

Introduction

Problem Statement: Prospects of a Restaurant in Singapore

I am a foodie and I really like Mexican food. I am currently living in Singapore, which has one of the highest population densities in the world of 8,358 per km². There are about 54 Zones in Singapore including the surrounding industrial islands, and being such a crowded city, there are restaurants everywhere and I found it hard to believe that the closest Mexican restaurant from where I live is a 20-minute car ride away which is considered pretty far on this island. If I were to open a restaurant, it makes sense for restaurant investors to prefer somewhere that has lower rental rates and less competition. The other aspect to consider is the type of customers you would want to target. Hence, this gave me the idea of exploring the prospects of opening a Mexican restaurant in the Zone I am living in - Bedok.

Target Audience for this project are:

1. Non-downtown office workers and students to find a reasonable lunch/dinner place.
2. Families with children to find a place for a great night out.
3. Other Foodies like me who plans to have their own restaurant.
4. Investors who want to invest or open a restaurant.

When all these factors are considered, we can create a map and information chart where each neighborhood is clustered to the venue density.

Data Acquisition and Cleaning

Data Sources

1. The data for the names of the Zones and subzones/neighborhoods can be found in this [Wikipedia page](#). The list of places in Singapore are first segregated by region, then further separated into zones and neighborhoods. There are a total of 54 zones in Singapore.
2. The scraped data from Wikipedia will then be used to retrieve the next set of data – the coordinates of each neighborhood in Singapore. The coordinates will be retrieved using [Geopy](#).
3. From there, Foursquare API can be utilized to explore locations within a specified radius of those coordinates. The number of nearby restaurants, the categories of these restaurants can be extracted using the API.

Data Cleaning

The names of the zones and neighborhoods were formatted into a table where each row contains one(1) neighborhood. Some of the scraped zone and neighborhood names contain parenthesis and square brackets. The brackets and the texts they contained were removed before inserted into the table. This is because only the name of the neighborhood is required for the coordinate's retrieval.

The coordinates of each neighborhood are then retrieved using Geopy. There are some neighborhoods where there are no coordinates, those neighborhoods were removed from the table before moving to the next stage.

Methodology

Feature Selection

Using the coordinates and Foursquare API, we narrowed the search to Bedok Zone and its neighborhoods. Nearby restaurants within a 750m radius of each neighborhood were extracted and merged into a table. Foursquare API return json files of each location which includes a whole list of features as listed [here](#). The features were narrowed down to the Venue name, Venue coordinates and Venue category.

Exploratory Data Analysis

Out of the 8 neighborhoods within Bedok, 207 restaurants of 39 unique categories were found. The distribution of restaurants in each neighborhood is displayed in the bar chart below.

