

# Applied Data Science Capstone Project

## Food Location in Bangkok Feel Like Home

Sira Haruethaipree

July 22, 2021

### 1. Introduction/Business problem

Bangkok, the capital of Thailand, is one of the world's top tourist destination cities. Bangkok is a truly cosmopolitan city, home to people from around the world. Consequently, Bangkok is various in culture and cuisine. You can meet people of different nationalities who live or work here. That has resulted in many ethnically diverse restaurants. that make the foreign do not need to much adjust eating Bangkok. Because there are many types of food on this city.

For This project, I would like to recommend it to tourists or those who must transfer to work in Bangkok for a long time. this block will help visitor for finding the national restaurants location as you prefer by identifying the most type of food in which district and it is part of the decision. that help you choose to stay in which district with a restaurant in the way that you are need.

Let us see how many districts there are and how they are distributed:



Credit: <https://maps-bangkok.com/bangkok-district-map>

## 2.Description of the data

As a request from the Applied Data Science Capstone Project assignment task uses the data from foursquare for scraping data restaurants in Bangkok. Foursquare is an American technology company. The company's location platform. It provides local search and discovery that help the developer discover a place in a location. I also use the data from Wikipedia:

[https://en.wikipedia.org/wiki/List\\_of\\_districts\\_of\\_Bangkok](https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok) that provide some information about latitude and longitude of Bangkok districts.

### Collect the data from Wikipedia

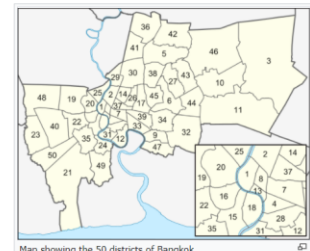
First, I need to know how many districts contain in Bangkok. In Wikipedia website. there are provide the information about the area of Bangkok such as s districts, population, latitude\longitude etc.

#### List of districts of Bangkok

From Wikipedia, the free encyclopedia

Bangkok is subdivided into 50 districts (*khet*, เขต, pronounced [kʰɛ́t], also sometimes wrongly called *amphoe* as in the other provinces, derived from Pali *khetṭa*, cognate to Sanskrit *kṣetra*), which are further subdivided into 180 subdistricts (*khwaeng*, แขวง, pronounced [kʰwǎɯ̯ŋ]), roughly equivalent to *tambon* in the other provinces.<sup>[1]</sup>

District ( <i>Khet</i> )	Map Nr	Post- code	Thai	Popu- lation	No. of Subdis- tricts <i>Khwaeng</i>	Latitude	Longitude
Bang Bon	50	10150	บางบอน	105,161	4	13.6592	100.3991
Bang Kapi	6	10240	บางกะปิ	148,465	2	13.765833	100.647778
Bang Khae	40	10160	บางแค	191,781	4	13.696111	100.409444
Bang Khen	5	10220	บางเขน	189,539	2	13.873889	100.596389
Bang Kho Laem	31	10120	บางคอแหลม	94,956	3	13.693333	100.5025
Bang Khun Thian	21	10150	บางขุนเทียน	165,491	2	13.660833	100.435833
Bang Na	47	10260	บางนา	95,912	2	13.680081	100.5918
Bang Phlat	25	10700	บางพลัด	99,273	4	13.793889	100.505
Bang Rak	4	10500	บางรัก	45,875	5	13.730833	100.524167
Bang Sue	29	10800	บางซื่อ	132,234	2	13.809722	100.537222
Bangkok Noi	20	10700	บางกอกน้อย	117,793	5	13.770867	100.467933
Bangkok Yai	16	10600	บางกอกใหญ่	72,321	2	13.722778	100.476389
Buang Kum	27	10240	บึงกุ่ม	145,830	3	13.785278	100.669167
Chatuchak	30	10900	จตุจักร	160,906	5	13.828611	100.559722



Map showing the 50 districts of Bangkok

Credit: [https://en.wikipedia.org/wiki/List\\_of\\_districts\\_of\\_Bangkok](https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok)

### Collect the data from foursquare

Next, I use the foursquare API for scraping the data restaurants in area latitude, longitude, and importantly category of restaurants that describe it what kind of restaurants and it helps to identify and group the data.

