THE BATTLE OF NEIGHBOURHOODS – KUALA LUMPUR

APPLIED DATA SCIENCE CAPSTONE PROJECT

BY

WK NG

INTRODUCTION: PROBLEM DESCRIPTION

As one of the developing countries in the Asia Pacific region, Malaysia is well known for its blooming travelling industry. In 2019, the capital state of Malaysia, which is Kuala Lumpur, ranked 10th in destination popularity worldwide for tourists, with an approximate amount of 14.07 million tourists arriving. In the same year, Malaysia also ranked 22nd in worldwide tourism popularity with a total of 26 million tourists visiting the country.

The high popularity amongst tourists and backpackers eventually resulted in the rise of hotels and hostels in Kuala Lumpur. In 2019, there were a total amount of 4,826 of hotels in the capital state. With the increasing amount of hotels, it was expected there will be more people who are interested in starting a hotel. This project is mainly aimed to help owners to find the optimal location for business, which includes analysis such as:

- What is the best location to start a new hotel in Kuala Lumpur?

- Which area has more facilities around?

DATA REQUIREMENTS

- A LIST OF KL DISTRICTS, WITH THEIR POSTCODES

SOURCE (CREDITS TO ZAEPHAER): HTTPS://GITHUB.COM/ZAEPHAER/CAPSTONEPROJECT/BLOB/MASTER/KL_DISRICT.CSV

- FOURSQUARE API, TO LOCATE THE LONGITUDE AND LATITUDE OF THE AREAS, AS WELL AS LISTING THE NAME AND CATEGORY OF EACH AREA

SOURCE: HTTPS://FOURSQUARE.COM/DEVELOPERS/APPS

- GEOCODER PACKAGE, TO TRACK DOWN THE LONGITUDE AND LATITUDE OF THE AREAS

METHODOLOGY

- THE DATASET PROVIDED CONSISTS OF 66 AREAS AND HAD BEEN CLASSIFIED INTO 11 DISTRICTS. BY USING APIS SUCH AS POSITIONSTACK, WE ARE ABLE TO OBTAIN THE LONGITUDE AND LATITUDE OF THE AREAS IN KUALA LUMPUR.

- BY USING FOURSQUARE API, WE ARE ABLE TO DETERMINE THE MOST POPULAR VENUES IN EACH AREAS. 2001 VENUES ARE DETECTED, BEING CLASSIFIED INTO 262 UNIQUE CATEGORIES.

- AFTER PERFORMING ONE HOT ENCODING ON THE DATASET, WE OBTAIN THE 10 MOST COMMON VENUE CATEGORY IN EACH AREA. CLUSTERING IS THEN PERFORMED FOR K-NEAREST NEIGHBOR CLUSTERING. ELBOW METHOD IS CONDUCTED BEFOREHAND TO GAIN THE OPTIMAL VALUE OF K (5).

- THROUGH THE CLUSTERS ON THE MAP, WE ARE ABLE TO ANALYZE AND DETERMINE THE MAJOR BUSINESS TYPES IN EACH CITY, HENCE DETERMINING THE BEST PLACE TO START A HOTEL BUSINESS.

