



# IBM Data Science

## Capstone Project

**Picking a suitable location for a new restaurant**  
in Western Singapore

## Table of Contents

01. Introduction .....	2
1.1 The problem and background of the study .....	2
1.2 Audience .....	2
02. Data .....	2
03. Methodology.....	3
3. 1 Data Collection and preparation .....	3
3.2 K-means Clustering .....	4
04. Analysis .....	5
05. Results.....	9
06. Discussion and Recommendations .....	10
07. Limitations.....	10
08. Conclusions .....	10

## 01. Introduction

A restaurant is a business that prepares and serves food and drinks to customers. Meals are generally served and eaten on the premises, but many restaurants also offer take-out and food delivery services. Restaurants vary greatly in appearance and offerings, including a wide variety of cuisines and service models ranging from inexpensive fast-food restaurants and cafeterias, to mid-priced family restaurants, to high-priced luxury establishments.

Restaurants can be also considered as places where people can get together, have discussions, while eating food. So, restaurants play a major role in the society, and it greatly affects the lifestyle of people.

### 1.1 The problem and background of the study

In the project I have focused on a business scenario where the owner of a high-end restaurant chain is deciding to open a new restaurant in the Western Region of Singapore since he is not having a restaurant in Singapore. But he is uncertain about the city which he will choose to open the restaurant. The intention of the study is to find an optimal location in the western region for his restaurant.

### 1.2 Audience

Here we are assuming a random business owner to whom we are addressing this report, but in the real world this restaurant owner can be treated as a person who is interested in opening a restaurant in the western region of Singapore, thus this analysis could be useful for a group of restaurant owners.

## 02. Data

To perform the analysis, the following data will be used.

1. List of cities in Singapore.
2. Geo-coordinates of each city in Singapore
3. Top venues in the cities in western region.

The list of cities along with their coordinates has been obtained from the web site, <https://simplemaps.com/data/sg-cities>

Top venues have been obtained with the help of Foursquare API.

## 03. Methodology

### 3. 1 Data Collection and preparation

The list of cities in Singapore along with their coordinates has been obtained from the web site, <https://simplemaps.com/data/sg-cities>. Due to the limitations in the website in collecting the data, the details of only 75 cities were obtained. Thus, going forward this list of 75 cities will be used in the analysis.

The data obtained from the csv file were then transformed into a pandas DataFrame in the Jupyter Notebook for the analysis. A snapshot of the DataFrame obtained is as follows.

	City	Latitude	Longitude	Region
0	Singapore	1.3000	103.8000	Central Singapore
1	Keat Hong Village	1.3778	103.7442	South West
2	Choa Chu Kang New Town	1.3833	103.7500	South West
3	Bukit Panjang Estate	1.3817	103.7525	South West
4	Jalan Kayu	1.3972	103.8719	Central Singapore
5	Bukit Sembawang Estate	1.3956	103.8692	Central Singapore
6	Kampung Serangoon Kechil	1.3900	103.9072	North East
7	Kampung Cutforth	1.3819	103.7364	South West
8	Mok Peng Hiang Estate	1.4456	103.7969	North West
9	Woodlands New Town	1.4363	103.7867	North West
10	Kampung Pinang	1.3811	103.8942	Central Singapore
11	Serangoon	1.3756	103.8972	Central Singapore
12	Kampung Sungai Tengah	1.3833	103.9000	Central Singapore
13	Kampung Tongkang Pechah	1.3900	103.8764	Central Singapore
14	Kampung Teban	1.3786	103.9078	North East
15	Kangkar	1.3761	103.9017	North East
16	Jurong Town	1.3342	103.7228	South West
17	Kampung Ulu Jurong	1.3500	103.7167	South West
18	Seletar Hills Estate	1.3844	103.8681	Central Singapore

