian Restaurant	Dim Sum Restaurant	Food Court	Seafood Restaurant
2	Ang Mo Kio	Coffee Shop	Dessert Shop	Food Court	Supermarket	Japanese Restaurant	Bubble Tea Shop	Ramen Restaurant	Snack Place	Sushi Restaurant	Modern European Restaurant
3	Bartley	Bus Station	Noodle House	Pet Store	Café	Bus Stop	Food Truck	Indian Restaurant	Metro Station	Frozen Yogurt Shop	Fried Chicken Joint
4	Bayfront	Hotel	Boutique	Bar	Theater	Bridge	Lounge	Japanese Restaurant	Tea Room	Italian Restaurant	Waterfront

First 5 rows of the station_venues_sorted dataframe showing top 10 venue category for each station

KMeans Clustering – finding number of clusters

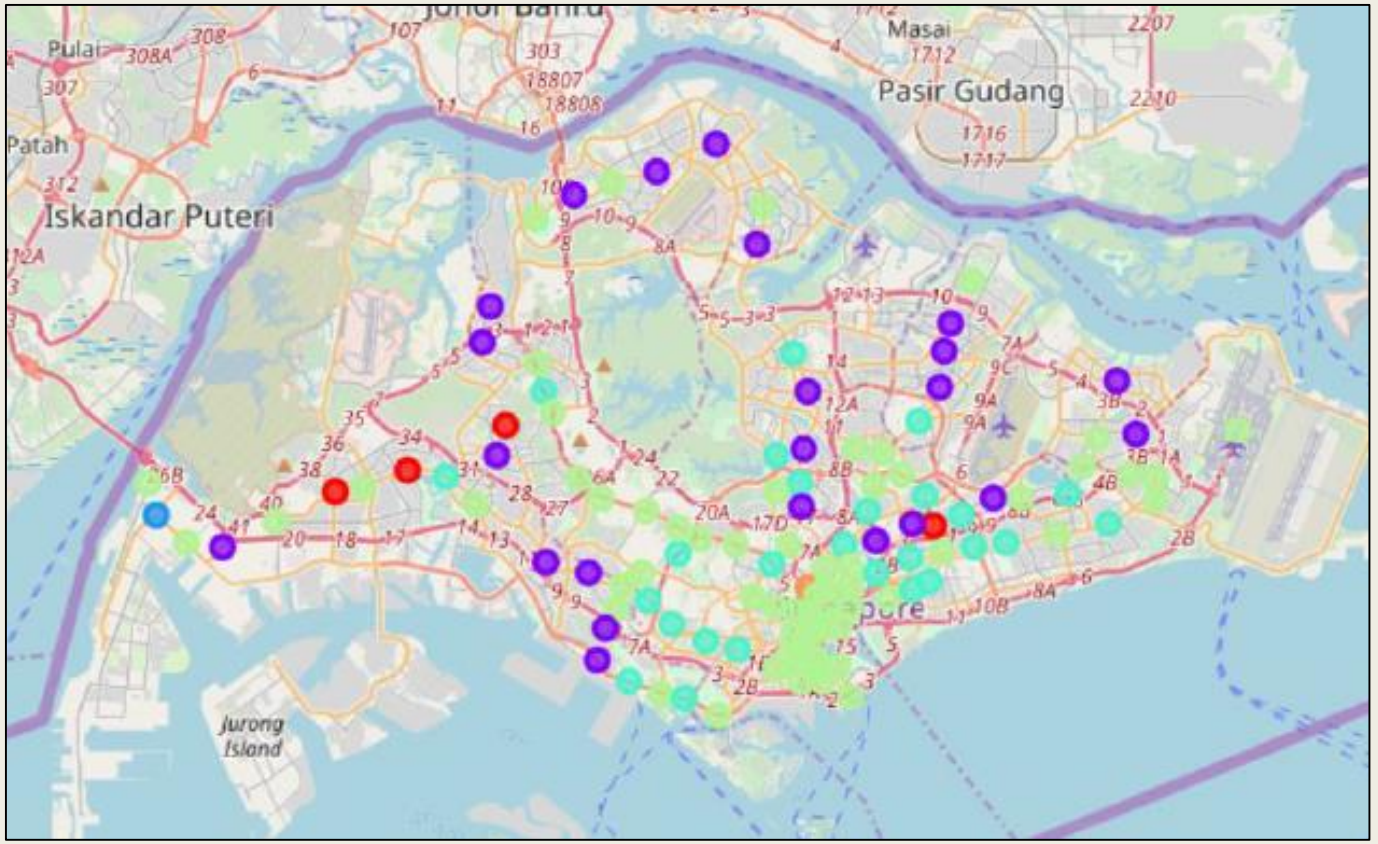
- KMeans clustering was used to analyze the station_grouped dataframe
- Number of clusters was estimated using the Elbow method
- A cluster size of 6 was selected

```
In [32]: station_grouped_clustering = station_grouped.drop('Station', 1)
         Error =[]
         for i in range(2, 8):
             kmeans = KMeans(n_clusters = i).fit(station_grouped_clustering)
             kmeans.fit(station_grouped_clustering)
             Error.append(kmeans.inertia_)
         import matplotlib.pyplot as plt
         plt.plot(range(2, 8), Error)
         plt.title('Elbow method')
         plt.xlabel('No of clusters')
         plt.ylabel('Error')
         plt.show()
```



