Applied Data Science Capstone Project

(IBM Coursera)

Places that you should not miss when you are visiting Bangkok

By

Nattapol Patarapureeruk

18 January 2021

Introduction and Problem Statement

Background

Bangkok; the capital city of Thailand; is one of the world's top destination for traveler for a decade. Over 45 million people fly to Thailand annually. There are a lot of perfect places in Bangkok for family, couple, and friends to spend time together.

Problem

Unfortunately, as the Covid-19 situation is not getting better in a coming day, we do not have enough time to explore all fascinating places in Bangkok due to the social distance restriction.

This project aims to create and analysis of similar places comparing between each area in Bangkok such as restaurants, groceries, street foods, and tourist attractions. It will help traveler to find all places that they should not miss when they are visiting Bangkok with limited time during the pandemic.

Data Acquisition and Cleaning

Data Sources

Based on definition of the problem, factors that will be likely to influence the result and recommendation are:

Districts and coordinates of each venues in Bangkok

First, we need all information about district in Bangkok which includes district names, coordinate (longitude and latitude), and postal code which can be collected from Query Data World (https://query.data.world/s/o7qvkaesgxjg4hy4ilceeyubqch4fy)

Places, traffic, and venues' coordinate in Bangkok

To explore all places in Bangkok, we need data about venues in each area around the city. All information such as names of the venue, location (coordinates), and venues' category were retrieved from Foursquare API.

Data Cleaning

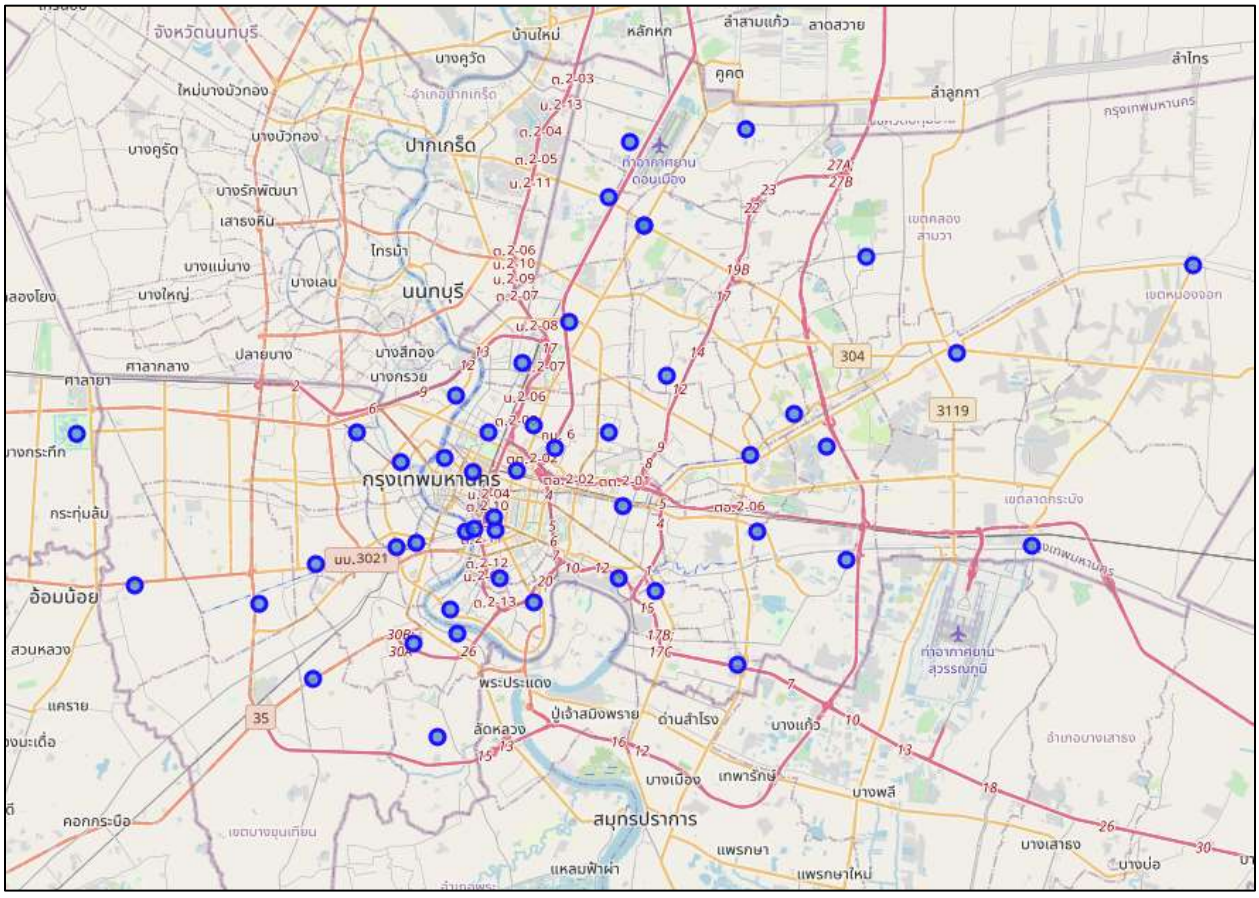
Let's start with importing the latitude, longitude, and its names for all districts covering all Bangkok area. Then cleaning and formatting data, especially for translation of transliterated Thai words to English if it is available.

