

ANALYZING NEIGHBOURHOODS OF KUALA LUMPUR FOR NEW BAKERY MARKETING STRATEGY

IBM APPLIED DATA SCIENCE CAPSTONE PROJECT

BY

N. VIGNESWARAN

DECEMBER 2020

INTRODUCTION

Kuala Lumpur, or KL, is the federal capital and most populous city of Malaysia. Kuala Lumpur is the fastest growing metropolitan area of the country. It's the economic, financial and cultural city of the country and one of three Malaysian Federal Territories. Greater Kuala Lumpur, or the Klang Valley, is a fundamental urban aggregation with an approximation population of 7.2 million in 2016 with a population density that is nearly equal to that of the city proper. It's 2020 population is now estimated at 7,996,830.

The city is mainly a mix of Malays, Chinese, and Indians, although there are many cultures in the city such as Eurasians, Kadazans, Ibans and indigenous peoples from East Malaysia and Peninsular Malaysia. The diversity among the population create a path for variety of foods. A report notes that the Malaysian market has seen positive growth of baked goods as more products are consumed at meal times in place of rice and noodles. So, in this project, we will study the neighbourhood and make recommendation accordingly.

BUSINESS PROBLEM.

The main objective of this project is to analyze the neighbourhood of Kuala Lumpur to find strategic location for opening new bakery. Bakeries can categorize as conventional stand-alone bakeries, commercial bakeries , bakery counters, and home-based bakers providing a handmade breads and cakes. So, we would like to extract and analyse a various data about neighbourhood of Kuala Lumpur using data science methodology and machine learning techniques such as clustering to provide a right solution.

DESCRIPTION OF DATA

To solve the problem, we will need the following data:

- List of Kuala Lumpur neighbourhood
- Geographical coordinates of neighbourhood
- Venue data from Foursquare

Neighbourhood data

The data extracted from Wikipedia page using web scrapping with BeautifulSoup library in Python.

[illegible]

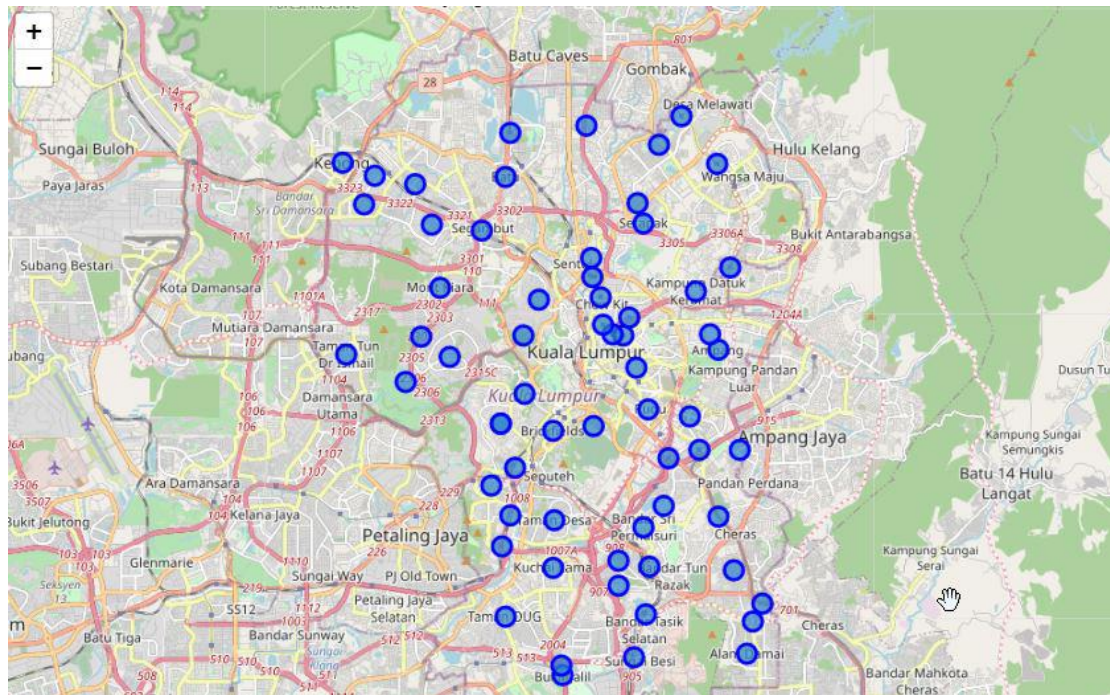
Geographical coordinates of neighbourhood

Using Geopy geocoders, the geographical coordinates of various suburbs in Kuala Lumpur were extracted. Additionally, two columns were included into data frame for latitude and longitude.

