Final Report: Battle of Neighbouhoods

Opening a new shopping mall in Kuala Lumpur, Malaysia

Introduction

For many shoppers, visiting shopping malls is a great way to relax and enjoy themselves during weekends and holidays. They can do grocery shopping, dine at restaurants, shop at the various fashion outlets, watch movies and perform many more activities. Shopping malls are like a one-stop destination for all types of shoppers. For retailers, the central location and the large crowd at the shopping malls provides a great distribution channel to market their products and services. Property developers are also taking advantage of this trend to build more shopping malls to cater to the demand. As a result, there are many shopping malls in the city of Kuala Lumpur and many more are being built. Opening shopping malls allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Kuala Lumpur, Malaysia to open a new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Kuala Lumpur, Malaysia, if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

Target Audience of this Project

This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in the capital city of Malaysia i.e. Kuala Lumpur. This project is timely as the city is currently suffering from oversupply of shopping malls. Data from the National Property Information Centre (NAPIC) released last year showed that an additional 15 per cent will be added to existing mall space, and the agency predicted that total occupancy may dip below 86 per cent. The local newspaper The Malay Mail also reported in March last year that the true occupancy rates in malls may be as low as 40 per cent in some areas, quoting a Financial Times (FT) article cataloguing the country's continued obsession with building more shopping space despite chronic oversupply.

Data

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

• List of neighbourhoods in Kuala Lumpur. This defines the scope of this project which is confined to the city of Kuala Lumpur, the capital city of the country of Malaysia in South East Asia.

• Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.

• Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur (https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur)) contains a list of neighbourhoods in Kuala Lumpur, with a total of 70 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers.Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.

Methodoloy

Opening a New Shopping Mall in Kuala Lumpur, Malaysia

- Build a dataframe of neighborhoods in Kuala Lumpur, Malaysia by web scraping the data from Wikipedia page
- Get the geographical coordinates of the neighborhoods
- Obtain the venue data for the neighborhoods from Foursquare API
- Explore and cluster the neighborhoods
- Select the best cluster to open a new shopping mall

---

1. Import Libraries

In [5]:

```
!pip install geocoder
```

```
Collecting geocoder
  Downloading https://files.pythonhosted.org/packages/4f/6b/13166c909ad2f2d76b929a4227c952630eb
af0d729f6317eb09cbceccbab/geocoder-1.38.1-py2.py3-none-any.whl (98kB)
     |████████████████████████████████| 102kB 4.6MB/s ta 0:00:01
Requirement already satisfied: requests in /opt/conda/envs/Python36/lib/python3.6/site-packages
(from geocoder) (2.21.0)
Requirement already satisfied: click in /opt/conda/envs/Python36/lib/python3.6/site-packages (f
rom geocoder) (7.0)
Requirement already satisfied: future in /opt/conda/envs/Python36/lib/python3.6/site-packages (
from geocoder) (0.17.1)
Requirement already satisfied: six in /opt/conda/envs/Python36/lib/python3.6/site-packages (fro
m geocoder) (1.12.0)
Collecting ratelim (from geocoder)
  Downloading https://files.pythonhosted.org/packages/f2/98/7e6d147fd16a10a5f821db6e25f192265d6
ecca3d82957a4fdd592cad49c/ratelim-0.1.6-py2.py3-none-any.whl
Requirement already satisfied: urllib3<1.25,>=1.21.1 in /opt/conda/envs/Python36/lib/python3.6/
site-packages (from requests->geocoder) (1.24.1)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /opt/conda/envs/Python36/lib/python3.6/
site-packages (from requests->geocoder) (3.0.4)
Requirement already satisfied: idna<2.9,>=2.5 in /opt/conda/envs/Python36/lib/python3.6/site-pa
ckages (from requests->geocoder) (2.8)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/envs/Python36/lib/python3.6/sit
e-packages (from requests->geocoder) (2019.6.16)
Requirement already satisfied: decorator in /opt/conda/envs/Python36/lib/python3.6/site-package
s (from ratelim->geocoder) (4.3.2)
Installing collected packages: ratelim, geocoder
Successfully installed geocoder-1.38.1 ratelim-0.1.6
```

