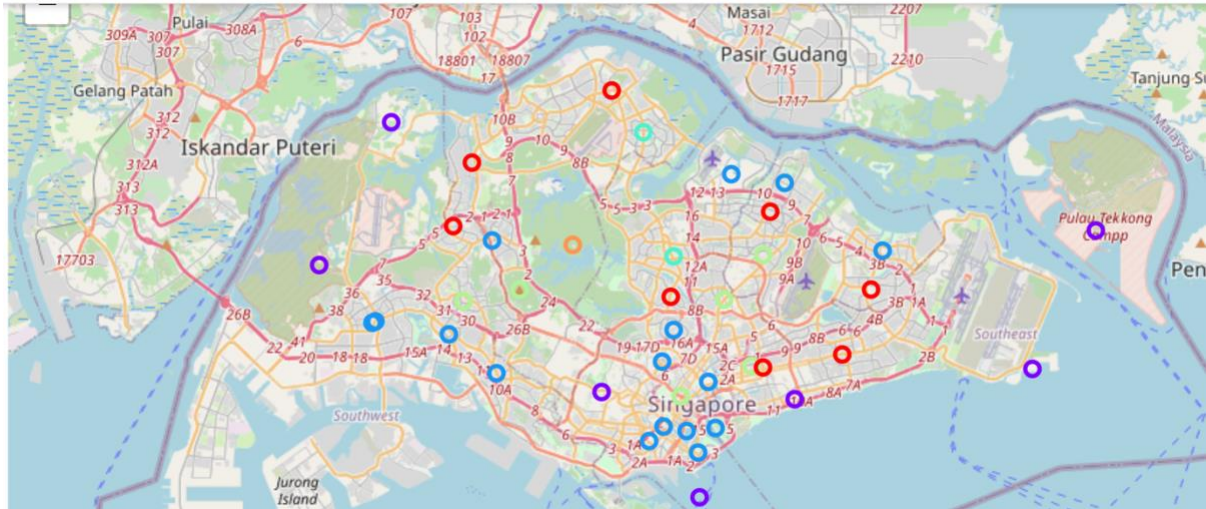


The Battle of Neighborhoods- Opening a Thai Restaurant in Singapore

Minerva Singh
(IBM Data Science Capstone)



1. INTRODUCTION- BUSINESS PROBLEM

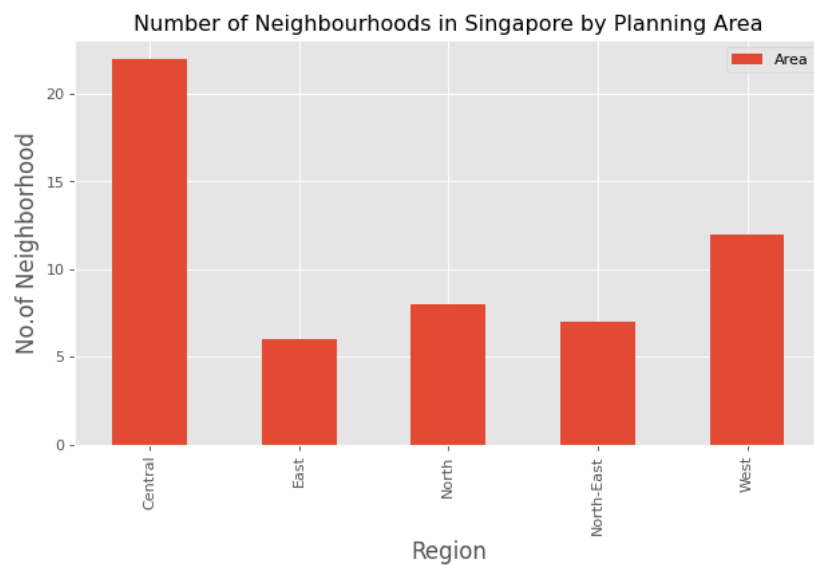
The city state of Singapore is a vibrant and culturally diverse economic powerhouse in SE Asia. The city state has an excellent social scene, especially a very diverse and dynamic food scene. Singapore's diverse eating out scene is the result of its multicultural heritage, colonial legacy and the fact that Singapore is a major financial hub that attracts professionals from across the globe. Hence opening a restaurant in Singapore can arguably be both rewarding and challenging at the same time. With one of the highest population densities in the world resulting in steep real estate prices, the choice and location of the restaurant can make or break the business.

