

Introduction / Business Problem

Location is one of the most important attribute when setting up a new restaurant. A good location would help tremendously in the early stages and the future of a new restaurant startup. Therefore, it would be advantageous for the owner to locate potential locations before investing to start a new restaurant. Owners who wish to open a new restaurant would preferably choose a location that has similar type of restaurants nearby because it is a strong indication that there is demand for that particular kind of restaurant.

Data

Data Source

The location for areas in Singapore shall be extracted from wikipedia page using the link below:

https://en.wikipedia.org/wiki/Postal_codes_in_Singapore

The locations shall be split into individual location. Nearby venues for each location shall be explored using Foursquare. All nearby venues for each location shall then be consolidated and transformed for machine learning analysis. The clustering algorithm called Kmeans shall be used to cluster the location into clusters based on how common the venues are for that particular location.

Once the most common clusters are identified for all the locations, the owner can then decide on the most suitable location to open his/her new restaurant.

Data Cleaning

The data obtained from the wikipedia page is shown below:

	Postal district	General location
0	1	Raffles Place, Cecil, Marina, People's Park
1	2	Anson, Tanjong Pagar
2	3	Bukit Merah, Queenstown, Tiong Bahru
3	4	Telok Blangah, Harbourfront
4	5	Pasir Panjang, Hong Leong Garden, Clementi New...
5	6	High Street, Beach Road (part)
6	7	Middle Road, Golden Mile
7	8	Little India, Farrer Park, Jalan Besar, Lavender
8	9	Orchard, Cairnhill, River Valley
9	10	Ardmore, Bukit Timah, Holland Road, Tanglin
10	11	Watten Estate, Novena, Thomson
11	12	Balestier, Toa Payoh, Serangoon
12	13	Macpherson, Braddell
13	14	Geylang, Eunos
14	15	Katong, Joo Chiat, Amber Road
15	16	Bedok, Upper East Coast, Eastwood, Kew Drive
16	17	Loyang, Changi
17	18	Simei, Tampines, Pasir Ris
18	19	Serangoon Garden, Hougang, Punggol
19	20	Bishan, Ang Mo Kio
20	21	Upper Bukit Timah, Clementi Park, Ulu Pandan
21	22	Jurong, Tuas
22	23	Hillview, Dairy Farm, Bukit Panjang, Choa Chu ...
23	24	Lim Chu Kang, Tengah
24	25	Kranji, Woodgrove, Woodlands
25	26	Upper Thomson, Springleaf
26	27	Yishun, Sembawang
27	28	Seletar

Figure 1: Dataframe of Location in Singapore Extracted from Wikipedia

As can be seen from the table above, multiple locations are combined together within a single postal district. This will interfere with the machine learning model. The locations will be split into single location for every row. Postal district will be removed as it served very little purpose for the machine learning model.

The first 10 rows of the cleaned dataframe is shown below:

Location	
0	Raffles Place
1	Cecil
2	Marina
3	People's Park
4	Anson
5	Tanjong Pagar
6	Bukit Merah
7	Queenstown
8	Tiong Bahru
9	Telok Blangah
10	Harbourfront
11	Pasir Panjang
12	Hong Leong Garden
13	Clementi New Town
14	High Street
15	Beach Road (part)
16	Middle Road
17	Golden Mile
18	Little India
19	Farrer Park
20	Jalan Besar

Figure 2: Cleaned Dataframe of Location in Singapore

The location in Singapore shall be used to extract useful info via the Foursquare API.

Methodology

Exploratory Data Analysis

Using the Geocoder Python Package, the latitude and longitude of each of the location in the dataframe can be identified.

Using the Foursquare API, nearby venues for each location can be retrieved. In this project, the number of venues for each location shall be limited to 100 venues per location.