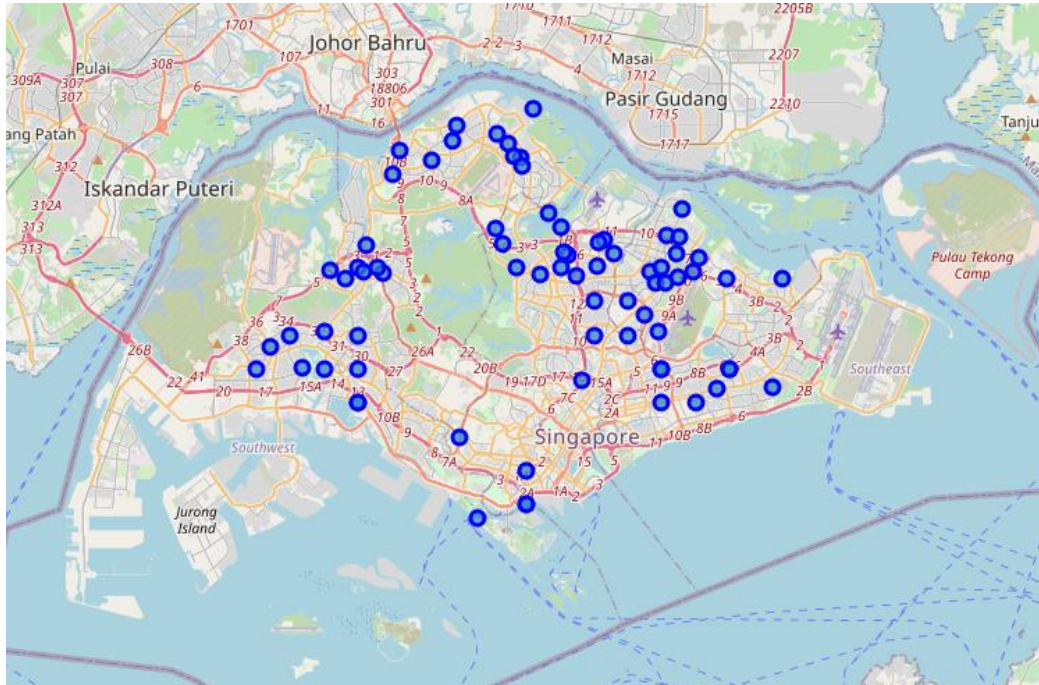
### 3.2 K-means Clustering

K-means clustering algorithm is the most selected technique to cluster data. K-means is a nonhierarchical clustering and use looping to group data into K groups. The K-means clustering start the iterative process by finding the initial centroid, or central point, of each group by randomly selecting representative data from raw data to be a centroid in each K data groups. Then assign each data to the closest group by calculating the Euclidean distance between each data record to each centroid to allocate the data record to the nearest group. After that each cluster will find new centroid to replace the initial one and repeat steps of Euclidean distance computation to group data members and send each member to group of the nearest centroid. The process will stop when each group has stable centroid and members do not change their groups.

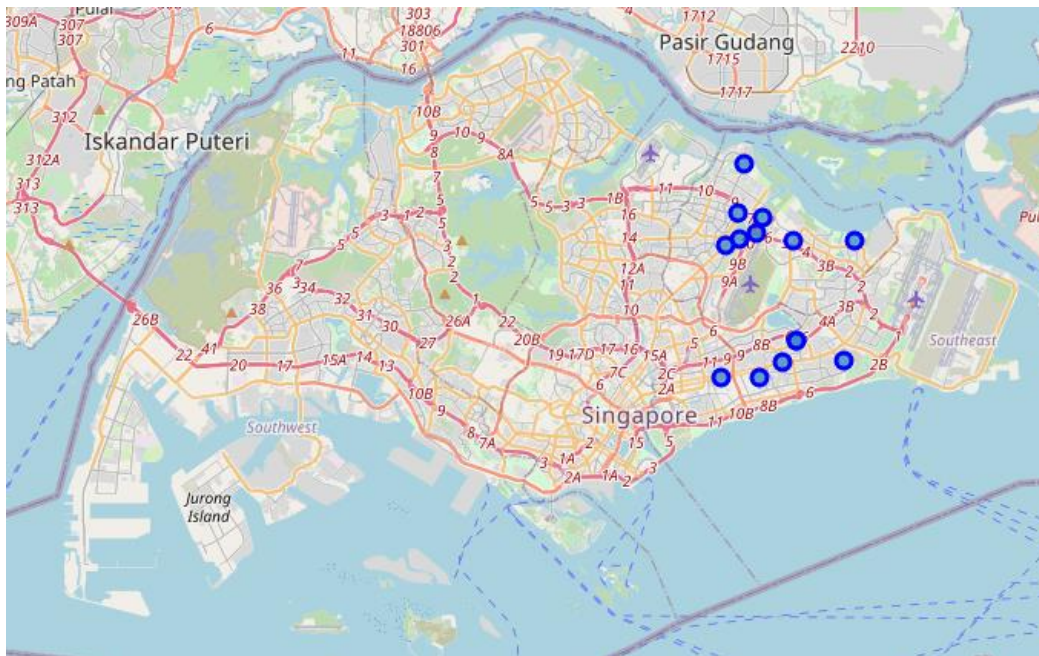
To achieve our objective K-means clustering has been used, with the optimal number of K clusters.

