

Capstone Project – The Battle of Neighborhoods

Location analysis for launching new hotel in Bangkok

Kuntida Uampranee

Business Problem

Bangkok is the capital of Thailand. Before the pandemic began, Bangkok was one of the world's top tourist destination cities. Each year approximately 22.7 million international visitors arrive in Bangkok. Also, MasterCard ranked Bangkok as the top destination for global travelers. Till now, there are millions of people have been vaccinated around the world. The travelling will be return soon. To serve the travelers around the world are looking forward to visiting Bangkok, the lodging is dedicated to perfecting the travel experience. Therefore, if we can help decision making on finding the potential location to construct new hotel. It can reduce the risk of investors.

This project tries to understand the preferences of each district by leveraging venue data from Foursquare's "Places API" and "K-means clustering" machine learning algorithm.

Data Acquisition

1.1 Data sources

- The list of districts get from this https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok link
- Venue data from Foursquare API

1.2 Data extraction

pandas.io.html is web scraping technique which use to extract the data from Wikipedia page https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok to get Bangkok districts

```

from pandas.io.html import read_html
url='https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok'
wikitable = read_html(url, attrs={"class":"wikitable"},header =0)
wiki_df = pd.DataFrame(wikitable[0])
wiki_df= wiki_df.rename(columns={'District(Khet)': 'DistrictEng'})
wiki_df= wiki_df.rename(columns={'MapNr': 'code'})
wiki_df= wiki_df.rename(columns={'Post-code': 'Postcode'})
wiki_df= wiki_df.rename(columns={'Thai': 'District'})
wiki_df= wiki_df.rename(columns={'Popu-lation': 'Population'})
wiki_df= wiki_df.rename(columns={'No. ofSubdis-tractsKhwaeng': 'No_of_Subdistricts'})
wiki_df.head()

```

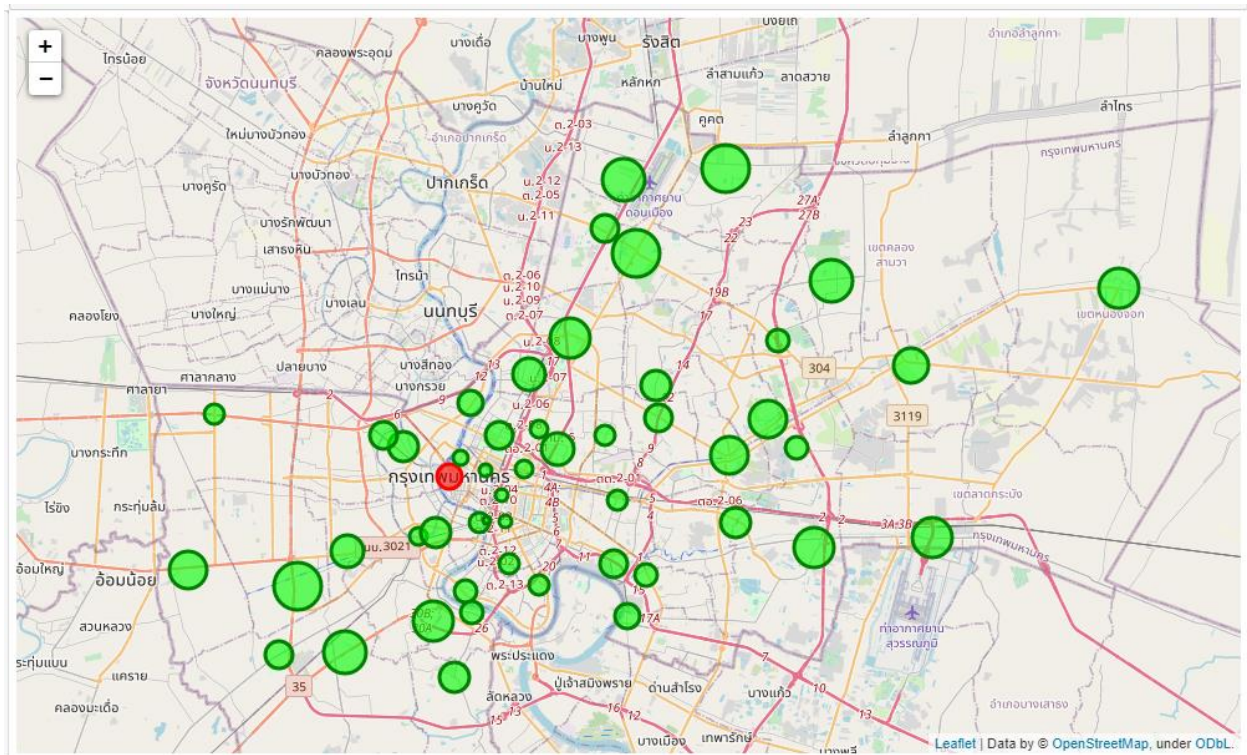
Below is the sample of Bangkok districts data from Wikipedia which contain district name in both English and local language, district code, postcode, population, latitude, and longitude