Kmeans Clustering

- KMeans clustering was run after dropping the first column ‘Station’ as it had a non-numerical value
- The resulting dataframe was then merged with the original *df_MRT* to include the latitude and longitude data.
- Clusters were then visualised in a Folium map



Folium map showing the 6 clusters obtained by KMeans Clustering

Examining the clusters – 1 & 2

- Each of the clusters (1-6) were examined for commonalities

```
In [56]: #Examine the clusters - cluster 1
station_merged.loc[station_merged['Cluster Labels'] == 0, station_merged.columns[[0] + list(range(5, station_merged.shape[1]))]]
```

Out[56]:

	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
2	Bukit Gombak	Food Court	Vegetarian / Vegan Restaurant	Stadium	Ice Cream Shop	Steakhouse	Supermarket	Flea Market	Fast Food Restaurant	Chinese Restaurant	Sandwich Place
31	Pioneer	Gym / Fitness Center	Bus Station	Stadium	Convenience Store	Pool	Fast Food Restaurant	Shopping Mall	Food Court	Snack Place	Bus Line
33	Lakeside	Food Court	Convenience Store	Trail	Snack Place	Food & Drink Shop	Field	Filipino Restaurant	Fish & Chips Shop	Flea Market	Flower Shop
79	MacPherson	Food Court	Climbing Gym	Basketball Court	Hobby Shop	BBQ Joint	Thai Restaurant	Food Truck	Asian Restaurant	Bakery	Office

Cluster 1

```
In [58]: #Examine the clusters - cluster 2
station_merged.loc[station_merged['Cluster Labels'] == 1, station_merged.columns[[0] + list(range(5, station_merged.shape[1]))]]
```

Out[58]:

	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	Chinese Restaurant	Frozen Yogurt Shop	Bus Station	Bowling Alley	Multiplex	Café	Shopping Mall
3	Choa Chu Kang	Coffee Shop	Fast Food Restaurant	Ice Cream Shop	Smoke Shop	Gym	Thai Restaurant	Sushi Restaurant	Supermarket	Bubble Tea Shop	Food Court
4	Yew Tee	Fast Food Restaurant	Japanese Restaurant	Pool	Electronics Store	Café	Sandwich Place	Coffee Shop	Food Court	Diner	Shopping Mall
6	Marsiling	Coffee Shop	Flea Market	Steakhouse	Grocery Store	BBQ Joint	Asian Restaurant	Hainan Restaurant	Paintball Field	Track	Trail
8	Admiralty	Supermarket	Coffee Shop	Park	Sushi Restaurant	Optical Shop	Snack Place	Food Court	Food & Drink Shop	Filipino Restaurant	Fish & Chips Shop

Cluster 2 (first 5 shown here)

Examining the clusters - 3 and 4

In [59]:
```
#Examine the clusters - cluster 3
station_merged.loc[station_merged['Cluster Labels'] == 2, station_merged.columns[[0] + list(range(5, station_merged.shape[1]))]]
```

Out[59]:

	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	Bus Station	Fast Food Restaurant	Filipino Restaurant	Fish & Chips Shop	Flea Market	Flower Shop	Food	Food & Drink Shop	Food Court	Food Stand

In [60]:
```
#Examine the clusters - cluster 4
station_merged.loc[station_merged['Cluster Labels'] == 3, station_merged.columns[[0] + list(range(5, station_merged.shape[1]))]]
```