	Location	Location Latitude	Location Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Raffles Place	1.284408	103.85139	CITY Hot Pot Shabu shabu	1.284173	103.851585	Hotpot Restaurant
1	Raffles Place	1.284408	103.85139	The Fullerton Bay Hotel	1.283878	103.853314	Hotel
2	Raffles Place	1.284408	103.85139	CULINARYON	1.284876	103.850933	Comfort Food Restaurant
3	Raffles Place	1.284408	103.85139	The Salad Shop	1.285523	103.851177	Salad Place
4	Raffles Place	1.284408	103.85139	Virgin Active	1.284608	103.850815	Gym / Fitness Center

Figure 3: Example of 5 venues extracted from Foursquare for Raffles Place

The number of venues of each location shall be combined for each location for data analysis. The sorted highest number of venues identified for each location by Foursquare is plotted in the graph below:

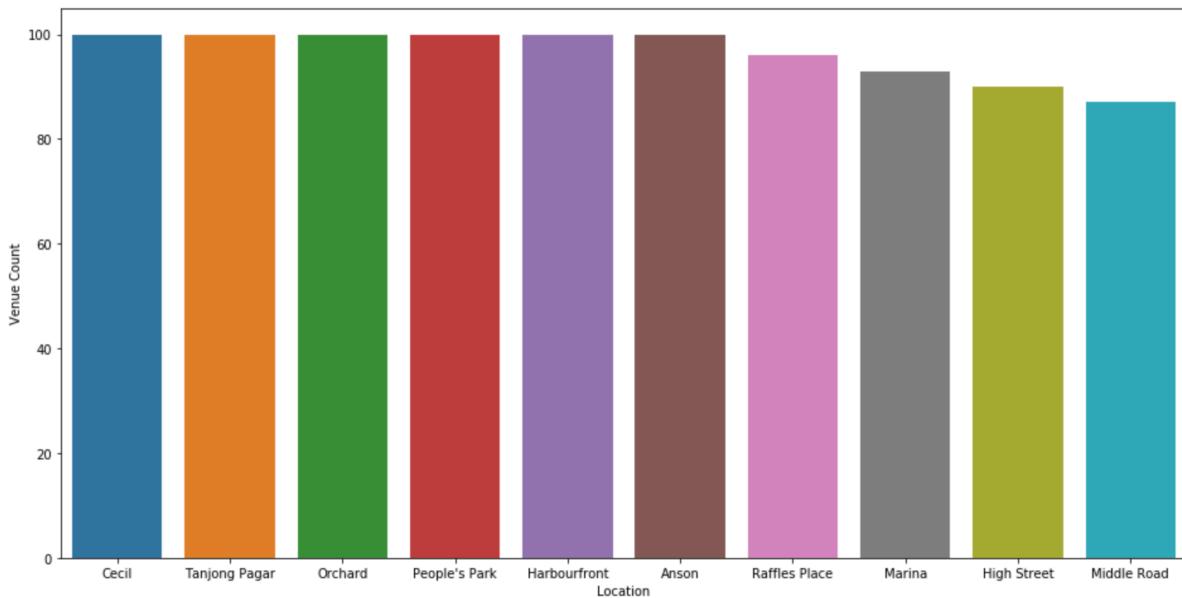


Figure 4: Location with the Highest Number of Venues by Foursquare

The location shown above are mainly city centers and central business districts. Therefore, there are a lot of venues identified by Foursquare which can be used for analysis. These locations should be the main area of interests in opening a new restaurant.

The sorted lowest number of venues identified for each location by Foursquare is plotted in the graph below:

