

Bangkok Tourists Recommender System

Soravis Prommas

January 17, 2021

1. Introduction

Bangkok is one of the world's most visited cities. The city is considerably famous among tourists from every corner of the world. There are large number of tourists that have frequently been to Bangkok and visited almost every of its neighborhoods. While another number of tourists also desire to explore this amazing experience, but they do not get any idea about where to visit and what to do in Bangkok.

The purpose of this project is to help those new tourists in exploring the venues, activities and facilities around the city. Since each area has different tone and characteristic, it will help people making smart and efficient decisions on visiting some great neighborhoods matching their interests and lifestyle out of the numbers of other neighborhoods in Bangkok.

This project aims to make an analysis and categorize the neighborhoods (districts) in Bangkok. The neighborhoods sharing some common groups of venues and places will be defined as the same category.

2. Data Acquisition

2.1 Data Requirement

Based on definition of our problem, information that needed to be collected included:

- 1) List of neighborhoods in Bangkok
- 2) Location (latitude, longitude) of each neighborhood
- 3) Existing venues in each neighborhood
- 4) Venue categories

2.2 Data Sources

The neighborhood names and their locations (latitudes and longitudes) will be scraped from an existing Wikipedia page https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok. Then, the Foursquare API will be used to search for the venues in each district. Using an API's specific endpoint, information retrieved from the API will be a list of venue's names and their category around the given radius of the input latitude and longitude.

3. Methodology

According to the problem definition, we want to group the similar neighborhoods together. To compare the similarities of neighborhoods, we explore neighborhoods, count the number of venues in each category and group them into clusters to find similar neighborhoods in Bangkok. To be able to do that, we needed to cluster data by using a machine learning algorithm *K-means Clustering*.

4. Results

By using Scikit-learn to perform K-means clustering algorithm, we ended up with 4 clusters of neighborhood. Figure 1-4 showed some parts of the results.

| Cluster Labels | | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|----------------|---|-----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | 0 | Bang Kapi | Coffee Shop | Japanese Restaurant | Som Tum Restaurant | Noodle House | Thai Restaurant |
| 3 | 0 | Bang Khen | Convenience Store | Coffee Shop | Fast Food Restaurant | Bookstore | Steakhouse |
| 4 | 0 | Bang Kho Laem | Coffee Shop | Noodle House | Convenience Store | Thai Restaurant | Asian Restaurant |
| 5 | 0 | Bang Khun Thian | Thai Restaurant | Coffee Shop | Japanese Restaurant | Ice Cream Shop | Asian Restaurant |
| 13 | 0 | Chatuchak | Coffee Shop | Thai Restaurant | Noodle House | Fast Food Restaurant | Japanese Restaurant |
| 19 | 0 | Khan Na Yao | Coffee Shop | Japanese Restaurant | Noodle House | Ice Cream Shop | Bakery |

Figure 1. Some parts from cluster 0 result.

| | Cluster Labels | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | |
|--|----------------|--------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|---------------------|
| | 8 | 1 | Bang Rak | Noodle House | Thai Restaurant | Hotel | Chinese Restaurant | Café |
| | 21 | 1 | Khlong San | Coffee Shop | Noodle House | Chinese Restaurant | Dessert Shop | Café |
| | 31 | 1 | Phaya Thai | Bar | Thai Restaurant | Café | Coffee Shop | Japanese Restaurant |
| | 33 | 1 | Phra Nakhon | Thai Restaurant | Bar | Hotel | Café | Noodle House |
| | 37 | 1 | Ratchathewi | Noodle House | Coffee Shop | Hotel | Café | Steakhouse |

Figure 2 - Some parts from cluster 1 result

| | Cluster Labels | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|----|----------------|--------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 0 | 2 | Bang Bon | Noodle House | Convenience Store | Thai Restaurant | Coffee Shop | Shopping Mall |
| 2 | 2 | Bang Khae | Convenience Store | Noodle House | Japanese Restaurant | Shopping Mall | Shop & Service |
| 6 | 2 | Bang Na | Thai Restaurant | Noodle House | Pier | Café | Asian Restaurant |
| 7 | 2 | Bang Phlat | Convenience Store | Coffee Shop | Fast Food Restaurant | Hotpot Restaurant | Japanese Restaurant |
| 10 | 2 | Bangkok Noi | Noodle House | Som Tum Restaurant | Asian Restaurant | Convenience Store | Café |

Figure 3 - Some parts from cluster 2 result

| | Cluster Labels | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|----|----------------|--------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 9 | 3 | Bang Sue | Noodle House | Thai Restaurant | Convenience Store | Train Station | Coffee Shop |
| 11 | 3 | Bangkok Yai | Noodle House | Asian Restaurant | Convenience Store | Farmers Market | BBQ Joint |
| 17 | 3 | Dusit | Noodle House | Thai Restaurant | Convenience Store | Asian Restaurant | Café |
| 18 | 3 | Huai Khwang | Thai Restaurant | Noodle House | Convenience Store | Japanese Restaurant | Dessert Shop |
| 29 | 3 | Pathum Wan | Noodle House | Asian Restaurant | Seafood Restaurant | Hotel | Thai Restaurant |

Figure 4 - Some parts from cluster 3 result

We also visualized the results as a cluster map. Figure 5 showed the clusters of neighborhood marked on a Bangkok map. The red, purple, blue and yellow dots on the map indicated cluster 0, 1, 2 and 3 respectively.

