

## Battle of Neighborhoods in Bangkok



# 1. Introduction

Bangkok is one of the most visiting city in the world. When the visitors, especially international visitors, are planning to visit, they need to explore the places and try to fetch as much information as possible regarding the city

This can be the neighborhoods, locality, market, restaurant including neighborhoods analysis.

It would be valuable to have an application which could make it easy by considering a comparative analysis between the neighborhoods with provided factors.

This project helps the end users to achieve the results by recommending the top 5 venues in each neighborhood of Bangkok.

## 2.Data Sets and APIs:

- **Bangkok Neighborhoods Data**

Beautiful soup web scraping is used to scrap the list of 50 districts of Bangkok from the folling url:[https://en.wikipedia.org/wiki/List\\_of\\_districts\\_of\\_Bangkok](https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok)

- **Foursquare API:**

This API has a database of more than 105 million places. This project would use Four-square API as its prime data gathering source. Many organizations are using to geo-tag their photos with detailed info about a destination, while also serving up contextually relevant locations for those who are searching for a place to eat, drink or explore. This API provides the ability to perform location search, location sharing and details about a business. Foursquare users can also use photos, tips and reviews in many productive ways to add value to the results.

- **Folium**

Python visualization library would be used to visualize the neighborhoods cluster distribution of Chicago city over an interactive leaflet map.Extensive comparative analysis of two randomly picked neighborhoods world be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.

- **Python packages**

- Pandas - Library for Data Analysis
- NumPy – Library to handle data in a vectorized manner
- JSON – Library to handle JSON files
- Geopy – To retrieve Location Data
- Requests – Library to handle http requests
- Matplotlib – Python Plotting Module
- Sklearn – Python machine learning Library
- Folium – Map rendering Library