The crux of the business problem was to identify the most suitable restaurant a food entrepreneur could open in Singapore and selecting the most optimum location for the same. The primary target audience for this report are restaurateurs and food entrepreneurs who are either interested in opening a new restaurant in Singapore or expanding their business. The secondary target audience are investors looking to invest in Singapore's booming hospitality sphere. In order to support the needs of the target audience I identified the most common social venues in Singapore as arguably these are already crowded. Then I identified the attributes of the selected restaurant type that the target audience could open (Thai restaurants) by identifying their distribution across the different neighborhoods of Singapore, average ratings and their spatial distribution. This assessment will help them identify the nature of the existing

Thai restaurants (competitor analysis) and select the most optimal location for opening a new Thai restaurant in Singapore.

2. DATA REQUIREMENTS

Singapore is divided into 5 planning areas- Central, North, East, West and North West. These in turn are divided into several neighborhoods. I obtained the names of the planning area and their neighborhoods via Wikipedia by employing the Beautiful Soup package of Python programming language. I plotted the distribution of neighborhoods within the different planning areas



The Central region has the highest number of neighborhoods (20+) while East has the fewest neighborhoods. I geocoded the locations of these different neighborhoods using the GeoPy package. Next, I obtained the details of venues in each neighborhood namely Venue, Venue Latitude, Venue Longitude, Venue Category using the Foursquare API. Additionally, I also obtained the details of the ratings, tips and likes (for the Thai restaurants).

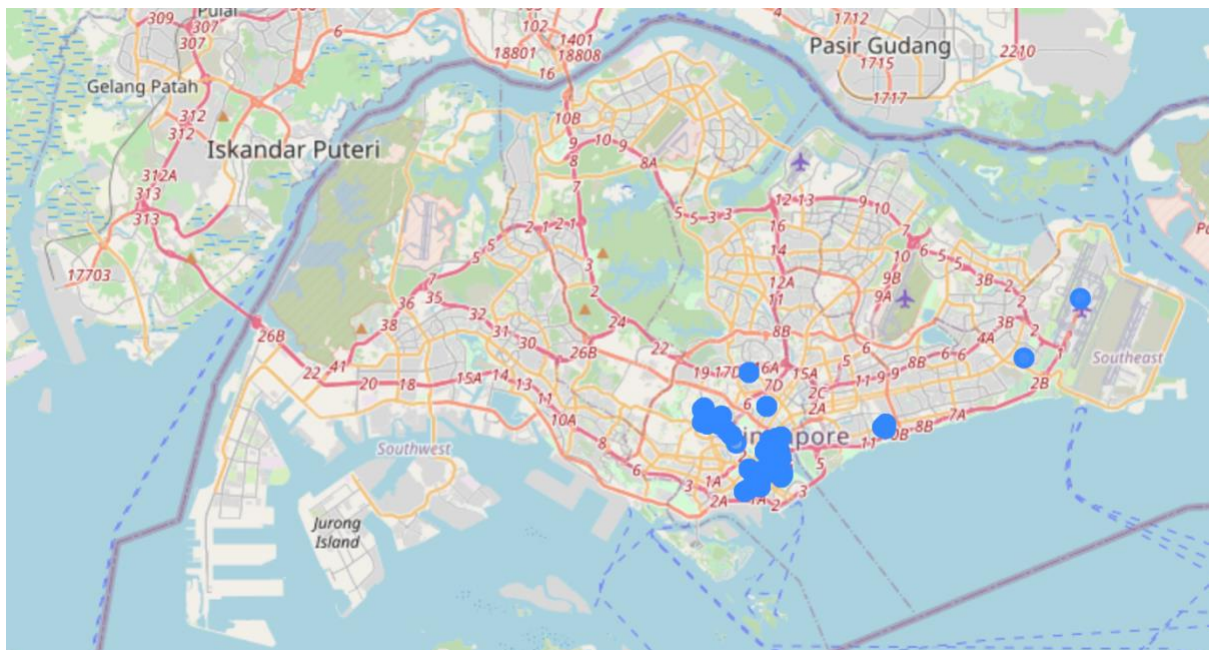
3. METHODOLOGY

My analysis depends on two major aspects- details of Singapore's planning regions (along with their geolocations) and details of social venues obtained from the Foursquare API. I obtained details of the different foursquare venues corresponding to the different neighborhoods. I carried out exploratory data analysis to identify the most common social venues in Singapore and their spatial concentration. As a part of

