## ANALYZING NEIGHBOURHOODS OF KUALA LUMPUR FOR STARTING A NEW BAKERY

# 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 select a best location for a opening of 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.

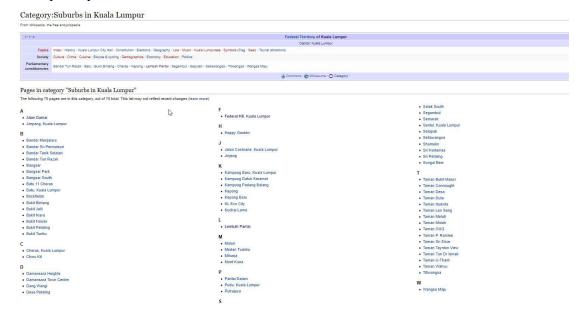
#### **DESCRIPTION OF DATA**

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.

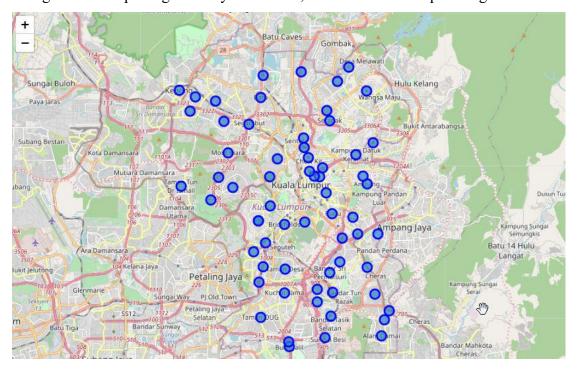


#### Geographical coordinates of neighbourhood

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
	322	2.25	82
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	Titiwangsa	101.695933	3.173145
69	Wangsa Maju	101.731908	3.205667

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.

Category	Longitude	Latitude	Name	Longitude	Latitude	Neighbourhood	
Supplement Shop	101.740696	3.061235	Pengedar Shaklee Kuala Lumpur	101.740735	3.064708	Alam Damai	0
Other Great Outdoors	101.740265	3.066424	Alam Damai Recreational Park		0	Alam Damai	1
Noodle House	101.748718	3.065749	Ivy Sekinchan Seafood Noodle House 适耕莊特制魚丸海鮮面	101.740735	3.084708	Alam Damai	2
Food & Drink Shop	101.748560	3.058397	Jo Deli 警察食坊	101.740735	3.064708	Alam Damai	3
Park	101.744512	3.069329	Alam Damai, Damai Jasa	101.740735	3.064708	Alam Damai	4
Noodle House	101.746635	3.057695	Machi Noodle 妈子面	101.740735	3.064708	Alam Damai	5
Trail	101.733300	3.073725	Bukit Besi Hiking Trail	101.740735	3.064708	Alam Damai	6
Chinese Restaurant	101.739140	3.077381	Big Tree Lin Kee Chan Sow Lin Steam Fish Head	101.740735	3.064708	Alam Damai	7
Bubble Tea Shop	101.749815	3.070599	Tealive	101.740735	3.064708	Alam Damai	8
Chinese Breakfast Place	101.745881	3.074038	【Lensen 点心】茶餐室	101.740735	3.064708	Alam Damai	9

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	Vietnar	mese Restaurant Vo	folleyball Court	Warehouse Store	Whisky Bar	Wine Bar	Wings Joint	Women's Store	Yakitori Restaurant	Yoga Stud	io Zoo
0	Alam Damai	0	0	0	0	0	0	0	0	0	77	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	30		0 0
4	Alam Damai	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0		0 0

