

Neighborhoods Segmentation of Bangkok, Thailand

Introduction/Business Problem

Bangkok is the capital city of Thailand and also the most crowd city in the country. From 2010 census conducted by the **National Statistic Office of Thailand**[1], the population in the city was over **8** millions**. **In 2019, it is estimated to home over **10** millions people making it 32th in the most populated city ranking, just a single place behind its neighbor **Ho Chi Minh City** in **Vietname** [2].

Even though the city is home to over **13%** of the country's total population, it is also an important international business areas as well as it prvides facilities to more than 120,000 asian people and notably over 50,000 western people. Bangkok is considered a **primate city**, or one that serves as the population, political and financial center of a country with no other rival city[3].

It is clear that we can find great diversity in terms of businesses, people, and places in this city; however, are those diverse categories mingling together equally in all areas perfectly rendering the city into an ideal mixture or are there still several segments that we can tell apart? Do all areas in Bangkok play equal roles in term of businesses or living places? **Knowing that there are a lot of opportunities to be discovered in this well-populated place, to be able to clearly identifying where we should start some certain businesses will be a great idea and this is the problem that we are going to solve here.**



Data

We will use data from two main sources in this analysis

1. Location data

This will tell us about the name of each district area in Bangkok together with their Latitude and Longitude as Geographic coordinate that will be used to fetch the important places within each area from FourSquare API. In this study, the data will be scraped from geonames using the following site. (https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok)

District (Khet)	Code	Thai	Population	No. of Subdistricts (Khwaeng)	Latitude	Longitude
Bang Bon	50	บางบอน	105,161	4	NA	NA
Bang Kapi	06	บางกะปิ	148,465	2	13.765833	100.647778
Bang Khae	40	บางแค	191,781	4	13.696111	100.409444
Bang Khen	05	บางเขน	189,539	2	13.873889	100.596389
Bang Kho Laem	31	บางคอแหลม	94,956	3	13.693333	100.5025
Bang Khun Thian	21	บางขุนเทียน	165,491	2	13.660833	100.435833
Bang Na	47	บางนา	95,912	2	13.680081	100.5918
Bang Phlat	25	บางพลัด	99,273	4	13.793889	100.505
Bang Rak	04	บางรัก	45,875	5	13.730833	100.524167
Bang Sue	29	บางซื่อ	132,234	2	13.809722	100.537222
Bangkok Noi	20	บางกอกน้อย	117,793	5	13.770867	100.467933
Bangkok Yai	16	บางกอกใหญ่	72,321	2	13.722778	100.476389
Bueng Kum	27	บึงกุ่ม	145,830	3	13.785278	100.669167
Chatuchak	30	จตุจักร	160,906	5	13.828611	100.559722
Chom Thong	35	จอมทอง	158,005	4	13.677222	100.484722

Figure 1: List of Districts in Bangkok with Latitude and Longitude

From the webpage, you can clearly see that we have access to

- Neighborhood name from District column
- The Latitude and Longitude of each district in Latitude and Longitude columns

2. Popular places in each area

Since we will cluster areas in Bangkok based on popular places in each area, we will rely on FourSquare API "Explore" request.

The request url will be in the form of <https://api.foursquare.com/v2/venues/explore?> following by user credentials and some parameters regarding the explore request. In this study, the parameters are as follows:

- ll: Latitude, Longitude
- radius: the area that we are going to observe. In this study, we will use 2,000 m radius to be consistent with the general size of each district
- limit: let's limit our search to top 80 places in each area

3. Income and expense data in each area

To obtain purchasing power of people in each area, income and expense data will be obtained using data collected by National Statistics Office of Thailand.

