# Relocation Analysis

Vittawat Dolbandarnchoke

July 12, 2020

# 1.Introduction

## 1.1. Background

I am currently in the process of relocating from the UK to Bangkok, Thailand and would like to find out which neighborhoods or districts should I choose to live in Bangkok that share some similarity to my current neighborhood in Birmingham, UK.

I used to live in Thailand a long time ago and has frequently visited the country, however, my knowledge about neighborhoods in Bangkok is limited. The last thing I want to happen is to choose a neighborhood to live in and find out later that it does not suit my lifestyle or divert too much from what I enjoy in my current neighborhood in the UK.

Thus, it is of my interests to explore available data and find some insights that could help me choose a neighborhood in Bangkok that suits what I prefer.

## 1.2. Problem

Data that I need to perform this analysis are neighborhoods/districts in Bangkok and relevant coordinates, coordinate of my current neighborhood - Edgbaston, Birmingham, UK, and venues data obtainable from Foursquare. I will also set the following requirements of venues I would like to have in the area that are within a set distance as well:

- Gym: within 2km
- Convenience Store/Grocery Store/Supermarket: within 2km
- Coffee Shop: within 1km

## 1.3. Interest

This problem seems to be heavily focused on my current life event, but I do think that the structure of analysis should be helpful for anyone who is moving to a new, unfamiliar area. The analysis should be adaptable to suit the needs of any user. The user will simply have to look out

for neighborhoods/districts data along with the coordinates to perform the same analysis on different locations.

# 2.Data Acquisition and Cleaning

## 2.1. Data Sources

Neighborhoods/Districts data for Bangkok along with relevant coordinates can be found on Wikipedia here. As for my current neighborhood, Nominatim from geopy.geocoders was used to obtain coordinates of the neighborhood based on my post code, which is B5 7SU. Lastly, a Foursquare API was used to obtain venues or point of interests within a specific radius from the coordinates.

## 2.2. Data Cleaning

Data scraped from Wikipedia for neighborhoods/districts in Bangkok was already formatted and ready to use, thus, there wasn't much required to be carried out. Pandas package was used as a tool to scrape this table from Wikipedia. Table below displays an example of the data from Wikipedia.

| District (*Khet*) | Map Nr | Post-code | Thai | Popu-lation | No. of Subdis-tricts *Khwaeng* | 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 |

Since not all features are needed for analysis, the following were dropped from the table for simplicity: Map Nr, Thai, Population, and No. of Subdistricts *Khwaeng*. The final DataFrame for Bangkok Data looks like the following.

| | District | Post Code | Latitude | Longitude |
|---|---|---|---|---|
| 0 | Bang Bon | 10150 | 13.659200 | 100.399100 |
| 1 | Bang Kapi | 10240 | 13.765833 | 100.647778 |
| 2 | Bang Khae | 10160 | 13.696111 | 100.409444 |
| 3 | Bang Khen | 10220 | 13.873889 | 100.596389 |
| 4 | Bang Kho Laem | 10120 | 13.693333 | 100.502500 |

As mentioned in 2.1, Nominatim was used to obtain latitude and longitude of my current neighborhood. These data are stored in another DataFrame in the similar format as Bangkok Data. The following shows the DataFrame storing my current neighborhood's data.

| | District | Post Code | Latitude | Longitude |
|---|---|---|---|---|
| 0 | Edgbaston | B5 7SU | 52.454128 | -1.905193 |

Lastly, venues/point of interests data were obtained through Foursquare API with a limit of 100 venues and radius of 3,000 meters. I chose 3,000 meters as a radius because I happened to currently live in a residential area with few places around, and Bangkok is a large metropolitan with hotspots of venues clustered together. A search through Foursquare with a smaller radius than 3,000 meters yielded significantly fewer results than preferred. Venue data for Bangkok and my neighborhood were then concatenated to a single DataFrame, which were transformed to a final DataFrame that listed out 10 most common venues for each neighborhood. An example of this table is shown below.

| | District | 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 | Bang Bon | Convenience Store | Coffee Shop | Noodle House | Fast Food Restaurant | Asian Restaurant | Thai Restaurant | Flea Market | Badminton Court | Hotpot Restaurant | Market |
| 1 | Bang Kapi | Noodle House | Coffee Shop | Thai Restaurant | Japanese Restaurant | Dessert Shop | Som Tum Restaurant | Fast Food Restaurant | Steakhouse | Clothing Store | Soccer Field |
| 2 | Bang Khae | Noodle House | Coffee Shop | Fast Food Restaurant | Convenience Store | Thai Restaurant | Dessert Shop | Asian Restaurant | Café | Shopping Mall | Department Store |
| 3 | Bang Khen | Coffee Shop | Noodle House | Convenience Store | Hotpot Restaurant | Asian Restaurant | Thai Restaurant | Som Tum Restaurant | Golf Course | Bookstore | Vietnamese Restaurant |
| 4 | Bang Kho Laem | Noodle House | Thai Restaurant | Coffee Shop | Hotel | Asian Restaurant | Pub | Chinese Restaurant | BBQ Joint | Bar | Massage Studio |

