The Battle of Neighborhoods

Which district in 4 cluster from 9 main economic center district in Bangkok new entrepreneur can open coffee shop

CAPSTONE PROJECT - FINAL REPOR
IBM – COURSERA

Rattapon Choogrotoud 2019

Table of content:

I. Introduction:

This report is for the final course of IBM Data Science the courses series created by IBM, hosted on Coursera platform. The problem and the analysis approach are left for the learner to decide, with a requirement of leveraging the Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve.

The main goal will be exploring the neighborhoods in 9 main economic center district in Bangkok for recommend new entrepreneur can open coffee shop and still have opportunity escape STARBUCK and other Coffee Shop from dense area.

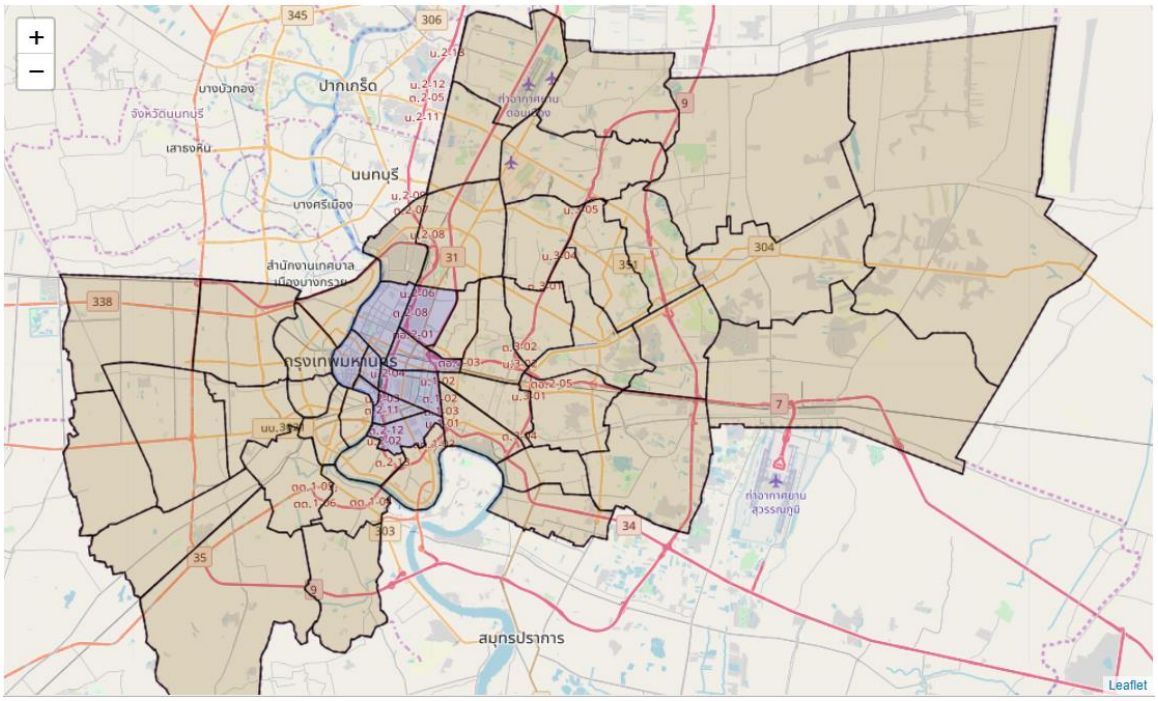

Figure 1 – Bangkok with 9 blue CBD district

List of 9 economic center district in Bangkok

1. Sathon(เขตสาทร)
2. Pathumwan(เขตปทุมวัน)
3. Ratchathewi(เขตราชเทวี)
4. Phaya Thai(เขตพญาไท)
5. phra nakhon(เขตพระนคร)
6. Sompanthawong (เขตสัมพันธวงศ์)
7. Pomprap (เขตป้อมปราบศัตรูพ่าย)
8. Bangruk (เขตบางรัก)
9. Dusit (เขตดุสิต)

the idea is in a trend of drink coffee are coming so many people want to open a new coffee shop but how they know where is a district in Bangkok suitable for open new business by not is the competitor with world largest coffee company and exist coffee shop dense in area. I will use Foursquare venues category locations with local district location to select district from 4 clusters from 9 district in Bangkok that have an opportunity to open a new coffee shop. Which location should be suggested to the new entrepreneur?

