The Battle of Neighborhoods

IBM DAta Science capstone Project

Niranjan Appaji

2021

Contents

[**1.** **Introduction** 2](#_Toc74258700)

[**2.** **Problem Description** 2](#_Toc74258701)

[**3.** **Target Audience** 2](#_Toc74258702)

[**4.** **Data Download** 3](#_Toc74258703)

[**5.** **Approach** 3](#_Toc74258704)

[**6.** **Step by Step Explanation:** 4](#_Toc74258705)

[**7.** **Conclusion** 11](#_Toc74258706)

The Battle of Neighborhoods

IBM Data Science Capstone Project

June 10, 2021

1. **Introduction**

Singapore with an estimated 2019 population of 5.70 million, is the second-densest sovereign state in the world, after the microstate Monaco. Singapore is multiracial and multicultural country with ethnic Chinese (76.2% of the citizen population), Malays (15.0%), and ethnic Indians (7.4%).

Singapore has marketed itself as a food paradise and is home of some of Asia's best restaurants. Singapore food culture include the cuisines of the Malays, the Chinese, and the Indians as well as Indonesian, and Western traditions (particularly English and Portuguese-influenced Eurasian, known as Kristang). Influences from neighboring regions such as Japan, Korea, and Thailand are also majorly present.

This means the restaurant business in Singapore is highly competitive as there are variety of cuisines available to the customers. Thus, any new business venture needs to be analyzed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market.

1. **Problem Description**

It is pretty evident that to survive in a competitive market, it is very important to strategically plan. The objective of this project is to recommend the best neighborhood in Singapore to open a Thai Restaurant.

1. **Target Audience**

The objective is to locate and recommend which neighborhood of Singapore will be best choice to start a Thai restaurant. The business owners also expect to understand the rationale of the recommendations made.

This analysis will help:

* the restaurant owners who wants to expand their business
* to those who are looking to open a new restaurant

1. **Data Download**

To get the data, I made use of Airbnb data that was publicly shared on the internet. Download the listings.csv file from the below link:

<http://insideairbnb.com/get-the-data.html>

The data contains Neighborhood\_Groups, Neighborhoods, Latitude and Longitude data of Singapore along with other columns too. But I have made use of only these 4 columns. The Singapore Coordinates data in the file "Singapore\_coordinates.csv" is imported in the project.

To get the venue data, use Foursquare API using the link:

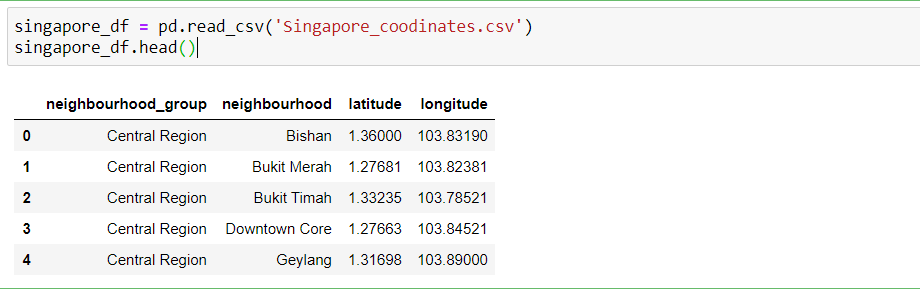
<https://foursquare.com/developers/apps>