## 3.Methodology

### 3.1 Exploratory Analysis

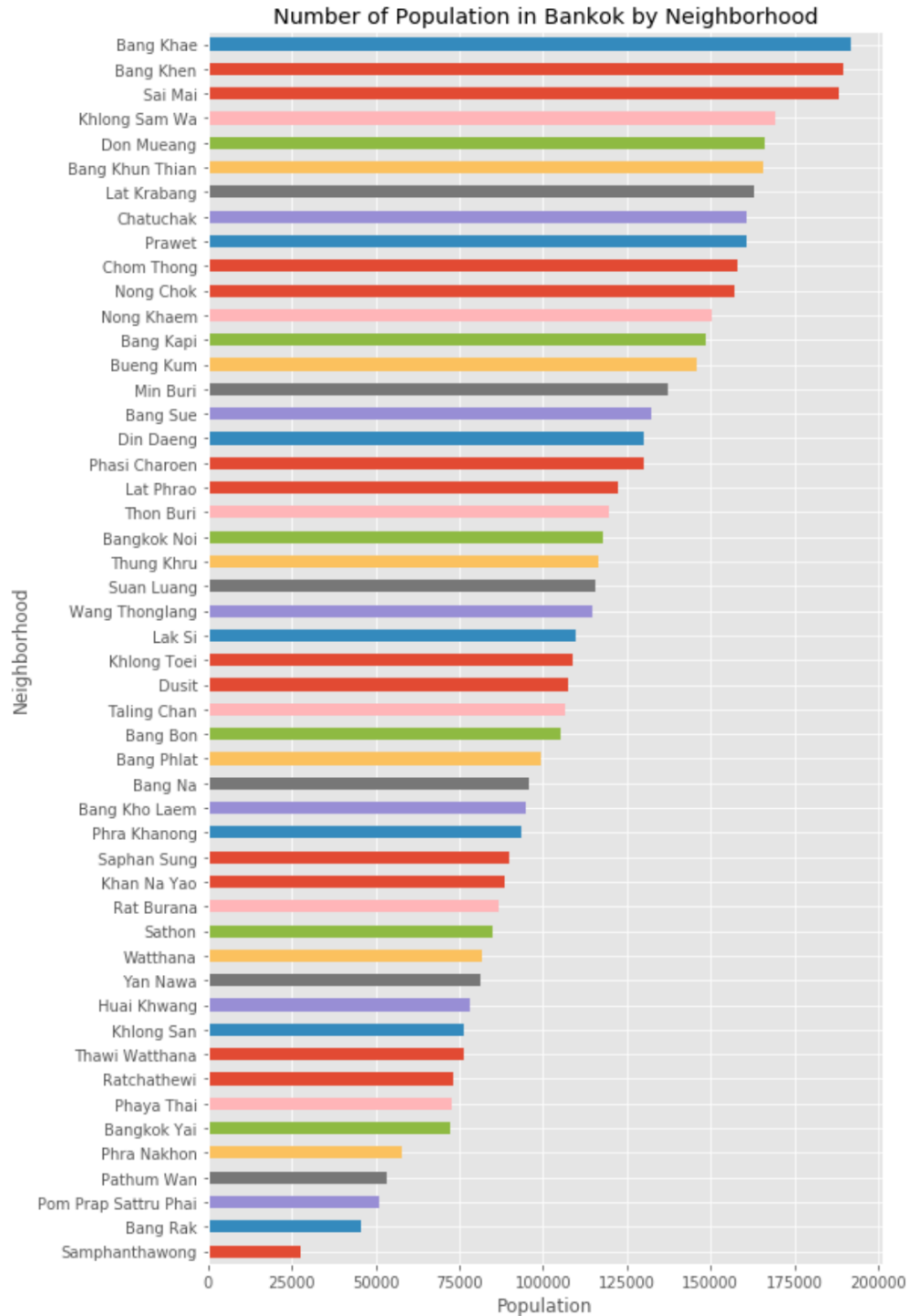


Figure 1. Number of Population in Bangkok of each neighborhood

## Comparing Number of Population in Each District

As illustrated by the Figure 1, Bang Khae has the highest population compared to the other, with almost 200,000 people living in the area. This is followed by Bang Khen, Sai Mai and Khlong Sam Wa, with the population of 189,000, 188,000 and 169,000 respectively.

For the district that has the least number of population is Samphanthawong accounting for 27,000, followed by Bang Rak, Pom Prap Sattru Phai and Pathum Wan, which compose 48,000, 51,000 and 53,000.

## Comparing Number of Sub districts

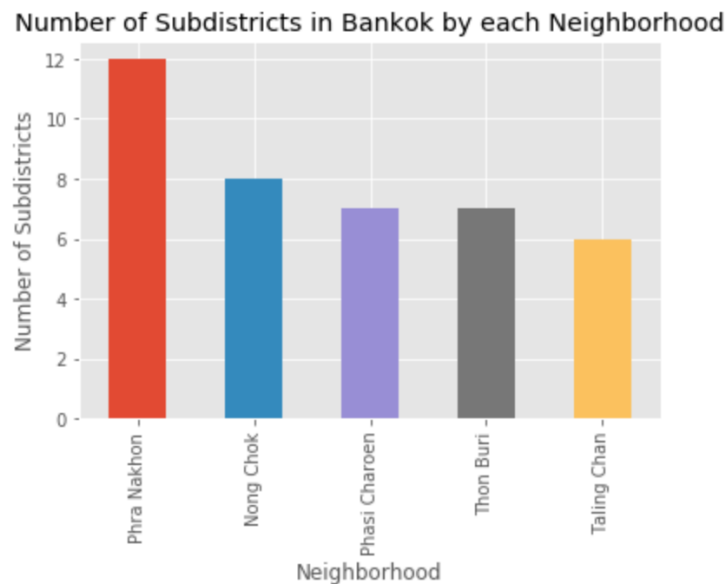


Figure 2: Top Five Districts by Number of Sub Districts

Figure 2 presents the top five districts that composes the highest number of sub districts. First, Phra Nakhon has the highest sub districts which accounts for 12. The second highest number is Nong Chok, Phasi Chareon and Thon Buri.