(<http://service.nso.go.th/nso/web/income/income52.html>)

จังหวัด กรุงเทพมหานคร

ตาราง 1 ตัวชี้วัดในแผนที่ยกย่อง พ.ศ. 2552 : มิติรายได้

จังหวัด/เขต	แขวง	รายได้เฉลี่ยต่อคนต่อเดือน						สัดส่วนคนจน						ช่องว่าง ความยากจน (ร้อยละ)	ความรุนแรง ความยากจน (ร้อยละ)	สัมประสิทธิ์ความไม่เท่าเทียม					
		บาท	SE	เทียบกับค่าเฉลี่ยของ			ร้อยละ	SE	เทียบกับค่าเฉลี่ยของ			ร้อยละ	SE			เทียบกับค่าเฉลี่ยของ					
				จังหวัด	เขต	แขวง			จังหวัด	เขต	แขวง					จังหวัด	เขต	แขวง			
กรุงเทพมหานคร พระนคร		11,829	282	0	0	0	0.8	0.1	0	0	0	0.2	0.0	37.6	0.6	0	0	0			
		10,955	604	0	0	0	1.9	0.7	(-)	(-)	0	0.4	0.1	37.7	1.5	0	0	0			
	วังบูรพาภิรมย์	13,641	1,281	(+)	(+)	0	0.2	0.3	(+)	(+)	0	0.0	0.0	33.6	2.4	(+)	(+)	0			
	วัดราชบพิธ	12,301	1,251	0	0	0	0.3	0.4	0	0	0	0.1	0.0	34.4	2.4	0	0	0			
	สำราญราษฎร์	10,060	794	(-)	0	0	0.9	0.8	0	0	0	0.2	0.0	34.0	2.6	0	0	0			
	ศาลเจ้าพ่อเสือ	7,473	1,013	(-)	(-)	0	3.5	2.4	(-)	(-)	0	0.7	0.2	35.9	3.3	0	0	0			
	เสาชิงช้า	11,619	1,211	0	0	0	0.9	0.8	0	0	0	0.2	0.1	36.8	2.9	0	0	0			
	บวรนิเวศ	6,952	1,328	(-)	(-)	0	5.9	3.4	(-)	(-)	0	1.3	0.5	37.6	4.1	0	0	0			
	ตลาดยอด	13,612	1,580	(+)	(+)	0	0.1	0.2	(+)	(+)	0	0.0	0.0	30.8	3.0	(+)	(+)	0			
	ชนะสงคราม	4,920	747	(-)	(-)	0	16.0	6.7	(-)	(-)	0	3.7	1.3	36.2	3.5	0	0	0			
	บ้านพานถม	10,999	908	0	0	0	1.0	0.6	0	0	0	0.2	0.0	36.9	2.2	0	0	0			
	บางขุนพรหม	11,145	1,142	0	0	0	0.4	0.4	0	0	0	0.1	0.0	34.3	2.1	0	0	0			
	วัดสามพระยา	14,073	1,749	(+)	(+)	0	0.1	0.2	(+)	(+)	0	0.0	0.0	32.7	3.1	(+)	(+)	0			

Figure 2: Income and Expense Data from National Statistics Office of Thailand

With the data from this page, we can have access to income and expense of people, on average, in each area.

4. Area boundary

For plotting purpose, the boundary of each neighborhood in Bangkok will be obtained via geojson file from WorldMap organization (https://worldmap.harvard.edu/data/geonode:adm2_greaterBK_hD4)

Methodology

The methodology to perform the analysis is as follows:

1. Data Scraping:

The location data including the neighborhood name, Latitude and Longitude of each place will be scraped from Wikipedia. For other sources that the data are in .pdf form, the author manually extract useful information from those sources and compile everything into an excel sheet which will be imported later.

2. Data Cleaning:

Even though, Wikipedia has collected useful information about the district name and geological coordinator of each area for us, there are still some districts without coordinator (those with NA in the Latitude and Longitude columns). For those districts, we will use GoogleMap to obtain representative geological locator.

Moreover, we still need to consider that the same district can be spelled in English in several ways; therefore, the integrity of the district names from each data sources must be checked and corrected, if needed.

3. Obtaining popular places from FourSquare:

With FourSquare API, 80 popular places from each districted will be fetched. Then, the category of each place will be determined and grouped together. In order to obtain the feature set for modelling, frequency of each place category for each district will be calculated. For example, if district A has 40 popular Asian restaurants and 40 Hotels in the area as the popular places, the frequency for Asian restaurant and hotel will be 0.5 each.

