Capstone Project - The Battle of Neighborhoods

Housing Sales Prices & Venues Data Analysis of Bangkok

Introduction

A.1. Description & Discussion of the Background

Bangkok is the capital and most populous city of Thailand. It is known in Thai as Krung Thep Maha Nakhon or simply Krung Thep. The city occupies 1,568.7 square kilometers (605.7 sq mi) in the Chao Phraya River delta in central Thailand, and has a population of over eight million, or 12.6 percent of the country's population. Over fourteen million people (22.2 percent) lived within the surrounding Bangkok Metropolitan Region at the 2010 census, making Bangkok the nation's primate city, significantly dwarfing Thailand's other urban centers in both size and importance to the national economy .The city of Bangkok has a population of 8,305,218 according to the 2010 census, or 12.6 percent of the national population. In 2018, the population has been estimated to be about 10 million. The city is divided into 50 districts in to total. Bangkok city's population density is an average 5,300 per square kilometer. When we think of it by the investor, we expect from them to prefer the districts where there is a lower real estate cost and the type of business, they want to install is less intense because Bangkok is the economic center of Thailand, and the heart of the country's investment and development. However, it is difficult to obtain information that will guide investors in direction. When we consider all these problems, we can create a map and information chart where the real estate index is placed on Bangkok and each district is clustered according to the venue density.

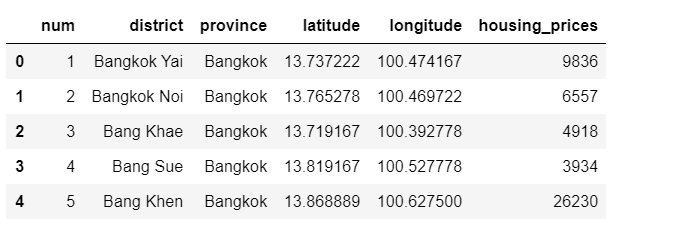
A.2. Data Description

To consider the problem we can list the data as below: We found geographic coordinate of Bangkok in website https://www.geonames.org/ and convert degrees minutes seconded to decimal degrees. We found Housing Sales Prices in website https://www.treasury.go.th/th/summary-of-land-valuation/ collected latest per square meter Housing Sales Price (HSP) Averages for each Borough of Bangkok housing retail. Cleaned the data and reduced it to city of Bangkok where we used it to create choropleth map of Housing sales Price Index of Bangkok. I used Foursquare API to get the most common venues of given Borough of Bangkok. There are not too many public data related to demographic and social parameters for the city of Bangkok. You must set-up your own data tables in most cases and used Google Map, ‘Search Nearby’ option to get the center coordinates of each borough.

A.3 Data Research and Preparation

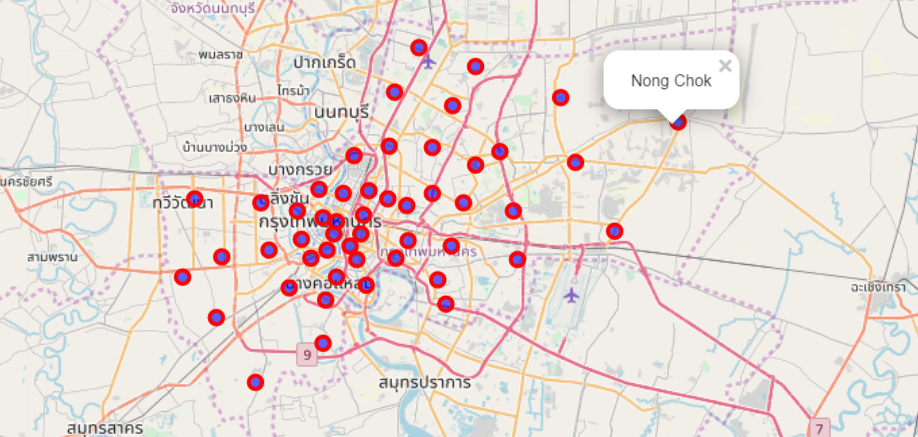
Bangkok is divided into 50 Arrondissements and initially looking to get this data by scraping the relevant Wikipedia page (https://en.wikipedia.org/wiki/Bangkok), fortunately, after much research, this data is available on the web and can be manipulated and cleansed to provide a meaningful dataset to use.

Table 1.1 50 Columns , 6 Row



B.1 Methodology

As a database, I used GitHub repository in my study. Let's download and import the data on police department incidents using pandas read\_csv() method. Now that we reduced the data a little bit, let's visualize (used python folium library) where these Housing Sales Prices took place in the city of Bangkok. We will use the default style and we will initialize the zoom level to 12. Now let's superimpose the locations of the Housing sale prices onto the map. The way to do that in Folium is to create a feature group with its own features and style and then add it to the Bangkok map.