# 3.Exploratory Data Analysis

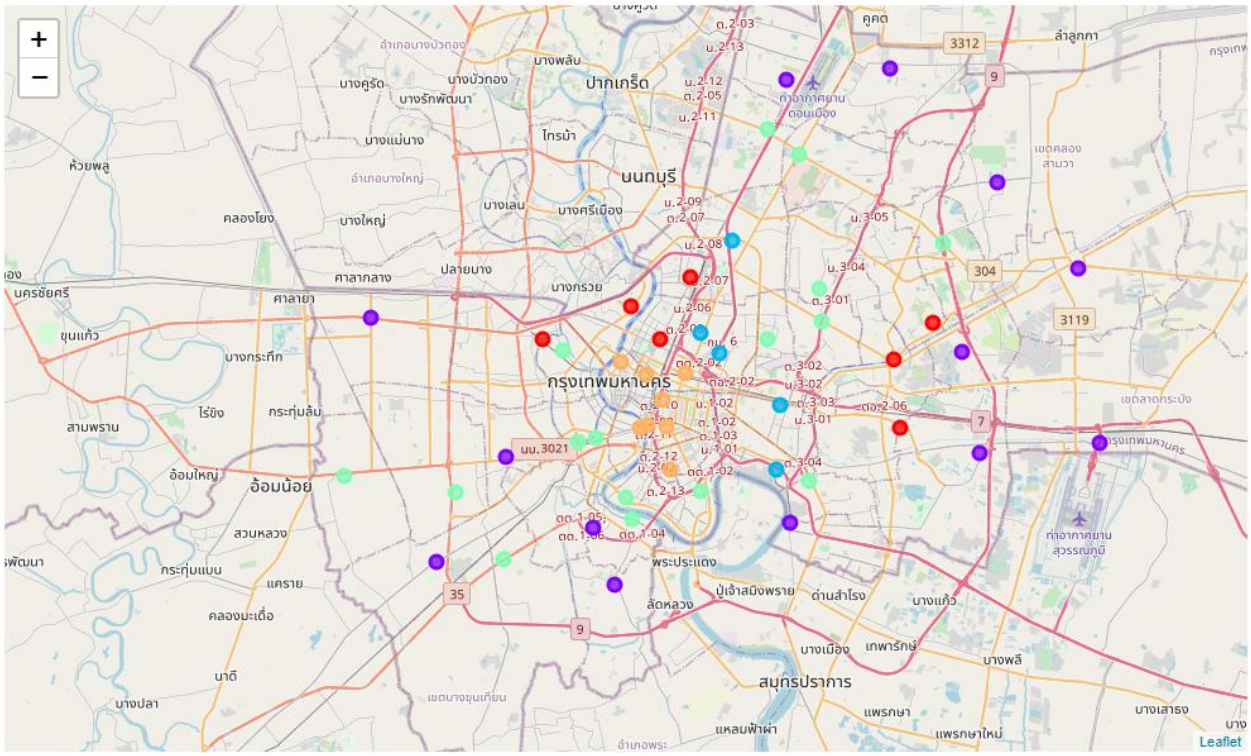
## 3.1. k-Means Clustering

Neighborhoods in Bangkok and my current neighborhood were segregated into different clusters using the k-Means method. The neighborhoods were grouped and labelled together based on the similarity they share from the 10 most common venues per the table above. The number of clusters for k-Means is set to 5, which was chosen after trials with different numbers of clusters and 5 seems to yield a good balance for neighborhoods per cluster. A number smaller or larger than 5 yielded results with few clusters having most neighborhoods and the remaining clusters having very few neighborhoods.

## 3.2. Mapping Neighborhoods

Results from k-Means clustering were plotted on a map of Bangkok via Folium package. A colored circular marker is assigned for each neighborhood based on the cluster label from k-Means. Below is the result map with output indicating the color of the marker for neighborhoods that share similarity to my current neighborhood and the cluster label associated with them. In this case, the neighborhoods with "deepskyblue" as a color and cluster label 2 are those that have similarity with my current neighborhood.

