

Capstone Project-The Battle of Neighborhoods (Week 2)

Muang Chiang Mai District , New Cafe Locations

By Kittipong Wongsupa



Figure 1: Chiang Mai,Thailand from (<https://bangkokattractions.com/bangkok-to-chiang-mai-flight-bus-train>)

top view photo illustrates the architecture of Chiang Mai and main mounten Doi Suthep.

1. Introduction: Business Problem

Chiang Mai is the favorite traveler city in the north of Thailand; there are 7-9 millions travelers per year from around the world. Because, Chiang Mai has many things special such as Nature, Culture, People etc. A few years ago, the social media made Chiang Mai has many people who want to visit by review from youtuber or vlogger. The place that they visit is a café, there are many cafés built in 4-5 year. But some places were successful, so others failed.

This study will focus on the location of the café vs the location of the accommodation based on the assumption the main customer of the café is a traveler and they will visit the café that is nearby their hotel. In the real situation might be more features that affect the successions of the café but this part will be the main focus for the location.

2. Data acquisition and cleaning

2.1 Data sources

There are two databases. First data from open source data in the link below that provide latitude and longitude of Chiang Mai sub-districts were in Muang Chiang Mai district. There are 15 districts including Phra Sing, Haiya, Chang Moi, Pa Tan, Si Phum, San Phi Suea, Nong Pa Khrang, Chang Khlan, Nong Hoi, Fa Ham, Wat Ket, Tha Sala, Pa Daet, Chang Phueak and Suthep. Other data from Foursquare which is API data for identifying hotel, restaurant,etc location near the location of sub-districts from the first data set. The data and map in Figure 2

Link for open source https://opendata.data.go.th/dataset/item_c6d42e1b-3219-47e1-b6b7-dfe914f27910

2.2 Data cleaning

The data from the open source for district data included the district of the whole country; the first cleaning was the district in Chiang Mai province. Second, Filter specifically in the sub-district in Muang Chiang Mai district only the center of Chiang Mai the traveler visits is Muang Chiang Mai district. The database was in the excel file and then imported to the jupyter by CSV form.