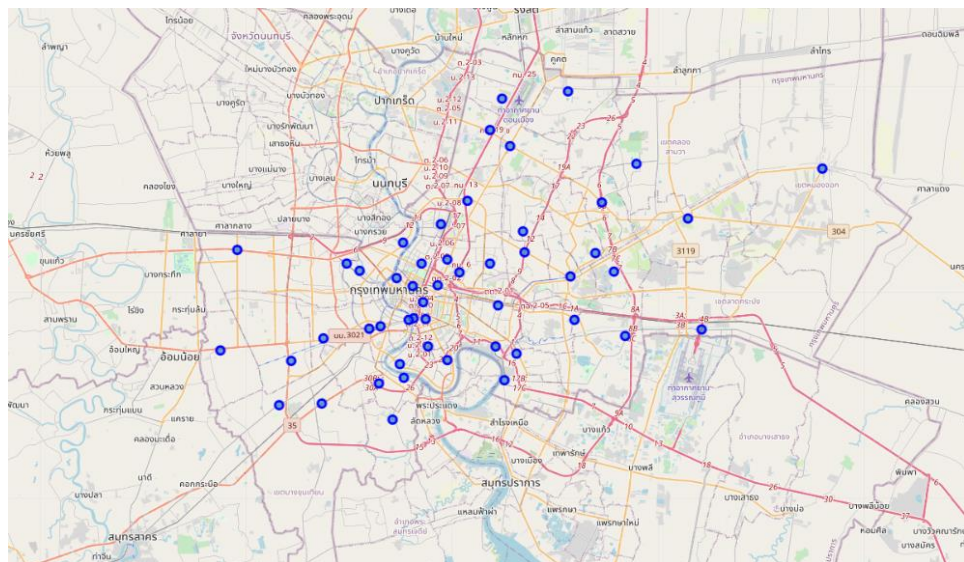
	name	categories	lat	lng
0	Recreational Bangkok Biking	Bike Rental / Bike Share	13.698552	100.544010
1	Tawandang German Brewery (โรงเบียร์เยอรมันดะว้...	Brewery	13.696742	100.545958
2	Hachiban Ramen (ชะชิบัง ราเมน)	Ramen Restaurant	13.697124	100.540969
3	Fuji (ฟูจิ)	Japanese Restaurant	13.697371	100.541175
4	Moo-Hun-Song-Pee-Nong (หมูหันสองพี่น้อง)	Chinese Restaurant	13.698668	100.542877

### 3.Methodology

In this section, I will describe the data analysis and how I used the data to yield the results. Firstly, I scraped data from Wikipedia to create a dataframe with the city districts of Bangkok scarping by beautiful soup library and clean the dataframe by using pandas. The result had to clean data frame in terms of unnecessary information or data that could not be handled in a data frame. After I already cleaning the data from Wikipedia the dataframe as show below.

	District	Map	Postcode	Thai	Population	Latitude	Longitude
0	Bang Bon	50	10150	บางบอน	105161	13.6592	100.3991
1	Bang Kapi	6	10240	บางกะปิ	148465	13.765833	100.647778
2	Bang Khae	40	10160	บางแค	191781	13.696111	100.409444
3	Bang Khen	5	10220	บางเขน	189539	13.873889	100.596389
4	Bang Kho Laem	31	10120	บางคอแหลม	94956	13.693333	100.5025
5	Bang Khun Thian	21	10150	บางขุนเทียน	165491	13.660833	100.435833
6	Bang Na	47	10260	บางนา	95912	13.680081	100.5918
7	Bang Phlat	25	10700	บางพลัด	99273	13.793889	100.505
8	Bang Rak	4	10500	บางรัก	45875	13.730833	100.524167
9	Bang Sue	29	10800	บางซื่อ	132234	13.809722	100.537222
10	Bangkok Noi	20	10700	บางกอกน้อย	117793	13.770867	100.467933