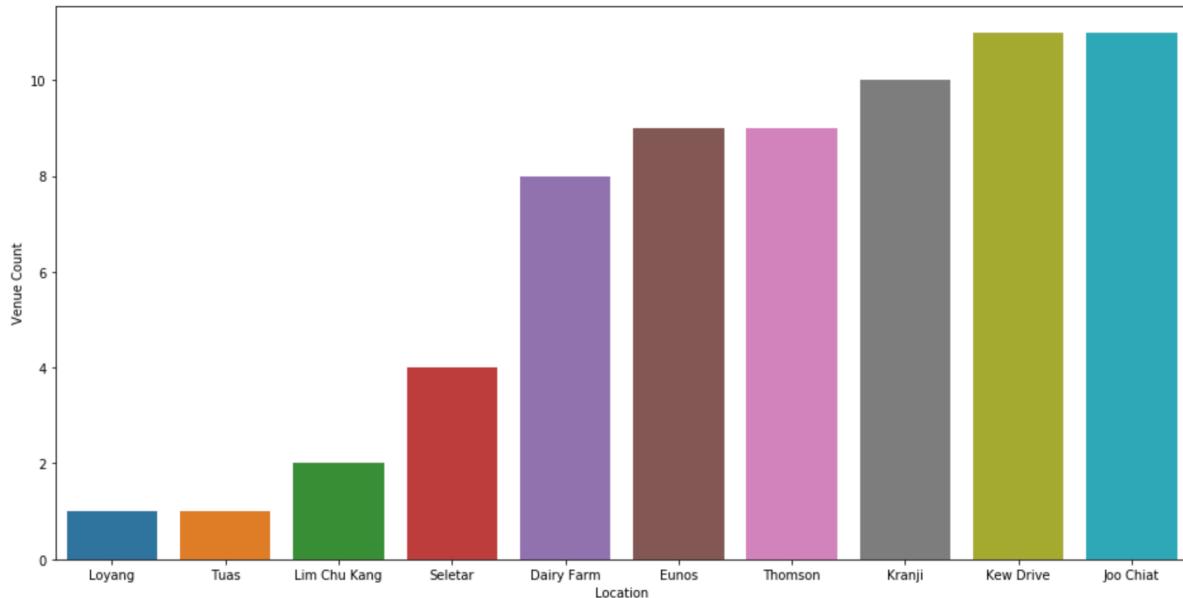


Figure 5: Location with the Lowest Number of Venues by Foursquare

The locations shown above are mainly industrial areas and residential housing areas. Therefore, very few venues are identified by Foursquare which can be used for analysis. It is recommended to **avoid** areas mentioned above in the final analysis to identify suitable location to open a restaurant because there would not be a strong demand.

It is also found that there are 257 unique categories such as hotel, salad place, gym/fitness center, etc. The unique categories identified by Foursquare shall be used in the machine learning model to find the most suitable location for the owner to open a new restaurant.

Machine Learning Model

An unsupervised machine learning algorithm called K-Means shall be used to cluster locations based on similarities of aggregated venue categories. By clustering locations which are similar to each other, we can identify group of locations which are suitable for the owner to open a restaurant. This will help narrow down the possible best locations and saves time and energy in the long run.

The venues for each location shall be aggregated and sorted based on the most common venue. The first few location showing the most common venue is tabulated below:

	Location	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	Amber Road	Hotel	Café	Japanese Restaurant	Indian Restaurant	Chinese Restaurant	Bar	Italian Restaurant	Noodle House	Bakery	Pool Hall
1	Ang Mo Kio	Coffee Shop	Dessert Shop	Food Court	Supermarket	Japanese Restaurant	Bubble Tea Shop	Multiplex	Miscellaneous Shop	Modern European Restaurant	Frozen Yogurt Shop
2	Anson	Coffee Shop	Japanese Restaurant	Bakery	Ramen Restaurant	Cocktail Bar	Hotel	Salad Place	Asian Restaurant	Gym / Fitness Center	Restaurant
3	Ardmore	Hotel	Japanese Restaurant	Spa	Café	American Restaurant	Chinese Restaurant	Department Store	French Restaurant	Miscellaneous Shop	Ramen Restaurant
4	Balestier	Chinese Restaurant	Hotel	Food Court	Fast Food Restaurant	Noodle House	Asian Restaurant	Dessert Shop	Middle Eastern Restaurant	Sandwich Place	Monument / Landmark

Figure 6: Most Common Venues based on Location

The K-Means algorithm shall be initiated with 5 clusters. The results will return 5 different clusters of location where the best cluster will be a group of similar locations which are suitable to open a new restaurant.

Results

The K-Means algorithm has clustered the locations into 5 different clusters. The clusters are then appended into the dataframe with the most common venues for reference.