4. Modelling and Visualization:

For the clustering purpose, K-Means clustering algorithm will be used with the number of clusters of 3 which is obtained from the elbow technique. The features, in this study, will be the frequency that each category of popular places happen in each district. The plot of clusters on the Bangkok map will be provided.

5. Analyzing:

Each cluster of neighbors will be analyzed in term of the most popular type of places in the area. The choropleth map of purchasing power will be created to identify opportunities in each area for new businesses.

Results and Discussion

Here is what the result from the clustering looks like. In Figure 3, it seems like the cluster is grouped, at some certain level, by location which is not the feature in our modelling process. Ones might hypothesize that the regions that are located together might be pretty similar; however, this is not the case as the outcome doesn't look like all groups are consisting of neighbors that are close to one another. In fact, we have a group of neighborhoods sticking together in the middle of the city (**purple dot**), another group surrounding the first group and spreading around the city (**light-green dot**) and the last one at the edge of the city (**red dot**).

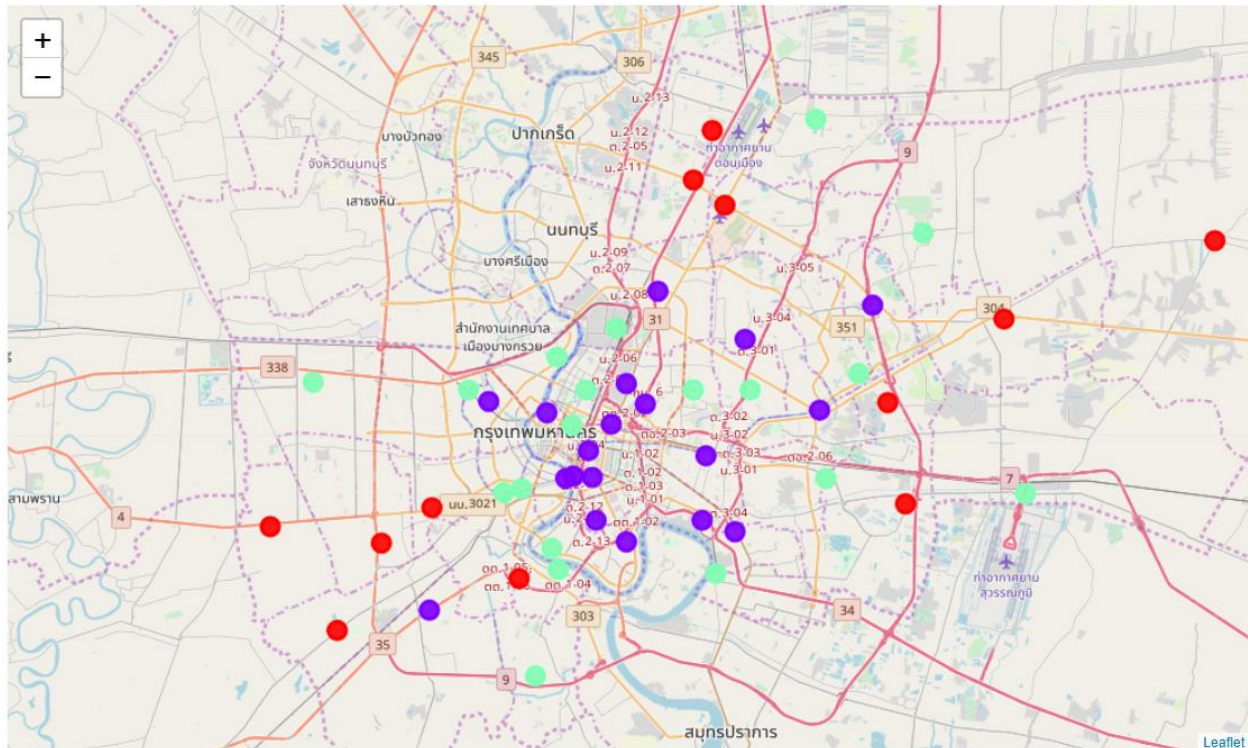


Figure 3: Bangkok Segmentation from K-means Clustering ($K = 3$)

Can we make some senses out of this segmentation? Why there are a lot of same group of businesses emerging at the center of the city? Let's explore the relationship of these clusters with the population density in each district first.