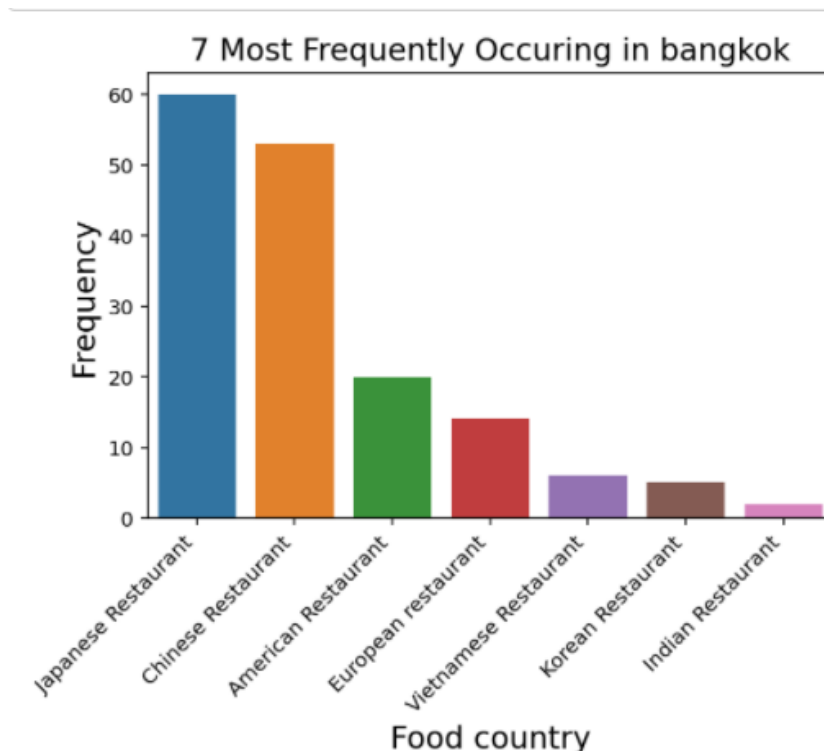
After I already created the dataframe district in Bangkok I receive the 50 districts and another information latitude longitude. then We will use folium to create a visualization of the Bangkok district.



Next, I will detect the data that indicates the nationality in the category of the restaurant in the Bangkok district. By using foursquare API for scarping. The result was list 338 restaurants come from 30 restaurant categories in Bangkok. However, there are some categories that need to modify for support analysis. There are 3 factors that need to be modified.

- Thai restaurant needs to remove because this purpose of the project needs to recommend the foreign that would like to find food as familiar their country.
- Remove some restaurants cannot identify the country of food.
- Grouping data. If there are some categories have similar with another type.

After I already cleaning and extract the data to match with the purpose. The result after preparing the data before the analysis section. It has 7 categories. I plot the bar chart for visualizing the frequency of restaurants. We can see that Japanese, Chinese restaurants are the most frequently occurring restaurants in Bangkok, which seems reasonable, as Thailand has a relatively high proportion of people with Japanese and Chinese.



To find clusters of restaurant types in the different city districts, I first transformed the data frame with the restaurant venues, associated to city districts, by one-hot encoding (0/1) binary type, as seen in the picture below.

	Neighborhood	American Restaurant	Chinese Restaurant	European restaurant	Indian Restaurant	Japanese Restaurant	Korean Restaurant	Vietnamese Restaurant
2	Phra Nakhon	0	0	0	0	1	0	0
3	Phra Nakhon	0	1	0	0	0	0	0
6	Phra Nakhon	0	0	0	1	0	0	0
7	Phra Nakhon	0	1	0	0	0	0	0
15	Bang Rak	0	1	0	0	0	0	0

Next, I used grouping to show the frequency of each category of restaurants in each city district.