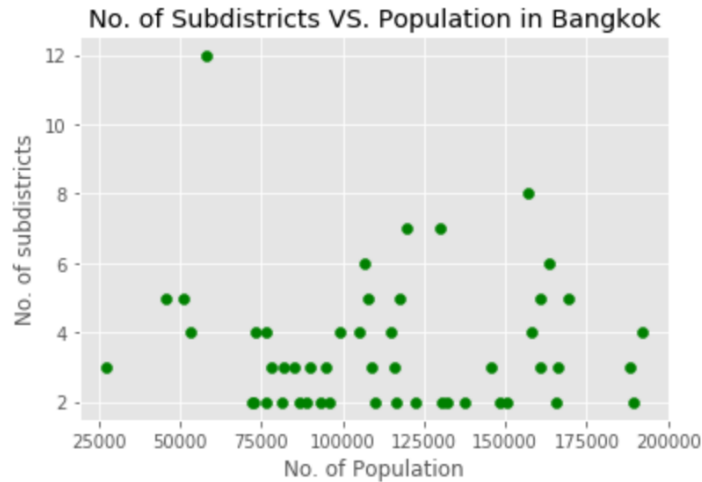


Figure 3: Correlation of Sub districts and Population

Figure 3 describes the correlation of number of sub districts and number of population. As you can see from the picture above, there does not exist the correlation between number sub districts and population in Bangkok. In other word, the areas that has more number of population do not constitute more sub districts and vice versa.

### 3.2 K-Means Clustering

This project applies K-mean clustering unsupervised machine learning algorithm to cluster the venues based on the place category such as restaurants, park, coffee shop, gym, clubs etc. This would give a better understanding of the similarities and dissimilarities between the chosen neighborhoods to retrieve more insights and to conclude with ease which neighborhood wins over other.

By the way, before we start applying K-Mean Clustering, we need to determine of number of Kth.

### 3.2.1 Elbow Analysis

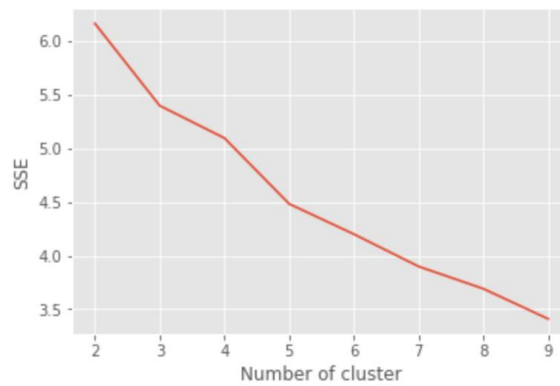


Figure 4: Elbow Analysis

Figure 4 shows the Elbow Analysis applied to find the optimal number of Kth for K-Mean. However, as presented, there is no obvious elbow point. It might be concluded that Elbow Analysis might not be the appropriate method for the dataset.