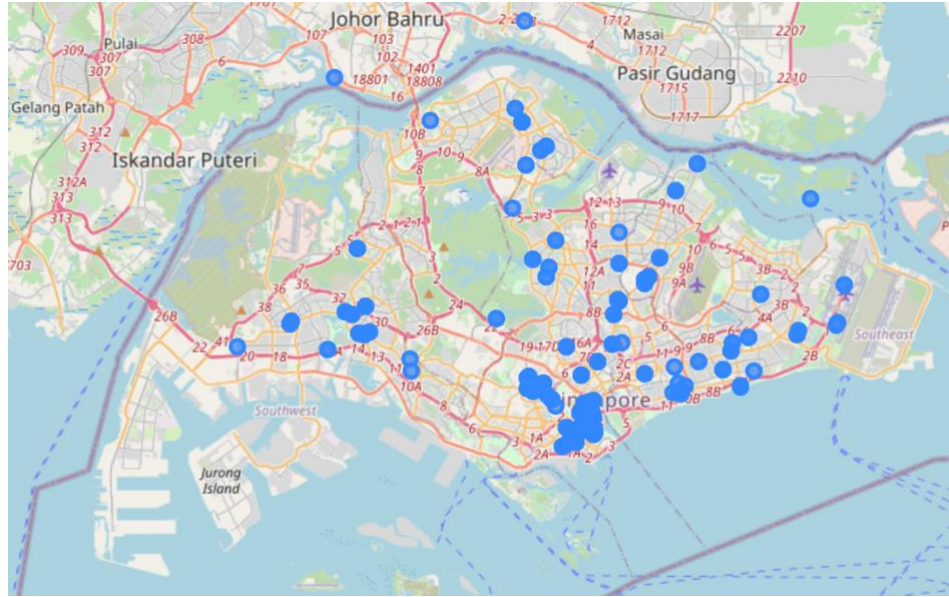
the EDA, I extracted the reviews of most popular existing Thai restaurants in Singapore as a way of scoping out locations to avoid. We can perform one hot encoding on the obtained data set and use it find the 10 most common venue category in each neighborhood. Then clustering can be performed on the dataset. Here K - Nearest Neighbor clustering technique have been used. To find the optimal number of clusters silhouette score metric technique is used. The clusters obtained can be analyzed to find the major type of venue categories in each cluster. This data can be used to suggest business people, suitable locations based on the category.

4. ANALYSIS

Exploratory data analysis revealed that the most common venue types in Singapore are hotels and Chinese restaurants. An examination of the spatial distribution of these revealed that while the hotels are clustered in Central Singapore, the Chinese restaurants, apart from being congregated in central Singapore are also distributed across the city-state