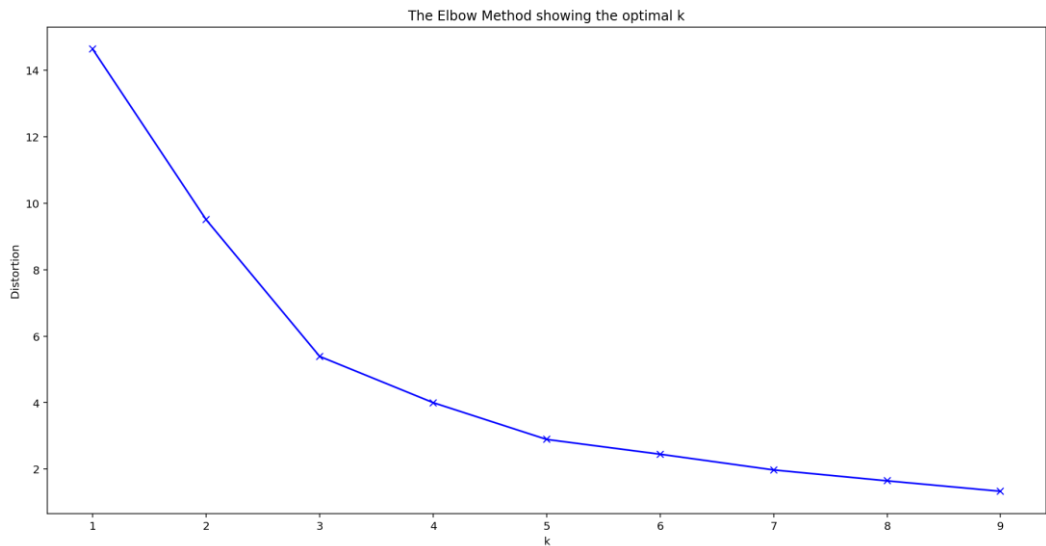
	Neighborhood	American Restaurant	Chinese Restaurant	European restaurant	Indian Restaurant	Japanese Restaurant	Korean Restaurant	Vietnamese Restaurant
0	Bang Bon	0.000000	0.000000	0.000000	0.0000	1.000000	0.000000	0.000000
1	Bang Kapi	0.000000	0.000000	0.000000	0.0000	0.666667	0.333333	0.000000
2	Bang Khae	0.500000	0.000000	0.000000	0.0000	0.500000	0.000000	0.000000
3	Bang Khen	0.000000	0.000000	0.000000	0.0000	0.000000	0.000000	1.000000
4	Bang Kho Laem	0.250000	0.500000	0.000000	0.0000	0.000000	0.000000	0.250000
5	Bang Khun Thian	0.166667	0.166667	0.000000	0.0000	0.666667	0.000000	0.000000
6	Bang Phlat	1.000000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000
7	Bang Rak	0.000000	1.000000	0.000000	0.0000	0.000000	0.000000	0.000000
8	Bangkok Noi	0.000000	1.000000	0.000000	0.0000	0.000000	0.000000	0.000000
9	Bangkok Yai	0.000000	1.000000	0.000000	0.0000	0.000000	0.000000	0.000000
10	Buang Kum	1.000000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000

I used this information to create a data frame in which you can see the most common restaurant venue types for each city district.

	Neighborhood	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	Bang Bon	Japanese Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant	American Restaurant
1	Bang Kapi	Japanese Restaurant	Korean Restaurant	Vietnamese Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant	American Restaurant
2	Bang Khae	Japanese Restaurant	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant
3	Bang Khen	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant	American Restaurant
4	Bang Kho Laem	Chinese Restaurant	Vietnamese Restaurant	American Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant
5	Bang Khun Thian	Japanese Restaurant	Chinese Restaurant	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant
6	Bang Phlat	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant
7	Bang Rak	Chinese Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	American Restaurant
8	Bangkok Noi	Chinese Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	American Restaurant
9	Bangkok Yai	Chinese Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	American Restaurant
10	Buang Kum	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant



Now the data is already to perform unsupervised learning as K-mean clustering specifically, a k-means clustering algorithm from the scikit-learn package. But we have 1 variable that needs to know before clustering. What K value with a suitable value for clustering model. I use the elbow method technique for finding K value.