We utilized the Foursquare API to explore the boroughs and segment them. We designed the limit as 100 venue and the radius 750 meter for each borough from given latitude and longitude information. We clean the json and structure it into a pandas data frame



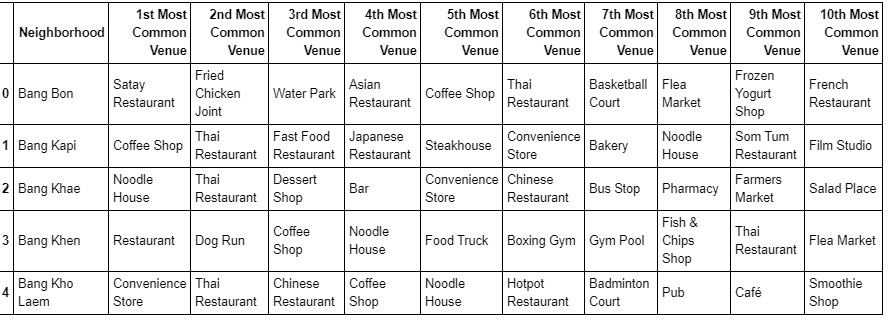
In summary of this data 17 venues were returned by Foursquare. Here is merged table of borough and vanues.



We can see that Bang Rak,Pathum Wan, Phra Nakhon, Ratchathewi and Watthana how reached the 100 limit pf venues. In our given coordinates with Latitude and Longitude, in below table 1.2

The result doesn’t mean that inquiry run all the possible results in boroughs. Actually, it depends on given Latitude and Longitude informations and here is we just run single Latitude and Longitude pair for each borough. We can increase the possibilities with Neighborhood informations with more Latitude and Longitude informations.

In summary of this graph 208 unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each borough in below table.



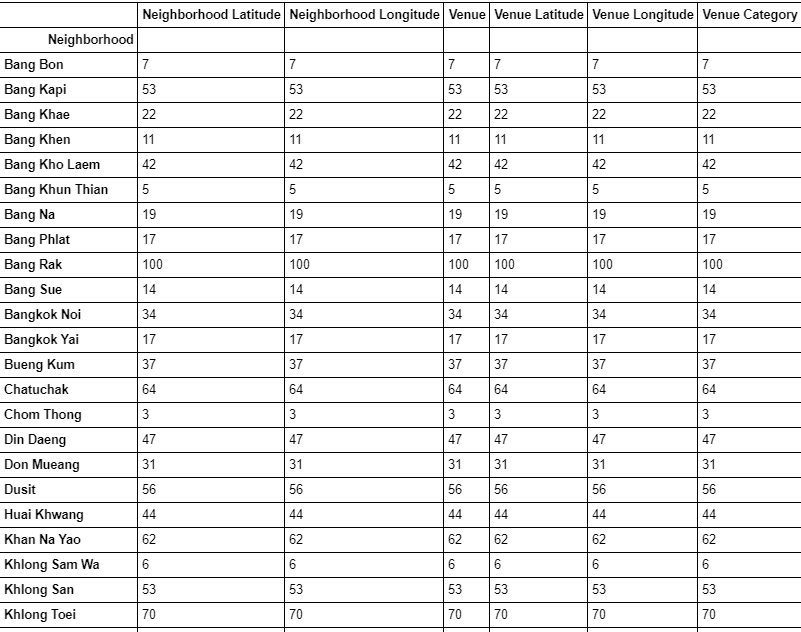
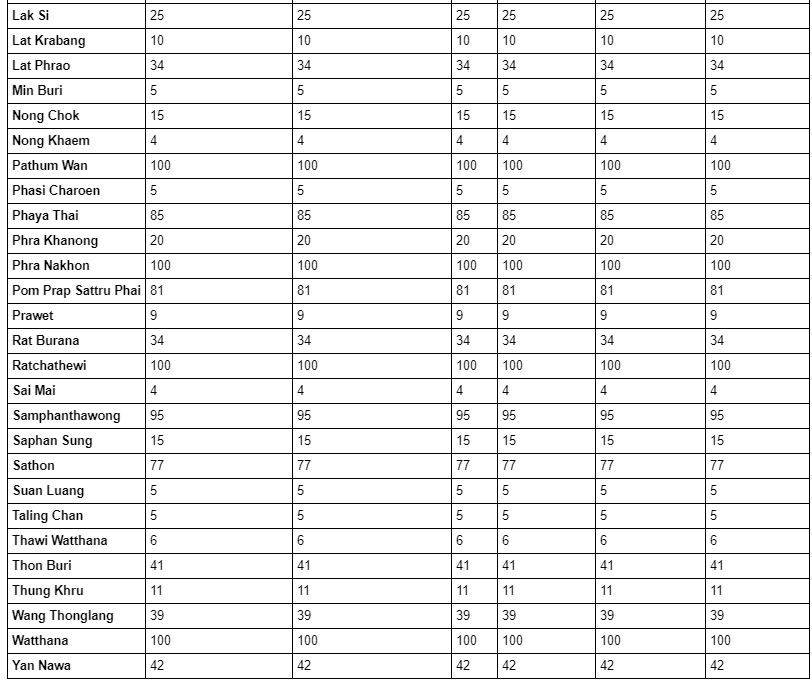
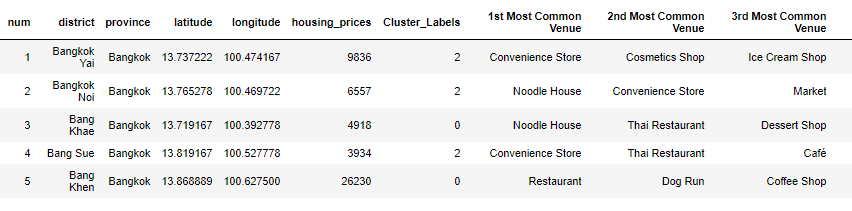
 