### 3.2.2 Silhouette Analysis

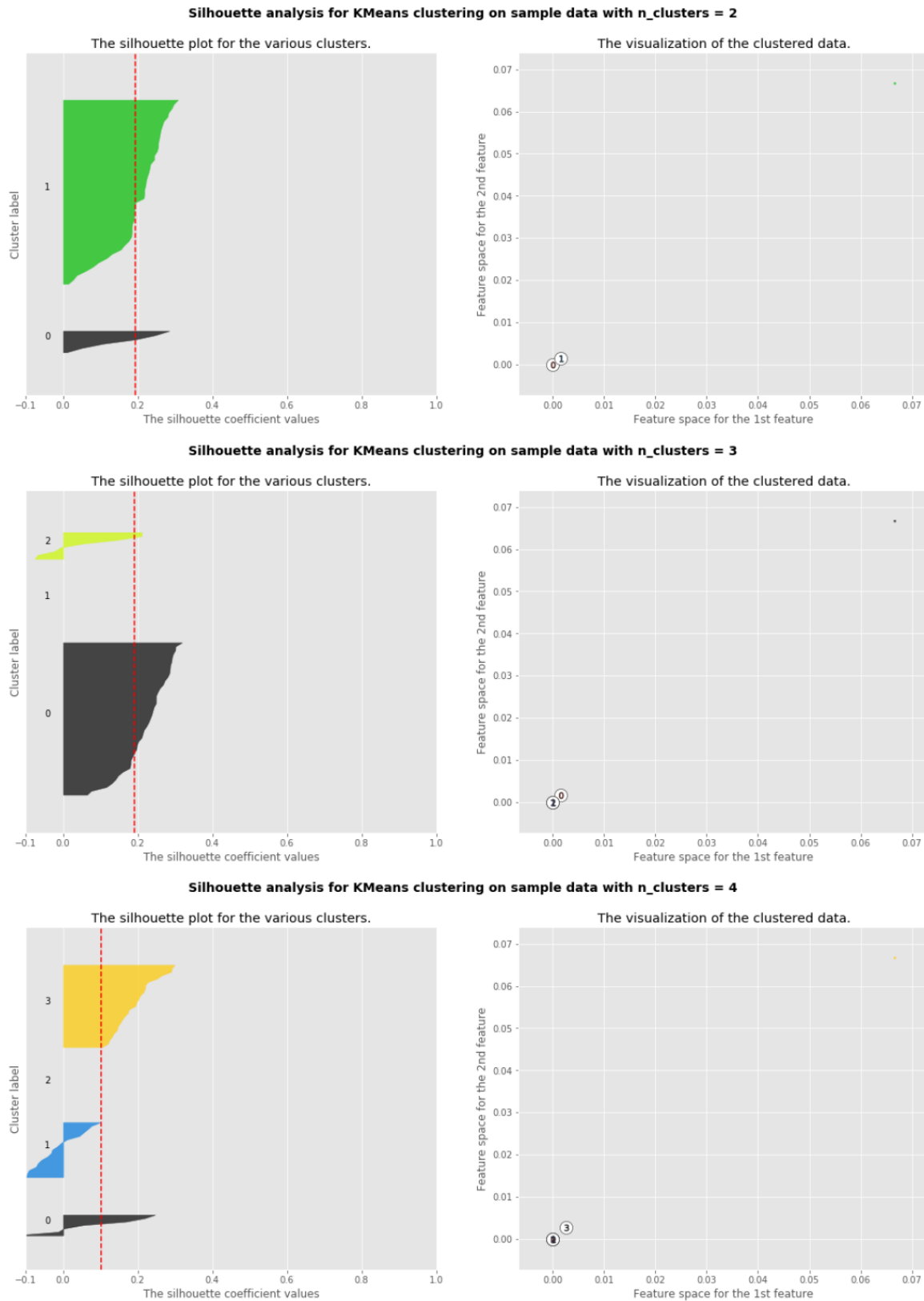
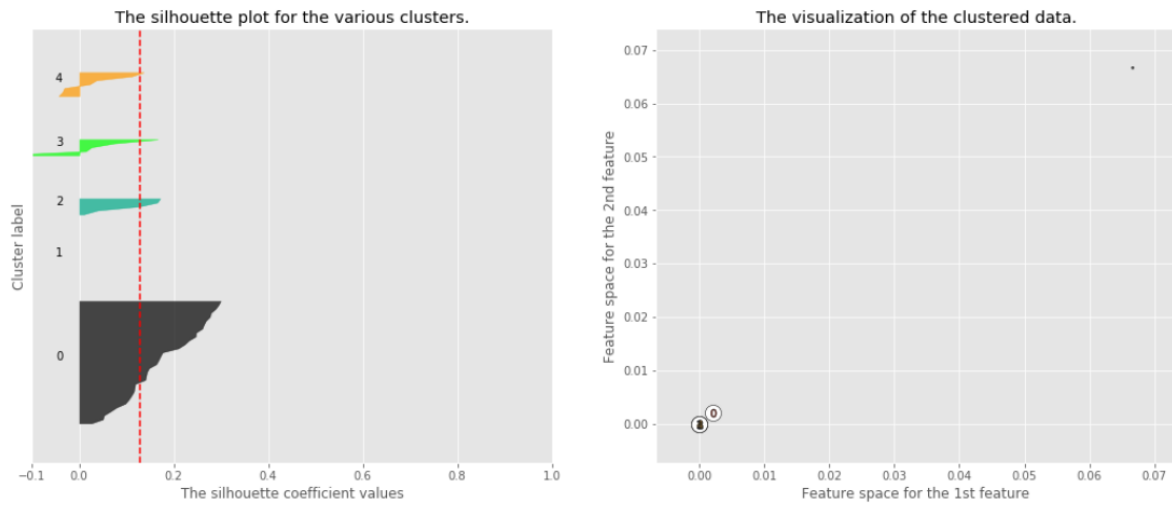