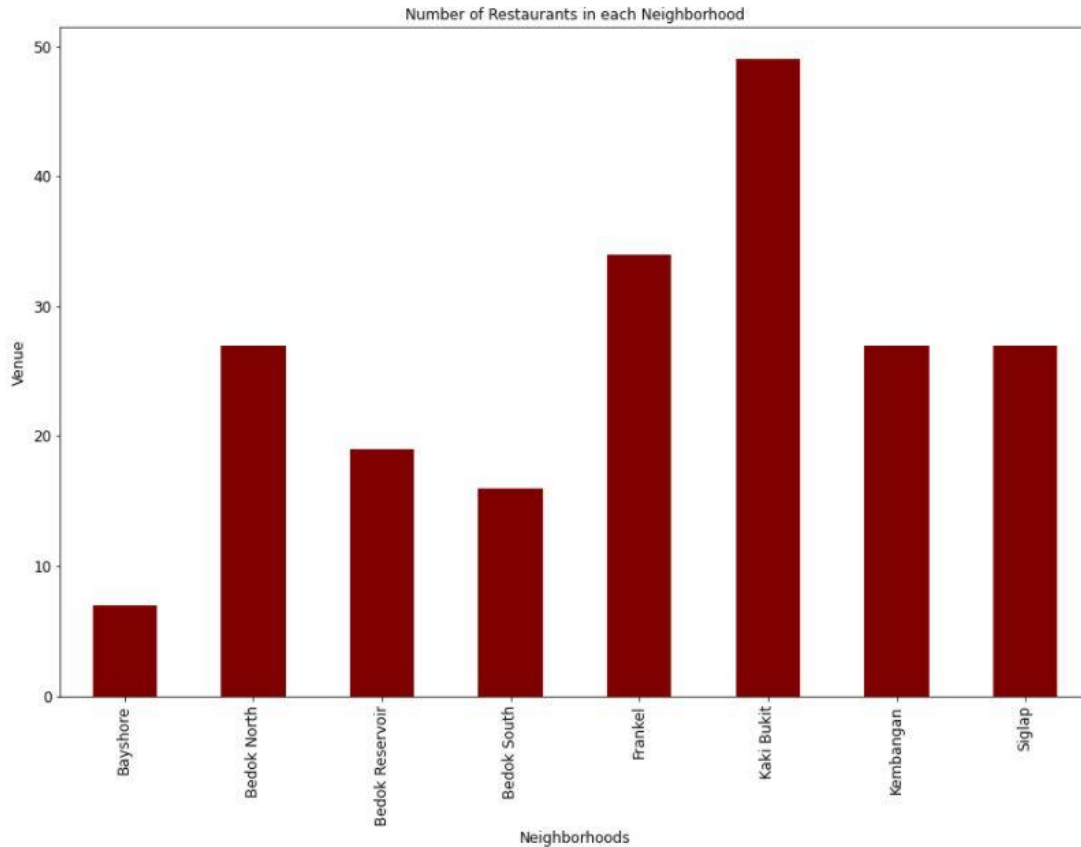


Diagram 1: Distribution of restaurants in each neighborhood.

Each neighborhood is analyzed and visualized below.

The graph below displays the Top 10 most common type of restaurant in Bedok and a table showing the top 10 most common type of restaurant in each neighborhood.

Table 1: Top 10 most common type of restaurant in each neighborhood.

	Neighborhoods	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
0	Bayshore	Asian Restaurant	Chinese Restaurant	Diner	Korean Restaurant	Pizza Place	Middle Eastern Restaurant	Vegetarian / Vegan Restaurant
1	Bedok North	Food Court	Fast Food Restaurant	Indian Restaurant	Asian Restaurant	Chinese Restaurant	Malay Restaurant	Café
2	Bedok Reservoir	Asian Restaurant	BBQ Joint	Food Court	Café	Vegetarian / Vegan Restaurant	Gastropub	Thai Restaurant
3	Bedok South	Food Court	Asian Restaurant	Seafood Restaurant	Malay Restaurant	Noodle House	Hainan Restaurant	Halal Restaurant
4	Frankel	Chinese Restaurant	Asian Restaurant	Indian Restaurant	American Restaurant	Gastropub	Japanese Restaurant	Bakery
5	Kaki Bukit	Food Court	Chinese Restaurant	Asian Restaurant	Thai Restaurant	Café	Restaurant	Vegetarian / Vegan Restaurant
6	Kembangan	Chinese Restaurant	Indian Restaurant	Food Court	Asian Restaurant	Malay Restaurant	Noodle House	Diner
7	Siglap	Chinese Restaurant	Indian Restaurant	Asian Restaurant	American Restaurant	Bakery	Café	Pizza Place