Out[60]:

	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
12	Yio Chu Kang	Food Court	Chinese Restaurant	Gym	Dance Studio	College Auditorium	Bus Stop	Cafeteria	Fast Food Restaurant	Tennis Court	Seafood Restaurant
15	Braddell	Noodle House	Chinese Restaurant	Food Court	Café	Hakka Restaurant	Asian Restaurant	Seafood Restaurant	Fast Food Restaurant	Bakery	Thai Restaurant
18	Newton	Chinese Restaurant	Seafood Restaurant	Hotel Bar	Italian Restaurant	Gym / Fitness Center	Convenience Store	Pool	Pizza Place	Café	Food Court
34	Chinese Garden	Chinese Restaurant	Coffee Shop	Bus Station	Train Station	Café	Pizza Place	Asian Restaurant	Food Court	Indian Restaurant	Food Truck
38	Commonwealth	Chinese Restaurant	Indian Restaurant	Asian Restaurant	Coffee Shop	Vegetarian / Vegan Restaurant	Noodle House	Diner	Fast Food Restaurant	Paper / Office Supplies Store	Food Court
39	Queenstown	Noodle House	Food Court	Chinese Restaurant	BBQ Joint	Stadium	Spa	Food & Drink Shop	Café	Seafood Restaurant	Train Station

Examining the clusters – 5 & 6

```
In [61]: #Examine the clusters - cluster 5
station_merged.loc[station_merged['Cluster Labels'] == 4, station_merged.columns[[0] + list(range(5, station_merged.shape[1]))]]
```

Out[61]:

	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	Food Court	Shopping Mall	Steakhouse	Korean Restaurant	Multiplex	Department Store	Café
5	Kranji	Bus Line	Noodle House	Go Kart Track	Dessert Shop	Stadium	Racetrack	Food Stand	Fish & Chips Shop	Flea Market	Flower Shop
7	Woodlands	Japanese Restaurant	Café	Coffee Shop	Clothing Store	Shopping Mall	Chinese Restaurant	Asian Restaurant	Frozen Yogurt Shop	Fast Food Restaurant	Electronics Store
10	Yishun	Chinese Restaurant	Food Court	Supermarket	Park	Coffee Shop	Italian Restaurant	Hainan Restaurant	Fried Chicken Joint	Bubble Tea Shop	Café
17	Novena	Café	Coffee Shop	Japanese Restaurant	Hotel	Chinese Restaurant	Italian Restaurant	Ramen Restaurant	Dessert Shop	Sandwich Place	Restaurant

Cluster 5

```
In [62]: #Examine the clusters - cluster 6
station_merged.loc[station_merged['Cluster Labels'] == 5, station_merged.columns[[0] + list(range(5, station_merged.shape[1]))]]
```

Out[62]:

	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
61	Little India	Indian Restaurant	Vegetarian / Vegan Restaurant	Coffee Shop	General College & University	Bakery	Playground	Music Venue	Restaurant	Hotel	Motel

Cluster 6

Determining the Station with the Maximum Number of a Venue Category

- Which station has the maximum number of a venue category, and what category is it?
- Applying the **groupby** and **size** functions

```
In [63]: #group by station and find number of venue category of each type
         stn_grp = station_venues.groupby(['Station', 'Venue_Category']).size()
         stn_grp
```

```
Out[63]: Station    Venue_Category
         Admiralty  Coffee Shop          1
                    Food Court           1
                    Optical Shop         1
                    Park                 1
                    Snack Place          1
                    Supermarket          2
                    Sushi Restaurant     1
         Aljunied   Asian Restaurant     4
                    BBQ Joint            1
                    Badminton Court      1
                    Basketball Court     1
                    Boarding House       1
                    Breakfast Spot       1
                    Bus Station          2
                    Café                 1
```

Output showing station_venues dataframe grouped by Station and number of venue category of each type

Determining the Station with the Maximum Number of a Venue Category

- The **idxmax** function was used to identify the station with the maximum number of a venue category

```
In [64]: #find station with max number of venue category
         stn_grp.idxmax()

Out[64]: ('Jalan Besar', 'Indian Restaurant')
```