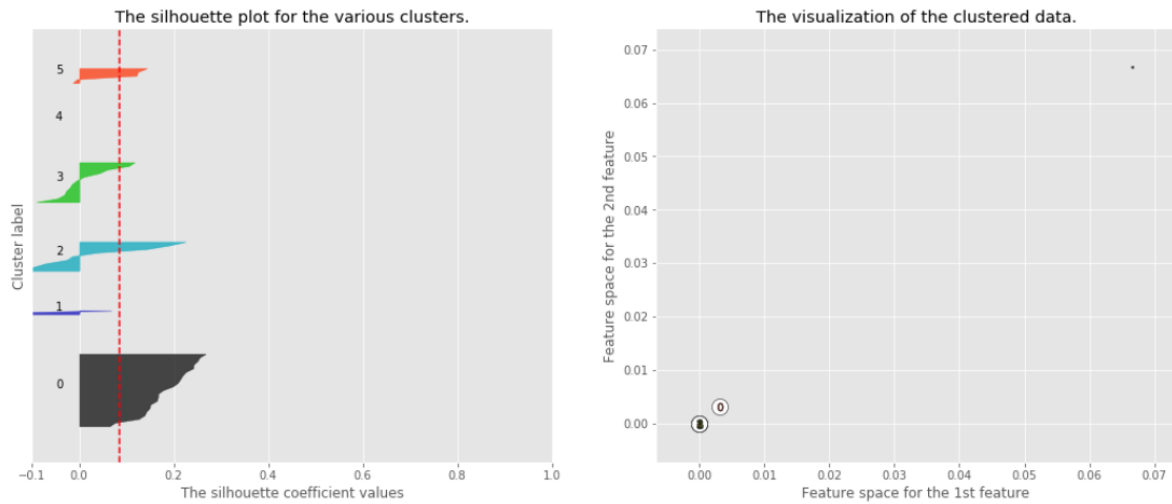
Figure 5: Silhouette Analysis 1 out of 2