1. **Approach**

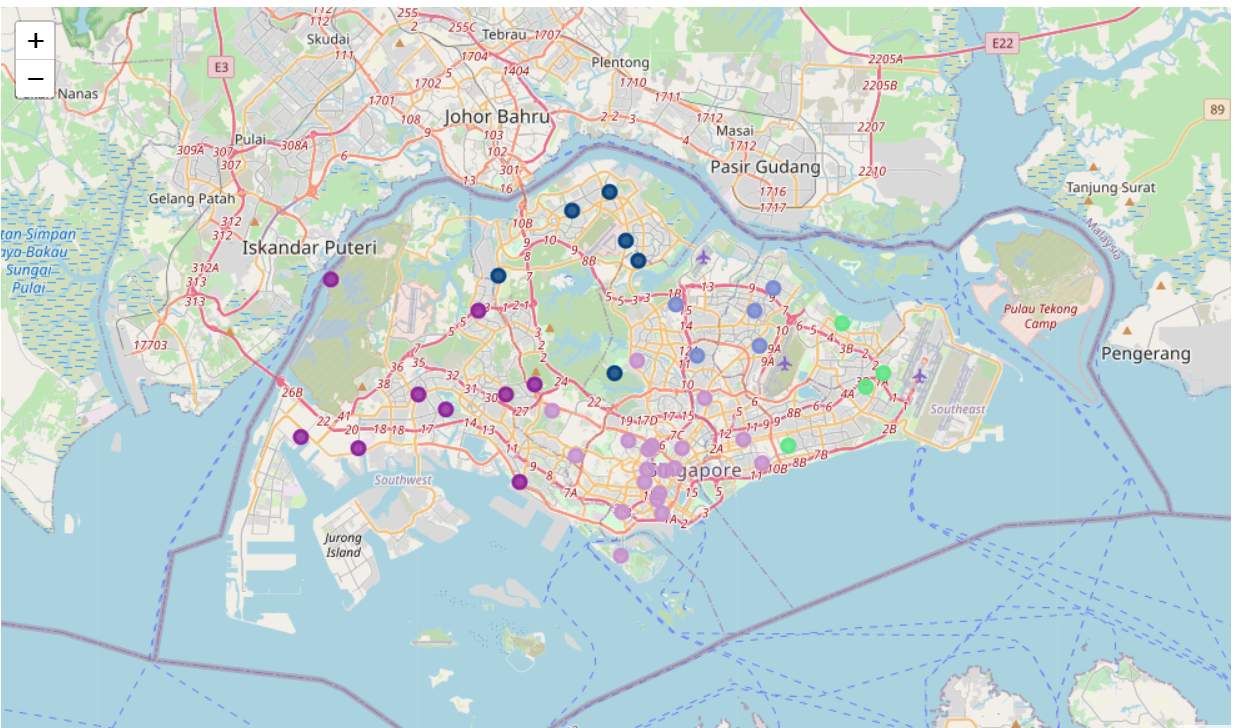
* Import the Singapore Neighborhood Geospatial Data
* Using Foursquare API we will find all the nearby venues for each neighborhood
* Filter out all venues with category of “Restaurant” in it.
* Use elbow method to find the optimal K value
* Cluster the data using KMeans
* Recommend the best neighborhoods to open a Thai Restaurant

1. **Step by Step Explanation:**

* Import the data of Singapore having geocoordinates using pandas library. Below is the sample of imported data.



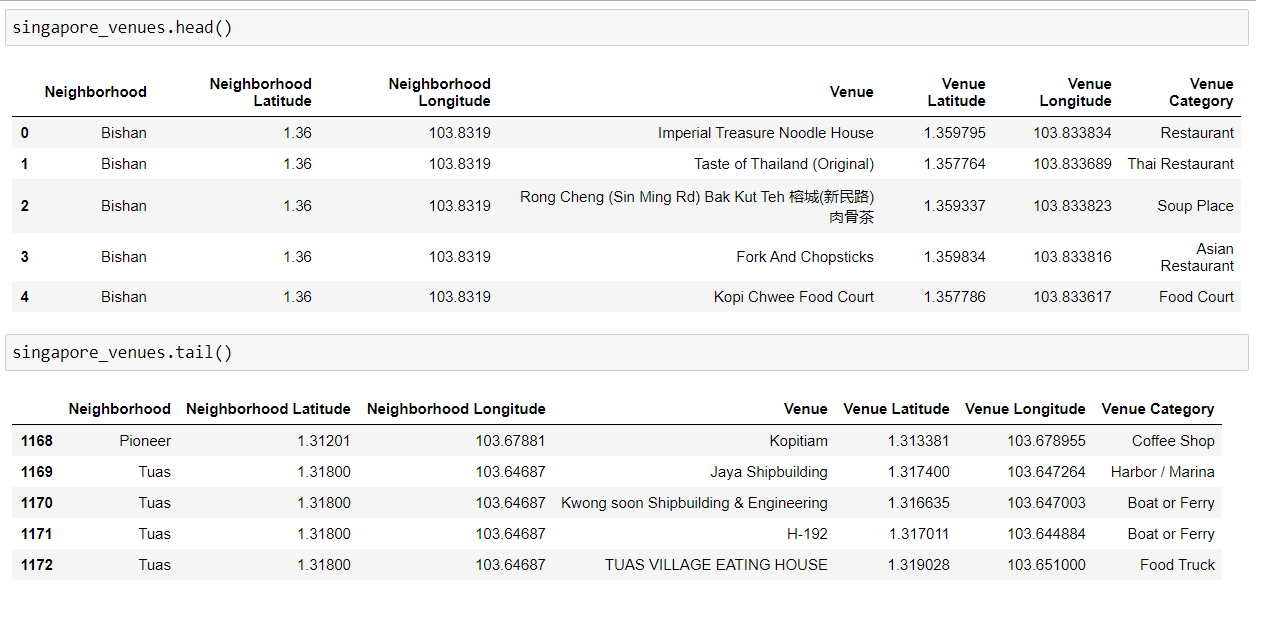
* Illustrate the above data on map using Folium library by plotting all the neighborhoods as below.