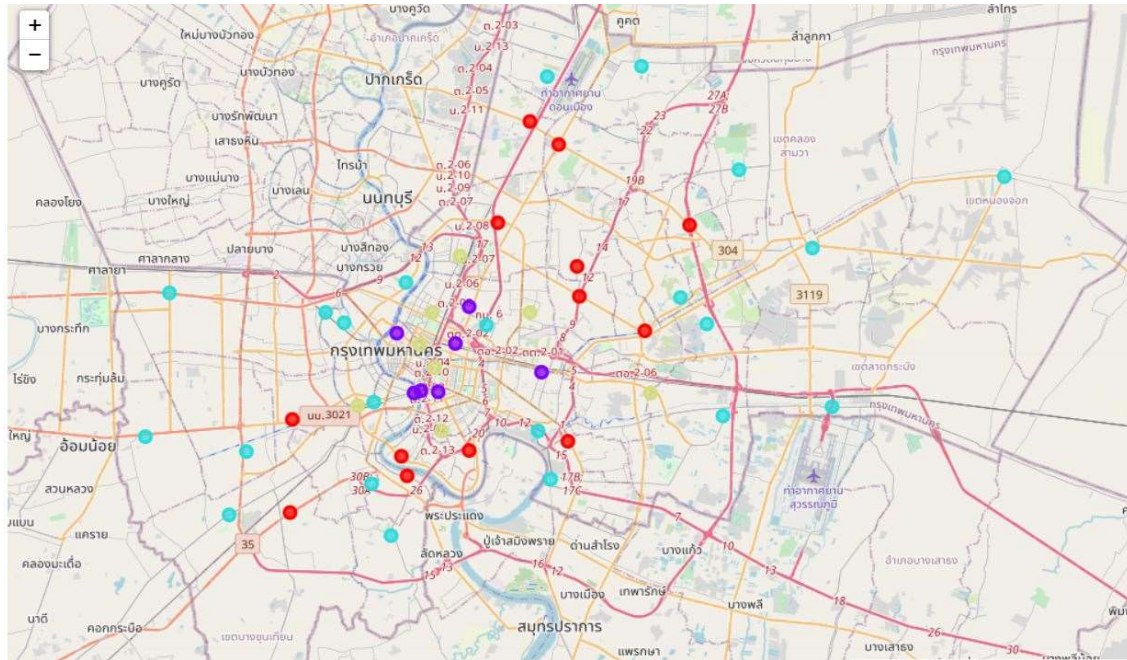


Figure 5 – Map of clusters

5. Discussion

Looking at the results of 4 clusters, we may define each cluster as follows:

Cluster 0: Obviously, this cluster is full of convenience stores, restaurants and cafe. People who love shopping are recommended to visit these neighborhoods. Every branded product in Thailand is available here.

Cluster 1: Neighborhoods in this cluster seem to be the most well-known by the tourists especially Phra Nakhon which is the location of Khaosan Road, a famous place to drink and enjoy night life. This cluster is full of pubs, bars and hotels. So, it's suitable for young backpackers who love party and entertainment. Not recommended for kids and elders.

Cluster 2: This is the suburb of Bangkok. Although there may be lots of restaurant, but most of them all local street food stalls. Either way, the convenience stores here are not that good and most of them are small stores. You may not find every things you want here. Actually, many areas in this cluster are not tourists' destination even for Thai tourists.

Cluster 3: This is the central Bangkok. Completely opposite to cluster 2. Many premium restaurants and five-star hotels are located here. This is also the center of business areas. It also provides the best facilities and infrastructure in Bangkok. Surely this cluster is suitable for tourists who love to try premium Thai and Asian foods and love to enjoy luxury life, not for the low-cost travelers at all. Aside from tourism, it is the best place to run any business.

It is to be noted that the definition above is from the result of our analysis using venue information from Foursquare alone, which may not provide information as much as it does for neighborhoods in the U.S. or any other countries.

6. Conclusion

Purpose of this project is to roughly define the characteristic of each neighborhood in Bangkok for the ease of tourists to establish their destinations and to create their traveling plans. We gathered information about venues in the neighborhoods from Foursquare, then clustered the neighborhoods in regarding with their categories of the ten most common venues using K-mean algorithm. From the final result, we got four clusters and defined the characteristic for each cluster.

For the further development of this project, more information should be gathered from sources other than Foursquare alone which may provide a better analysis.