II. Data description:

According to the question, following data to resolve the problem. Geographic coordinate of Bangkok and Disticct. i can find location of bangkok and district from

http://nominatim.openstreetmap.org/

then use openstreetmap to get polygon of district before create locations area before use Foresqure to get coffee venues in Bangkok and plot to map

http://polygons.openstreetmap.fr/index.py

Bangkok district data

https://nominatim.openstreetmap.org/details.php?place_id=198742921

Local name	*OSM*	*Address rank*	*Admin level*	*Distance*
วัฒนา	2938036	12	6	~22.9 km
เขตสาทร	3147425	12	6	~28.7 km
เขตดุสิต	92069	12	6	~30.5 km
เขตบางนา	3146412	12	6	~19.9 km
เขตบางแค	2946804	12	6	~42.4 km
เขตพญาไท	92068	12	6	~27.8 km
เขตจอมทอง	3147007	12	6	~36.2 km
เขตดินแดง	2938031	12	6	~26.3 km
เขตธนบุรี	3147281	12	6	~34.5 km
เขตบางบอน	3147312	12	6	~40.7 km
เขตบางรัก	92061	12	6	~29.5 km
เขตบางเขน	2938030	12	6	~18.7 km
เขตประเวศ	3146414	12	6	~11.9 km
เขตพระนคร	92053	12	6	~33.2 km
เขตสายไหม	2938035	12	6	~19.5 km
เขตคลองสาน	3147280	12	6	~32.9 km
เขตคลองเตย	2938033	12	6	~23.8 km
เขตจตุจักร	1639286	12	6	~25.2 km
เขตทุ่งครุ	3147104	12	6	~33.9 km
เขตบางกะปิ	2939952	12	6	~14.8 km

เขตบางซื่อ	1639285	12	6	~30.1 km
เขตบางพลัด	2942281	12	6	~32.8 km
เขตบึงกุ่ม	2946805	12	6	~15.5 km
เขตปทุมวัน	92063	12	6	~28.7 km
เขตพระโขนง	3146642	12	6	~20.4 km
เขตมีนบุรี	3146413	12	6	~4.7 km
เขตยานนาวา	3147456	12	6	~28.7 km
เขตราชเทวี	92064	12	6	~27.2 km
เขตสวนหลวง	3146643	12	6	~17 km
เขตหนองจอก	3147501	12	6	~2.3 km
เขตหนองแขม	2942283	12	6	~47.4 km
เขตหลักสี่	1639288	12	6	~27.4 km
เขตคันนายาว	2946806	12	6	~13 km
เขตดอนเมือง	1639287	12	6	~27.4 km
เขตดลิ่งชัน	2939954	12	6	~37.7 km
เขตทวีวัฒนา	2939955	12	6	~44.1 km
เขตลาดพร้าว	2938034	12	6	~21.8 km
เขตสะพานสูง	3146415	12	6	~11 km
เขตห้วยขวาง	2938032	12	6	~22.9 km
เขตคลองสามวา	2997412	12	6	~10.1 km
เขตบางคอแหลม	3147455	12	6	~31.4 km
เขตภาษีเจริญ	2946807	12	6	~37.4 km
เขตลาดกระบัง	3147500	12	6	~2.3 km
เขตบางกอกน้อย	2942282	12	6	~35.1 km
เขตบางกอกใหญ่	2939953	12	6	~35.1 km
เขตวังทองหลาง	2939956	12	6	~20.1 km
เขตบางขุนเทียน	3147313	12	6	~38.8 km
เขตราษฎร์บูรณะ	3147008	12	6	~32.8 km
เขตสัมพันธวงศ์	92056	12	6	~32.5 km
เขตป้อมปราบศัตรูพ่าย	92058	12	6	~32.3 km

Table 2 – List of All District on Bangkok

The process of collecting and clean data:

- For each neighborhood, pass the obtained coordinates to FourSquare API. The “explore” endpoint will return a list of surrounding venues in a pre-defined radius via use loop for each category for get more information out of limit of free API request and go to Document of FourSquare to get Category ID that we interested

'CoffeeShop':'4bf58dd8d48988d1e0931735',

'Theater':'4bf58dd8d48988d17f941735',

'Museum':'4bf58dd8d48988d181941735',

'Sportclub':'4d4b7105d754a06377d81259',

'AreaUniversity':'4d4b7105d754a06372d81259',

'Plaza':'4bf58dd8d48988d164941735',

'School':'4bf58dd8d48988d13b941735'

- Count the occurrence of each venue in each category and matching with district that venue occurred

	CBD	Num_of_CoffeeShop	Num_of_Theater	Num_of_Museum	Num_of_Sportclub	Num_of_AreaUniversity	Num_of_Plaza	Num_of_School
0	ปทุมวัน	72	21.0	8	50	67	16	60
1	พระนคร	67	0.0	25	27	49	15	28
2	พญาไท	61	0.0	2	15	9	5	16
3	บางรัก	53	2.0	1	15	6	10	39
4	ราชเทวี	39	1.0	1	17	39	15	22
5	สาทร	25	3.0	2	5	14	3	12
6	ดุสิต	25	0.0	13	18	65	1	38
7	ป้อมปราบศัตรูพ่าย	17	0.0	1	3	6	3	5
8	สัมพันธวงศ์	13	1.0	3	1	2	2	3

Figure 2 - Final dataset

The dataset has 1157 samples from 9 district and 7 Category ID from FourSquare. The number of features is use loop to run and get data from API because free has limit to retrieve data.