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

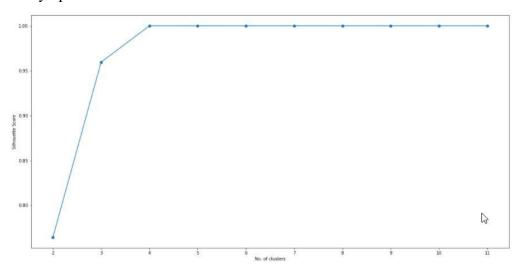
0	Alam Damai	0.0	0.0	0.1	0.0	0.00	0.00	0.00	0.02	0.0	0)	1	0.00	0.00	0.0	0.0	0.00	0.0	0.0	0.0
A	mpang, Kuala Lumpur	0.0	0.0	0.1	0.0	0.04	0.00	0.00	0.00	0.0	. 0	1	0.00	0.00	0.0	0.0	0.00	0.0	0.0	0.0
2	Bandar Menjalara	0.0	0.0	0.1	0.0	0.00	0.00	0.00	0.08	0.0	0.	1	0.00	0.00	0.0	0.0	0.00	0.0	0.0	0.0
3	Bandar Sri Permaisuri	0.0	0.0	0.0	0.0	0.02	0.00	0.00	0.10	0.0	. 0.	0	0.00	0.00	0.0	0.0	0.00	0.0	0.0	0.0
1	Bandar Tasik Selatan	0.0	0.0	0.1	0.0	0.00	0.00	0.00	0.12	0.0	. 0)	1	0.00	0.00	0.0	0.0	0.00	0.0	0.0	0.0
-																				
4	Taman Tun Dr Ismail	0.0	0.0	0.0	0.0	0.00	0.00	0.02	0.00	0.0	. 0)	)	0.00	0.00	0.0	0.0	0.00	0.0	0.0	0.0
5	Taman U-Thans	0.0	0.0	A 0.1	0.0	0.02	0.00	0.00	0.02	0.0	. 0)		0.00	0.00	0.0	0.0	0.00	0.0	0.0	0.0
6	Taman Wahyu	0.0	0.0	0.1	0.0	0.00	0.00	0.00	0.14	0.0	0)		0.00	0.00	0.0	0.0	0.00	0.0	0.0	0.0
7	Titwangsa	0.0	0.0	0.1	0.0	0.00	0.00	0.00	0.08	0.0	. 0.	1	0.02	0.02	0.0	0.0	0.00	0.0	0.0	0.0
13	Wangsa Maju	0.0	0.0	0.1	0.0	0.00	0.02	0.00	0.02	0.0	. 0	)	0.00	0.00	0.0	0.0	0.02	0.0	0.0	0.0