Figure 1: Data scraping from data world consisting district names, coordinates, and postal code in Bangkok

	district	province	lat	long	postal code
0	Bang Kapi	Bangkok	13.76583	100.64778	10240
1	Bang Khae	Bangkok	13.69611	100.40944	10160
2	Bang Khen	Bangkok	13.87389	100.59639	10220
3	Bang Kho Laem	Bangkok	13.69333	100.50250	10120
4	Bang Khun Thian	Bangkok	13.66083	100.43583	10150

To find out what Bangkok looks like, Geopy, matplotlib, and folium have been used together to create and display all area of Bangkok by using latitude and longitude. The result shown in Figure 2.

Figure 2: Map of Bangkok generated by Geopy, Matplotlib, and Folium



After that, we used coordinated data from previous step to find all venues located in each district from Foursquare API with the condition that the distance between venue and the city center must not be more than 500 meters.

Figure 3: Data collected from Foursquare API

	district	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bang Kapi	13.76583	100.64778	ตลาดนัดแฟลตคลองจั่น	13.770105	100.649124	Flea Market
1	Bang Kapi	13.76583	100.64778	ชั้งโภชนา ลูกชิ้นปลาทำเอง	13.761873	100.648290	Noodle House
2	Bang Kapi	13.76583	100.64778	Suan Son (สวนสน)	13.762940	100.648895	Neighborhood
3	Bang Kapi	13.76583	100.64778	Bang Kapi Market (ตลาดบางกะปิ)	13.766543	100.646203	Market
4	Bang Kapi	13.76583	100.64778	ทรงกิจลูกชิ้นเนื้อวัว	13.763036	100.648485	Noodle House
...	...	...	...	...	...	...	...
1063	Yan Nawa	13.69694	100.54306	Hermes Spa	13.698675	100.543722	Spa
1064	Yan Nawa	13.69694	100.54306	Lavita Coffee	13.698132	100.540867	Coffee Shop
1065	Yan Nawa	13.69694	100.54306	Acer Service Center	13.700052	100.545886	Electronics Store
1066	Yan Nawa	13.69694	100.54306	The Deck	13.700613	100.542068	Breakfast Spot
1067	Yan Nawa	13.69694	100.54306	Hunan Lamei (หูหนานล่าเม่)	13.700283	100.540212	Chinese Restaurant

Lastly, we modified data to find the most type of venue in each district in order to know that what is likely the best activity that we should not miss.