Spatial Distribution of Singapore's Hotels

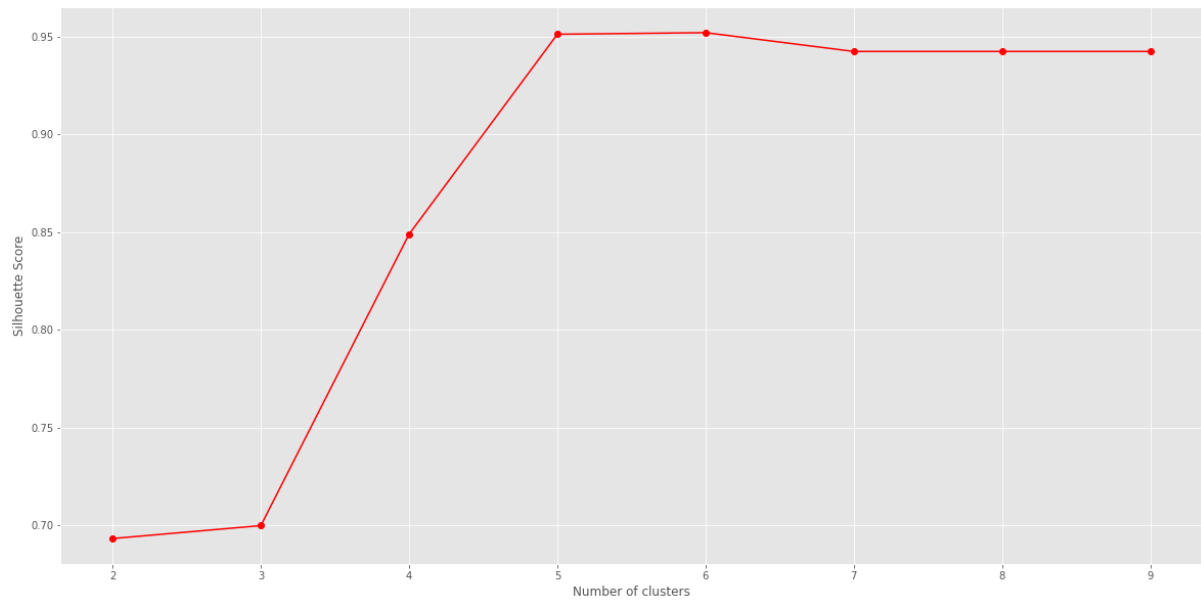


Spatial Distribution of Singapore's Chinese Restaurants

Given the popularity of hotels and Chinese restaurants, Thai restaurants are a good option for a new venture. Next, we will perform one hot encoding on the filtered data to obtain the venue categories in each neighborhood. From this we extract the top 5 most common venues for each of the neighborhoods.

Area	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Ang Mo Kio	Chinese Restaurant	Park	Thai Restaurant	Coffee Shop	Café
Bedok	Coffee Shop	Chinese Restaurant	Asian Restaurant	Dessert Shop	Beach
Bishan	Chinese Restaurant	Food Court	Park	Supermarket	Bakery
Boon Lay	Chinese Restaurant	Food Court	Coffee Shop	Japanese Restaurant	Fast Food Restaurant
Bukit Batok	Chinese Restaurant	Indian Restaurant	Japanese Restaurant	Bakery	Nature Preserve
Bukit Merah	Lake	Zoo Exhibit	English Restaurant	Event Space	Exhibit
Bukit Panjang	Zoo Exhibit	Park	Coffee Shop	Supermarket	Nature Preserve
Bukit Timah	Italian Restaurant	Chinese Restaurant	Bakery	Nature Preserve	Park

This dataset can be used for the clustering algorithm. Here, the K-means unsupervised clustering algorithm is used. It is an unsupervised machine learning technique that clusters the given data into K number of clusters. For optimal result we need to select the best value for K. Here, the silhouette score is used to find the best value for K. A range of values from 2 to 10 was considered, k-means clustering was performed on the dataset and the silhouette score was calculated and plotted on a line plot as shown in the figure.



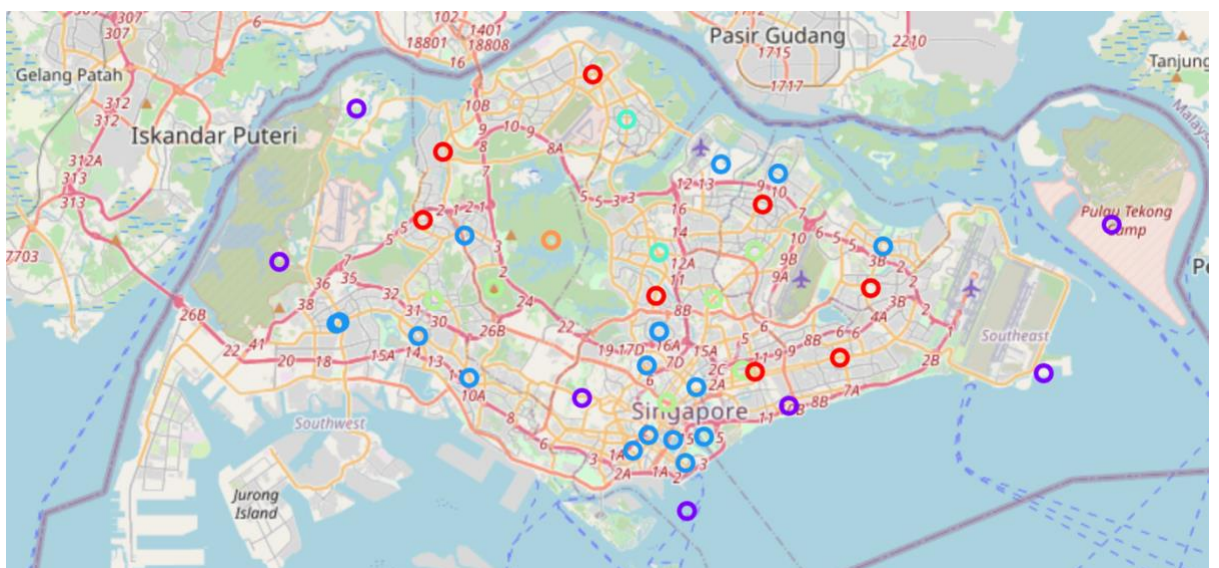
Silhouette Score for different Number of Clusters