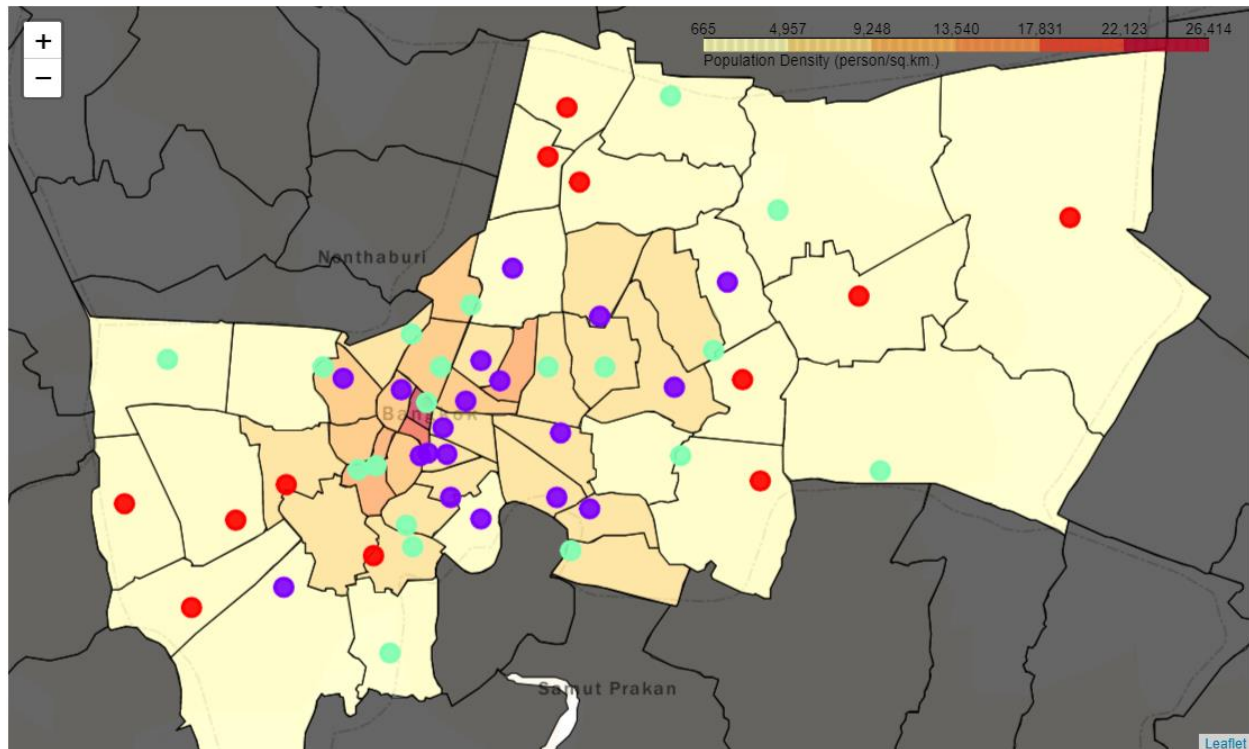


Figure 4: Bangkok Segmentation with Choropleth Map of Population Density

The figure above shows the clustering result on top of the Choropleth Map of the population density data created using geojson file. Note that the district boundary from this source is a little bit off compared with Bangkok area of the base map. You can see that even though Bangkok is the home to over 10% of population of Thailand, we are not living in an equally distributed fashion. A lot of people live close together in the small district at the heart of the city as occupied by **purple group**.