## 04. Analysis

In the first step, a visual representation of how the cities are situated in Singapore has been created.



Since our objective is to find an optimal location in the western region in Singapore for the restaurant, the data has been filtered and only the cities in Western Region has been obtained.



In the next step of the analysis, the cities were explored in greater detail. The venues were collected for each city via Foursquare API. After arranging the data there were up to 100 venues for each city. Venues are collected within a radius of 1000 meters from the point of city coordinates. The following table shows some venues from the first city.

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Kampong Serangoon Kechil	1.39	103.9072	Ice Talk Xpress	1.394152	103.912813	Bubble Tea Shop
1	Kampong Serangoon Kechil	1.39	103.9072	Subway	1.392211	103.904955	Sandwich Place
2	Kampong Serangoon Kechil	1.39	103.9072	KFC	1.392545	103.904667	Fast Food Restaurant
3	Kampong Serangoon Kechil	1.39	103.9072	Nihon Mura Express	1.392104	103.905114	Japanese Restaurant
4	Kampong Serangoon Kechil	1.39	103.9072	McDonald's	1.393536	103.913326	Fast Food Restaurant
5	Kampong Serangoon Kechil	1.39	103.9072	Master Crab @ Kopitiam	1.396699	103.908607	Chinese Restaurant
6	Kampong Serangoon Kechil	1.39	103.9072	Foodpark	1.388895	103.909468	Coffee Shop
7	Kampong Serangoon Kechil	1.39	103.9072	McDonald's	1.392176	103.905051	Fast Food Restaurant
8	Kampong Serangoon Kechil	1.39	103.9072	Heavenly Wang	1.385736	103.901558	Breakfast Spot
9	Kampong Serangoon Kechil	1.39	103.9072	Long John Silver's	1.392182	103.905093	Fast Food Restaurant

The number of Venues for each city in the western region has also been obtained.

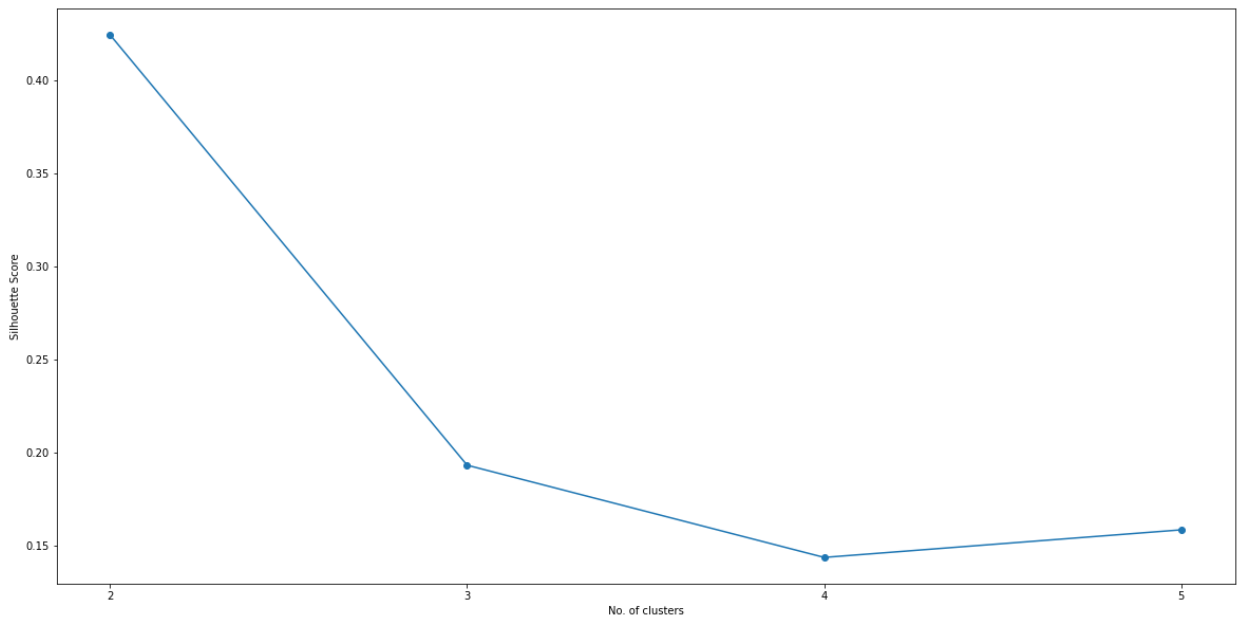
City	Venue
Bedok New Town	89
Bedok Ville	16
Kampong Beremban	6
Kampong Chai Chee	98
Kampong Loyang	44
Kampong Pachitan	71
Kampong Pasir Ris	33
Kampong Punggol	39
Kampong Serangoon Kechil	26
Kampong Sungai Blukar	22
Kampong Teban	14
Kampong Ubi	100
Kangkar	48
Ulu Bedok	98