MAP OF KUALA LUMPUR (BEFORE CLUSTERING)
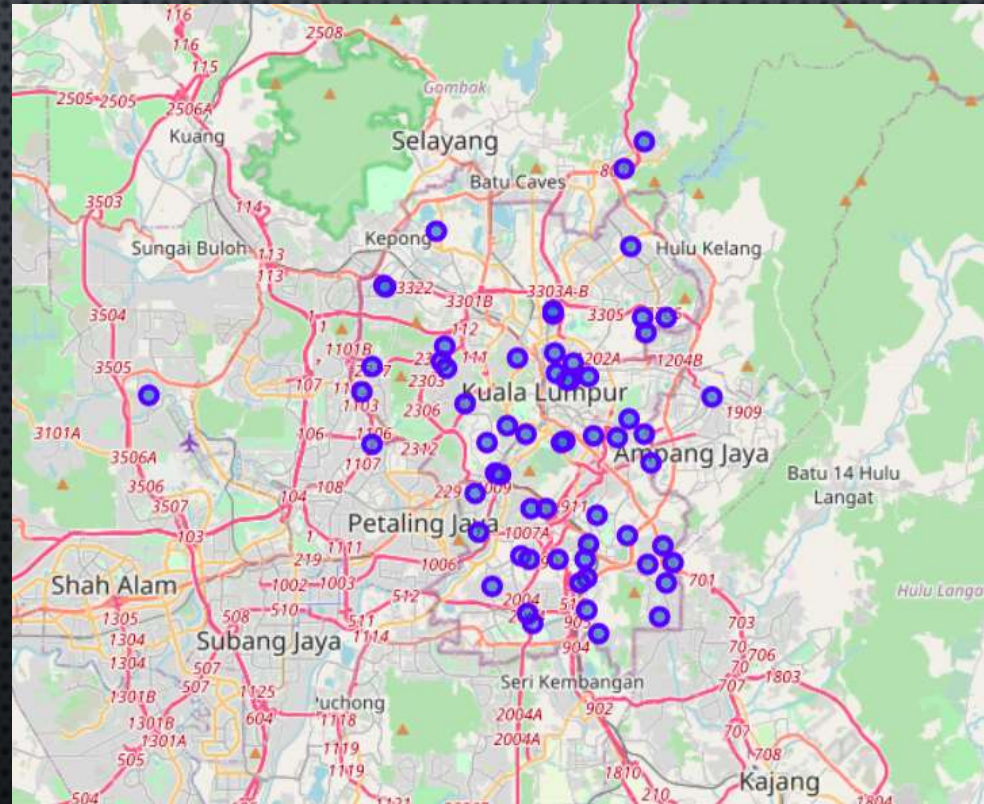
MAP OF KUALA LUMPUR (AFTER CLUSTERING)
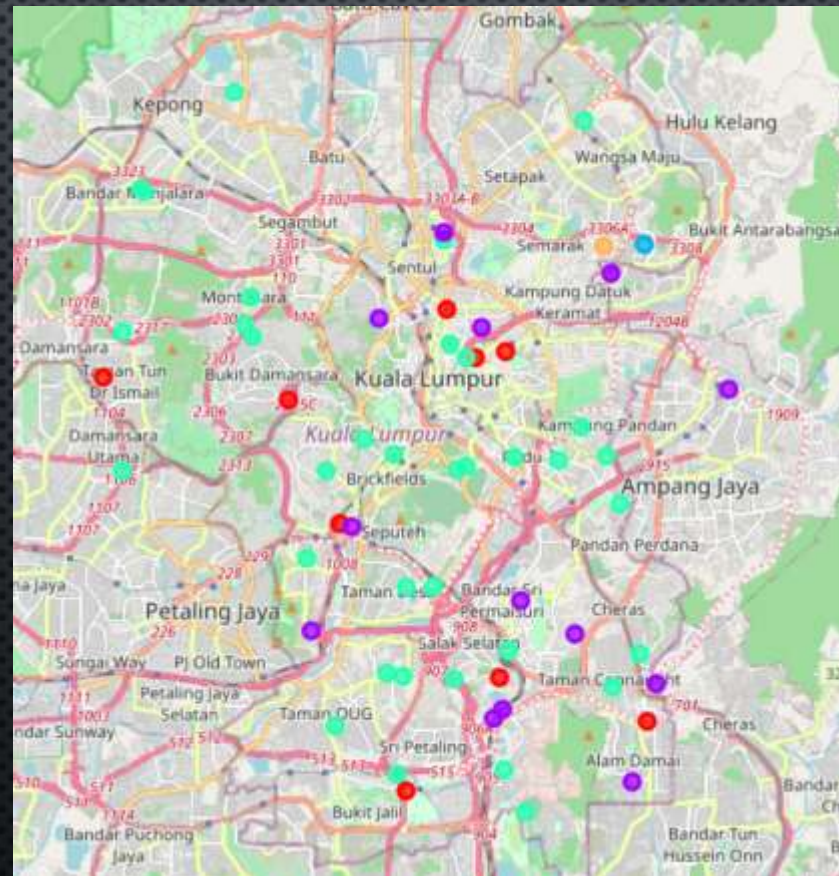
TOP 3 VENUES OF CLUSTER 0