	Location	Latitude	Longitude	Cluster Labels	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	Raffles Place	1.284408	103.851390	1	Café	Japanese Restaurant	Coffee Shop	Cocktail Bar	Salad Place	Gym / Fitness Center	Chinese Restaurant	Hotel	Italian Restaurant	Lounge
1	Cecil	1.282645	103.850787	1	Café	Japanese Restaurant	Hotel	Coffee Shop	Salad Place	Chinese Restaurant	Korean Restaurant	Wine Bar	Cocktail Bar	Gym / Fitness Center
2	Marina	1.290475	103.852036	1	Hotel	Café	Cocktail Bar	Bakery	Restaurant	Italian Restaurant	Bar	Coffee Shop	Concert Hall	Shopping Mall
3	People's Park	1.285810	103.844160	1	Hostel	Chinese Restaurant	Hotel	Food Court	Café	Japanese Restaurant	Seafood Restaurant	Spa	Vegetarian / Vegan Restaurant	Nightclub
4	Anson	1.275815	103.846491	1	Coffee Shop	Japanese Restaurant	Bakery	Ramen Restaurant	Cocktail Bar	Hotel	Salad Place	Asian Restaurant	Gym / Fitness Center	Restaurant

Figure 7: Appended Cluster Labels onto Dataframe

Using the information of Location, Latitude, Longitude and Cluster Labels from the dataframe, a visual map of Singapore with the location and clusters can be created as below:

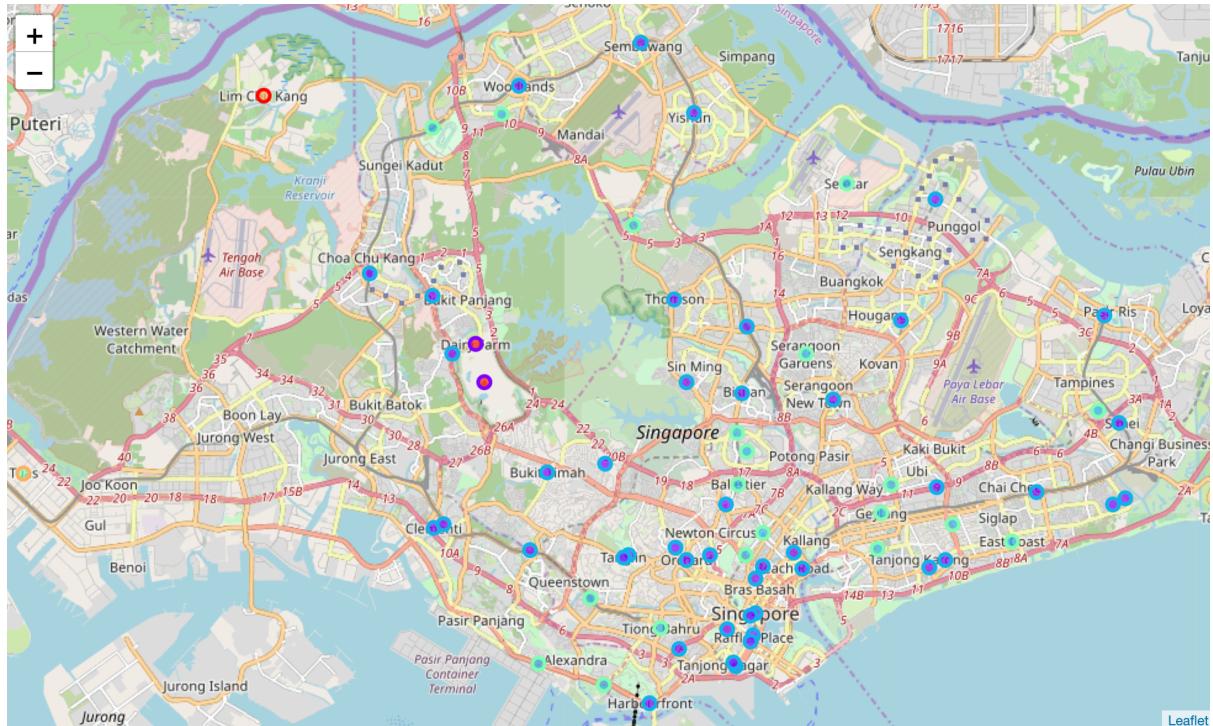


Figure 8: Singapore Map with Information on Location and Clusters

As can be observed from the map above, cluster 1 (blue) and cluster 2 (green) are the most common clusters in Singapore.