- Surprisingly, Jalan Besar had the highest number of Indian restaurants
- A total of 21 Indian restaurants were returned
- The Indian restaurants were queried from the station_venues dataframe

```
In [65]: # which indian restaurants
         station_venues.query('Station == ["Jalan Besar"] and Venue_Category == ["Indian Restaurant"]')
```

Output showing Indian Restaurants around Jalan Besar

Out[65]:

	Station	Station_Latitude	Station_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Category
4489	Jalan Besar	1.305551	103.855443	Bismillah Biryani	1.304956	103.853602	Indian Restaurant
4495	Jalan Besar	1.305551	103.855443	Murugan Idli Shop	1.308842	103.856380	Indian Restaurant
4502	Jalan Besar	1.305551	103.855443	Azmi Restaurant	1.308256	103.853075	Indian Restaurant
4503	Jalan Besar	1.305551	103.855443	Sakunthala's Restaurant	1.306000	103.852169	Indian Restaurant
4509	Jalan Besar	1.305551	103.855443	Komala Vilas (Buffalo Rd)	1.306308	103.851158	Indian Restaurant
4512	Jalan Besar	1.305551	103.855443	Khansama Tandoori Restaurant	1.308251	103.853122	Indian Restaurant
4522	Jalan Besar	1.305551	103.855443	Kailash Parbat	1.308039	103.852660	Indian Restaurant
4523	Jalan Besar	1.305551	103.855443	Lagnaa Barefoot Dining	1.306472	103.852298	Indian Restaurant
4531	Jalan Besar	1.305551	103.855443	Veeras Curry Restaurant @ Hindoo Rd	1.308650	103.853515	Indian Restaurant
4535	Jalan Besar	1.305551	103.855443	Sakunthala's Restaurant	1.309475	103.855717	Indian Restaurant
4542	Jalan Besar	1.305551	103.855443	Premaas Cuisine	1.305094	103.851980	Indian Restaurant
4552	Jalan Besar	1.305551	103.855443	Taste Of India	1.307900	103.852089	Indian Restaurant
4555	Jalan Besar	1.305551	103.855443	The Banana Leaf Apolo	1.305348	103.851698	Indian Restaurant
4557	Jalan Besar	1.305551	103.855443	Islamic Restaurant	1.303075	103.859172	Indian Restaurant
4559	Jalan Besar	1.305551	103.855443	Kebabs 'n Curries	1.309157	103.856456	Indian Restaurant
4562	Jalan Besar	1.305551	103.855443	Allauddin's Briyani	1.305799	103.851180	Indian Restaurant
4563	Jalan Besar	1.305551	103.855443	Ma Raj Restaurant	1.309817	103.855970	Indian Restaurant
4565	Jalan Besar	1.305551	103.855443	Saravanaa Bhavan	1.309272	103.855971	Indian Restaurant
4567	Jalan Besar	1.305551	103.855443	Copper Chimney	1.309712	103.855447	Indian Restaurant
4571	Jalan Besar	1.305551	103.855443	Ananda Bhavan Restaurant	1.309665	103.855614	Indian Restaurant
4576	Jalan Besar	1.305551	103.855443	Yakader	1.305794	103.851108	Indian Restaurant

Top 10 Venue Category for all Stations

- What are the top ten most common venue category for all MRT stations?
- The stn_grp object in the earlier analysis was converted to a dataframe, grouped by Venue Category and the counts for each venue category summed up
- This list was then sorted in descending order

Out[67]:

Venue_Category	
ATM	1
Accessories Store	4
Airport	2
Airport Lounge	5
American Restaurant	21
Arcade	3
Art Gallery	21
Art Museum	2
Arts & Crafts Store	9
Arts & Entertainment	1
Asian Restaurant	127
Athletics & Sports	2
Australian Restaurant	5
BBQ Joint	44
Baby Store	1
Badminton Court	1
Bagel Shop	2
Bakery	118
Bank	2
Bar	43
Basketball Court	9