Considering typical businesses in this area, we found that other than some normal restaurants we have lots of **hotels, museums, theaters, bars, and spas**. Yes, these are **Business centers**, the luxury districts with high financial value, and this is the reason why the population density is really high. The districts in this category include **Ratchathewi, Sathon, and Bang Rak**.

The second group is the **light-green** one. These are normal **living areas** of the city including **Thon Buri, Bang Na, and Thawi Watthana**. The common businesses in these areas are **Dessert Shop, Gym, and restaurants from other countries** which can thrive in the living areas.

The last one at the edge can be easily explained as the **developing areas** that will get more opportunity in the future when main public transportation systems of Bangkok like skytrain and subway are fully developed. Until then, the population density will be on the low side and the common businesses that thrive will be limited to those required for simple life like **convenience store, coffee shop and Thai restaurant**. The following box plot summarizes what explained above.

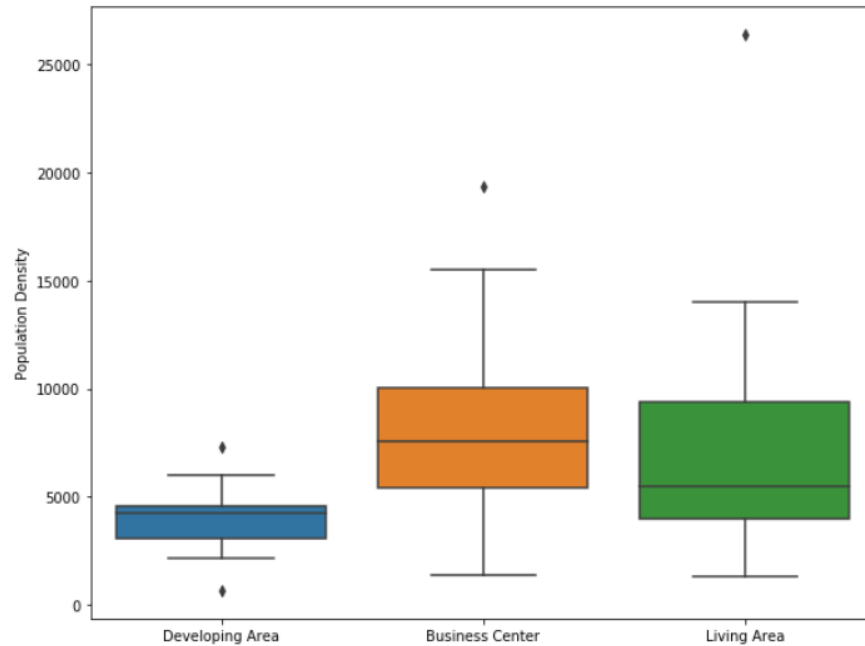


Figure 5: Box Plot Representing Population Density in Each Group

Now, let's answer another of our questions. Where should we focus on starting our own businesses considering the purchasing power of people in the area? By plotting the Choropleth map of the saving (Income – Expense) of people in each area, we obtain the following figure.