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

10th Most Common Venue	9th Most Common Venue	8th Most Common Venue	7th Most Common Venue	6th Most Common Venue	5th Most Common Venue	4th Most Common Venue	3rd Most Common Venue	2nd Most Common Venue	1st Most Common Venue	Neighbourhood	
Breakfast Spot	Convenience Store	Cantonese Restaurant	Indian Restaurant	Restaurant	Coffee Shop	Noosle House	Café	Malay Restaurant	Chinese Restaurant	Alam Damai	0
Resort	Art Gallery	Golf Course	Middle Eastern Restaurant	Coffee Shop	Juice Bar	Breakfast Spot	Asian Restaurant	Malay Restaurant	Hotel	Ampang, Kuala Lumpur	1
Mountain	Mosque	Grocery Store	Restaurant	Park	Vegetarian / Vegan Restaurant	Asian Restaurant	Japanese Restaurant	Coffee Shop	Chinese Restaurant	Bandar Menjalara	2
Lake	Bakery	Fast Food Restaurant	Pizza Place	Seafood Restaurant	Gas Station	Noodle House	Indonesian Restaurant	Asian Restaurant	Chinese Restaurant	Bandar Sri Permaisuri	3
Comfort Food Restaurant	Gym Pool	Cantonese Restaurant	Sandwich Place	Convenience Store	Ice Cream Shop	Malay Restaurant	Chinese Restaurant	Burger Joint	Asian Restaurant	Bandar Tasik Selatan	4
Market	Lake	Gas Station	Burger Joint	Malay Restaurant	Bakery	Flea Market	Asian Restaurant	Indonesian Restaurant	Chinese Restaurant	Bandar Tun Razak	5
Boutique	Steakhouse	Coffee Shop	Shopping Mall	Spa	Clothing Store	Burger Joint	Ice Cream Shop	Bar	Indian Restaurant	Bangsar	6
Boutique	Steakhouse	Coffee Shop	Shopping Mall	Spa	Clothing Store	Burger Joint	Ice Cream Shop	Bar	Indian Restaurant	Bangsar Park	7
Juice Bar	Chinese Restaurant	Shopping Mall	Clothing Store	Movie Theater	Malay Restaurant	Coffee Shop	Restaurant	Hotel	Japanese Restaurant	Bangsar South	8
Cantonese Restaurant	Café	Steakhouse	Dim Sum Restaurant	Lake	Food Truck	Coffee Shop	Malay Restaurant	Asian Restaurant	Chinese Restaurant	Batu 11 Cheras	9
Spa	Noodle House	Flea Market	Motorcycle Shop	Restaurant	Food Truck	Asian Restaurant	Seafood Restaurant	Malay Restaurant	Chinese Restaurant	Batu, Kuala Lumpur	10
South Indian Restaurant	Hotel Bar	Cocktail Bar	Coffee Shop	Shopping Mail	Café	French Restaurant	Clothing Store	Indian Restaurant	Hotel	Brickfields	11
Science Museum	Hainan Restaurant	Clothing Store	Japanese Restaurant	Boutique	Shopping Mall	Italian Restaurant	Spa	Café	Hotel	Bukit Bintang	12
Restaurant	Vegetarian / Vegan Restaurant	loe Cream Shop	Golf Course	Middle Eastern Restaurant	Massage Studio	Asian Restaurant	Japanese Restaurant	Café	Coffee Shop	Bukit Jalii	13
Bakery	Golf Course	Japanese Restaurant	Lounge	Malay Restaurant	Food Truck	Ice Cream Shop	Indian Restaurant	Asian Restaurant	Skate Park	Bukit Kiara	14
Dance Studio	Soshi Restaurant	Breakfast Spot	Monument / Landmark	Boutique	Building	Cosmetics Shop	Bar	Cocktail Bar	Hotel	Bukit Nanas	15
Performing Arts Venue	Burnto Place	Convenience Store	Coffee Shop	South Indian Restaurant	Juice Bar	Indian Restaurant	Hotel Bar	Café	Hotel	Bukit Petaling	16
Hotel	Spa	Noodle House	Bistro	Restaurant	Art Gallery	Café	Pool	Coffee Shop	Japanese Restaurant	Bukit Tunku	17
Nightolub	Market	Pizza Place	Noodle House	Food Truck	Seafood Restaurant	Indonesian Restaurant	Asian Restaurant	Malay Restaurant	Chinese Restaurant	Cheras, Kuala Lumpur	18
IT Services	Chinese Restaurant	Theater	Thai Restaurant	Seafood Restaurant	Asian Restaurant	Clothing Store	Coffee Shop	Malay Restaurant	Hotel	Chow Kit	19
Bakery	Yoga Studio	Grocery Store	Asian Restaurant	Korean Restaurant	Indian Restaurant	Ice Cream Shop	Italian Restaurant	Café	Japanese Restaurant	Damansara Heights	20
Electronics Store	Roof Deck	Convenience Store	BBQ Joint	Shopping Mall	Malay Restaurant	Coffee Shop	Racetrack	Chinese Restaurant	Seafood Restaurant	Damansara Town Centre	21
Gym	Multiplex	Building	Monument / Landmark	Breakfast Spot	Sushi Restaurant	Chettinad Restaurant	Cooktail Bar	Bar	Hotel	Dang Wangi	22
Café	Coffee Shop	Malay Restaurant	Bakery	Gym	Japanese Restaurant	Hotpot Restaurant	Burger Joint	Asian Restaurant	Chinese Restaurant	Desa Petaling	23
Hotel Bar	Shopping Mall	Spa	Café	Arts & Crafts Store	Ice Cream Shop	Steakhouse	French Restaurant	Boutique	Hotel	Federal Hill, Kuala Lumpur	24
Zoo	Gas Station	Garden Center	Garden	Furniture / Home Store	Gay Bar	Border Crossing	Chinese Restaurant	Outdoors & Recreation	Asian Restaurant	Happy Garden	25
Noodle House	Coffee Shop	Shopping Mall	Chinese Restaurant	Korean Restaurant	Café	Burger Joint	Furniture / Home Store	Nightolub	Asian Restaurant	lalan Cochrane, Kuala Lumpur	26
Japanese Restaurant	Coffee Shop	Shopping Mail	Movie Theater	Juice Bar	Clothing Store	Bakery	Department Store	Ice Cream Shop	Cosmetics Shop	KL Eco City	27
The state of the s	********	C		F		Contact Co.	Carrio Pierra	Charles Com	1 hand	Vancous Rev. Works I	20

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.