In [7]:

```
!pip install folium
```

```
Collecting folium
  Downloading https://files.pythonhosted.org/packages/72/ff/004bfe344150a064e558cb2aedeaa02ecbf
75e60e148a55a9198f0c41765/folium-0.10.0-py2.py3-none-any.whl (91kB)
     |████████████████████████████████| 92kB 4.1MB/s eta 0:00:01
Requirement already satisfied: requests in /opt/conda/envs/Python36/lib/python3.6/site-packages
(from folium) (2.21.0)
Requirement already satisfied: jinja2>=2.9 in /opt/conda/envs/Python36/lib/python3.6/site-packa
ges (from folium) (2.10)
Collecting branca>=0.3.0 (from folium)
  Downloading https://files.pythonhosted.org/packages/63/36/1c93318e9653f4e414a2e0c3b98fc898b49
70e939afeedeee6075dd3b703/branca-0.3.1-py3-none-any.whl
Requirement already satisfied: numpy in /opt/conda/envs/Python36/lib/python3.6/site-packages (f
rom folium) (1.15.4)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /opt/conda/envs/Python36/lib/python3.6/
site-packages (from requests->folium) (3.0.4)
Requirement already satisfied: urllib3<1.25,>=1.21.1 in /opt/conda/envs/Python36/lib/python3.6/
site-packages (from requests->folium) (1.24.1)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/envs/Python36/lib/python3.6/sit
e-packages (from requests->folium) (2019.6.16)
Requirement already satisfied: idna<2.9,>=2.5 in /opt/conda/envs/Python36/lib/python3.6/site-pa
ckages (from requests->folium) (2.8)
Requirement already satisfied: MarkupSafe>=0.23 in /opt/conda/envs/Python36/lib/python3.6/site-
packages (from jinja2>=2.9->folium) (1.1.0)
Requirement already satisfied: six in /opt/conda/envs/Python36/lib/python3.6/site-packages (fro
m branca>=0.3.0->folium) (1.12.0)
Installing collected packages: branca, folium
Successfully installed branca-0.3.1 folium-0.10.0
```

In [8]:

```python
import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analsysis
pd.set_option("display.max_columns", None)
pd.set_option("display.max_rows", None)

import json # library to handle JSON files

from geopy.geocoders import Nominatim # convert an address into latitude and longitude values
import geocoder # to get coordinates

import requests # library to handle requests
from bs4 import BeautifulSoup # library to parse HTML and XML documents

from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

import folium # map rendering library

print("Libraries imported.")
```

```
Libraries imported.
```

2. Scrap data from Wikipedia page into a DataFrame

In [9]:

```python
# send the GET request
data = requests.get("https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur").text
```

In [10]:

```python
# parse data from the html into a beautifulsoup object
soup = BeautifulSoup(data, 'html.parser')
```

In [11]:

```
# create a list to store neighborhood data
neighborhoodList = []
```

In [12]:

```
# append the data into the list
for row in soup.find_all("div", class_="mw-category")[0].findAll("li"):
    neighborhoodList.append(row.text)
```

In [13]:

```
# create a new DataFrame from the list
kl_df = pd.DataFrame({"Neighborhood": neighborhoodList})

kl_df.head()
```

Out[13]:

	Neighborhood
0	Alam Damai
1	Ampang, Kuala Lumpur
2	Bandar Menjalara
3	Bandar Sri Permaisuri
4	Bandar Tasik Selatan

In [14]:

```
# print the number of rows of the dataframe
kl_df.shape
```

Out[14]:

```
(71, 1)
```

3. Get the geographical coordinates

In [15]:

```
# define a function to get coordinates
def get_latlng(neighborhood):
    # initialize your variable to None
    lat_lng_coords = None
    # loop until you get the coordinates
    while(lat_lng_coords is None):
        g = geocoder.arcgis('{}, Kuala Lumpur, Malaysia'.format(neighborhood))
        lat_lng_coords = g.latlng
    return lat_lng_coords
```

In [16]:

```
# call the function to get the coordinates, store in a new list using list comprehension
coords = [ get_latlng(neighborhood) for neighborhood in kl_df["Neighborhood"].tolist() ]
```