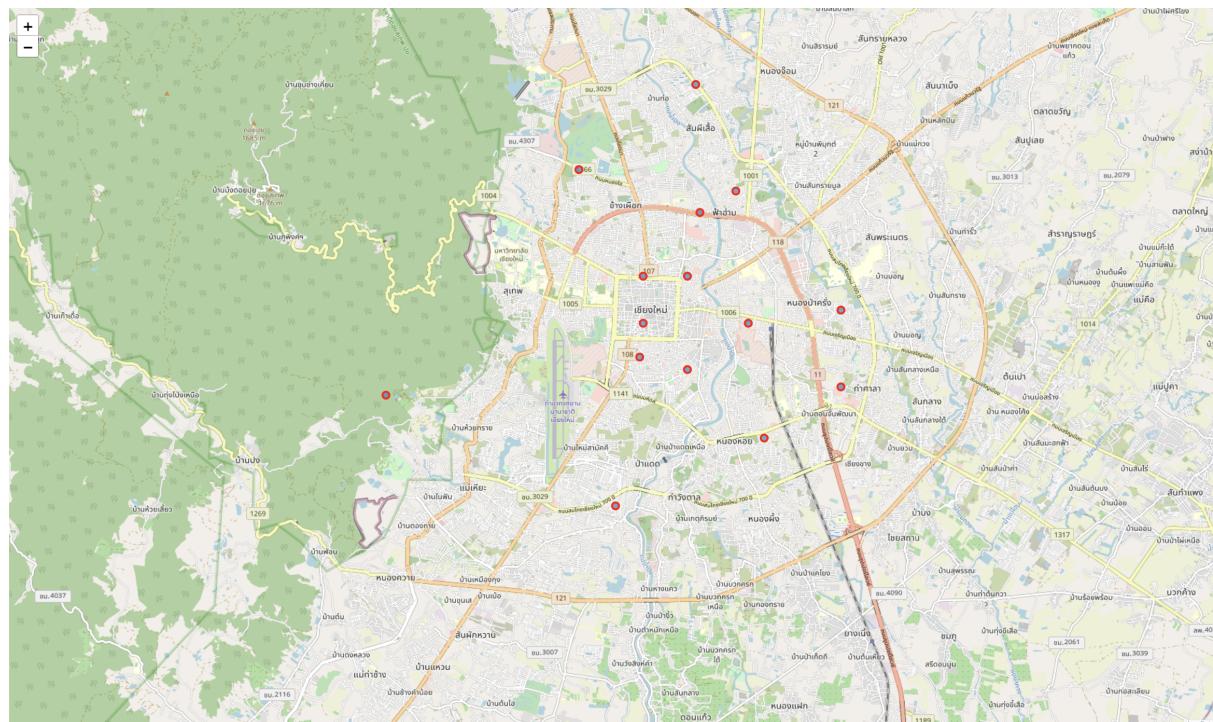


Figure 2 The map illustrate the location of sub district around center of Chiang Mai city show in red dots (Foursquare)

AD_LEVEL	TA_ID	TAMBON_T	TAMBON_E	AM_ID	AMPHOE_T	AMPHOE_E	CH_ID	CHANGWAT_T	CHANGWAT_E	LAT	LONG
0	4	500102	ต. พระสิงห์	Phra Sing	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.785	98.985
1	4	500103	ต. หาดใหญ่	Hai Ya	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.777	98.984
2	4	500104	ต. ช้างม่อย	Chang Moi	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.796	98.996
3	4	500115	ต. ป่าตัน	Pa Tan	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.811	98.999
4	4	500101	ต. ศรีภูมิ	Si Phum	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.796	98.985
5	4	500116	ต. สันมีเสือ	San Phisuea	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.841	98.998
6	4	500113	ต. หนองบัวครัง	Nong Pa Khrang	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.788	99.034
7	4	500105	ต. ช้างคลาน	Chang Khlan	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.774	98.996
8	4	500111	ต. หนองหอย	Nong Hoi	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.758	99.015
9	4	500114	ต. พ้าส่าน	Fa Ham	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.816	99.008
10	4	500106	ต. วัดเกต	Wat Ket	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.785	99.011
11	4	500112	ต. ท่าศาลา	Tha Sala	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.770	99.034
12	4	500110	ต. ป่าแดด	Pa Daet	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.742	98.978
13	4	500107	ต. ช้างเมือก	Chang Phueak	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.821	98.969
14	4	500109	ต. สุเทพ	NaN	5001 อ. เมืองเชียงใหม่	Mueang Chiang Mai	50	จ. เชียงใหม่	Chiang Mai	18.768	98.921

Table 1 The data set sub district in Muang Chiang Mai district includes 15 sub districts and others information latitude, longitude and etc. The data display from Jupiter as pandas liberty.

The API hotel, restaurants etc were imported 500 meter of each sub-district location, the example shown in Figure 3. For using them for analysis and identifying the good location for a new opening cafe. The analysis was used unsupervised machine learning (K-mean clustering)