For analysis in the cities, the focus is on venue categories. Therefore, we use one hot encoding to create dummy variables for the categories so the data set could be used to apply machine learning techniques.

Then the following table with the 10 most common venues in each city has been obtained. (The table only displays the first 5 rows)

	City	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	Bedok New Town	Coffee Shop	Chinese Restaurant	Café	Food Court	Noodle House	Supermarket	Sandwich Place	Fast Food Restaurant	Bakery	Indian Restaurant
1	Bedok Ville	Noodle House	Halal Restaurant	Food Court	Malay Restaurant	Smoke Shop	Chinese Restaurant	Japanese Restaurant	Basketball Court	Supermarket	Golf Course
2	Kampung Beremban	Airport	Pet Store	Farm	Golf Course	River	Bus Line	Food Truck	Dive Bar	Dog Run	Drugstore
3	Kampung Chai Chee	Coffee Shop	Chinese Restaurant	Food Court	Noodle House	Café	Asian Restaurant	Bus Station	Bakery	Thai Restaurant	Supermarket
4	Kampung Loyang	Coffee Shop	Fast Food Restaurant	Bus Station	Chinese Restaurant	Park	Thai Restaurant	Resort	Food Court	Miscellaneous Shop	Diner

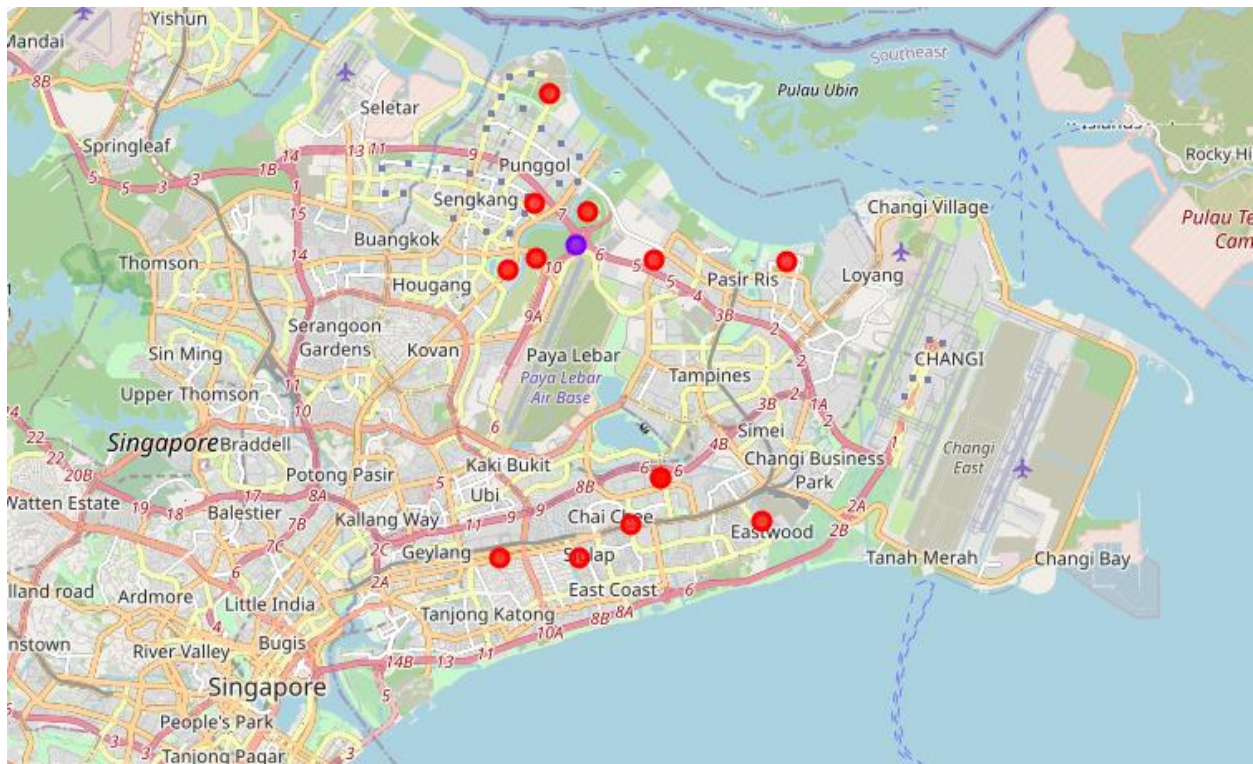
In the clustering process, the K-Means algorithm has been used. To identify the optimal number of K, the Silhouette Scoring method has been used, and the graph obtained were as follows. The optimal number of K's was 2. Therefore, 2 clusters have been used.