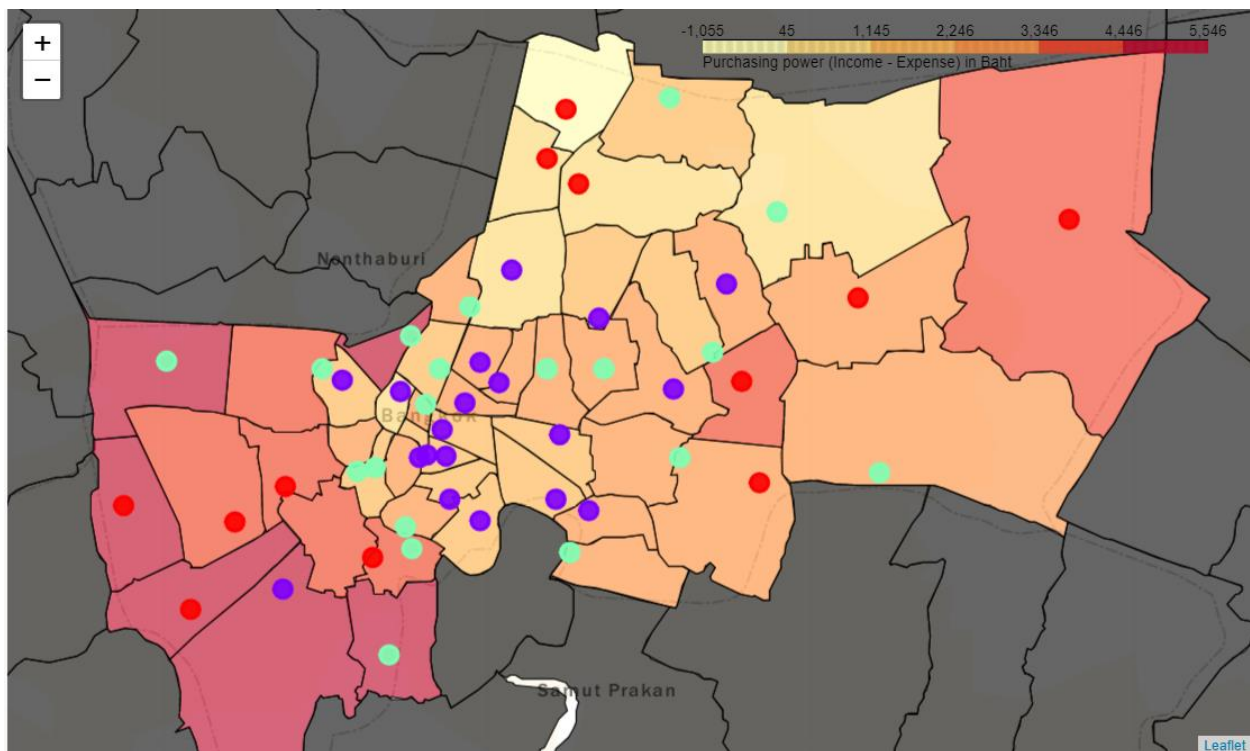


Figure 6: Bangkok Segmentation with Choropleth Map of Purchasing Power in Baht

As one can see, people in the developing area have the most savings per month on average. While the financial opportunity and career opportunity might be greater in the center area, the cost of living in the remote area is much lower resulting in more money left unspent.

If you are going to start your journey at the middle of the town, introducing entirely new business might not be a good idea since, on average, people there have not much saving to pay for other items already. However, we still cannot neglect the fact that people from outside of the town will come here for working in offices as well so we can connect to much more people by having our business in this area.

On the other hand, while opening new type of business to remote area, which might not have that kind of business before, might sound like a good idea since people there still have money left, you will have less customers since they live far apart. However, Bangkok Metropolitan Administration has a solid plan to expand the public rail system to several developing districts including **Lak Si** and **Bang Khae** within the next 2-3 year [4]. Therefore, we can definitely expect to see more people in that region. Probably, considering how living area model currently is, opening some foreign restaurants in these areas might be a great idea for the future.

Conclusion

In this study, we deployed K-means clustering algorithm to segment Bangkok neighborhoods (districts) into three main clusters; the business centers locating at the heart of the city, living area where diverse businesses relating to living aspect thrive, and developing area with hidden future prosperity. We then visualized clusters on Bangkok map and discuss their relationship with different social properties of the area including population density and the gap between income and expense of local people.

Future improvement of the analysis

The population data including income and expense are pretty obsolete as the National Statistic Office of Thailand doesn't collect the data every year. What we have now is the data from several years ago. Moreover, more accurate geojson file is needed for better Choropleth map.

References

- [1] National Statistic Office of Thailand (<http://web.nso.go.th/>)
- [2] Capital City Population Ranking (<http://www.citymayors.com/statistics/largest-cities-population-125.html>)
- [3] Bangkok as the Major City of Thailand (<http://worldpopulationreview.com/world-cities/bangkok-population/>)
- [4] Bangkok public transportation system plan (https://en.wikipedia.org/wiki/Mass_Rapid_Transit_Master_Plan_in_Bangkok_Metropolitan_Region)