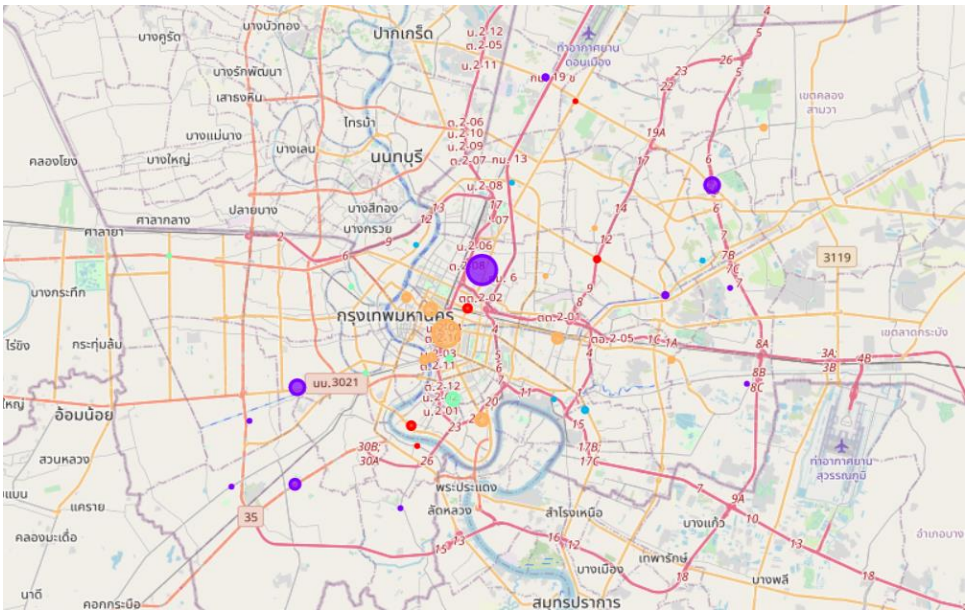
From the plot up here, the best of K value is 3 But it seems a little value So I use the second value is 5.

#### 4.Results

This section shows the result after use K-mean clustering that indicate on cluster label column.

	Neighborhood	Postcode	Thai	Population	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Map												
1	Phra Nakhon	10200	พระนคร	57876	13.764444	100.499167	4	Chinese Restaurant	Japanese Restaurant	Indian Restaurant	Vietnamese Restaurant	Korean Restaurant
4	Bang Rak	10500	บางรัก	45875	13.730833	100.524167	3	Chinese Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant
5	Bang Khen	10220	บางเขน	189539	13.873889	100.596389	0	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant
6	Bang Kapi	10240	บางกะปิ	148465	13.765833	100.647778	1	Japanese Restaurant	Korean Restaurant	Vietnamese Restaurant	Indian Restaurant	European restaurant
7	Pathum Wan	10330	ปทุมวัน	53263	13.744942	100.522200	4	Chinese Restaurant	Japanese Restaurant	Korean Restaurant	Indian Restaurant	Vietnamese Restaurant
8	Pom Prap Sattru Phai	10100	ป้อมปราบศัตรูพ่าย	51006	13.758056	100.513056	4	Chinese Restaurant	Japanese Restaurant	European restaurant	American Restaurant	Vietnamese Restaurant
9	Phra Khanong	10260	พระโขนง	93482	13.702222	100.601667	2	American Restaurant	European restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant

We see in the table are the city districts and their most common venues, and they now have been assigned five different cluster labels from 0 to 4. We can now use the cluster labels to show the city districts marked with a cluster-specific color on a map, again using folium:



You will see bubbles for the city districts in Bangkok, with five different colors for the five different clusters. Now, what is the final result of this exercise? We can show five clusters of restaurant type concentrations for the city of Bangkok, which I named according to the restaurant concentration the data shows.

### Cluster 1 - The Vietnamese Food Cluster

	Neighborhood	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
Map									
5	Bang Khen	0	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant	American Restaurant
24	Rat Burana	0	Vietnamese Restaurant	Chinese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	American Restaurant
31	Bang Kho Laem	0	Chinese Restaurant	Vietnamese Restaurant	American Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant
37	Ratchathewi	0	European restaurant	Vietnamese Restaurant	Korean Restaurant	Chinese Restaurant	American Restaurant	Japanese Restaurant	Indian Restaurant
45	Wang Thonglang	0	Vietnamese Restaurant	Japanese Restaurant	American Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant