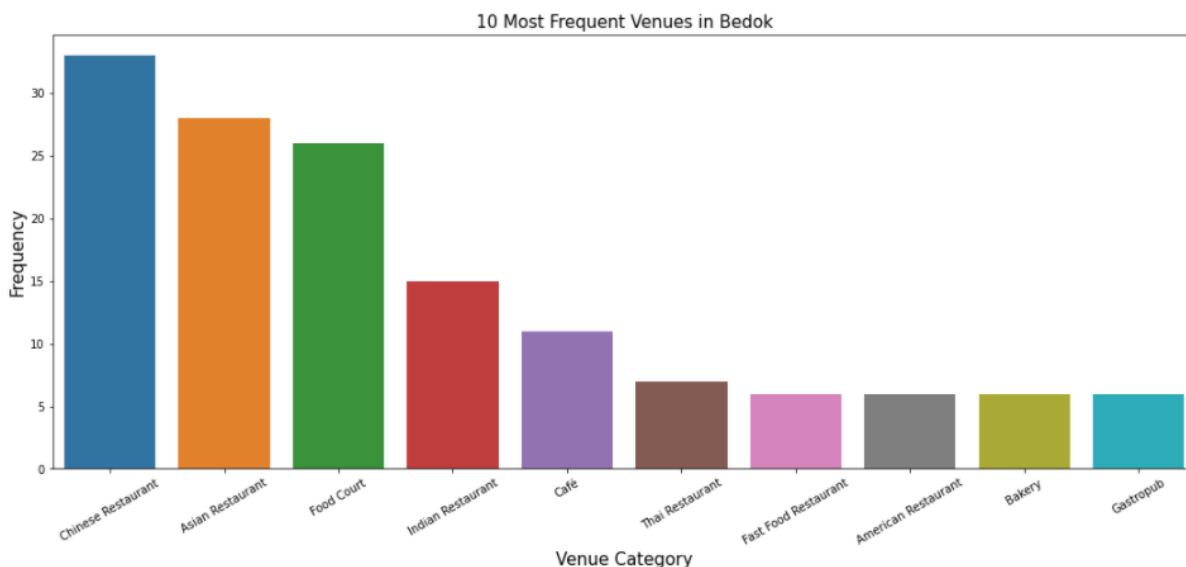


Diagram 2: Top 10 most common type of restaurant in Bedok.

As you can see from the table, there are some common venue categories in each neighborhood. For this reason, K-means cluster is very helpful in clustering these neighborhoods. K-means algorithm is one of the most common clustering methods in unsupervised learning. It starts by randomly selecting k centroids and assigning the points to the closest cluster, then it updates each centroid with the mean of all points in the cluster.

Here, in order to define the right k , the Elbow Method is used: plotting the variance as a function of the number of clusters and picking the k that flats the curve or shows a bend in the curve like an Elbow. The graph below shows the outcome of the elbow method.

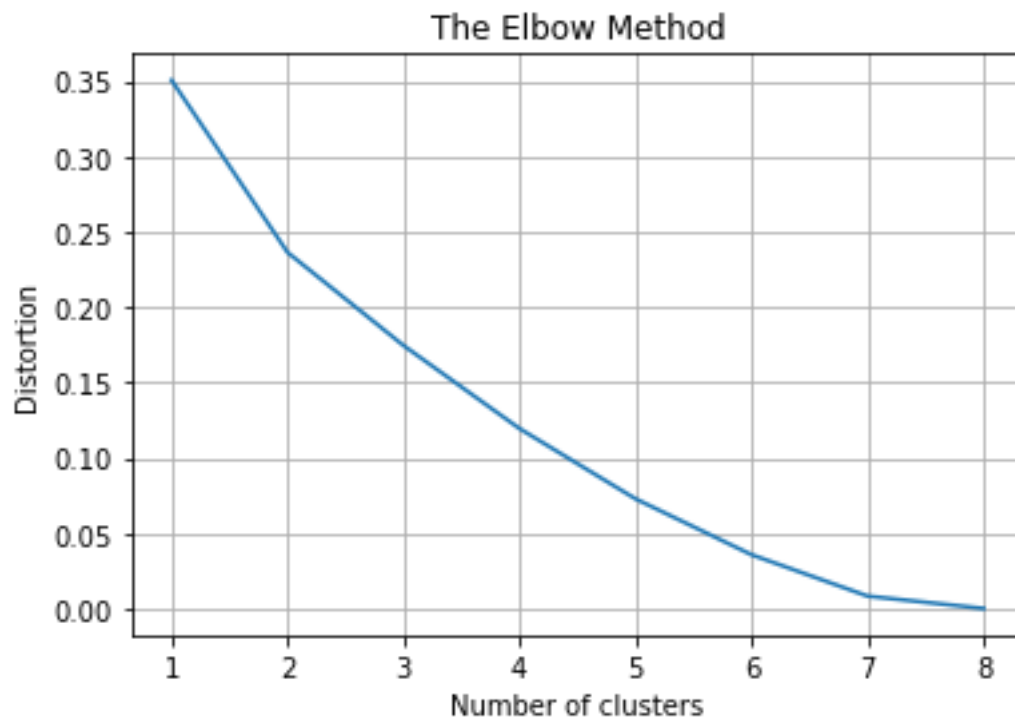


Diagram 3: Elbow method results

The k is determined to be 3. $k = 3$ is used to cluster the neighborhoods. The output is shown in the table below.