III. Methodology:

The assumption is district to open new coffees shop is dependent how much exsist coffee shop in area and dense in area include with other number of other place like school,Plaza,Sport area that are one of factor to supoort which district have opportunity to open new coffee shop compare with other district.

K-means Cluster will apply to solve this model because we need to group district that similaly with 4 group and recommend to new entrepreneur.

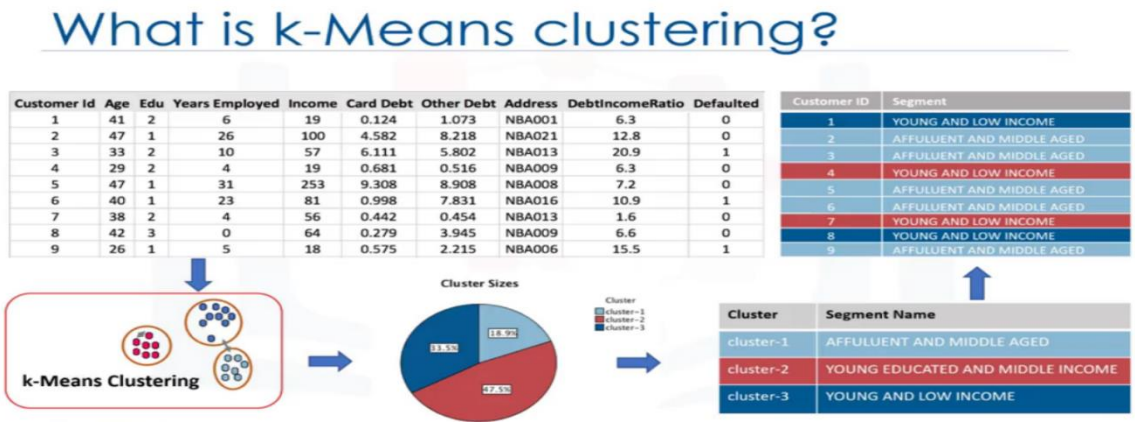

Figure 3 – what is k-Means Clustering

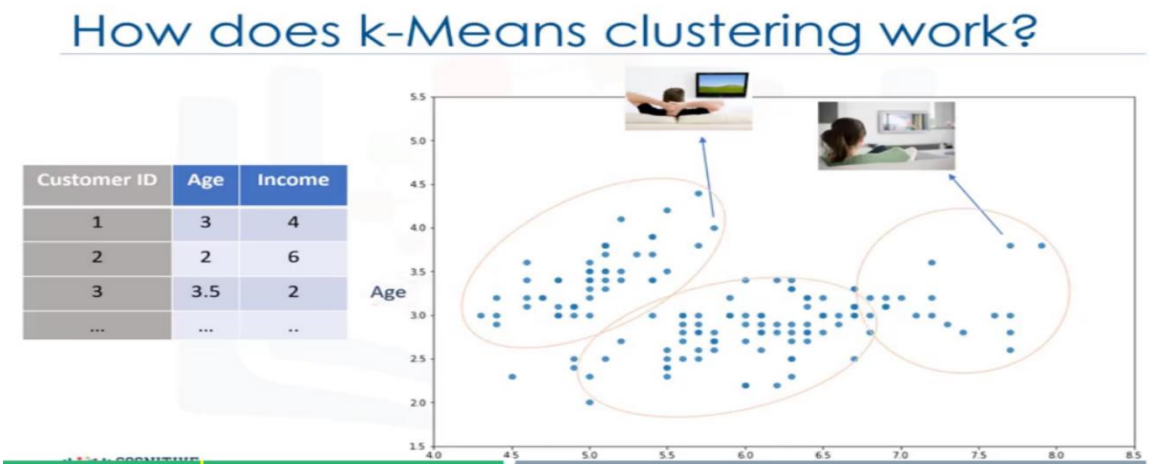

Figure 4 – How Does k-Means Clustering work

Combine with FourSquare API on counting how many different venues Coffee Shop,Theater,Museum,Sport,University,Plaza,School The k-mean cluster can be applied on possible district locations with venues information to generate a ranking result. The descriptive models can be recommended to new entrepreneur.

