

Bangkok Property Investment Analysis

Author : Chittawat Pourpongpan

Introduction

Bangkok is the capital city of Thailand which could be divided into 50 districts. There are currently more than 10 million people living in Bangkok. With the size of the metropolis, it is difficult to determine the suitable living area for those who want to invest in the property.

The goal of this project is to analyse the suitable area for the property investment within Bangkok. The target audiences of this analysis are a stakeholder who is looking to make a property investment in one of the major cities in South East Asia. The analysis will be based on the location data obtained using the foursquare API.

The analyse is developed in Jupyter notebook using Python 3 with aids of multiple libraries. One of the important libraries is Panda which its functionality is mainly involving data. The second package is NumPy allowing advanced mathematical process possible in Python program. Scikitlearn is also a relevant library using in implementing the machine learning model. Next is beautiful soup. Beautiful soup is responsible for processing the webpage and output as an HTML code. Final library is Folium which was used to visualise the cluster on the map.

Data Collection

The first stage of the data collection is to find the information related to districts in Bangkok. This data can be found in from Wikipedia list of districts in Bangkok containing the name, postcode, population, number of subdistricts, latitude and longitude. The location information of each district gained is used to search for the interested venues with foursquare API. The location of interest obtain from the API for the analysis are :

- Number of schools in each district
- Number of supermarkets
- Public transport availability (based on sky train station and river pier)
- Number of hospital or medical centre
- Number of green areas

Methodology

To develop the analysis for the topic, first, the relevant data related to Bangkok districts to be collected. The district data from the Wikipedia was collected using the beautiful soup library which helps transfer webpage to the notebook in the form of HTML. Next, the code will be developed to gather the data from the list of districts and store it in the form of lists representing data in columns including name of district, Postcodes, number of subdistricts, Latitude, Longitude. Finally, data is stored in the data frame of relevant district information.

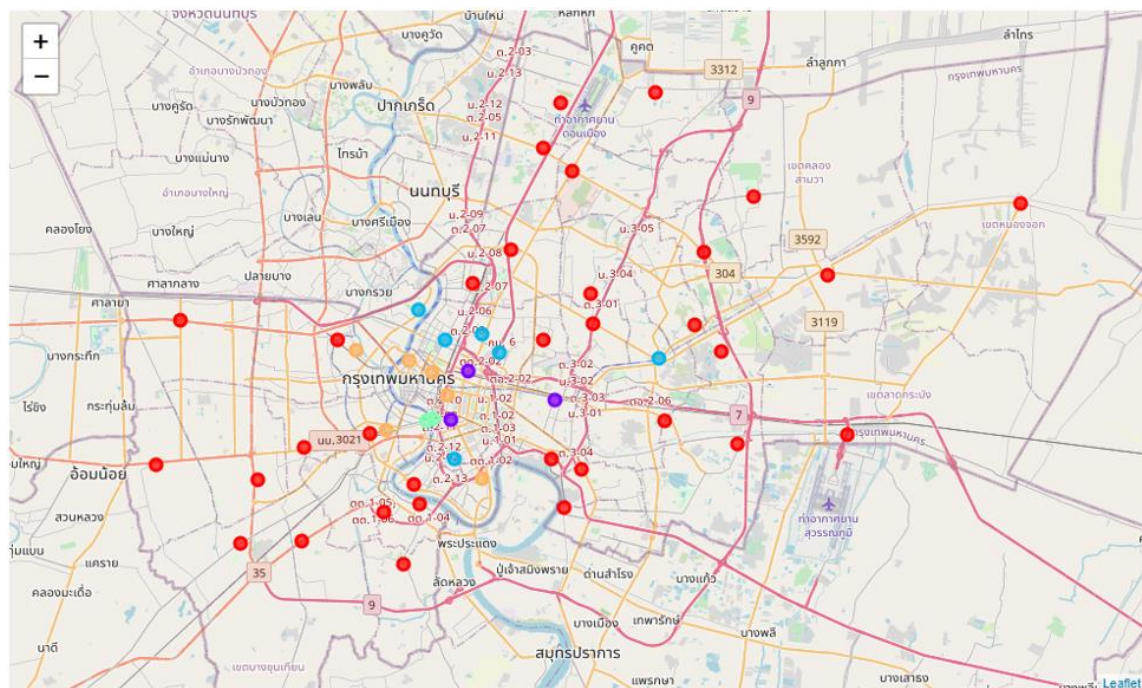
With the district data successfully collected, the location information (Name, latitude, longitude) is use in collecting venue data specified in the previous section using four square API. After collecting each of the venue category, the data is filtered to remove the unrelated venue and remove the duplicated data. After all the data for each category is collected and filtered , all the location data are combined and sorted into each category.

After completing all data collection, the data are pre-processed. This stage involving convert the data frame into one hot state and group the data based on the districts. The data are transformed to produce the number of each venue category for each district. This allows to understand the amount of facility within the district which are a factor of becoming a suitable district for property investment.

Finally, the pre-processed data are modelled using K-mean clustering , allowing the algorithm to produce the cluster which rate the district based on number of facilities available. The number of clusters is set to 5 which rate the district from 1 – 5. The cluster label is included in the main table which include the district information and number of venue available. All the location is visualised on the map representing Bangkok metropolis using folium library.

Result

Figure 1 represented the visualisation of clusters of districts within Bangkok metropolis where each colour representing the suitability rating for property investment. Most of the districts in the outer area of the cities are classified as unsuitable where the suitability of living is rated differently for the district within the inner zone.



Unsuitable

Suitable



Figure 1 The map visualising the district rating analyse by k-mean clustering

Discussion

The visualise map in the previous section shows that most of the districts in the outer zone of the metropolis are considered not suitable for the investment. This is due that most of the district in the outer zone has lack of the venue available when compared to the area in the inner zone. The most suitable area for property investment is including Bangkok Noi, Pathum Wan, Phra Nakorn, Pom Prap Sattru Pai, Thon Buri and Yannawa. These districts have a balance in all venues within the focus categories.

With the analysis, there are some aspects which might cause the accuracy of the analysis. Firstly, the API used in this analysis limited only 100 venues per API calls. This makes some of the data could not be fully collected. Another problem is that the amount of search radius is fixed to 1500 meters from the centre of the district. This makes some of the venues outside the search radius was not observed.

Conclusion

The most suitable district in Bangkok for the property investment are Bangkok Noi, Pathum Wan, Phra Nakorn, Pom Prap Sattru Pai, Thon Buri and Yannawa. These districts have a suitable number and a balanced number of venues within the focus category. Most of the district which has a suitable facility is mostly located in the inner area of Bangkok

To increase the accuracy for suitable property investment analysis, many factors will be considered in the future work including population density and average property price in each area. As the current result did not show all venues in the district due to the API limitation, it is possible to increase the search of nearby venue by using the other API such as google map place API. The searching radius could be set dynamically in the order to adjust the search for each district area suitability.