Table 1.2

B.2 Methodology

We have some common venue categories in boroughs. In this reason I used unsupervised learning K-means algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster methods of unsupervised learning. First, I will run K-Means to cluster the boroughs into 3 clusters and merged table with cluster labels for each borough.



We have examined each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, you can then assign a name to each cluster.

Cluster 0



Cluster 1



Cluster 2



When we examine above table, we can label each cluster as follows:

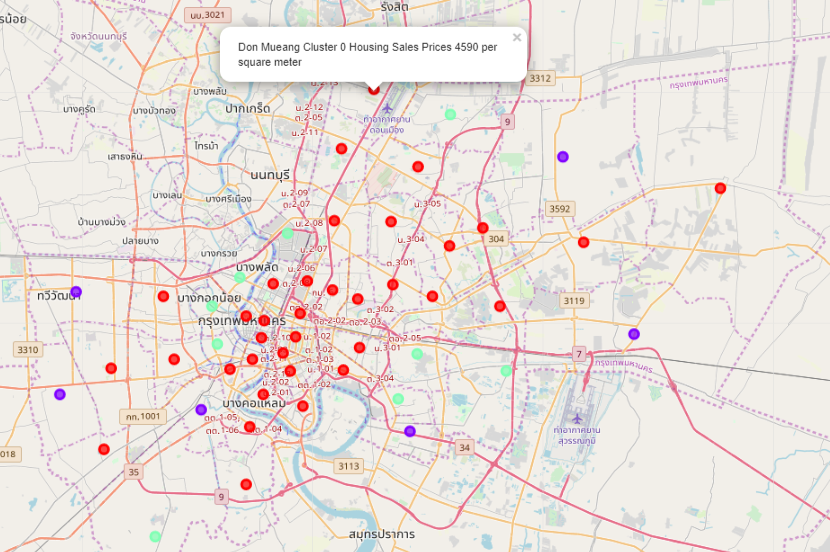
Cluster 0: “Noodle House & Coffee Shop Venues”, Red Cycle.

Cluster 1: “Thai Restaurant Venues”, Violet Cycle.

Cluster 2: “Convenience Store Venues”, Green Cycle.

C. 1 Results

You can also see a clustered map borough of Bangkok in the below.



D. 1 Discussion

As We mentioned before, 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. As there is such a complexity, very different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high-quality results for this metropole.

We used the K-means algorithm as part of this clustering study. However, only 50 district coordinates were used. For more detailed and accurate guidance, the data set can be expanded and the details of the neighborhood or street can also be drilled.

We also performed data analysis through this information by adding the coordinates of districts and home sales price averages as static data on GitHub. In future studies, these data can also be accessed dynamically from specific platforms or packages.

We ended the study by visualizing the data and clustering information on the Bangkok map. In future studies, web or telephone applications can be carried out to direct investors.

F. Conclusion

As a result, people are turning to big cities to start a business or work. For this reason, people can achieve better outcomes through their access to the platforms where such information is provided.

Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.

To the future,

Wittawut.T