All locations grouped into the corresponding clusters are shown below:

Cluster 0

	Location	Cluster Labels	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
24	Bukit Timah	0	Trail	Rest Area	Mountain	Scenic Lookout	Food	Fast Food Restaurant	Filipino Restaurant	Fish & Chips Shop	Flea Market	Flower Shop
60	Dairy Farm	0	Trail	Parking	Nature Preserve	Bike Trail	Rest Area	Zoo Exhibit	Filipino Restaurant	Fish & Chips Shop	Flea Market	Flower Shop

Figure 9: Most Common Venues for Cluster 0

Cluster 1

	Location	Cluster Labels	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	Raffles Place	1	Café	Japanese Restaurant	Coffee Shop	Cocktail Bar	Salad Place	Gym / Fitness Center	Chinese Restaurant	Hotel	Italian Restaurant	Lounge
1	Cecil	1	Café	Japanese Restaurant	Hotel	Coffee Shop	Salad Place	Chinese Restaurant	Korean Restaurant	Wine Bar	Cocktail Bar	Gym / Fitness Center
2	Marina	1	Hotel	Café	Cocktail Bar	Bakery	Restaurant	Italian Restaurant	Bar	Coffee Shop	Concert Hall	Shopping Mall
3	People's Park	1	Hostel	Chinese Restaurant	Hotel	Food Court	Café	Japanese Restaurant	Seafood Restaurant	Spa	Vegetarian / Vegan Restaurant	Nightclub
4	Anson	1	Coffee Shop	Japanese Restaurant	Bakery	Ramen Restaurant	Cocktail Bar	Hotel	Salad Place	Asian Restaurant	Gym / Fitness Center	Restaurant
5	Tanjong Pagar	1	Coffee Shop	Japanese Restaurant	Bakery	Hotel	Italian Restaurant	Café	Ramen Restaurant	Cocktail Bar	Gym / Fitness Center	Salad Place
6	Bukit Merah	1	Café	Chinese Restaurant	Asian Restaurant	Coffee Shop	Noodle House	Bookstore	Japanese Restaurant	Clothing Store	Spa	Soup Place
10	Harbourfront	1	Chinese Restaurant	Japanese Restaurant	Clothing Store	Toy / Game Store	Fast Food Restaurant	Bakery	Coffee Shop	Multiplex	Office	Asian Restaurant
12	Clementi New Town	1	Food Court	Coffee Shop	Dessert Shop	Asian Restaurant	Noodle House	Gym	Snack Place	Shopping Mall	Fried Chicken Joint	BBQ Joint
13	High Street	1	Japanese Restaurant	Cocktail Bar	Italian Restaurant	Café	Bar	Shopping Mall	Concert Hall	Lounge	Yoga Studio	Hotel
14	Middle Road	1	Café	Japanese Restaurant	Hotel	Chinese Restaurant	Food Court	Coffee Shop	Bookstore	Bakery	Asian Restaurant	Ice Cream Shop
15	Golden Mile	1	Thai Restaurant	French Restaurant	Vietnamese Restaurant	BBQ Joint	Food Court	Coffee Shop	Indonesian Restaurant	Hotel	Burger Joint	Performing Arts Venue
18	Jalan Besar	1	Indian Restaurant	Café	Vegetarian / Vegan Restaurant	Dessert Shop	Bakery	Japanese Restaurant	Shopping Mall	Hostel	Food Court	Chinese Restaurant
19	Lavender	1	Coffee Shop	Restaurant	Café	Hotel	Bed & Breakfast	Food Court	Fast Food Restaurant	Supermarket	Bus Station	BBQ Joint
20	Orchard	1	Boutique	Hotel	Bakery	Shopping Mall	Chinese Restaurant	Sushi Restaurant	Coffee Shop	Cosmetics Shop	Japanese Restaurant	Asian Restaurant
21	Cairnhill	1	Hotel	Chinese Restaurant	Wine Bar	Other Great Outdoors	Pub	Hotel Bar	Café	Filipino Restaurant	Flower Shop	Fast Food Restaurant
23	Ardmore	1	Hotel	Japanese Restaurant	Spa	Café	American Restaurant	Chinese Restaurant	Department Store	French Restaurant	Miscellaneous Shop	Ramen Restaurant
25	Holland Road	1	Park	French Restaurant	Café	Lake	Trail	Museum	Modern European Restaurant	Garden	Cafeteria	Lounge
26	Tanglin	1	Park	French Restaurant	Thai Restaurant	Lake	Brewery	Trail	Lounge	Gastropub	Bar	Bakery
27	Watten Estate	1	Italian Restaurant	Bakery	Japanese Restaurant	Modern European Restaurant	Cheese Shop	Gym	Chinese Restaurant	Shabu-Shabu Restaurant	Bistro	Seafood Restaurant