Count for each venue category for all MRT stations

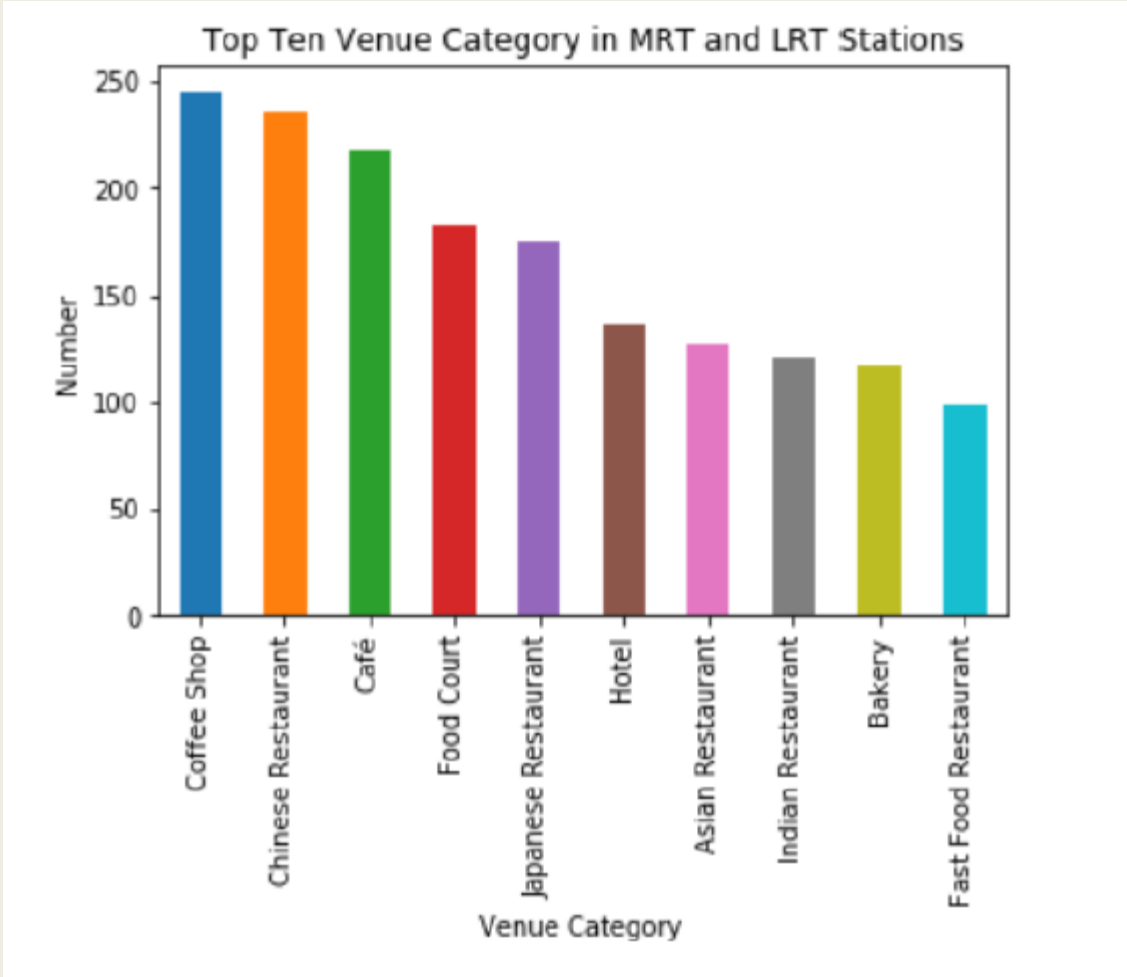
Top 10 Venue Category for all Stations

- Top ten venue categories displayed using the **head(10)** command.

```
In [68]: Alist.sort_values(ascending = False, inplace = True) #sort venue category in descending order
         Top_10 = Alist.head(10) #find top 10 venue category
         Top_10

Out[68]: Venue_Category
         Coffee Shop              245
         Chinese Restaurant       236
         Café                     218
         Food Court               182
         Japanese Restaurant      175
         Hotel                    136
         Asian Restaurant         127
         Indian Restaurant        121
         Bakery                   118
         Fast Food Restaurant      98
         Name: Counts, dtype: int64
```

- To visualise the results, Matplotlib bar chart was used to plot the top 10 categories



Bar chart showing top 10 venue categories for all MRT stations

Determining the Station with the Maximum Number of a Venue

- A similar analysis was done as the venue category analysis, but this time the stations were grouped by venue instead