Figure 4: The most common venue for each district

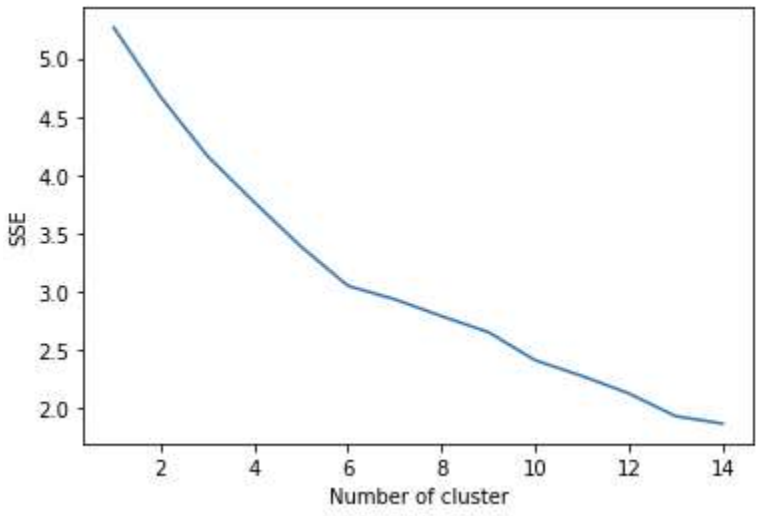
	district	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bang Kapi	Noodle House	Convenience Store	Flea Market	Neighborhood	Multiplex
1	Bang Khae	Convenience Store	Noodle House	Supermarket	Asian Restaurant	Shopping Mall
2	Bang Khen	Asian Restaurant	Convenience Store	Som Tum Restaurant	Thai Restaurant	Garden
3	Bang Kho Laem	Noodle House	Thai Restaurant	Chinese Restaurant	Coffee Shop	Supermarket
4	Bang Khun Thian	Japanese Restaurant	Thai Restaurant	Bakery	Restaurant	Gym / Fitness Center

Methodology

To find which districts were similar, we used K-Means Clustering as a main tool to segment and group them in to the same cluster.

However, limitation of K-Means Clustering is what is the value of K should be used. In this case, we generated multiple value of K (1-15), then calculated Standard Error of Estimate (SSE) for each case. The result shown in Figure 5.

Figure 5: SSE for each value of K



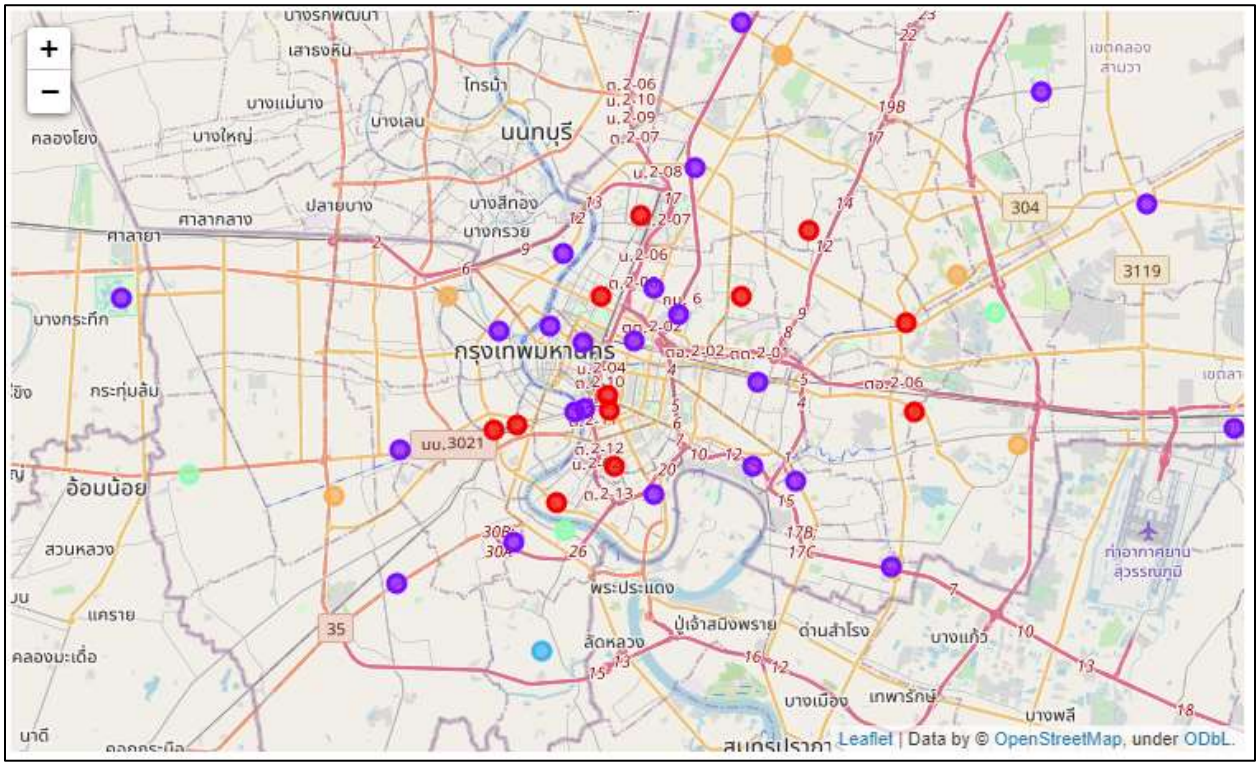
It can be interpreted that the higher number of cluster (K), the better model performance. Unfortunately, it did not simple as that. As we have analyzed the output of each cluster, we found that some cluster was just an outlier with only one data point for each cluster. Those had happened when the value of K was higher than 5, Consequently, we decided to put the value of K equaling to 5 to avoid those problem.