Table 2: Neighborhoods grouped by their clusters

Zones	Neighborhoods	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
Bedok	Bedok North	1.33167	103.918	0	Food Court	Fast Food Restaurant	Indian Restaurant	Asian Restaurant	Chinese Restaurant	Malay Restaurant	Café	Restaurant	Pizza Place	Seafood Restaurant
Bedok	Bedok South	1.31729	103.95	0	Food Court	Asian Restaurant	Seafood Restaurant	Malay Restaurant	Noodle House	Hainan Restaurant	Halal Restaurant	Indian Restaurant	Chinese Restaurant	Food Truck
Bedok	Kaki Bukit	1.33603	103.902	0	Food Court	Chinese Restaurant	Asian Restaurant	Thai Restaurant	Café	Restaurant	Vegetarian / Vegan Restaurant	Bakery	Cafeteria	Chinese Breakfast Place
Bedok	Kembangan	1.32103	103.913	0	Chinese Restaurant	Indian Restaurant	Food Court	Asian Restaurant	Malay Restaurant	Noodle House	Diner	Café	Middle Eastern Restaurant	Mediterranean Restaurant
Bedok	Bayshore	1.31281	103.939	1	Asian Restaurant	Chinese Restaurant	Diner	Korean Restaurant	Pizza Place	Middle Eastern Restaurant	Vegetarian / Vegan Restaurant	Food Truck	Food Court	Food
Bedok	Bedok Reservoir	1.34251	103.925	2	Asian Restaurant	BBQ Joint	Food Court	Café	Vegetarian / Vegan Restaurant	Gastropub	Thai Restaurant	Indian Restaurant	Fast Food Restaurant	Chinese Restaurant
Bedok	Frankel	1.31201	103.92	2	Chinese Restaurant	Asian Restaurant	Indian Restaurant	American Restaurant	Gastropub	Japanese Restaurant	Bakery	Café	Comfort Food Restaurant	Food Court
Bedok	Siglap	1.317	103.92	2	Chinese Restaurant	Indian Restaurant	Asian Restaurant	American Restaurant	Bakery	Café	Pizza Place	Gastropub	Sandwich Place	South Indian Restaurant

Results & Discussion

Examining the clusters, each cluster can be labeled as follows:

Cluster 0: Food Court Venues

Cluster 1: Least number of restaurants compared to other neighborhoods

Cluster 2: Asian, Chinese and Indian Restaurant Venues

Based on the cluster results, we can confirm that Asian and Chinese restaurants are the most frequented types of restaurants in Bedok zone in general. Therefore, there will not be much intra-cuisine competition if a person were to open a Mexican restaurant. In this case, restaurant owners can consider a few options.

Option 1: Opening the restaurant in venues where Food Courts are more frequented. The rental rates are lower compared to independent spaces. Suitable for restauraters who are just starting out and testing the waters. Neighborhoods in Cluster 0 would be the best place to start.

Option 2: Although there aren't many Mexican restaurants to compete with, there are neighborhoods with high number of other restaurants. Therefore, Restauraters can consider opening a restaurant where there are less restaurants in the area. Cluster 1 would be the place to start.

Adding more pointers to option 1, Food courts generally have a lower rental rate than independent spaces but they include hawker centers and food courts within shopping malls where the ones located in shopping malls will have higher rental rates. This should be considered but will require more data for further analysis in future studies.

Conclusion

Based on the results of this analysis, I think if I do make the decision to a restaurant, I would start at a food court in one of the neighborhoods of Cluster 0. Not only due to the lower rental rates but also food courts happen to be the most common venues in these neighborhoods.

References

- [1] https://en.wikipedia.org/wiki/List_of_places_in_Singapore
- [2] <https://geopy.readthedocs.io/en/stable/#module-geopy.geocoders>
- [3] <https://developer.foursquare.com/developer/>
- [4] <https://developer.foursquare.com/docs/api-reference/venues/search>