From the plot we can see that a K value of 6 provides the best score. This K value is used for the K-Means Clustering Technique.

Additionally I obtained the data pertaining to additional details relating to Thai restaurants such as Likes and Tips with the view of identifying the neighborhoods with popular Thai restaurants.

5. RESULTS

The results of the k-means clustering provided the spatial distribution of clusters of Thai restaurants, including the areas with the highest number of clusters.



Clusters of Thai Restaurants in Singapore

The 5th cluster just had 1 Thai restaurant while clusters 1 and 2 had the highest number of Thai restaurants.

1	18
2	16
0	8
4	7
3	2
5	1

Let's examine the 6 clusters and find the discriminating venue categories that distinguish each cluster. For this purpose, let's also look into the most common venue category in each cluster. For instance, coffee shops and Chinese restaurants are the most common venues in Cluster 1:

	Region	ClusterLab	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Central	0	Chinese Restaurant	Park	Food Court	Supermarket	Bakery
22	East	0	Chinese Restaurant	Coffee Shop	Dessert Shop	Asian Restaurant	Beach
25	East	0	Hotel	Coffee Shop	Café	Supermarket	Ice Cream Shop
26	East	0	Coffee Shop	Chinese Restaurant	Bakery	Dessert Shop	Park
30	North	0	Park	Chinese Restaurant	Seafood Restaurant	Coffee Shop	Café

I also obtained the ratings and likes data pertaining to Thai restaurants in the different regions and neighborhoods.

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
0	Central	Bishan	587a1563561ded2967529265	Tomyum Mama	17	8.7	7
1	Central	Bishan	4c1a0400838020a1cd83e661	E-Sarn Thai Corner	21	7.4	19
2	Central	Bishan	5602a497498eeb82cd6918ab	Taste of Thailand (Original)	9	7.4	3
3	Central	Bukit Timah	51df898a498e56e0e64c5409	Royal Thai	13	7.8	7
4	Central	Bukit Timah	4b83c01af964a520130f31e3	E-Sarn Thai Cuisine	8	8.0	15

Popular Thai Restaurants by Planning Region

The most popular Thai restaurants were located in the Central and East planning regions

	Borough	Average Rating
0	Central	8.393750
1	East	8.110000
2	North	7.757143
3	North-East	5.166667
4	West	0.000000

Average Ratings of Thai Restaurants By Planning Region

I also identified the average ratings of Thai restaurants by neighborhoods

	Neighborhood	Average Rating
10	Downtown Core	8.8
16	Marina East	8.8
17	Marina South	8.8
18	Outram	8.8
27	Singapore River	8.8

6. DISCUSSIONS

While cluster 1 and cluster 2 have the maximum number of restaurants, cluster 4 and cluster 3 have Thai restaurants among their most frequent locations. Cluster 5 has the lowest number of restaurants. So, areas such as Holland park which have the lowest number of Thai restaurants can be considered as viable area for opening a Thai restaurant owing to lower competition. Planning areas such as Changi and Serangoon road which have a high concentration of Thai restaurants can have a higher competition for a newer Thai establishment. Additionally both the Central and East Planning region have the most popular Thai restaurants. These too should be avoided. While I only considered Foursquare data, other datasets, such as those relating to

property prices and transport links can be included to better pinpoint optimal localities.