Result and Discussion

Based on the result of the model, all districts in Bangkok can be classified to 5 groups[1] which are

1. Eat all day: especially for one who loves noodles, Asian food, and dessert
2. Cafe hopping and night life: fun forever
3. Outliers: I was surprised how was this result came from, so I have considered this cluster as an outlier
4. Traditional Thai things
5. Outskirt: nothing special there unless you want to go shopping at seven-eleven (convenience store)

Figure 6: Map of clusters in Bangkok



So, if you are travelers, I will suggest you; based on the result of the model; that you should visit only 3 cluster which are cluster 1, 2, and 4. There are a lot of fascinating places in those areas especially foods, drinks, and traditional Thai venues.

However, in fact, the minority of the result contradicts what actually in Bangkok it is. After an investigation, I found that it must be because of Foursquare data. The majority of foursquare users in Thailand are foreigner so all information are highly reflected from foreigner's perspective not Thai people as it should be. This might be one of the weaknesses relying solely on Foursquare API.

[1] Detailed clusters can be found in Appendix

Conclusion

In this project, K-means clustering has been used to classify similar districts in Bangkok in terms of venues and interesting places. All information were supported by data world and Foursquare API. At the end, the model suggested to classify all districts into 5 groups as discussed in result section. Even though it is not perfect as it should be due to some limitation of input data, it can help traveler to explore Bangkok more effectively and fulfil their needs and expectations

Appendix

Cluster 1

	district	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bang Kapi	0	Noodle House	Convenience Store	Flea Market	Neighborhood	Multiplex
3	Bang Kho Laem	0	Noodle House	Thai Restaurant	Chinese Restaurant	Coffee Shop	Supermarket
6	Bangkok Yai	0	Noodle House	Farmers Market	Dessert Shop	Asian Restaurant	Coffee Shop
9	Bang Rak	0	Noodle House	Hotel	Chinese Restaurant	Thai Restaurant	Massage Studio
10	Bang Sue	0	Thai Restaurant	Noodle House	Coffee Shop	Bar	Seafood Restaurant
16	Dusit	0	Noodle House	Convenience Store	Coffee Shop	Asian Restaurant	Dessert Shop
17	Huai Khwang	0	Noodle House	Hotel	Som Tum Restaurant	Convenience Store	Hotpot Restaurant
25	Lat Phrao	0	Noodle House	Café	Som Tum Restaurant	Asian Restaurant	Coffee Shop
29	Pathum Wan	0	Noodle House	Thai Restaurant	Chinese Restaurant	Som Tum Restaurant	Steakhouse
41	Sathon	0	Noodle House	Asian Restaurant	Dessert Shop	Convenience Store	Chinese Restaurant
42	Suan Luang	0	Noodle House	Asian Restaurant	Massage Studio	Convenience Store	Coffee Shop
45	Thon Buri	0	Noodle House	Deli / Bodega	Train Station	Pizza Place	BBQ Joint