Adding the cluster labels to the dataset we get the following table, and the clusters are also shown in the map.

	City	Latitude	Longitude	Region	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
6	Kampung Serangoon Kechil	1.3900	103.9072	North East	0	Fast Food Restaurant	Food Court	Playground	Coffee Shop	Bus Station	Gym / Fitness Center	Basketball Court	Chinese Restaurant	Bus Stop	Sandwich Place
14	Kampung Teban	1.3786	103.9078	North East	0	Park	Food Court	New American Restaurant	Playground	Drugstore	Golf Course	Trail	Indonesian Restaurant	Bus Station	Supermarket
15	Kangkar	1.3761	103.9017	North East	0	Food Court	Coffee Shop	Bus Station	Supermarket	Playground	Chinese Restaurant	Café	Fast Food Restaurant	Asian Restaurant	Drugstore
22	Kampung Beremban	1.3814	103.9158	North East	1	Airport	Pet Store	Farm	Golf Course	River	Bus Line	Food Truck	Dive Bar	Dog Run	Drugstore
28	Kampung Chai Chee	1.3333	103.9333	South East	0	Coffee Shop	Chinese Restaurant	Food Court	Noodle House	Café	Asian Restaurant	Bus Station	Bakery	Thai Restaurant	Supermarket
29	Ulu Bedok	1.3333	103.9333	South East	0	Coffee Shop	Chinese Restaurant	Food Court	Noodle House	Café	Asian Restaurant	Bus Station	Bakery	Thai Restaurant	Supermarket
30	Kampung Pachitan	1.3167	103.9167	South East	0	Coffee Shop	Asian Restaurant	Indian Restaurant	Chinese Restaurant	Gastropub	Convenience Store	Pizza Place	Food Court	Restaurant	Fast Food Restaurant
31	Bedok New Town	1.3236	103.9273	South East	0	Coffee Shop	Chinese Restaurant	Café	Food Court	Noodle House	Supermarket	Sandwich Place	Fast Food Restaurant	Bakery	Indian Restaurant
33	Bedok Ville	1.3244	103.9544	South East	0	Noodle House	Halal Restaurant	Food Court	Malay Restaurant	Smoke Shop	Chinese Restaurant	Japanese Restaurant	Basketball Court	Supermarket	Golf Course
43	Kampung Ubi	1.3167	103.9000	South East	0	Asian Restaurant	Chinese Restaurant	Coffee Shop	Food Court	Noodle House	Fast Food Restaurant	Malay Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Restaurant



## 05. Results

By looking at the clustering data, we can clearly identify the cluster which we are interested in.

### Cluster 01

The 1<sup>st</sup> cluster is exactly what we are looking for. We can observe many coffee shops and restaurants in the cluster 1.