	Neighborhood	Longitude	Latitude
0	Alam Damai	101.740735	3.064708
1	Ampang, Kuala Lumpur	101.732250	3.152135
2	Bandar Menjalara	101.630030	3.193954
3	Bandar Sri Permaisuri	101.710872	3.101158
4	Bandar Tasik Selatan	101.711447	3.076097
...
65	Taman Tun Dr Ismail	101.624919	3.150808
66	Taman U-Thant	101.729820	3.156499
67	Taman Wahyu	101.672211	3.214504
68	Titivangsa	101.695933	3.173145
69	Wangsa Maju	101.731908	3.205667

70 rows × 3 columns

Using the folium package and my data frame, I then created a map of neighbourhood.



Venue data from Foursquare

Foursquare is a database of more than 105 million places worldwide and an API that enables location data for Apple, Samsung, Microsoft, Tencent, Snapchat, Twitter, Uber, and others. The venue data extracted using Foursquare API to study the various neighbourhood in Kuala Lumpur. The Data provides various locations of existing bakeries around the neighbourhood and its vital to analyse and draw conclusion of this project.

METHODOLOGY

At data description, I explained about the neighbourhood web scrapping, geographical coordinates and venue data from Foursquare. In this section, I'm going to elaborate about method and techniques which are use to extract and analysis data.

Firstly, the data for venues retrieved from Foursquare within 2000 meter by limit to 50. The result was a list of 3378 venues all over Kuala Lumpur neighbourhood.

	Neighbourhood	Latitude	Longitude	Name	Latitude	Longitude	Category
0	Alam Damai	3.064708	101.740735	Pengedar Shaklee Kuala Lumpur	3.061235	101.740696	Supplement Shop
1	Alam Damai	3.064708	101.740735	Alam Damai Recreational Park	3.066424	101.740265	Other Great Outdoors
2	Alam Damai	3.064708	101.740735	Ivy Sekinchan Seafood Noodle House 這排莊海鮮麵	3.065749	101.748718	Noodle House
3	Alam Damai	3.064708	101.740735	Jc Deli 壹喜食坊	3.058397	101.748560	Food & Drink Shop
4	Alam Damai	3.064708	101.740735	Alam Damai, Damai Jasa	3.069329	101.744512	Park
5	Alam Damai	3.064708	101.740735	Machi Noodle 妈子面	3.057695	101.746635	Noodle House
6	Alam Damai	3.064708	101.740735	Bukit Besi Hiking Trail	3.073725	101.733300	Trail
7	Alam Damai	3.064708	101.740735	Big Tree Lin Kee Chan Sow Lin Steam Fish Head ...	3.077381	101.739140	Chinese Restaurant
8	Alam Damai	3.064708	101.740735	Tealive	3.070599	101.749615	Bubble Tea Shop
9	Alam Damai	3.064708	101.740735	【Lensen 点心】茶餐室	3.074036	101.745661	Chinese Breakfast Place

To find clusters of bakeries at various neighbourhood, the data frame was transformed with the venues, associated to neighbourhood, by one-hot encoding (0/1), as seen in the picture below.

Neighbourhood	Accessories Store	African Restaurant	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Vietnamese Restaurant	Volleyball Court	Warehouse Store	Whisky Bar	Wine Bar	Wings Joint	Women's Store	Yakitori Restaurant	Yoga Studio	Zoo
0 Alam Damai	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1 Alam Damai	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2 Alam Damai	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 Alam Damai	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 Alam Damai	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

K means = 5437 iterations

Next, I group rows by neighbourhood and by taking the mean of the frequency of occurrence of each category.