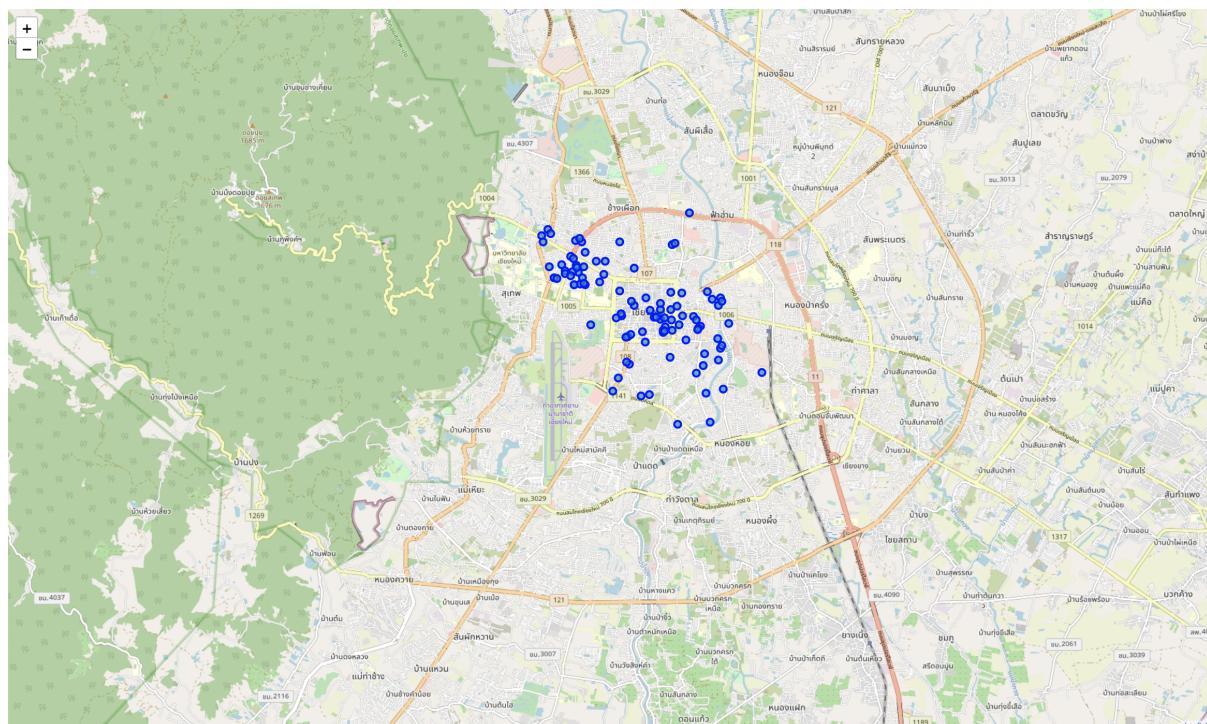


Figure 3 The example hotel that is located (shown in blue dots) around The geographical coordinate of Chiang Mai City (lat 18.7743599, long 98.9771944.)

3. Methodology

The methodology of this study was started by an import library that will be used, including numpy for mathematics calculation, pandas for evaluate and manage data, json for read data, matplotlib for data visualization, sklearn for machine learning and fulium for map plotting.

Then import CSV file data of sub districts and use padas liberty to manage and create data frames as shown in Figure 2. Reselection for only name and latitude and longitude of the sub district data in the name 'df'. Next use the geocode to find the latitude and longitude of Chiang Mai and assign it to 'latitude' and 'longitude'.

When the data and coordinate of Chiang Mai is ready, Sub district data will be plotted in the map that shows in Figure 2, The next step it uses the API data from Foursquare of the hotel data in 500 radian around each sub district location. The example was in Figure 3 (this is the data in only latitude and longitude of the center of the city not of each sub-districts location).

The analysis step is preparing the data set for machine learning (K-mean clustering) to identify sub district location of each cluster by the data of hotels nearby. The K-mean has chosen 5 clusters to separate this data.

4. Result and Discussion

After using API to search the hotel's location nearby, each sub-district found 155 for all places. The data showed Phra Sing sub district has the most nearby hotels number and is followed by Si Phum. On the other hand Fa Ham, Nong Pa Khrang and San Pri Suea have less nearby cafes. The all details in Table 2.

After run machine learning (K-mean clustering) the separate in 5 clusters. The first cluster includes Wat Ket and Pa Daet sub districts. The second cluster includes Phra Sing, Hai Ya, Chang Moi, and Sri Phum. The third cluster includes Pa Tan, Nong Pa Khrang, Chang Khlan, Nong Hoi and Fa Ham. The fourth cluster includes San Phi Suea. The last cluster includes Chang Phueak. There are clusters was separate and plot in the map in Figure 4

TAMBON_E	TAMBON_E Latitude	TAMBON_E Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
TAMBON_E						
Chang Khlan	7	7	7	7	7	7
Chang Moi	30	30	30	30	30	30
Chang Phueak	2	2	2	2	2	2
Fa Ham	1	1	1	1	1	1
Hai Ya	15	15	15	15	15	15
Nong Hoi	4	4	4	4	4	4
Nong Pa Khrang	1	1	1	1	1	1
Pa Daet	2	2	2	2	2	2
Pa Tan	3	3	3	3	3	3
Phra Sing	47	47	47	47	47	47
San Phisuea	1	1	1	1	1	1
Si Phum	40	40	40	40	40	40
Wat Ket	2	2	2	2	2	2