* Now let us explore the neighborhoods within 500 meters, by retrieving the venues nearby. Limit the venues to maximum number of 100 to each neighborhood.
* Below is the complete list of neighborhoods.

|  |  |  |
| --- | --- | --- |
| Bishan | Changi | Bukit Batok |
| Bukit Merah | Pasir Ris | Bukit Panjang |
| Bukit Timah | Central Water Catchment | Choa Chu Kang |
| Downtown Core | Mandai | Clementi |
| Geylang | Sembawang | Jurong East |
| Kallang | Sungei Kadut | Jurong West |
| Marine Parade | Woodlands | Pioneer |
| Museum | Yishun | Tuas |
| Newton | Ang Mo Kio | Western Water Catchment |
| Novena | Hougang | Toa Payoh |
| Orchard | Punggol | Tampines |
| Outram | Sengkang | Bedok |
| Queenstown | Serangoon | Southern Islands |
| River Valley | Rochor | Tanglin |
| Singapore River |  |  |

* Below is the sample list of local venues within the 500 meters vicinity of neighborhoods.



* Various venue categories that are retrieved around the neighborhood are as below.

['Restaurant' 'Thai Restaurant' 'Soup Place' 'Asian Restaurant'

'Food Court' 'Bubble Tea Shop' 'Bus Station' 'Grocery Store'

'Tennis Court' 'Coffee Shop' 'Café' 'Chinese Restaurant'

'Vegetarian / Vegan Restaurant' 'Cosmetics Shop' 'Convenience Store'

'Cafeteria' 'Italian Restaurant' 'Historic Site' 'Bus Stop'

'Gym / Fitness Center' 'Bistro' 'Soccer Field' 'Dessert Shop' 'Bakery'

'Hotel' 'Japanese Restaurant' 'Tapas Restaurant' 'Sushi Restaurant'

'Bagel Shop' 'Bar' 'Kebab Restaurant' 'Cocktail Bar' 'BBQ Joint'

'Spanish Restaurant' 'Ramen Restaurant' 'Salad Place' 'Bookstore'

'Wine Bar' 'Noodle House' 'Korean Restaurant' 'Ice Cream Shop'

'Yoga Studio' 'Tea Room' 'Dumpling Restaurant' 'Diner' 'Deli / Bodega'

'Mosque' 'Burrito Place' 'Mexican Restaurant' 'Spa' 'Nightclub'

'Sandwich Place' 'Music Venue' 'Discount Store' 'American Restaurant'

'French Restaurant' 'Fast Food Restaurant' 'Cantonese Restaurant'

'Basketball Court' 'Multiplex' 'Fried Chicken Joint'

'Vietnamese Restaurant' 'Breakfast Spot' 'Steakhouse'

'Seafood Restaurant' 'General Entertainment' 'Gym Pool' 'Gym' 'Building'

'Pool' 'Sports Club' 'Creperie' 'Pizza Place' 'Bridal Shop'

'Shopping Mall' 'Clothing Store' 'Indian Restaurant'

'Sporting Goods Shop' 'Hostel' 'Climbing Gym' 'Dim Sum Restaurant'

'North Indian Restaurant' 'Plaza' 'Frozen Yogurt Shop'

'Mediterranean Restaurant' 'Bed & Breakfast' 'Snack Place'

'Toy / Game Store' 'Malay Restaurant' 'Soccer Stadium'

'Fruit & Vegetable Store' 'Bowling Alley' 'Department Store' 'Pub'

'Karaoke Bar' 'Flower Shop' 'Australian Restaurant' 'Supermarket'

'Gift Shop' 'Movie Theater' 'Whisky Bar' 'Miscellaneous Shop'