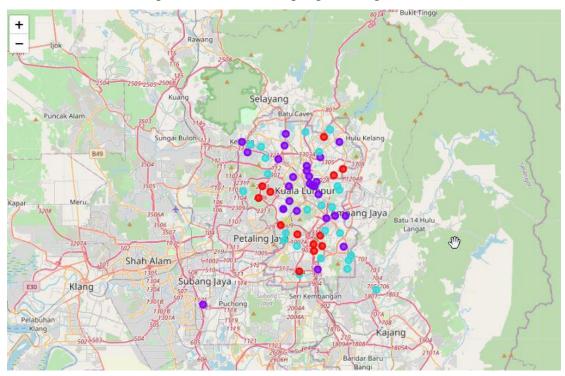
K means clustering algorithm was implemented in order to find out the similarities between found similarities between neighbourhood. Elbow method was use to identify optimal value of clusters. As result:



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

Neighb	borhood	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	th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Alan	m Darrai	101.740735	5.084708		Chinese Restaurant	Malay Restaurant	Café	Noodle House	Coffee Shop	Restaurant	Indian Restaurant	Cartonese Restaurant	Convenience Store	Breakfast Spot
Ampang, Kuala	Lumpur	101.732250	3.152135		Hotel	Malay Restaurant	Asian Restaurant	Breakfast Spot	Julge Bar	Coffee Shop	Middle Eastern Restaurant	Golf Course	Art Gallery	Resort
Bandar M	/enjalara	101,630030	3.193954	N 1	Chinese Restaurant	Coffee Shop	Japanese Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Park	Restaurant	Grocery Store	Mosque	Mountain
Bandar Sri Pe	emalsuri	101.710872	3.101158	r.c. (	Chinese Restaurant	Asian Restaurant	Indonesian Restaurant	Noodle House	Gas Station	Seafood Restaurant	Pizza Place	Fast Food Restaurant	Bakery	Lake
Bandar Tasik	k Selatan	101.711447	3.078097	2	Asian Restaurant	Burger Joint	Chinese Restaurant	Malay Restaurant	Ice Cream Shop	Convenience Store	Sandwich Place	Careonese Restaurant	Gym Pool	Comfort Food Restauran
Bandar Tu	an Razak	101.712467	3.089895		Chinese Restaurant	Indonesian Restaurant	Asian Restaurant	Flea Market	Bakary	Malay Restaurant	Burger Joint	Gas Station	Lake	Marke
	Bangsar	101.889440	3.130830		Indian Restaurant	Bar	loe Cream Shop	Burger Joint	Clothing Store	Spa	Shopping Mall	Coffee Shop	Steakhouse	Boutique
Bang	par Park	101.889440	3.135830		Indian Restaurant	Bar	Ice Cream Shop	Burger Joint	Clothing Store	Spa	Shopping Mall	Coffee Shop	Steakhouse	Boutque
Bangs	sar South	101.686729	3.112973	9	Japanese Restaurant	Hotel	Restaurant	Coffee Shop	Malay Restaurant	Movie Theater	Clothing Store	Shopping Mail	Chinese Restaurant	Juice Sa
Baou 11	1 Cheras	101.773369	3.041331		Chinese Restaurant	Asian Restaurant	Malay Restaurant	Coffee Shop	Food Truck	Lake	Dim Sum Restaurant	Steakhouse	Café	Cantonese Restauran
Batu, Kuala	Lumpur	101.671022	3.201823		Chinese Restaurant	Malay Restaurant	Seafood Restaurant	Asian Restaurant	Food Truck	Restaurant	Motorcycle Shop	Flea Market	Noodle House	Spo
Br	iriokfields	101.684553	3.128857		Hotel	Indian Restaurant	Clothing Store	French Restaurant	Café	Shopping Mall	Coffee Shop	Cooktail Bar	Hotel Bar	South Indian Restauran
Bukit	Birtang	101.708801	3.147107		Hotel	CaNi	Spa	Italian Restaurant	Shopping Mail	Boutique	Japanese Restaurant	Clothing Store	Hainan Rastaurant	Science Museur
8	Bukit Jalil	101.887439	3.058453		Coffee Shop	Calli	Japanese Restaurant	Asian Restaurant	Massage Studio	Middle Eastern Restaurant	Goff Course	Ice Cream Shop	Vegetarian / Wegan Restaurant	Restauran
Bu	okit Klara	101.642108	3.143000		Skate Park	Asian Restaurant	Indian Restaurant	Ice Cream Shop	Food Truck	Maley Restaurant	Lounge	Japanese Restaurant	Golf Course	Baker
