

# **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.

## 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.

Category:Suburbs in Kuala Lumpur

from Wikipedia, the free encyclopedia	
<span></span>	
<b>v · t · e</b>	<b>Federal Territory of Kuala Lumpur</b> <div>Capital: Kuala Lumpur</div>
<b>Topics</b>	Index · History · Kuala Lumpur City Hall · Constitution · Elections · Geography · Law · Music · Kuala Lumpurse · Symbols (Flag · Seal) · Tourist attractions
<b>Society</b>	Culture · Crime · Cuisine · Bicycle & cycling · Demographics · Economy · Education · Politics
<b>Parliamentary constituencies</b>	Bandar Tun Razak · Batu · Bukit Bintang · Cheras · Kajang · Lembah Pantai · Segambut · Sepuloh · Setiawangsa · Tuaran · Wangsa Maju
	<div><span><span></span></span> Commons<span> </span><span><span></span></span> Wikisource<span> </span><span><span></span></span> Category</div>

  

<h2>Pages in category "Suburbs in Kuala Lumpur"</h2> <p>The following 70 pages are in this category, out of 70 total. This list may not reflect recent changes (<a href="#">learn more</a>).</p>	
<p><b>A</b></p> <ul style="list-style-type: none"><li>Alam Damai</li> <li>Ampang, Kuala Lumpur</li></ul> <p><b>B</b></p> <ul style="list-style-type: none"><li>Bandar Menjalara</li> <li>Bandar Sri Permaisuri</li> <li>Bandar Tasik Selatan</li> <li>Bandar Tun Razak</li> <li>Bangsar</li> <li>Bangsar Park</li> <li>Bangsar South</li> <li>Batu 11 Cheras</li> <li>Batu, Kuala Lumpur</li> <li>Brickfields</li> <li>Bukit Bintang</li> <li>Bukit Jalil</li> <li>Bukit Kiara</li> <li>Bukit Nanas</li> <li>Bukit Petaling</li> <li>Bukit Tunku</li></ul> <p><b>C</b></p> <ul style="list-style-type: none"><li>Cheras, Kuala Lumpur</li> <li>Chow Kit</li></ul> <p><b>D</b></p> <ul style="list-style-type: none"><li>Damansara Heights</li> <li>Damansara Town Centre</li> <li>Dang Wangi</li> <li>Dessa Petaling</li></ul>	<ul style="list-style-type: none"><li>Salak South</li> <li>Segambut</li> <li>Semarak</li> <li>Senlis, Kuala Lumpur</li> <li>Selapak</li> <li>Setiawangsa</li> <li>Shamelin</li> <li>Sri Hartamas</li> <li>Sri Petaling</li> <li>Sungai Besi</li></ul> <p><b>T</b></p> <ul style="list-style-type: none"><li>Taman Bukit Maluri</li> <li>Taman Connaught</li> <li>Taman Desa</li> <li>Taman Duda</li> <li>Taman Ibu Kota</li> <li>Taman Len Seng</li> <li>Taman Melati</li> <li>Taman Melati</li> <li>Taman MUG</li> <li>Taman P. Ramlee</li> <li>Taman Sri Sinar</li> <li>Taman Tropicana View</li> <li>Taman Tun Dr Ismail</li> <li>Taman U-Thant</li> <li>Taman Wahyu</li> <li>TBwangsa</li></ul> <p><b>W</b></p> <ul style="list-style-type: none"><li>Wangsa Maju</li></ul>

**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  $\times$  3 columns

### **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.