	City	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
6	Kampung Serangoon Kechil	Fast Food Restaurant	Food Court	Playground	Coffee Shop	Bus Station	Gym / Fitness Center	Basketball Court	Chinese Restaurant	Bus Stop	Sandwich Place
14	Kampung Teban	Park	Food Court	New American Restaurant	Playground	Drugstore	Golf Course	Trail	Indonesian Restaurant	Bus Station	Supermarket
15	Kangkar	Food Court	Coffee Shop	Bus Station	Supermarket	Playground	Chinese Restaurant	Café	Fast Food Restaurant	Asian Restaurant	Drugstore
28	Kampung Chai Chee	Coffee Shop	Chinese Restaurant	Food Court	Noodle House	Café	Asian Restaurant	Bus Station	Bakery	Thai Restaurant	Supermarket
29	Ulu Bedok	Coffee Shop	Chinese Restaurant	Food Court	Noodle House	Café	Asian Restaurant	Bus Station	Bakery	Thai Restaurant	Supermarket
30	Kampung Pachitan	Coffee Shop	Asian Restaurant	Indian Restaurant	Chinese Restaurant	Gastropub	Convenience Store	Pizza Place	Food Court	Restaurant	Fast Food Restaurant
31	Bedok New Town	Coffee Shop	Chinese Restaurant	Café	Food Court	Noodle House	Supermarket	Sandwich Place	Fast Food Restaurant	Bakery	Indian Restaurant
33	Bedok Ville	Noodle House	Halal Restaurant	Food Court	Malay Restaurant	Smoke Shop	Chinese Restaurant	Japanese Restaurant	Basketball Court	Supermarket	Golf Course
43	Kampung Ubi	Asian Restaurant	Chinese Restaurant	Coffee Shop	Food Court	Noodle House	Fast Food Restaurant	Malay Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Restaurant
57	Kampung Loyang	Coffee Shop	Fast Food Restaurant	Bus Station	Chinese Restaurant	Park	Thai Restaurant	Resort	Food Court	Miscellaneous Shop	Diner
58	Kampung Sungai Bluar	Park	Fast Food Restaurant	Beer Garden	Department Store	Coffee Shop	Farm	Pool Hall	Fish & Chips Shop	Café	Fishing Spot
63	Kampung Punggol	Park	Chinese Restaurant	Seafood Restaurant	Bistro	Bike Rental / Bike Share	Thai Restaurant	Bus Stop	Fast Food Restaurant	Monument / Landmark	Modern European Restaurant
69	Kampung Pasir Ris	Pet Store	Bus Station	Grocery Store	Moving Target	Asian Restaurant	Furniture / Home Store	Italian Restaurant	Coffee Shop	Pharmacy	Dog Run

### Cluster 02

The cities which belong to cluster 2 are the ones the stakeholder should avoid considering to open the restaurant.

	City	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
22	Kampung Beremban	Airport	Pet Store	Farm	Golf Course	River	Bus Line	Food Truck	Dive Bar	Dog Run	Drugstore

## 06. Discussion and Recommendations

Based on the results we've obtained, we can advise the restaurant owner to consider the cities in the cluster 1, which are the perfect locations for opening a restaurant. These are the cities where coffee shops, restaurants are very frequent. And he should avoid the cities under cluster 2, where no type of restaurant is preferred going to.

## 07. Limitations

The analysis was performed only on 75 cities in Singapore, due to the limitations of the website of obtaining the data.

The analysis on performed on city level.

When collecting the venues, a 1000-meter radius was considered around the center coordinates of the cities, and the number of collected venues for each city was limited to 100.

In the clustering process only the top 10 venues in each city in western Region of Singapore was considered.

## 08. Conclusions

In the analysis we have discussed the process of coming up with an answer for a real-life business scenario. The analysis was performed based on the tools of data science and relied heavily on the use of Python and Python libraries such as Pandas, Scikit learn, Folium etc. Data were collected from different sources in various formats. For the clustering process the K-means clustering process has been used.

From the results obtained we can clearly identify, which cluster we should consider in opening a new restaurant in the western region of Singapore. From the analysis it was found out that only one city should not be considered, since restaurants are not in the top 10 venues in that city. So, the owner should choose a city from the cluster one.