	CBD	Num_of_CoffeeShop	Num_of_Theater	Num_of_Museum	Num_of_Sportclub	Num_of_AreaUniversity	Num_of_Plaza	Num_of_School	Cluster Labels
0	เขตปทุมวัน	72	21.0	8	50	67	16	60	2
1	เขตพระนคร	67	0.0	25	27	49	15	28	1
2	เขตพญาไท	61	0.0	2	15	9	5	16	3
3	เขตบางรัก	53	2.0	1	15	6	10	39	3
4	เขตราชเทวี	39	1.0	1	17	39	15	22	1
5	เขตสาทร	25	3.0	2	5	14	3	12	0
6	เขตดุสิต	25	0.0	13	18	65	1	38	1
7	เขต ป้อมปราบศัตรูพ่าย	17	0.0	1	3	6	3	5	0
8	เขตสัมพันธวงศ์	13	1.0	3	1	2	2	3	0

Figure 5 – Results Cluster

IV. Results:

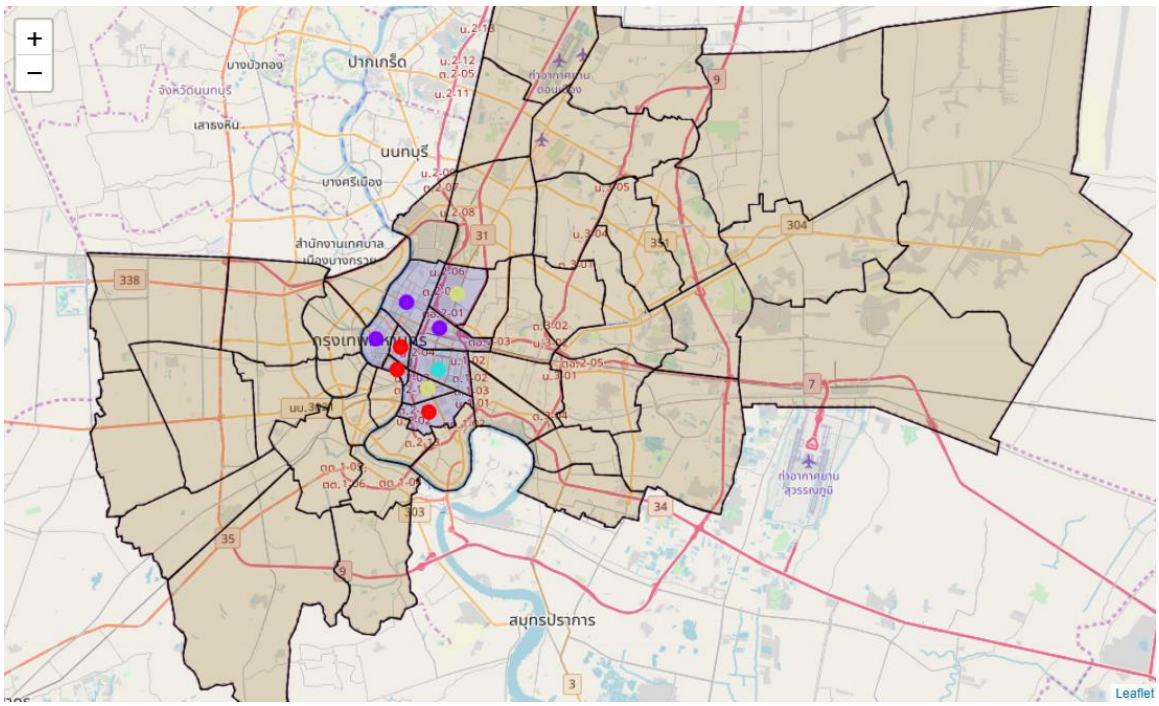

Figure 6 – Results Cluster Map on Bangkok

In K-mean cluster result is group distirct to 4 groups

- Group 3 Phaya Thai(เขตพญาไท),Bangruk (เขตบางรัก)
- Group 2 are Pathumwan(เขตปทุมวัน)
- Group 1 are phra nakhon(เขตพระนคร),Dusit (เขตดุสิต),Ratchathewi(เขตราชเทวี)
- Group 0 are Sathon(เขตสาทร),Sompanthawong (เขตสัมพันธวงศ์),Pomprap (เขตป้อมปราบศัตรูพ่าย)

So we recommend Group 1 for new entrepreneur for who want to open new coffee shop by district not dense and have many place location enough to support people to visit when open new coffee shop

V. Discussion:

The real challenge is constructing the dataset: - Data to consider should not in only FourSquere API because free version data can't get more enough to use to prediction on real life. Have many location that can't provide by FourSqure and have a lot of factor should get data to considering too but is not public so this project may can't use to refer on real life.

VI. Conclusion:

For this project is way to advance on future to solve problem by use geo locations.

Doing this project helps practicing in the specialization, and thus, equipping learners with Data Science methodology and tools using Python libraries and Other API. Also doing a real project certainly helps one learns so much more outside the curriculum, as well as realizes what more to research into after completing the program.

Thank you IBM and Coursera to this course that help me alot to up skill and can adapt to use in real life