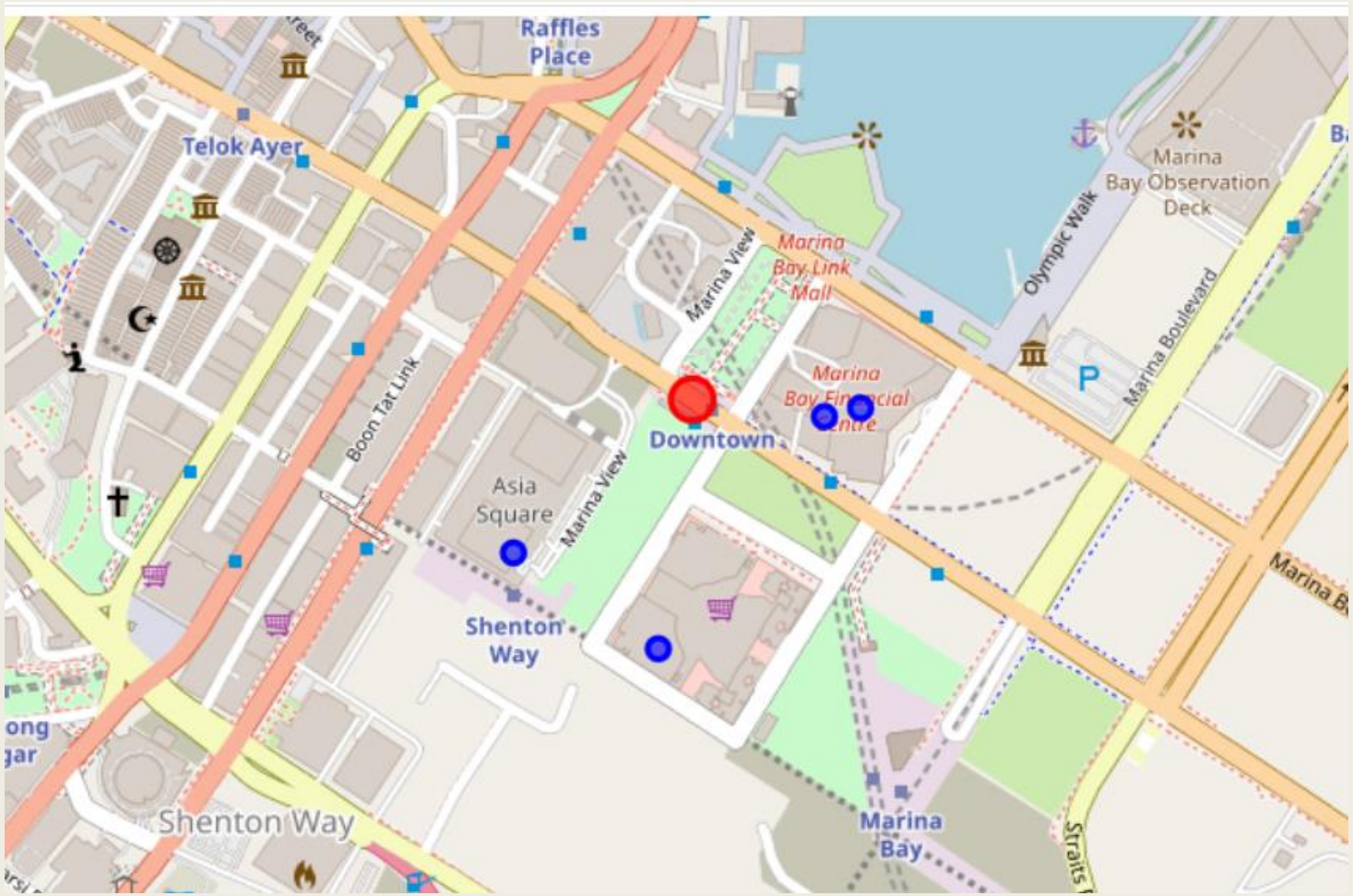
```
In [71]: stn_grp2.idxmax() #Find venue with max number and the corresponding station
Out[71]: ('Downtown', 'Starbucks')
```

- The result shows that Downtown station had the maximum number of Starbucks
- The location of the 4 Starbucks outlets around Downtown station were queried

Out[72]:

	Station	Station_Latitude	Station_Longitude	Venue	Venue_Latitude	Venue_Longitude
4110	Downtown	1.27949	103.852802	Starbucks	1.279335	103.854128
4142	Downtown	1.27949	103.852802	Starbucks	1.277949	103.850985
4161	Downtown	1.27949	103.852802	Starbucks	1.276988	103.852458
4176	Downtown	1.27949	103.852802	Starbucks	1.279422	103.854494