### Silhouette analysis for KMeans clustering on sample data with $n\_clusters = 5$



### Silhouette analysis for KMeans clustering on sample data with $n\_clusters = 6$



### Silhouette analysis for KMeans clustering on sample data with $n\_clusters = 7$

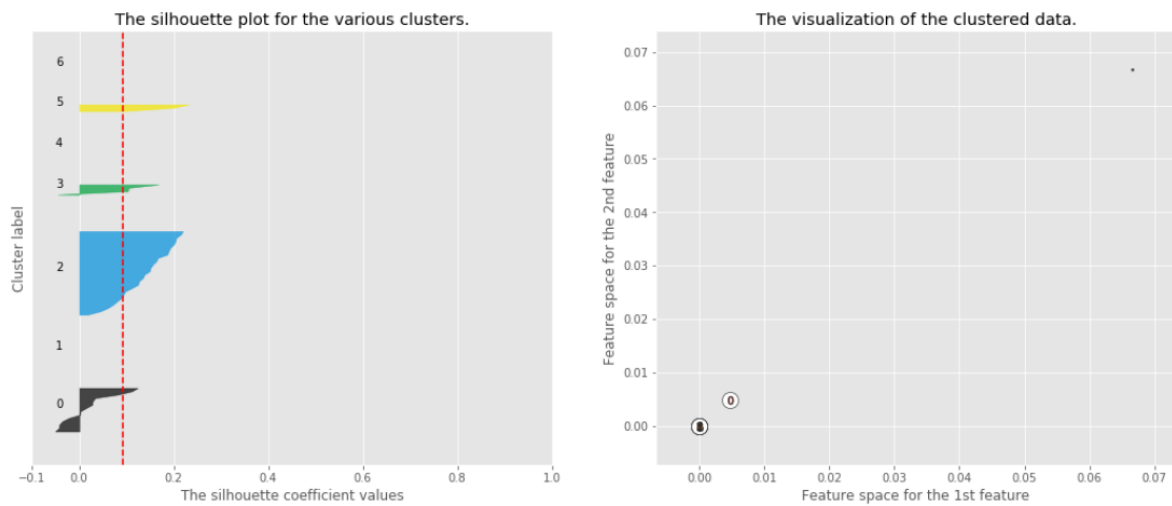


Figure 6: Silhouette Analysis 2 out of 2

As illustrated from the figure 5 and 6,  $n\_clusters = 7$  seems to be the best optimal number. This is due to the fact that the value of 2,3,4 and 5 have the wide fluctuations in the size of the silhouette plots. Whereas, the value of 7 are more or less similar thick ness and all of the clusters are above the average silhouette score.