Buk	kit Nanas	101,704833	3,155180		Hotel	Cooksil Bar	Qur.	Cosmetics Shop	Building	Boutique	Monument / Landmark	Breakfast Spot	Sushi Restaurant	Dance Studi
Bukit	Petaling	101.090380	3.130081	- 2	Hotel	Caté	Hotel Bar	Indian Restaurant	Juice Bar	South Indian Restaurant	Coffee Shop	Convenience Store	Burnto Place	Performing Arts Venu
Buk	kit Tunku	101.680668	3.100581	1	Japanese Restaurant	Coffee Shop	Pool	Café	Art Gallery	Restaurant	Bistro	Noodle House	Spa	Hote
Cheras, Kuala	Lumpur	101.718490	3.107178	2	Chinese Restaurant	Malay Restaurant	Asian Restaurant	Indonesian Restaurant	Seafood Restaurant	Food Truck	Neodle House	Pizza Place	Market	Nightolu
	Chow lift	101.898374	3.187408		Hotal	Malay Restaurant	Coffee Shop	Clothing Store	Asian Restaurant	Seafood Restaurant	Thai Restaurant	Theater	Chinese Restaurant	IT Service
Damansara	Heights	101.854907	3.140074		Japanese Restaurant	Callé	Italian Restaurant	Ice Cream Shop	Indian Restaurant	Korean Restaurant	Asian Restaurant	Grocery Stone	Yoga Studio	Baker
Damansara Town	in Centre	101.580388	3.024748		Seafood Restaurant	Chinese Restaurant	Racetrack	Coffee Shop	Maley Restaurant	Shopping Mail	BBQ Joint	Convenience Store	Roof Deck	Electronics Stor
Dan	ng Wangi	101.701975	3,150000		Hotel	Dat	Cocktail Bar	Chettinad Restaurant	Sushi Restaurant	Breakfast Spot	Monument / Landmark	Building	Multiplex	Oye
Desa	Petaling	101.703552	3.084185		Chinese Restaurant	Asian Restaurant	Burger Joint	Hotpot Restaurant	Japanese Restaurant	Gym	Bakery	Malay Restaurant	Coffee Shop	Caf
Federal Hill, Kuala	Lumpur	101.676423	3.139594		Hotel	Boutique	French Restaurant	Steakhouse	Ice Cream Shop	Arts & Crafts Store	Carlé	Spa	Shopping Mall	Hotel Ba
Нарру	y Garden	115.000413	4.738131		Asian Restaurant	Outdoors & Recreation	Chinese Restaurant	Border Crossing	Gay Bar	Furniture / Home Store	Garden	Garden Certer	Gas Station	Zo
Jalan Cochrane, Kuala	Lumpur	101.724003	3.133066	2	Asian Restaurant	Nightolub	Furniture / Home Store	Burger Joint	Callé	Korean Restaurant	Chinese Restaurant	Shopping Mall	Coffee Shop	Needle House
Kampung Baru, Kuala	Lumpur	101.708597	3.161326		Hotel	Clothing Store	Snack Place	Cocktail Bar	Bar	Cosmetics Shop	Accessories Store	Science Museum	Steakhouse	Buildin
Kampung Datuk I	Karamat	101.725970	3.168963		Hotel	Malay Restaurant	Italian Restaurant	Coffee Shop	Seafood Restaurant	Caté	Grocery Store	Art Gallery	French Restaurant	Hotel Ba
Kampung Padang	g Balang	101.094328	3.210092		Malay Restaurant	Asian Restaurant	Convenience Store	Steakhouse	Chinese Restaurant	Snack Place	Restaurant	Sushi Restaurant	Halal Restaurant	Oye
	Kepong	101.523711	3.200933		Chinese Restaurant	Caté	Noodle House	Asian Restaurant	Carsonese Restaurant	Food Truck	Vegetarian / Vegan Restaurant	Desset Shop	Restaurant	Miscellaneous Shop
Kepo	ong Baru	101.644962	3.199935		Chinese Restaurant	Asian Restaurant	Dessert Shop	Thai Restaurant	Food Truck	Caté 1	Vegetarian / Vegan Restaurant	Coffee Shop	Restaurant	Snack Place
KLI	Eco City	101.673764	3.118147		Cosmerics Shop	Ice Cream Shop	Department Store	Bakery	Clothing Store	Juice Bar	Movie Theater	Shopping Mall	Coffee Shop	Japanese Restauram

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.