### Cluster 2 - The Japanese Food Cluster

	Neighborhood	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
Map									
6	Bang Kapi	1	Japanese Restaurant	Korean Restaurant	Vietnamese Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant	American Restaurant
14	Phaya Thai	1	Japanese Restaurant	Chinese Restaurant	European restaurant	Vietnamese Restaurant	Korean Restaurant	American Restaurant	Indian Restaurant
21	Bang Khun Thian	1	Japanese Restaurant	Chinese Restaurant	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant
22	Phasi Charoen	1	Japanese Restaurant	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant
32	Prawet	1	Japanese Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant	American Restaurant
40	Bang Khae	1	Japanese Restaurant	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant
41	Lak Si	1	Japanese Restaurant	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant
43	Khan Na Yao	1	Japanese Restaurant	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant
44	Saphan Sung	1	Japanese Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant	American Restaurant
49	Thung Khru	1	Japanese Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant	American Restaurant
50	Bang Bon	1	Japanese Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant	American Restaurant

### Cluster 3 - The American Food Cluster

	Neighborhood	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
Map									
9	Phra Khanong	2	American Restaurant	European restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	Chinese Restaurant
25	Bang Phlat	2	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant
27	Bueng Kum	2	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant
30	Chatuchak	2	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant
33	Khlong Toei	2	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	Chinese Restaurant

### Cluster 4 - The Chinese Food Cluster

	Neighborhood	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
Map									
4	Bang Rak	3	Chinese Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	American Restaurant
16	Bangkok Yai	3	Chinese Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	American Restaurant
20	Bangkok Noi	3	Chinese Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	American Restaurant
28	Sathon	3	Chinese Restaurant	Japanese Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	American Restaurant
48	Thawi Watthana	3	Chinese Restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	European restaurant	American Restaurant

### Cluster 5 - The Chinese and Japanese Food Cluster

	Neighborhood	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
Map									
1	Phra Nakhon	4	Chinese Restaurant	Japanese Restaurant	Indian Restaurant	Vietnamese Restaurant	Korean Restaurant	European restaurant	American Restaurant
7	Pathum Wan	4	Chinese Restaurant	Japanese Restaurant	Korean Restaurant	Indian Restaurant	Vietnamese Restaurant	European restaurant	American Restaurant
8	Pom Prap Sattru Phai	4	Chinese Restaurant	Japanese Restaurant	European restaurant	American Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant
12	Yan Nawa	4	Japanese Restaurant	Chinese Restaurant	American Restaurant	Korean Restaurant	Vietnamese Restaurant	Indian Restaurant	European restaurant
13	Samphanthawong	4	Chinese Restaurant	Japanese Restaurant	European restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	American Restaurant
17	Huai Khwang	4	Japanese Restaurant	Chinese Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	American Restaurant
18	Khlong San	4	Japanese Restaurant	Chinese Restaurant	European restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	American Restaurant
34	Suan Luang	4	Chinese Restaurant	European restaurant	Vietnamese Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant	American Restaurant
38	Lat Phrao	4	Japanese Restaurant	Chinese Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	American Restaurant
39	Watthana	4	Japanese Restaurant	European restaurant	Chinese Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	American Restaurant
46	Khlong Sam Wa	4	Chinese Restaurant	Japanese Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	European restaurant	American Restaurant

Here we are at the end of the analysis, I tried to set up a realistic data-analysis scenario using several different ways such as: web scraping on Wikipedia, open data from public, checking the name of restaurant that match with categories and use some powerful python libraries such as Folium and Pandas, Foursquare API, etc...

So now we have the opportunity to make some argument about the clusters. Let's see what we have found:

1. The most common venues in Bangkok are Japanese restaurant.
2. The Japanese restaurant usually stay at outskirts area of Bangkok.
3. In the center of Bangkok city there is most common venues of Chinese restaurant.

However, I think the data from foursquare in Bangkok location. There are some data that should be fixed. Almost the data scarping from the foursquare API does not indicate the type of restaurant. It only indicates restaurants in categories that make the data was removed in this case. The number of restaurants was decreased.



## 5. Conclusion

The purpose of this project is to introduce the neighborhood to tourists or foreigners who must live in Bangkok and need to live in an area that there are foreign restaurants which have a type of restaurant nearly native domicile by using data from foursquare and then manage data in a format that meets the requirements and clustering each location to be appropriate.

However, the final decision on the optimal location will be made by the occupants. It may be based on specific characteristics taking into consideration additional factors like attractiveness, environment, travel convenience, and so on.