## 4. Results

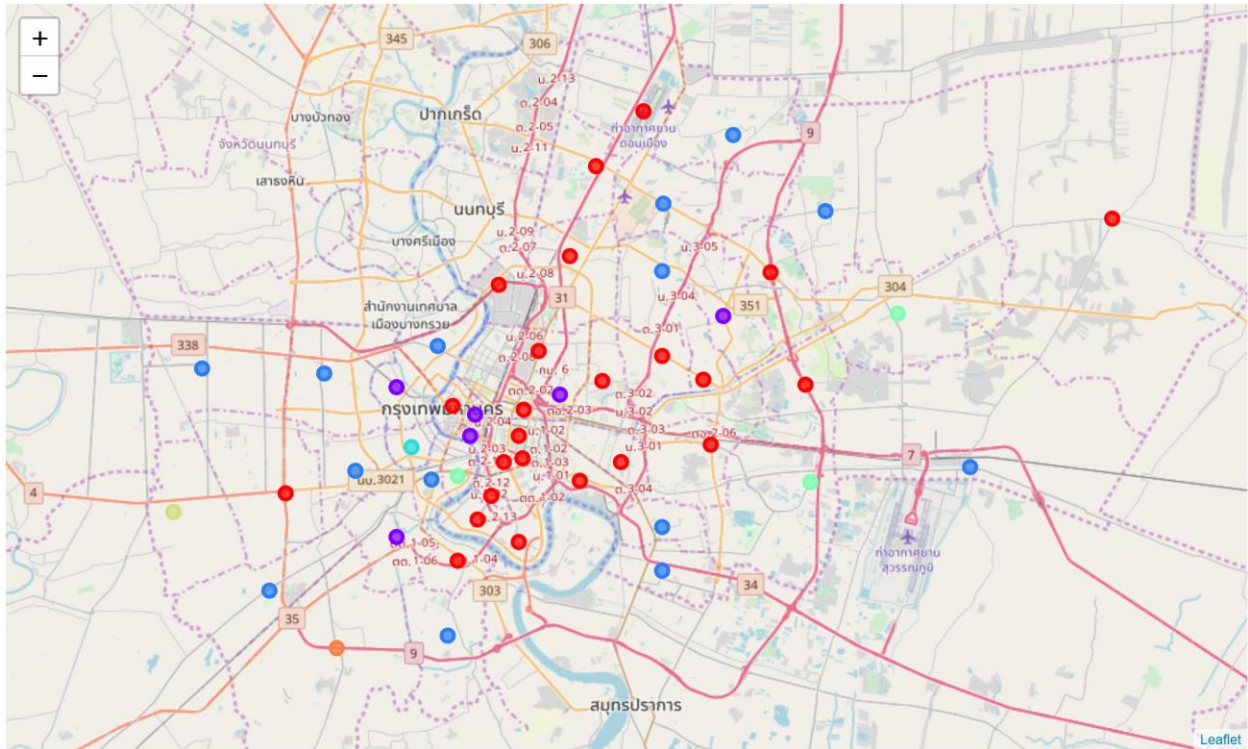


Figure 7: Map Visualization Result by Clusters

## Cluster 0

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Bang Kapi	0	Hotel	Hotpot Restaurant	Coffee Shop	Thai Restaurant	Som Tum Restaurant
2	Bang Khae	0	Dessert Shop	Coffee Shop	Japanese Restaurant	Bakery	Electronics Store
4	Bang Kho Laem	0	Convenience Store	Chinese Restaurant	BBQ Joint	Pub	Coffee Shop
8	Bang Rak	0	Hotel	Café	Restaurant	Chinese Restaurant	Japanese Restaurant
9	Bang Sue	0	Convenience Store	Chinese Restaurant	Pub	Road	Café
13	Chatuchak	0	Badminton Court	Coffee Shop	Pub	Shabu-Shabu Restaurant	Asian Restaurant
16	Don Mueang	0	Coffee Shop	Fast Food Restaurant	Airport Lounge	Furniture / Home Store	Donut Shop
17	Dusit	0	Coffee Shop	Ramen Restaurant	Japanese Restaurant	Spa	Thai Restaurant
18	Huai Khwang	0	Thai Restaurant	Theater	Hotpot Restaurant	Korean Restaurant	Convenience Store
19	Khan Na Yao	0	Japanese Restaurant	Thai Restaurant	Som Tum Restaurant	Multiplex	Ice Cream Shop
22	Khlong Toei	0	Chinese Restaurant	Thai Restaurant	Coffee Shop	BBQ Joint	Sporting Goods Shop
23	Lak Si	0	Coffee Shop	Convenience Store	Steakhouse	Japanese Restaurant	Hotpot Restaurant
27	Nong Chok	0	Convenience Store	Flea Market	Dessert Shop	Gym / Fitness Center	Other Repair Shop
29	Pathum Wan	0	Thai Restaurant	Coffee Shop	Japanese Restaurant	Hotel	Sports Club
31	Phaya Thai	0	Som Tum Restaurant	Noodle House	Café	Thai Restaurant	Coffee Shop
33	Phra Nakhon	0	Bar	Café	Asian Restaurant	Thai Restaurant	Hotel
36	Rat Burana	0	Noodle House	Hotpot Restaurant	Japanese Restaurant	Ramen Restaurant	Coffee Shop
37	Ratchathewi	0	Hotel	Convenience Store	Coffee Shop	Asian Restaurant	Buffet
40	Saphan Sung	0	Lake	Brewery	Convenience Store	Playground	Candy Store
41	Sathon	0	Convenience Store	Café	Bar	Chinese Restaurant	Hotpot Restaurant
42	Suan Luang	0	Thai Restaurant	Fast Food Restaurant	Bakery	Japanese Restaurant	Ice Cream Shop
47	Wang Thonglang	0	Thai Restaurant	Asian Restaurant	Fast Food Restaurant	Coffee Shop	Film Studio
48	Watthana	0	Café	Breakfast Spot	Thai Restaurant	Hotel	Restaurant
49	Yan Nawa	0	Coffee Shop	Thai Restaurant	Convenience Store	Performing Arts Venue	Noodle House

Considering Cluster 0, the 1<sup>st</sup> common venue consists of 4 Coffee Shops, 4 Thai Restaurants, 1 Som Tum Restaurant, 1 Chinese Restaurant, 1 Japanese Restaurant. As restaurants exist the most in the cluster, we can say that the Cluster 0 is the Cluster of Restaurant

### Cluster 1

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
10	Bangkok Noi	1	Noodle House	Car Wash	Asian Restaurant	Som Tum Restaurant	Yoshoku Restaurant
12	Bueng Kum	1	Noodle House	Coffee Shop	Vietnamese Restaurant	Hotpot Restaurant	Dim Sum Restaurant
14	Chom Thong	1	Noodle House	Seafood Restaurant	Yoshoku Restaurant	Dumpling Restaurant	Fast Food Restaurant
15	Din Daeng	1	Noodle House	Thai Restaurant	Hotpot Restaurant	Satay Restaurant	Soup Place
34	Pom Prap Sattru Phai	1	Noodle House	Asian Restaurant	Convenience Store	Coffee Shop	Chinese Restaurant
39	Samphanthawong	1	Noodle House	Chinese Restaurant	Dessert Shop	Coffee Shop	Seafood Restaurant

Considering Cluster 1, all of the 1<sup>st</sup> common venue is Noodle House, it can be concluded that this cluster is the cluster of Noodle House.