Table 2 this table show the number of hotels nearby each sub district

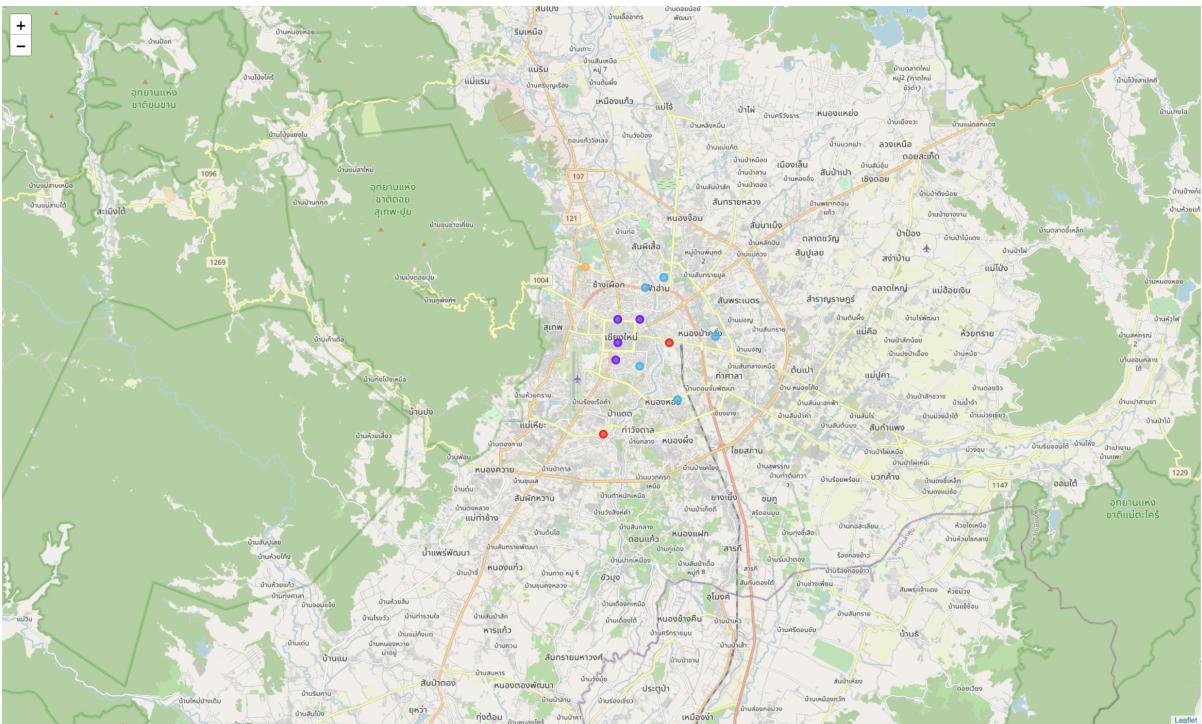


Figure 4 the map shows the cluster locations of sub districts located separate by color.

From the K-mean clustering analysis to identify the cluster for the data and group the data by using the nearby hotel information can be used for choosing the location for a new coffee shop.

5. Conclusion

The result of the K-mean cluster method can be used to identify Phra Singh, Hai Ya, Chang Moi, and Sri Phum is the best group location for open new coffee shops. The data confirmed by these locations was found in a

number of hotels nearby. This is based on the assumption the cafe in Chiang Mai was almost visited by travelers. By the way there still are more factors for the study have to add in the study for choose the best location which is the precise of renting and the supply of coffee or the comfortable for parking lot. All of this would be studied further.

In addition the hotel's data that exported from Foursquare might not relate to the real hotel that exists in reality, because there are many hotels that do not update in the web data. This study uses the prototype to guide the way to study. It still needs additional information and updated information.

The last comment: the location of the sub district is the point location at the center of the area that might not represent the area for use. The study should use the location of the center the zoom in to the exact area the might break down to smaller than sub district but the this would involve with economic area in Chiang Mai