'Gaming Cafe' 'Indonesian Restaurant' 'Other Repair Shop'

'Fish & Chips Shop' 'Museum' 'Hotpot Restaurant' 'Park'

'Cuban Restaurant' 'Concert Hall' 'Music Store' 'Halal Restaurant'

'Gastropub' 'Burger Joint' 'Arts & Crafts Store'

'Modern European Restaurant' 'Monument / Landmark' 'Mattress Store'

'Accessories Store' 'Garden' 'Hotel Bar' 'Shoe Store' 'Electronics Store'

'Nail Salon' 'Chocolate Shop' 'Boutique' 'Peking Duck Restaurant'

'Kids Store' 'Hainan Restaurant' 'Beer Garden' 'Buddhist Temple'

'Other Great Outdoors' 'Candy Store' 'Pharmacy' 'Salon / Barbershop'

'Metro Station' 'Pet Store' 'Playground' 'Athletics & Sports' 'Beer Bar'

'Indoor Play Area' 'Jewelry Store' 'Middle Eastern Restaurant'

'Event Space' 'Sports Bar' 'Buffet' 'Shabu-Shabu Restaurant' 'Art Museum'

'Dive Bar' 'College Classroom' 'Brewery' 'Waterfront' 'Shopping Plaza'

'River' 'Pet Café' 'Theme Park Ride / Attraction' 'Lounge'

'History Museum' 'Theme Park' 'Souvenir Shop' 'Resort' 'Dance Studio'

'Performing Arts Venue' 'Beach' 'Sculpture Garden' 'Theater' 'Water Park'

'Casino' 'Hotel Pool' 'Campground' 'Donut Shop' 'Lingerie Store'

'Library' 'Food' 'Moving Target' 'Thrift / Vintage Store'

'Recreation Center' 'Food & Drink Shop' 'Ski Chalet' 'Boat or Ferry'

'Furniture / Home Store' 'Fishing Spot' 'Train Station' 'Club House'

'Stadium' 'Lake' 'Bus Line' 'High School' 'Food Truck'

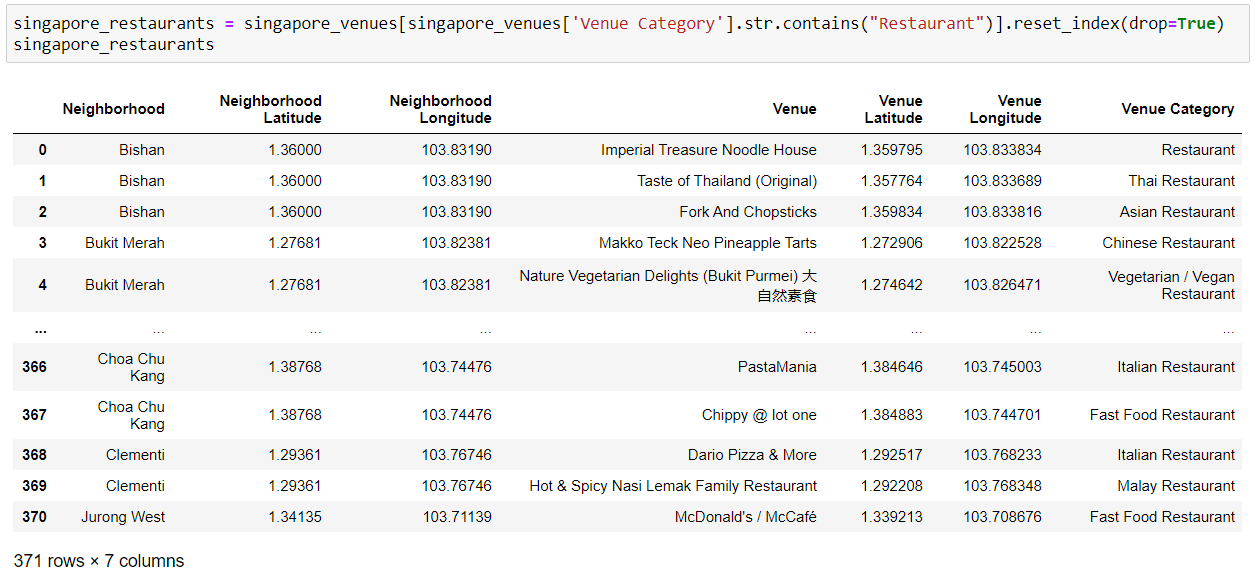
'Szechuan Restaurant' 'South Indian Restaurant'