## Cluster 2

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bang Bon	2	Convenience Store	Thai Restaurant	Shopping Mall	Coffee Shop	Arcade
3	Bang Khen	2	Noodle House	Thai Restaurant	Café	Shopping Mall	Dessert Shop
6	Bang Na	2	Thai Restaurant	Restaurant	Soccer Field	Hotel	Noodle House
7	Bang Phlat	2	Thai Restaurant	Convenience Store	Paper / Office Supplies Store	Noodle House	Bar
20	Khlong Sam Wa	2	Convenience Store	Thai Restaurant	Café	Department Store	Coffee Shop
24	Lat Krabang	2	Thai Restaurant	Convenience Store	Café	Noodle House	Steakhouse
25	Lat Phrao	2	Convenience Store	BBQ Joint	Noodle House	Pub	Thai Restaurant
30	Phasi Charoen	2	Music Venue	Thai Restaurant	Electronics Store	Flea Market	Film Studio
32	Phra Khanong	2	Café	Convenience Store	Thai Restaurant	Noodle House	Ice Cream Shop
38	Sai Mai	2	Thai Restaurant	Convenience Store	Water Park	Spa	Yoshoku Restaurant
43	Taling Chan	2	Noodle House	Other Nightlife	Thai Restaurant	Som Tum Restaurant	Yoshoku Restaurant
44	Thawi Watthana	2	Restaurant	Thai Restaurant	Café	Vineyard	Asian Restaurant
45	Thon Buri	2	Thai Restaurant	Convenience Store	Asian Restaurant	Coffee Shop	Bus Station
46	Thung Khru	2	Cupcake Shop	Music Store	Thai Restaurant	Lounge	Yoshoku Restaurant

Considering Cluster 2, the most common venue is Thai Restaurant. We can say that this is the cluster of Thai Restaurant.

## Cluster 3

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
11	Bangkok Yai	3	Asian Restaurant	Noodle House	Karaoke Bar	Yoshoku Restaurant	Electronics Store

Considering Cluster 3, there is only one neighborhood in this cluster. The most common venue is Asia Restaurant

#### Cluster 4

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
21	Khlong San	4	Convenience Store	Factory	Coffee Shop	Seafood Restaurant	Food Truck
26	Min Buri	4	Convenience Store	Coffee Shop	Thai Restaurant	Auto Workshop	Asian Restaurant
35	Prawet	4	Convenience Store	Halal Restaurant	Food Stand	Noodle House	Dessert Shop

Considering Cluster 4, all of the 1<sup>st</sup> common venue is Convenience Store, it can be concluded that this cluster is the cluster of Convenience Store.

#### Cluster 5

	Neighborhood	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	5	Pet Store	Yoshoku Restaurant	Food	Film Studio	Fast Food Restaurant

Considering Cluster 5, there is only one neighborhood in this cluster. The most common venue is Pet Store

#### Cluster 6

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Bang Khun Thian	6	Flea Market	Coffee Shop	Convenience Store	Deli / Bodega	Electronics Store

Considering Cluster 6, there is only one neighborhood in this cluster. The most common venue is Flea Market

## **5. Discussion**

In this study, Bangkok is a big city with a high population density in a narrow area. The total number of measurements and population densities of the 50 districts in total can vary. Different approaches of exploratory analysis are applied to find insights of the data.

Additionally, to find the best number of clusters. Elbow Analysis was first applied. However, there was no elbow point existed. This might be because Elbow Analysis is not appropriate for the data. Therefore, Silhouette Analysis also used afterwards and found that 7 is the optimal number of clusters.

The analysis ended up with visualizing the locations of each venue cluster, labeled by different colors for different clusters. Moreover, tables demonstrating the neighborhoods and its most common venues in each cluster are also included in the report.

## **6. Conclusion**

Bangkok is one of the cities that attract the considerable numbers of international visitors. One reason might be there are various kinds of tourist attractions, from Restaurants to Convenience Store. People can apply the findings to see the most common venues in each neighborhood. The findings are valuable for both visitors and business persons. The visitors can see the findings in order to find where to visit in each area. Business persons can find the gap to open their business in the neighborhood.