Cluster 2

	district	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Bang Khun Thian	1	Japanese Restaurant	Thai Restaurant	Bakery	Restaurant	Gym / Fitness Center
5	Bangkok Noi	1	Market	Dessert Shop	Convenience Store	Canal	Farmers Market
7	Bang Na	1	Convenience Store	Japanese Restaurant	Coffee Shop	Fast Food Restaurant	Noodle House
8	Bang Phlat	1	Som Tum Restaurant	Convenience Store	Bus Stop	Bed & Breakfast	Massage Studio
12	Chatuchak	1	Coffee Shop	Thai Restaurant	Flea Market	Beer Bar	Bed & Breakfast
13	Chom Thong	1	American Restaurant	Convenience Store	Fast Food Restaurant	Thai Restaurant	Toll Plaza
14	Din Daeng	1	Toll Plaza	Convenience Store	Soccer Field	Recreation Center	Sports Club
15	Don Mueang	1	Restaurant	Convenience Store	Thai Restaurant	RV Park	Café
19	Khlong Sam Wa	1	Pub	Japanese Restaurant	Chinese Restaurant	Thai Restaurant	Convenience Store
20	Khlong San	1	Coffee Shop	Café	Dessert Shop	Noodle House	Hotel Bar
21	Khlong Toei	1	Bar	Thai Restaurant	Canal	Food Court	Karaoke Bar
23	Lak Si	1	Coffee Shop	Thai Restaurant	Fast Food Restaurant	Japanese Restaurant	Steakhouse
24	Lat Krabang	1	Convenience Store	Café	Soup Place	Noodle House	Thai Restaurant
26	Min Buri	1	Intersection	Thai Restaurant	Massage Studio	Department Store	Asian Restaurant
30	Phasi Charoen	1	BBQ Joint	Coffee Shop	Japanese Restaurant	Fast Food Restaurant	Discount Store
31	Phaya Thai	1	Coffee Shop	Thai Restaurant	Café	Japanese Restaurant	Sushi Restaurant
32	Phra Khanong	1	Convenience Store	Fast Food Restaurant	Hotel	Ice Cream Shop	Italian Restaurant
33	Phra Nakhon	1	Café	Noodle House	Hostel	Massage Studio	Hotel
34	Pom Prap Sattru Phai	1	Noodle House	Café	Chinese Restaurant	Convenience Store	Dim Sum Restaurant
37	Ratchathewi	1	Coffee Shop	Hostel	Café	Hotel	Restaurant
39	Samphanthawong	1	Art Gallery	Hotel Bar	Coffee Shop	Chinese Restaurant	Hotel
44	Thawi Watthana	1	Park	History Museum	Spiritual Center	Health Food Store	Wings Joint
48	Watthana	1	Café	Coffee Shop	BBQ Joint	Cocktail Bar	Hotel
49	Yan Nawa	1	Convenience Store	Chinese Restaurant	Fast Food Restaurant	Hotpot Restaurant	Café

Cluster 3

	district	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
46	Thung Khru	2	Cupcake Shop	Lounge	Wings Joint	Flea Market	Fried Chicken Joint

Cluster 4

	district	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
28	Nong Khaem	3	Dessert Shop	Noodle House	Thai Restaurant	Market	Floating Market
36	Rat Burana	3	Thai Restaurant	Asian Restaurant	Som Tum Restaurant	Vietnamese Restaurant	Hotpot Restaurant
38	Sai Mai	3	Thai Restaurant	Bar	Asian Restaurant	Noodle House	Grocery Store
40	Saphan Sung	3	Thai Restaurant	Japanese Restaurant	Convenience Store	Stadium	Wings Joint

Cluster 5

	district	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Bang Khae	4	Convenience Store	Noodle House	Supermarket	Asian Restaurant	Shopping Mall
2	Bang Khen	4	Asian Restaurant	Convenience Store	Som Tum Restaurant	Thai Restaurant	Garden
11	Bueng Kum	4	Park	Market	Miscellaneous Shop	Convenience Store	Shop & Service
18	Khan Na Yao	4	Asian Restaurant	Convenience Store	Som Tum Restaurant	Thai Restaurant	Garden
22	Khong	4	Convenience Store	Food Court	Noodle House	Shopping Mall	Coffee Shop
27	Nong Chok	4	Park	Other Repair Shop	Asian Restaurant	Convenience Store	Shopping Mall
35	Prawet	4	Convenience Store	Halal Restaurant	Noodle House	Comfort Food Restaurant	Flea Market
43	Taling Chan	4	Seafood Restaurant	Convenience Store	Noodle House	Floating Market	Satay Restaurant

These are all the libraries which were used in this project:

Pandas:	Creating, calculating, and manipulating DataFrames.
Folium:	Map visualization
Matplotlib:	Map visualization
Scikit Learn:	K-Means Clustering
Geocoder:	Location data based on coordinates