Figure 10: Most Common Venues for Cluster 1

Cluster 2

	Location	Cluster Labels	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
7	Queenstown	2	Noodle House	Food Court	Chinese Restaurant	Spa	Train Station	Café	Seafood Restaurant	BBQ Joint	Stadium	Food & Drink Shop
8	Tiong Bahru	2	Chinese Restaurant	Coffee Shop	Food Court	Sushi Restaurant	Bakery	Japanese Restaurant	Hainan Restaurant	Noodle House	Café	Supermarket
9	Telok Blangah	2	Convenience Store	Food Court	Coffee Shop	Chinese Restaurant	Supermarket	Market	Golf Course	Metro Station	Noodle House	Pool
11	Pasir Panjang	2	Chinese Restaurant	Food Court	Thai Restaurant	Asian Restaurant	Metro Station	Indian Restaurant	Noodle House	Café	Monument / Landmark	Vietnamese Restaurant
16	Little India	2	Indian Restaurant	Vegetarian / Vegan Restaurant	Playground	Music Venue	Hotel	Hospital	Motel	Coffee Shop	Bakery	General College & University
17	Farrer Park	2	Indian Restaurant	Café	Chinese Restaurant	Hotel	Vegetarian / Vegan Restaurant	Asian Restaurant	Community Center	Hostel	Fried Chicken Joint	Snack Place
22	River Valley	2	Noodle House	Dessert Shop	Chinese Restaurant	Seafood Restaurant	Food Court	Asian Restaurant	BBQ Joint	Dim Sum Restaurant	Cantonese Restaurant	Bus Line
30	Balestier	2	Chinese Restaurant	Hotel	Food Court	Fast Food Restaurant	Noodle House	Asian Restaurant	Dessert Shop	Middle Eastern Restaurant	Sandwich Place	Monument / Landmark
31	Toa Payoh	2	Chinese Restaurant	Noodle House	Food Court	Coffee Shop	Snack Place	Steakhouse	Asian Restaurant	Bus Station	Thai Restaurant	Frozen Yogurt Shop
33	Macpherson	2	Food Court	Convenience Store	Asian Restaurant	Food Truck	Sporting Goods Shop	Metro Station	Office	Basketball Court	Thai Restaurant	Bakery
34	Braddell	2	Noodle House	Chinese Restaurant	Food Court	Café	Fast Food Restaurant	Bakery	Seafood Restaurant	Asian Restaurant	Hakka Restaurant	Thai Restaurant
35	Geylang	2	Chinese Restaurant	Noodle House	Food Court	Asian Restaurant	Vegetarian / Vegan Restaurant	Grocery Store	Dim Sum Restaurant	Cantonese Restaurant	Basketball Court	Steakhouse
38	Joo Chiat	2	Malay Restaurant	Pet Store	Seafood Restaurant	Chinese Restaurant	Noodle House	Coffee Shop	Food Court	Bar	Gas Station	Juice Bar
41	Upper East Coast	2	Coffee Shop	Indian Restaurant	Asian Restaurant	Chinese Restaurant	Dessert Shop	Gastropub	Supermarket	Massage Studio	American Restaurant	Convenience Store
47	Tampines	2	Coffee Shop	Indian Restaurant	Chinese Restaurant	Fast Food Restaurant	Supermarket	Malay Restaurant	Seafood Restaurant	Food Court	Bus Station	Noodle House
49	Serangoon Garden	2	Chinese Restaurant	Bakery	Noodle House	Italian Restaurant	Bus Station	Food Court	Japanese Restaurant	Pub	Beer Garden	South Indian Restaurant
65	Kranji	2	Bus Line	Racetrack	Dessert Shop	Go Kart Track	Stadium	Noodle House	Zoo Exhibit	Flea Market	Flower Shop	Food
66	Woodgrove	2	Food Court	Shopping Mall	Fast Food Restaurant	Convenience Store	Halal Restaurant	Supermarket	Basketball Court	Salon / Barbershop	Fried Chicken Joint	Pet Store
69	Springleaf	2	Chinese Restaurant	Asian Restaurant	Indian Restaurant	Café	Park	Bus Station	Thai Restaurant	Convenience Store	Art Gallery	Filipino Restaurant
72	Seletar	2	Snack Place	Harbor / Marina	Café	Food Court	Food & Drink Shop	Fish & Chips Shop	Flea Market	Flower Shop	Food	Zoo Exhibit