	DistrictEng	code	Postcode	District	Population	No_of_Subdistricts	Latitude	Longitude
0	Bang Bon	50	10150	บางบอน	105161	4	13.659200	100.399100
1	Bang Kapi	6	10240	บางกะปิ	148465	2	13.765833	100.647778
2	Bang Khae	40	10160	บางแค	191781	4	13.696111	100.409444
3	Bang Khen	5	10220	บางเขน	189539	2	13.873889	100.596389
4	Bang Kho Laem	31	10120	บางคอแหลม	94956	3	13.693333	100.502500

2. Exploratory data analysis

To extract the location of each district, Python Geocoder package is the technique use to plot into the map by using latitude and longitude from Wikipedia.

Red circle represents the central area of Bangkok. Size of green circle represent population in each district.



Sample hotel in each district from Foursquare

	District	Population	Code	Latitude	Longitude	VenueName	VenueId	VenueLatitude	VenueLongitude	VenueCategory
0	บางบอน	50	105161	13.659200	100.399100	ชาหนุมางหว่า	4e880a81f790e992e01d7284	13.657136	100.395230	Thai Restaurant
1	บางบอน	50	105161	13.659200	100.399100	ข้าวต้มม่วน	4ec7d62f0e6158bafef41324	13.666550	100.412108	Asian Restaurant
2	บางบอน	50	105161	13.659200	100.399100	ร้านต้นไม้ ริมถนนกาญจนาภิเษก	4bf8e392508c0f4796f13e31	13.654098	100.405054	Garden Center
3	บางบอน	50	105161	13.659200	100.399100	KFC	570f0a1bcd109301f16f70fa	13.670449	100.405502	Fast Food Restaurant
4	บางบอน	50	105161	13.659200	100.399100	Burger King (เบอร์เกอร์คิง)	5884172803e29a6e0f77197e	13.670830	100.405089	Fast Food Restaurant
...
4866	ยานนาวา	12	81521	13.696944	100.543056	Pullman Bangkok Hotel G (โรงแรมพูลแมน กรุงเทพฯ)	4b9b1d27f964a520b8f335e3	13.726129	100.525795	Hotel
4867	ยานนาวา	12	81521	13.696944	100.543056	Rengaya (เรนกายา)	4d96ae48af3d236acf1311c7	13.729172	100.533386	BBQ Joint
4868	ยานนาวา	12	81521	13.696944	100.543056	25 Degrees (ทเว็นตี้ไฟว์ ดีกรีส์)	4fb34c70e4b0d8fe963a2826	13.725711	100.525800	Burger Joint
4869	ยานนาวา	12	81521	13.696944	100.543056	สวนธรรมชาติ@สวนลุมพินี	50a7247de4b027fd39cdc095	13.730093	100.539104	Park
4870	ยานนาวา	12	81521	13.696944	100.543056	Tealicious Café (ทีแอลไอเค)	52a578c0498ef669db82cdfb	13.722545	100.516866	Thai Restaurant

4871 rows x 10 columns



The total number of hotels in each district shows as bar chart as below