Folium map of Starbucks locations around Downtown Station



Map showing location of the four Starbucks outlets around Downtown station

ATMs Around MRT Stations

- How many ATMs were located around the MRT stations?
- This was done by was filtering station_venues dataframe by ‘Venue Category’ == ‘ATM’

Out[74]:

	Station	Station_Latitude	Station_Longitude	Venue_Category	Venue	Venue_Latitude	Venue_Longitude
1400	Commonwealth	1.302439	103.798326	ATM	POSB ATM	1.300509	103.801128

- Only one station, Commonwealth, was returned by the analysis which was highly unusual

Results

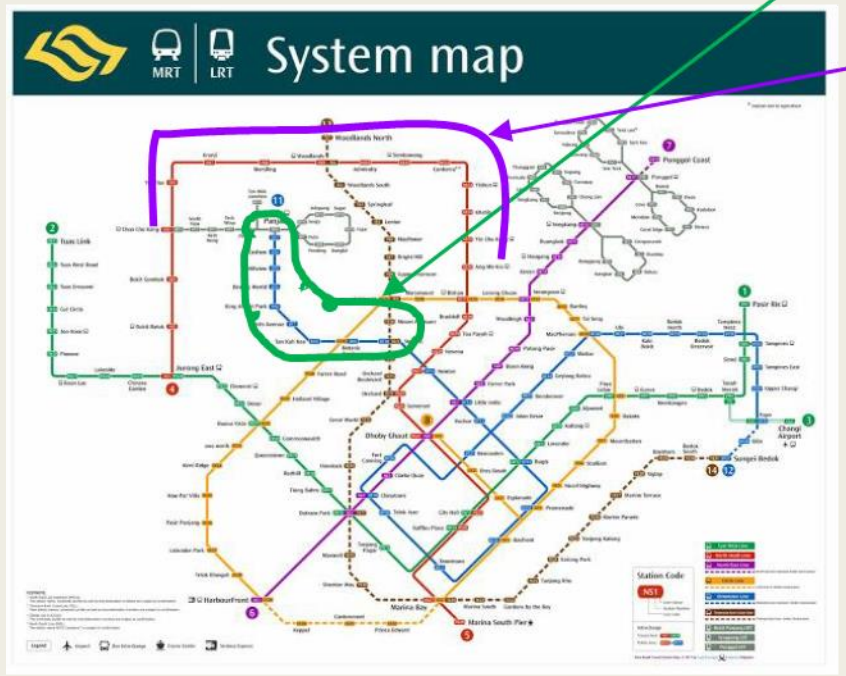
- 327 unique venue categories and 3226 unique venues were found around the 119 MRT stations
- KMeans clustering using a cluster size of 6 determined by Elbow method revealed the following commonalities in the clusters:
  - *Cluster 1 (4 stations) : had a food court and fitness venue*
  - *Cluster 2 (23 stations) : had a coffee shop*
  - *Cluster 3 (1 station) : had bus stop as top category*
  - *Cluster 4 (26 stations) : had Chinese Restaurant as top category*
  - *Cluster 5 (64 stations) : had retail (shopping mall, supermarket or flea markets) or hotels*
  - *Cluster 6 (1 station): had Indian restaurant and vegetarian restaurants*
- The maximum number (21) of one venue category was found in Jalan Besar, with the category Indian Restaurant
- The top 10 venue categories around the MRT stations are coffee shops, Chinese restaurants, café, food court, Japanese restaurant, hotel, Asian restaurant, bakery, Indian restaurant and fast food restaurant in that order