	Area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
3	Bukit Kiara	0	Café	Convenience Store	Burger Joint
12	Taman Tun Dr Ismail	0	Burger Joint	Coffee Shop	Restaurant
27	Bukit Nanas	0	Hotel	Café	Nightclub
29	Chow Kit	0	Chinese Restaurant	Pakistani Restaurant	Convenience Store
31	Desa Petaling	0	Chinese Restaurant	Convenience Store	Asian Restaurant
32	KL City Centre	0	Hotel	Coffee Shop	Clothing Store
40	Kg Abdullah Hukum	0	Ice Cream Shop	Bakery	Food Court
44	Bukit Jalil	0	Café	Racetrack	Athletics & Sports
58	Taman Len Seng	0	Chinese Restaurant	Café	Pool Hall

TOP 3 VENUES OF CLUSTER 1

	Area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
4	Bukit Tunku	1	Garden	Gym	Indie Theater
14	Sentul	1	Malay Restaurant	Chinese Restaurant	Building
18	Kampung Datuk Keramat	1	Malay Restaurant	Middle Eastern Restaurant	Convenience Store
21	Ampang	1	Chinese Restaurant	Asian Restaurant	Indian Restaurant
24	Kampung Baru	1	Malay Restaurant	Indonesian Restaurant	Thai Restaurant
41	Mid Valley City	1	Ice Cream Shop	Bakery	Department Store
42	Pantai Dalam	1	Malay Restaurant	Flea Market	Coffee Shop
51	Alam Damai	1	Convenience Store	Malay Restaurant	Food Court
55	Taman Cheras Hartamas	1	Sandwich Place	Malay Restaurant	Café
57	Taman Midah	1	Indian Restaurant	Chinese Restaurant	Malay Restaurant
60	Bandar Sri Permaisuri	1	Restaurant	Indian Restaurant	Malay Restaurant
61	Bandar Tasik Selatan	1	Convenience Store	Ice Cream Shop	Café
63	Kampung Malaysia	1	Bakery	Convenience Store	Bus Station

TOP 3 VENUES OF CLUSTER 2

	Area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
19	Semarak	2	Malay Restaurant	Café	Fast Food Restaurant

TOP 3 VENUES OF CLUSTER 3 (PART 1)

	Area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	
0	Jinjang	3	Vegetarian / Vegan Restaurant	Convenience Store	Chinese Restaurant	
1	Taman Bukit Maluri	3	Chinese Restaurant	Café	Vegetarian / Vegan Restaurant	
2	Bandar Menjalara	3	Chinese Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	
5	Damansara	3	Café	Japanese Restaurant	Korean Restaurant	
6	Damansara Town Centre	3	Dog Run	Plaza	Gym	
7	Jalan Duta	3	Café	Pizza Place	Spa	
8	Kampung Kasipillay	3	Malay Restaurant	Indonesian Restaurant	Bookstore	
9	Kampung Sungai Penchala	3	Malay Restaurant	Food Truck	Restaurant	
10	Mont Kiara	3	Japanese Restaurant	Korean Restaurant	Restaurant	
11	Sri Hartamas	3	Japanese Restaurant	Korean Restaurant	Café	
13	Bandar Baru Sentul	3	Building	Automotive Shop	Night Market	
15	Setapak	3	Asian Restaurant	Coffee Shop	Bakery	
16	Taman Melati	3	Thai Restaurant	Food Court	Malay Restaurant	
17	Taman P. Ramlee	3	Chinese Restaurant	Malay Restaurant	Asian Restaurant	
22	Jalan Cochrane	3	Korean Restaurant	Clothing Store	Café	
23	Maluri	3	Asian Restaurant	Convenience Store	Food Truck	
25	Kampung Pandan	3	Malay Restaurant	Convenience Store	Food Truck	
26	KL Sentral	3	Indian Restaurant	Hotel	Coffee Shop	S
28	Bukit Petaling	3	Malay Restaurant	Café	Building	
30	Dang Wangi	3	Hotel	Café	Nightclub	
33	Medan Tuanku	3	Malay Restaurant	Hotel	Asian Restaurant	
34	Pudu	3	Chinese Restaurant	Nightclub	Dessert Shop	
35	Salak South	3	Chinese Restaurant	Dumpling Restaurant	Café	