The marker in **deepskyblue** are neighborhoods that share similarity to my current neighborhood, which have Cluster Label 2




Below is also a summary table showing similar neighborhoods to my current one (only part of the table is shown - see more in the Notebook).

| | District | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 13 | Chatuchak | 2 | Coffee Shop | Thai Restaurant | Ice Cream Shop | Gym / Fitness Center | Noodle House | Som Tum Restaurant | Japanese Restaurant | Clothing Store | Hotpot Restaurant | Asian Restaurant |
| 15 | Din Daeng | 2 | Café | Hotel | Thai Restaurant | Coffee Shop | Som Tum Restaurant | Japanese Restaurant | Hostel | Bakery | Bar | Burger Joint |
| 22 | Khlong Toei | 2 | Coffee Shop | Japanese Restaurant | BBQ Joint | Thai Restaurant | Café | Spa | Hotel | Sushi Restaurant | Noodle House | Gym / Fitness Center |
| 31 | Phaya Thai | 2 | Coffee Shop | Thai Restaurant | Noodle House | Café | Japanese Restaurant | Som Tum Restaurant | Restaurant | Bar | Sushi Restaurant | Hotel |
| 48 | Watthana | 2 | Thai Restaurant | Japanese Restaurant | Café | Coffee Shop | Hotel | Bar | Spa | BBQ Joint | Noodle House | Cocktail Bar |
| 50 | Edgbaston | 2 | Pub | Coffee Shop | Indian Restaurant | Café | Bar | Restaurant | Hotel | Park | Indie Movie Theater | Middle Eastern Restaurant |

### 3.3. Narrowing Down the Search

Lastly, I set a set of requirements for the neighborhood to have based on the preference of my lifestyle. The following is the requirement for type of venues and the distance I want:

- Gym: within 2km
- Convenience Store/Grocery Store/Supermarket: within 2km
- Coffee Shop: within 1km

Distance (in kilometer) between each venue and neighborhood it belongs to was calculated using geopy.distance.distance package. The calculated distance is then appended to the

DataFrame, creating a table containing district and its coordinates, venue and its coordinates, and the distance to the venue from district's coordinate. Below is the sample of the table.

| | District | District Latitude | District Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category | Distance(km) |
|---|---|---|---|---|---|---|---|---|
| 0 | Chatuchak | 13.828611 | 100.559722 | Baan Sawasdee (Massage & Cafe) | 13.820331 | 100.560363 | Spa | 0.918676 |
| 1 | Chatuchak | 13.828611 | 100.559722 | Ratchavipha Badminton Court | 13.832558 | 100.560978 | Badminton Court | 0.457273 |
| 2 | Chatuchak | 13.828611 | 100.559722 | Supalai Fitness Club | 13.822622 | 100.564809 | Gym / Fitness Center | 0.861114 |
| 3 | Chatuchak | 13.828611 | 100.559722 | Sewana Paploen (เสวนาพาเพลิน) | 13.828686 | 100.569099 | Pub | 1.013841 |
| 4 | Chatuchak | 13.828611 | 100.559722 | หอจดหมายเหตุพุทธทาส อินทปัญโญ (BIA) Buddhadasa... | 13.816911 | 100.556254 | Library | 1.347610 |

A filter was then applied to the above table based on the requirements mentioned above. Groupby function was used on the resulted DataFrame to display the number of venues for each district. Below is the summary table of the analysis.

| | | Venue |
|---|---|---|
| District | Venue Category | |
| Chatuchak | Coffee Shop | 4 |
| | Gym / Fitness Center | 3 |
| | Supermarket | 2 |
| Din Daeng | Supermarket | 2 |
| Khlong Toei | Coffee Shop | 1 |
| | Convenience Store | 2 |
| | Gym / Fitness Center | 3 |
| Phaya Thai | Coffee Shop | 7 |
| | Gym / Fitness Center | 1 |
| | Supermarket | 1 |
| Watthana | Coffee Shop | 2 |
| | Supermarket | 2 |

The table has shown that the following districts fulfills the requirement I set:

1. Chatuchak
2. Khlong Toei
3. Phaya Thai

There is at least a coffee shop within 1km, a gym within 2km, and a supermarket within 2km from the coordinate of these three districts. The remaining districts do not fulfill all of the requirements.

# 4.Conclusion

In this study, I performed data analysis comparing neighborhoods in Bangkok to my current neighborhood (Edgbaston, Birmingham, UK) for the purpose of finding a similar neighborhood for myself to relocate to in the near future. I utilized data from Wikipedia to obtain a list of neighborhoods in Bangkok and relevant coordinates and use this information to obtain venues within a specific radius for comparison to my current neighborhood. k-Means clustering method was used to segregate neighborhoods based on the 10 most common venues and the results were mapped with color coded markers for each cluster label. Finally, distance between venues and neighborhood's coordinate was calculated to determine the neighborhood that fits my requirement of having a coffee shop within 1km, and a gym and a supermarket within 2km. The result has shown that there are three districts that share similarity in the 10 most common venues to my current neighborhood and has the required venues within the required distance. It seems my next task would be looking for an affordable house in these districts.