'Southern / Soul Food Restaurant' 'Roof Deck' 'Video Store'

'Rental Car Location' 'Gas Station' 'Nature Preserve' 'Scenic Lookout'

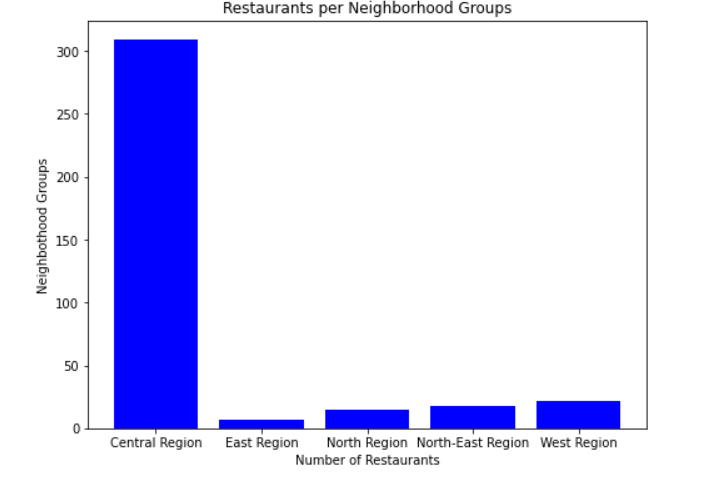
'Bike Trail' 'Trail' 'Road' 'Portuguese Restaurant' 'Harbor / Marina']

* Filter the data with Venue Category contains ‘Restaurant’.



* Before clustering the data, find out the number of restaurants per neighborhood group.

Clearly below figure shows that central region has a greater number of restaurants.



* Also plotted what type of restaurants are more in number using a word cloud. Chinese

Restaurants & Japanese restaurants are more in number around the neighborhoods. Our

interest type Thai Restaurant is not in big number.



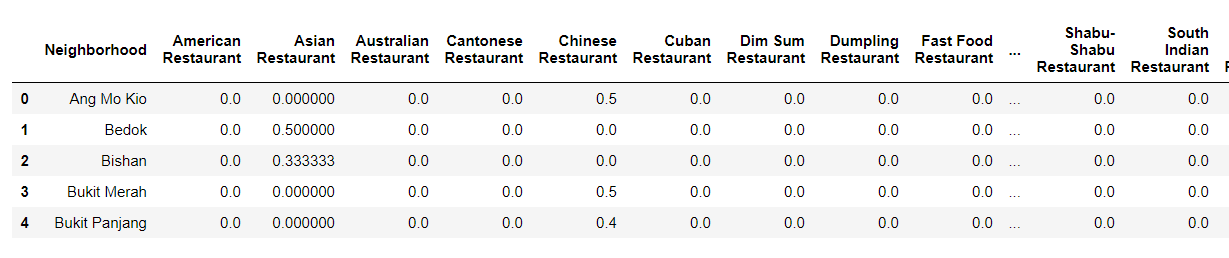
* Venue categories variable is a categorical column and to make use of this categorical

Variable convert it into dummy variables using get\_dummies() method. This method is

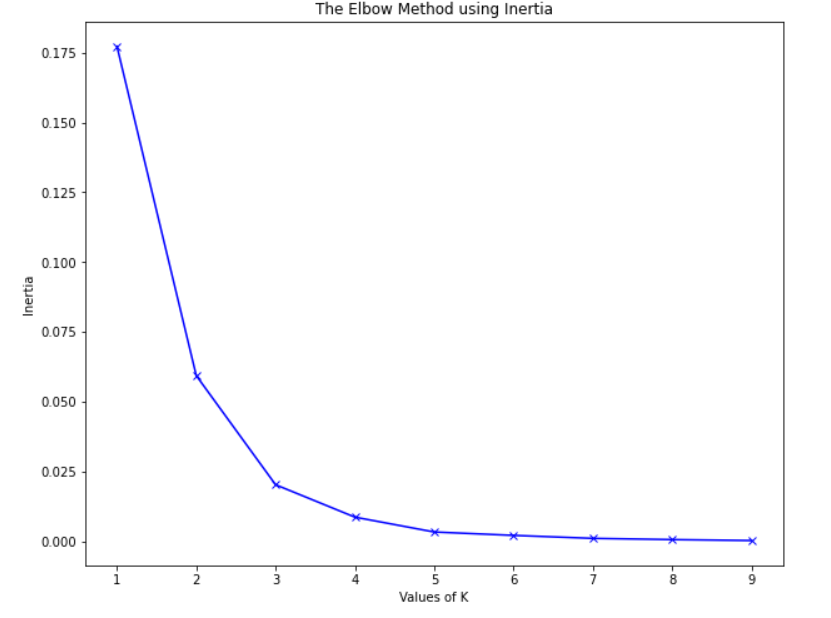
called one-hot encoding.