Neighbourhood	Accessories Store	African Restaurant	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Vietnamese Restaurant	Volleyball Court	Warehouse Store	Whisky Bar	Wine Bar	Wings Joint	Women's Store	Yakitori Restaurant	Yoga Studio	Zoo
0 Alam Damai	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1 Ampang, Kuala Lumpur	0.0	0.0	0.0	0.0	0.04	0.00	0.00	0.06	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2 Bandar Menjalara	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.06	0.0	0.0	0.00	0.00	0.0	0.0	0.00	0.0	0.0	0.0	0.0
3 Bandar Sri Permaisuri	0.0	0.0	0.0	0.0	0.02	0.00	0.00	0.10	0.0	0.0	0.0	0.00	0.0	0.0	0.00	0.0	0.0	0.0	0.0
4 Bandar Tasik Selatan	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.12	0.0	0.0	0.00	0.00	0.0	0.0	0.00	0.0	0.0	0.0	0.0
...
64 Taman Tun Dr Ismail	0.0	0.0	0.0	0.0	0.00	0.00	0.02	0.00	0.0	0.0	0.00	0.00	0.0	0.0	0.00	0.0	0.0	0.0	0.0
65 Taman U-Thant	0.0	0.0	0.0	0.0	0.02	0.00	0.00	0.02	0.0	0.0	0.00	0.00	0.0	0.0	0.00	0.0	0.0	0.0	0.0
66 Taman Wahyu	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.14	0.0	0.0	0.00	0.00	0.0	0.0	0.00	0.0	0.0	0.0	0.0
67 Titiwangsa	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.06	0.0	0.0	0.02	0.02	0.0	0.0	0.00	0.0	0.0	0.0	0.0
68 Wangsa Maju	0.0	0.0	0.0	0.0	0.00	0.02	0.00	0.02	0.0	0.0	0.00	0.00	0.0	0.0	0.02	0.0	0.0	0.0	0.0

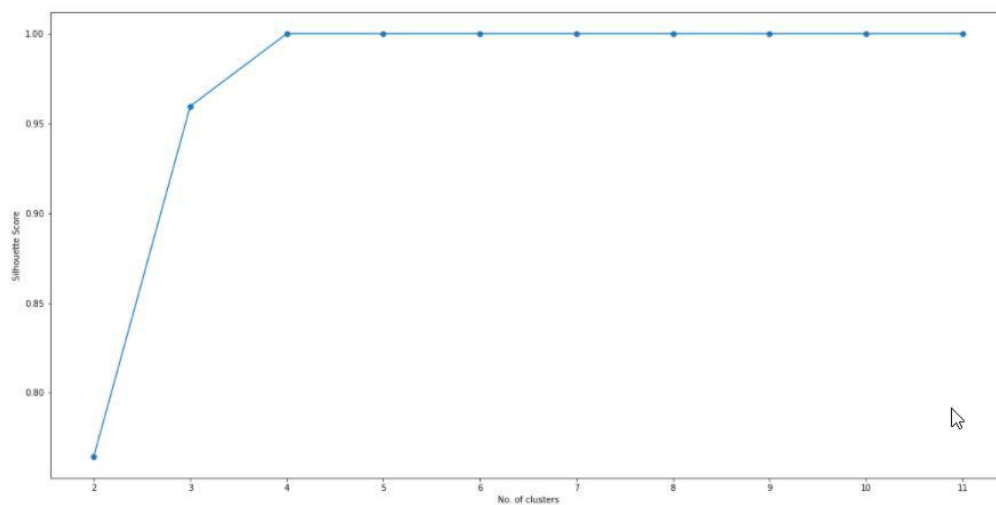
69 rows x 202 columns

I used this information to create a data frame in which you can see the top 10 most common venue types for each neighbourhood.

Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 Aiam Damai	Chinese Restaurant	Malay Restaurant	Cafe	Noodle House	Coffee Shop	Restaurant	Indian Restaurant	Cantonese Restaurant	Convenience Store	Breakfast Spot
1 Ampang Kuala Lumpur	Hotel	Malay Restaurant	Asian Restaurant	Breakfast Spot	Juice Bar	Coffee Shop	Indian Restaurant	Golf Course	Air Gallery	Razon
2 Bandar Meragaha	Chinese Restaurant	Coffee Shop	Japanese Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Park	Middle Eastern Restaurant	Restaurant	Grocery Store	Monsoon
3 Bandar Sri Permaisuri	Chinese Restaurant	Asian Restaurant	Indonesian Restaurant	Noodle House	Gas Station	Seafood Restaurant	Pizza Place	Fast Food Restaurant	Bakery	Lake
4 Bandar Tasik Selatan	Asian Restaurant	Burger Joint	Chinese Restaurant	Malay Restaurant	Ice Cream Shop	Convenience Store	Sandwich Place	Cantonese Restaurant	Gym Pool	Comfort Food Restaurant
5 Bandar Tun Razak	Chinese Restaurant	Indonesian Restaurant	Asian Restaurant	Flea Market	Bakery	Malay Restaurant	Burger Joint	Gas Station	Lake	Market
6 Bangsar	Indian Restaurant	Bar	Ice Cream Shop	Burger Joint	Clothing Store	Spa	Shopping Mall	Coffee Shop	Seashouse	Boutique
7 Bangsar Park	Indian Restaurant	Bar	Ice Cream Shop	Burger Joint	Clothing Store	Spa	Shopping Mall	Coffee Shop	Seashouse	Boutique
8 Bangsar South	Japanese Restaurant	Hotel	Restaurant	Coffee Shop	Malay Restaurant	Movie Theater	Clothing Store	Shopping Mall	Chinese Restaurant	Juice Bar
9 Batu 11 Cheras	Chinese Restaurant	Asian Restaurant	Malay Restaurant	Coffee Shop	Food Truck	Lake	Dim Sum Restaurant	Seashouse	Cafe	Cantonese Restaurant
10 Batu, Kuala Lumpur	Chinese Restaurant	Malay Restaurant	Seafood Restaurant	Asian Restaurant	Food Truck	Restaurant	Motorcycle Shop	Flea Market	Noodle House	Spa
11 Brickfields	Hotel	Indian Restaurant	Clothing Store	French Restaurant	Cafe	Shopping Mall	Coffee Shop	Cocktail Bar	Hotel Bar	South Indian Restaurant
12 Bukit Bintang	Hotel	Cafe	Spa	Asian Restaurant	Shopping Mall	Boutique	Japanese Restaurant	Clothing Store	Harbin Restaurant	Science Museum
13 Bukit Jalil	Coffee Shop	Cafe	Japanese Restaurant	Asian Restaurant	Massage Studio	Middle Eastern Restaurant	Out Course	Ice Cream Shop	Vegetarian / Vegan Restaurant	Restaurant
14 Bukit Kariak	State Park	Asian Restaurant	Indian Restaurant	Ice Cream Shop	Food Truck	Malay Restaurant	Lounge	Japanese Restaurant	Golf Course	Bakery
15 Bukit Nanas	Hotel	Cocktail Bar	Bar	Cosmetics Shop	Building	Boutique	Monument / Landmark	Breakfast Spot	Sushi Restaurant	Dance Studio
16 Bukit Petaling	Hotel	Cafe	Hotel Bar	Indian Restaurant	Juice Bar	South Indian Restaurant	Coffee Shop	Convenience Store	Buncho Place	Performing Arts Venue
17 Bukit Tunku	Japanese Restaurant	Coffee Shop	Pool	Cafe	Art Gallery	Restaurant	Bar	Noodle House	Spa	Hotel
18 Cheras, Kuala Lumpur	Chinese Restaurant	Malay Restaurant	Asian Restaurant	Indonesian Restaurant	Seafood Restaurant	Food Truck	Asian Restaurant	Noodle House	Pizza Place	Nightclub
19 Chow Kit	Hotel	Malay Restaurant	Coffee Shop	Clothing Store	Asian Restaurant	Seafood Restaurant	Thai Restaurant	Theater	Chinese Restaurant	IT Services
20 Damansara Heights	Japanese Restaurant	Cafe	Italian Restaurant	Ice Cream Shop	Indian Restaurant	Korean Restaurant	Asian Restaurant	Grocery Store	Yoga Studio	Bakery
21 Damansara Town Centre	Seafood Restaurant	Chinese Restaurant	Racetrack	Coffee Shop	Malay Restaurant	Shopping Mall	BBQ Joint	Convenience Store	Roof Deck	Electronics Store
22 Dang Wangi	Hotel	Bar	Cocktail Bar	Chinese Restaurant	Sushi Restaurant	Breakfast Spot	Monument / Landmark	Building	Multiplex	Gym
23 Desa Petaling	Chinese Restaurant	Asian Restaurant	Burger Joint	Hospital Restaurant	Japanese Restaurant	Gym	Bakery	Malay Restaurant	Coffee Shop	Cafe
24 Federal Hill, Kuala Lumpur	Hotel	Boutique	French Restaurant	Seashouse	Ice Cream Shop	Art & Crafts Store	Cafe	Spa	Shopping Mall	Hotel Bar
25 Happy Garden	Asian Restaurant	Outdoors & Recreation	Chinese Restaurant	Border Crossing	Day Bar	Furniture / Home Store	Garden	Garden Center	Gas Station	Zoo
26 Jalan Cochrane, Kuala Lumpur	Asian Restaurant	Nightclub	Furniture / Home Store	Burger Joint	Cafe	Korean Restaurant	Chinese Restaurant	Shopping Mall	Coffee Shop	Noodle House
27 KL Eco City	Cosmetics Shop	Ice Cream Shop	Department Store	Bakery	Clothing Store	Juice Bar	Movie Theater	Shopping Mall	Coffee Shop	Japanese Restaurant