TOP 3 VENUES OF CLUSTER 3 (PART 2)

36	Tun Razak Exchange	3	Gas Station	Flea Market	Light Rail Station
37	Bangsar	3	Indian Restaurant	Bar	Ice Cream Shop
38	Bangsar South	3	Café	Coffee Shop	Hotel
39	Federal Hill	3	Garden	Park	Yoga Studio
43	Bandar Malaysia	3	Malay Restaurant	Building	Falafel Restaurant
46	Happy Garden	3	Chinese Restaurant	Asian Restaurant	Café
47	Kuchai Lama	3	Coffee Shop	Chinese Restaurant	Café
48	Sri Petaling	3	Malay Restaurant	German Restaurant	Korean Restaurant
49	Taman Desa	3	Juice Bar	Vegetarian / Vegan Restaurant	Café
50	Taman OUG	3	Chinese Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant
54	Shamelin Perkasa	3	Café	Malay Restaurant	Chinese Restaurant
56	Taman Connaught	3	Chinese Restaurant	Asian Restaurant	Café
59	Taman Taynton View	3	Chinese Restaurant	Convenience Store	Spa
64	Lake Fields	3	Malay Restaurant	Indian Restaurant	Food Truck
65	Sungai Besi	3	Park	Food Court	Chinese Restaurant

TOP 3 VENUES OF CLUSTER 4

	Area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
20	Desa Tun Hussein Onn	4	Food Truck	Gym	Flower Shop

OBSERVATIONS

- 1. Hotels are mostly located in Cluster 0, 1 and 3. However, cluster 3 has 4 areas with hotels as the top 3 most common venue (KL Sentral, Dang Wangi, Medan Tuanku, Bangsar South), making Cluster 3 the most popular place the set up a hotel. Cluster 0, however, has the 2nd highest popularity with 2 areas having hotels as the most common venue (Bukit Nanas and KL City Center).

- 2. For the 2 areas in Cluster 0 which has hotel as their most popular venue (Bukit Nanas and KL City Center), both areas have food related shops nearby (Examples: Restaurant, Bar, Lounge)

- 3. The food industry is shown to be really popular in Kuala Lumpur, as a majority of the areas have either a restaurant or a food related shop as their most common venue. Most of the restaurants, however, are popular in Cluster 3, with a majority of areas having restaurants as their most popular venue.

- 4. Cluster 0 is also famous for their sports facilities, with places such as soccer stadiums and racetrack being listed as some of the most common venues.

CONCLUSION

After the analysis above, despite the high popularity in Cluster 3, it is deduced that areas in Cluster 1 is the most suitable to open a new hotel. This is due to it's above average popularity in tourism with two areas listing hotels as their most common venues (Bukit Nanas and KL City Center). Despite having the most amount of hotels, Cluster 3 may not be the best option as there could be too many competitions around the area. Cluster 0 was also surrounded with a variety of multinational restaurants, giving tourists a variety of choices when it comes to food.

DRAWBACKS

Despite the conclusion being made, a few suggestions are provided to further increase the accuracy of the analysis. Firstly, the amount of areas need to be reduced into a smaller amount. As an example, only the 2 districts with the most amount of areas (Segambut and Bukit Bintang) are analyzed to simplify the analysis process. Other factors, such as tourist hotspots, competitions and price must also be taken account to further analyze the suitability of the place.

THANK YOU!