* A dummy variable is a binary variable that indicates whether a separate categorical variable takes on a specific value. Take the average of the restaurant categories grouping

neighborhoods.

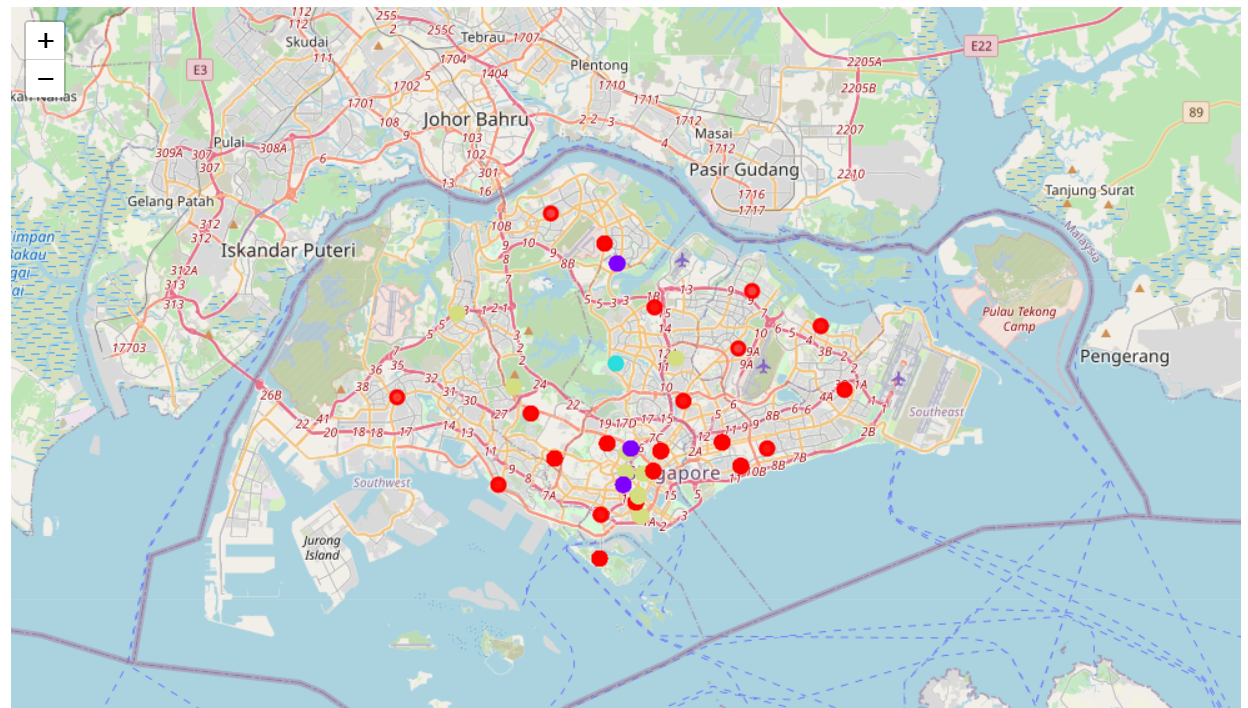


* Now, let us analyze which is the best neighborhood to open a Thai Restaurant using KMeans clustering. To do this, first we will have to find out the optimal K-value to cluster.
* To find the optimal K-value, we use the elbow method. Plot the graph. As per the graph, the best K- value is found to be 4.



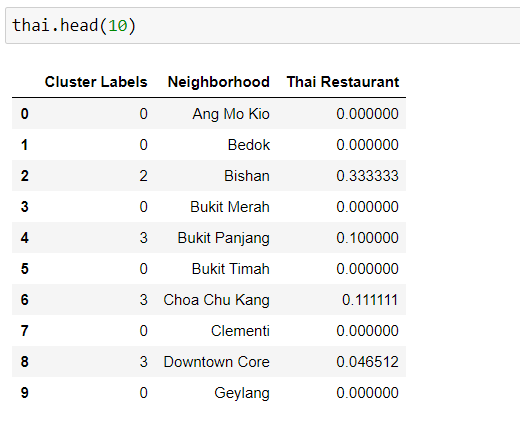
* Cluster the data using sklearn KMeans library with 4 clusters. Plot the clustered data on

Singapore Map to visualize the clustering process.