Now, with all this data, I could finally run an unsupervised machine learning algorithm, more specifically, a k-means clustering algorithm from the scikit-learn package.

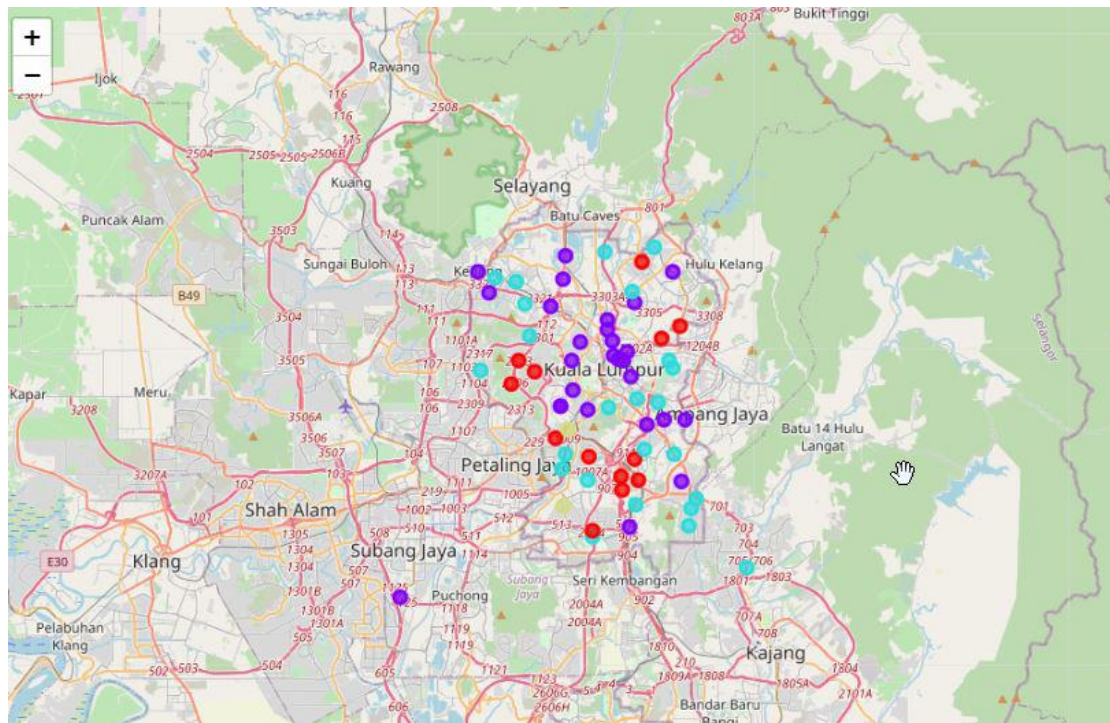
K means clustering algorithm was implemented in order to find out the similarities between found similarities between neighbourhood. Elbow method was use to identify optimal value of clusters. As result:



The plot show that the optimal number of clusters is 4. so, the final version of data frame containing the cluster labels as below:

	Neighbourhood	Longitude	Latitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
8	Asian Centre	101.747138	3.024155	2	Chinese Restaurant	Many Restaurant	Asian Restaurant	Cafe	Breadfruit Spot	Coffee Shop	Restaurant	Indian Restaurant	Convenience Store	Breakfast Spot
9	Amang, Kuala Lumpur	101.732252	3.101131	3	Chinese Restaurant	Many Restaurant	Asian Restaurant	Cafe	Breadfruit Spot	Coffee Shop	Restaurant	Indian Restaurant	Convenience Store	Breakfast Spot
10	Bandar Sri Permaisuri	101.719372	3.011058	0	Chinese Restaurant	Many Restaurant	Asian Restaurant	Cafe	Breadfruit Spot	Coffee Shop	Restaurant	Indian Restaurant	Convenience Store	Breakfast Spot
11	Bandar Tun Razak	101.714467	3.070907	2	Asian Restaurant	Burger Joint	Chinese Restaurant	Many Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Pan	Restaurant	Grocery Store	Market
12	Bandar Tun Razak	101.712487	3.058655	0	Chinese Restaurant	Many Restaurant	Asian Restaurant	Cafe	Breadfruit Spot	Coffee Shop	Restaurant	Indian Restaurant	Convenience Store	Breakfast Spot
13	Bangar	101.854440	3.133550	1	Indian Restaurant	Bar	Ice Cream Shop	Burger Joint	Clothing Store	Coffee Shop	Restaurant	Indian Restaurant	Convenience Store	Breakfast Spot
14	Bangar Park	101.854440	3.133550	1	Indian Restaurant	Bar	Ice Cream Shop	Burger Joint	Clothing Store	Coffee Shop	Restaurant	Indian Restaurant	Convenience Store	Breakfast Spot
15	Bangar South	101.858739	3.113273	0	Japanese Restaurant	Hotel	Restaurant	Coffee Shop	Many Restaurant	Movie Theater	Clothing Store	Shopping Mall	Chinese Restaurant	Juice Bar
16	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
17	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
18	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
19	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
20	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
21	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
22	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
23	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
24	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
25	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
26	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
27	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
28	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
29	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
30	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
31	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
32	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot
33	Batu 11 Cheras	101.773269	3.041231	2	Chinese Restaurant	Asian Restaurant	Many Restaurant	Coffee Shop	Food Truck	Dim Sum Restaurant	Steakhouse	Cafe	Convenience Store	Breakfast Spot

The above table shows the most common venues which are assigned into 4 different cluster labels from 0 to 3. now use the cluster labels to show the neighbourhood marked with a cluster-specific color on a map, again using folium:



After examining the all 4 individual clusters, some important conclusion was derived. The neighbourhood that have most number of bakeries is Cluster 1.

DISCUSSION

As said earlier, the most suitable neighbourhood to start a bakery are present in Cluster 1. The places such as Bandar Sri Permaisuri, Taman Desa, Sri Hartamas, Desa Petaling, Damansara Heights and Bukit Kiara are most suitable with less competition. As per report, Kepong, Bukit Jalil and Desa Petaling are most affordable areas for low cost properties around Kuala Lumpur neighbourhood. Since the Desa Petaling present in Cluster 1, so it will be more suitable location to start 'Cloud Bakery' or bakery without dining area. The decision do not depend on other factors such as availability and legal requirements that are out of scope of this project.

CONCLUSION

Data science and machine learning techniques used in this project were very helpful to derived the results. We achieved the goal presented and the entrepreneurs or business peoples can see in the results which neighbourhood best match for their bakeries.