Figure 11: Most Common Venues for Cluster 2

Cluster 3

	Location	Cluster Labels	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
58	Tuas	3	Coffee Shop	Zoo Exhibit	Food Court	Filipino Restaurant	Fish & Chips Shop	Flea Market	Flower Shop	Food	Food & Drink Shop	Food Stand

Figure 12: Most Common Venues for Cluster 3

Cluster 4

Location	Cluster Labels	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	
63	Lim Chu Kang	4	Zoo Exhibit	Farm	Fruit & Vegetable Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food Stand	Food Court	Food & Drink Shop

Figure 13: Most Common Venues for Cluster 4

Discussion

Observations

Cluster 0 indicates that the majority of the neighbourhood in this cluster has a park and playground nearby. Cluster 0 is not a suitable location to open a new restaurant.

Cluster 1 indicates that the majority of the neighbourhood in this cluster has Japanese, French, cafe and coffee shops. Locations in Cluster 1 are recommended to open a new restaurant. If the owner's new restaurant genre revolves around these theme, the owner should look at locations clustered in Cluster 1.

Cluster 2 indicates that the majority of the neighbourhood in this cluster has oriental restaurant and food court. Locations in Cluster 2 are recommended to open a new restaurant. If the owner's new restaurant genre revolves around these theme, the owner should look at locations clustered in Cluster 2.

Cluster 3 is a standalone cluster far from the city with coffee shop and zoo exhibit. Cluster 3 is not a suitable location to open a new restaurant.

Cluster 4 is a standalone cluster far from the city with zoo exhibit and farm. Cluster 4 is not a suitable location to open a new restaurant.

Recommendations

Locations in Cluster 1 and Cluster 2 are suitable to open a new restaurant. The differences are:

- Locations in Cluster 1 are suitable for Japanese, French, cafe and coffee shops.
- Locations in Cluster 2 are suitable for oriental restaurant and food court.

The owner will have the freedom to choose their location either in Cluster 1 or Cluster 2 depending on their restaurant's theme.

Conclusion

It is concluded that based on the data extracted from Foursquare API for locations in Singapore and the unsupervised machine learning algorithm called K-Means, suitable locations for a new restaurant has been clustered successfully.

Two main clusters has been identified by the machine learning algorithm to be:

- Locations in Cluster 1 are suitable for Japanese, French, cafe and coffee shops.
- Locations in Cluster 2 are suitable for oriental restaurant and food court.

The owner will have the freedom to choose their location either in Cluster 1 or Cluster 2 depending on their restaurant's theme.