* Our interest type i.e. Thai Restaurant after clustering and taking mean of it per

neighborhood is as below.



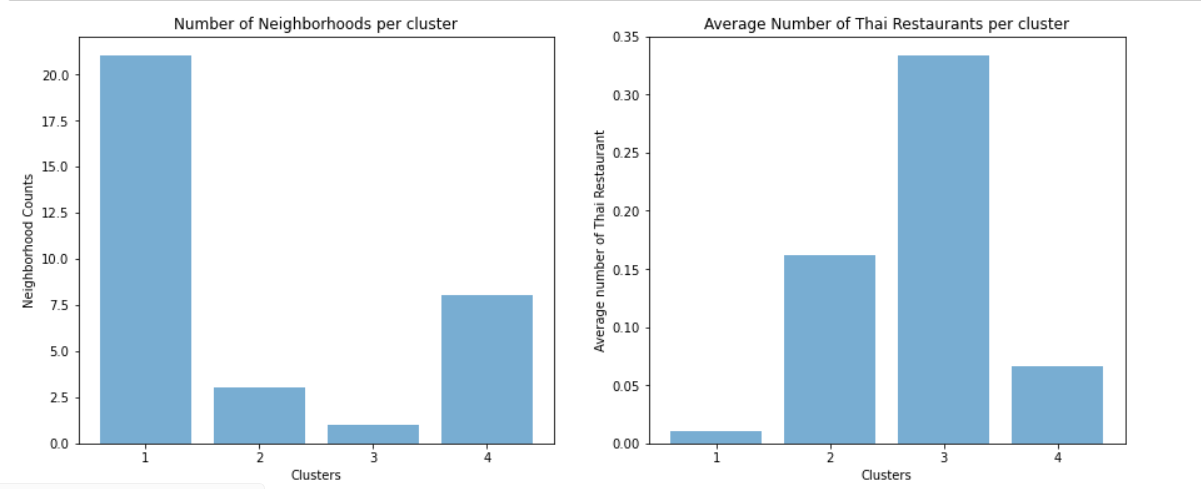
* Firstly, let us analyze the number of neighborhoods per cluster by plotting the clusters and

Count of neighborhoods in the cluster.

* Secondly, we will understand the average number of Thai restaurants per cluster, so that

we will get a fair idea on which neighborhood belonging to the cluster is the best place to

start a Thai restaurant.



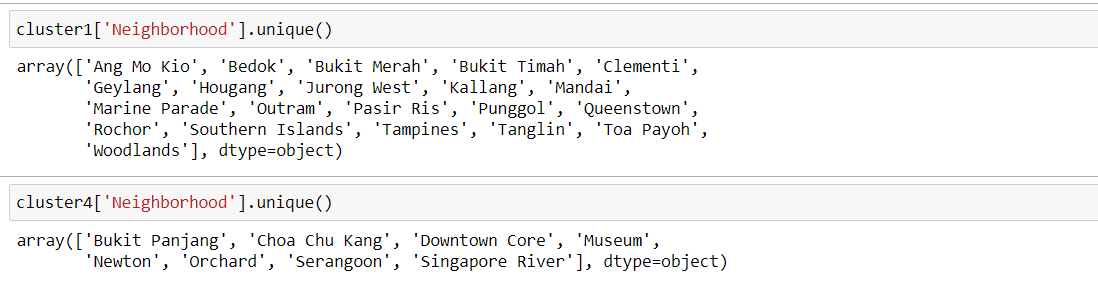
1. **Conclusion**

* Based on our analysis, cluster 1 & 4 is having a greater number of neighborhoods. But the

number of Thai restaurants in both the clusters are very less.

* So, neighborhoods coming under these two clusters are more favorable for opening a Thai

cuisine restaurant.



* Whereas, the number of neighborhoods in cluster 3 are low and the number of Thai

restaurants are way more compared to other clusters.

* Hence the neighborhoods in this cluster will have high competition and is risky to open a

Thai restaurant.