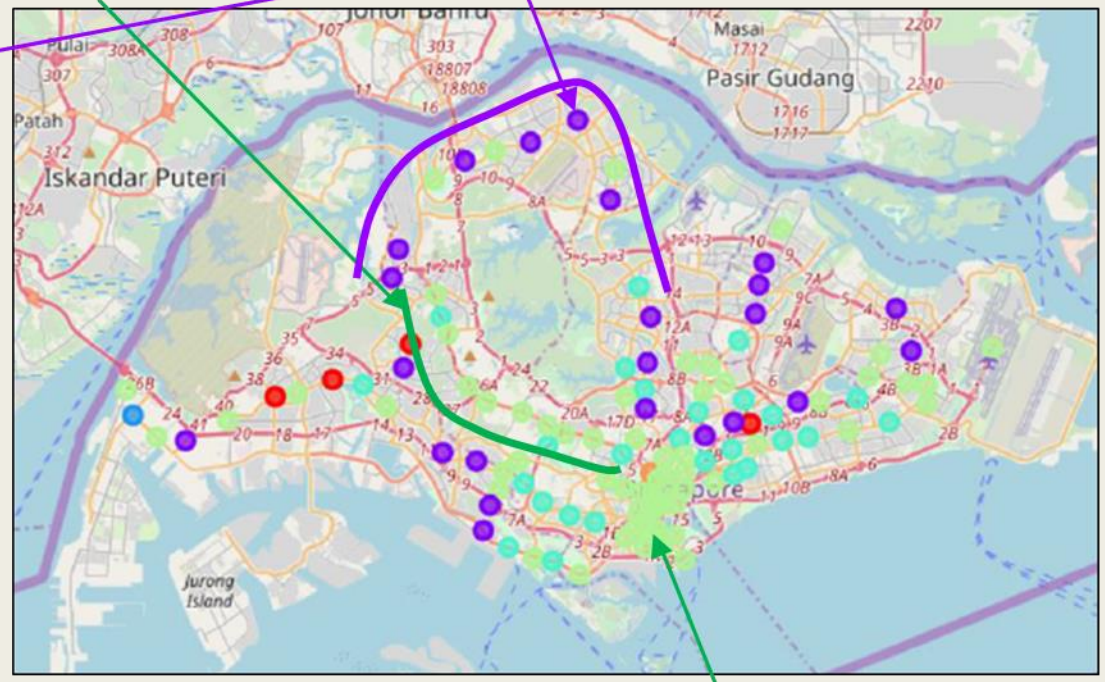
Results

- An overwhelming 9 out of the 10 top venue categories were associated with food
- The most common venue in any MRT station was Starbucks, which had 4 outlets around Downtown station
- Only one ATM was found at Commonwealth station – an abnormality which would be discussed in the following section

Discussion

- The surprisingly large number of venue categories (327) could be a result of the same venue being classified as different categories by Foursquare users, e.g. Toast box could be classified as a breakfast place or a café
- KMeans Clustering results found some clusters which appeared to follow certain sections of the MRT lines:





Discussion

- Cluster 4 was dominated by stations at the central town regions around Singapore river such as Outram, Raffles Place, Chinatown and Clarke Quay, as well as a section of the downtown line from Stevens station all the way towards Bukit Panjang
- Cluster 1 appeared to trace the north-south line including Choa Chu Kang, Yew Tee, Admiralty to Ang Mo Kio
- It is possible that these sections of the MRT network, being built around the same time frame or surrounded by housing estates of a certain age, had similar characteristics or amenities
- The surprising concentration of 21 Indian restaurants in Jalan Besar, even though it was not part of the Little Indian district, showed how the siting of restaurants may evolve organically, and differ from common expectation.
- Workers in the downtown station area could be of an expatriate profile, since there are four Starbucks in that locale alone
- The single ATM result could be due to the nature of Foursquare data, which is biased towards cool and hip venues. Users are unlikely to check into a mundane place such as an ATM unless it is difficult to find at that location, as a tip for other users.

Conclusion

- This analysis shows both the power and limitations of using Foursquare API location data to analyze venues around a location
- Such analysis can lead to interesting insights e.g. :
  - the correlation between certain sections of the MRT line with parts of different clusters
  - The unexpectedly high concentration of Indian restaurants around Jalan Besar
  - The existence of 4 Starbucks around Downtown station
- However, for mundane, utility-type venues such as ATMs, it is better to leverage on location data from the service provider, e.g. banks
- This is due to the nature of Foursquare data, which is dependent on user check-ins, which are in general, biased towards cool restaurants and hip cafes
- Foursquare location data remains a powerful and invaluable tool to shed new insights on data, especially in an increasingly socially-connected world

References

1. https://en.wikipedia.org/wiki/Mass_Rapid_Transit_(Singapore)
2. https://en.wikipedia.org/wiki/Light_Rail_Transit_(Singapore)
3. https://www.kaggle.com/yxlee245/singapore-train-station-coordinates
4. https://www.findlatitudeandlongitude.com/l/Admiralty+mrt/1214628/
5. Foursquare API
6. SMRT Network Map