In [17]:

```
coords
```

Out[17]:

```
[[3.0576900000000364, 101.74388000000005],
 [3.1531525006886456, 101.70041313127312],
 [3.190350000000236, 101.62545000000006],
 [3.1039100000000417, 101.71226000000007],
 [3.072620000000029, 101.71471000000008],
 [3.08280000000002, 101.72281000000004],
 [3.1292000000000257, 101.67844000000008],
 [3.1347800000000348, 101.67262000000005],
 [3.111020000000053, 101.66283000000004],
 [3.09898000000004, 101.73499000000004],
 [3.1357600000000616, 101.70837000000006],
 [3.12916000000007, 101.68406000000004],
 [3.147770000000037, 101.70855000000006],
 [3.057800000000043, 101.68965000000009],
 [3.1434800000000678, 101.64433000000008],
 [3.152017197420035, 101.70102760046613],
 [3.129290000000026, 101.69892000000004],
 [3.17381000000006, 101.68276000000009],
 [3.061870000000056, 101.74675000000008],
 [3.163590000000056, 101.69811000000004],
 [3.147980000000075, 101.66798000000006],
 [3.1387586696676304, 101.6840455304707],
 [3.1387586696676304, 101.6840455304707],
 [3.156685250705136, 101.6980764971795],
 [3.08331000000004, 101.70438000000007],
 [3.1363700000000563, 101.68564000000003],
 [3.201630000000023, 101.72107000000005],
 [3.132977301586902, 101.72466893310815],
 [3.1801625281028416, 101.67788025972793],
 [3.209500000000048, 101.65874000000008],
 [3.165460000000053, 101.71028000000007],
 [3.166400000000067, 101.73046000000005],
 [3.2094300000000544, 101.69318000000004],
 [3.2175000000000296, 101.63763000000006],
 [3.117140000000063, 101.67389000000009],
 [3.0907400000000393, 101.67733000000004],
 [3.1212017325659525, 101.6638990103096],
 [3.147890000000075, 101.69405000000006],
 [3.1592600000000743, 101.69834000000009],
 [3.147890000000075, 101.69405000000006],
 [3.165320000000065, 101.65243000000004],
 [3.094760000000065, 101.66747000000004],
 [3.133540000000039, 101.71307000000007],
 [3.125861531801437, 101.7186241383456],
 [3.0810200000000236, 101.69724000000008],
 [3.186390000000074, 101.66810000000004],
 [3.1799266000286424, 101.72144192942527],
 [3.1874305003813963, 101.69145318264098],
 [3.188160000000039, 101.70415000000008],
 [3.191802843003158, 101.74007037312064],
 [3.124580000000037, 101.73597000000007],
 [3.162200000000041, 101.65036000000003],
 [3.0726000000000226, 101.68252000000007],
 [3.050640000000044, 101.70613000000009],
 [3.2006600000000276, 101.63337000000007],
 [3.0826300000000515, 101.74671000000006],
 [3.0826900000000705, 101.73689000000007],
 [3.1029700000000275, 101.68471000000005],
 [3.21216000000004, 101.71540000000005],
 [3.069080000000042, 101.74287000000004],
 [3.223570000000052, 101.72399000000007],
 [3.093590000000063, 101.72837000000004],
 [3.2100497379521316, 101.63450794879562],
 [3.1936000000000604, 101.70598000000007],
 [3.1900700000000484, 101.65293000000008],
 [3.0870700000000397, 101.73681000000005],
 [3.1528300000000513, 101.62271000000004],
 [3.157700000000034, 101.72452000000004],
 [3.22240000000005, 101.67173000000008],
 [3.1806700000000205, 101.70322000000004],
 [3.2039100000000644, 101.73719000000006]]
```

In [18]:

```python
# create temporary dataframe to populate the coordinates into Latitude and Longitude
df_coords = pd.DataFrame(coords, columns=['Latitude', 'Longitude'])
```

In [19]:

```
# merge the coordinates into the original dataframe
kl_df['Latitude'] = df_coords['Latitude']
kl_df['Longitude'] = df_coords['Longitude']
```

In [20]:

```
# check the neighborhoods and the coordinates
print(kl_df.shape)
kl_df
```

(71, 3)

Out[20]:

	Neighborhood	Latitude	Longitude
0	Alam Damai	3.057690	101.743880
1	Ampang, Kuala Lumpur	3.153153	101.700413
2	Bandar Menjalara	3.190350	101.625450
3	Bandar Sri Permaisuri	3.103910	101.712260
4	Bandar Tasik Selatan	3.072620	101.714710
5	Bandar Tun Razak	3.082800	101.722810
6	Bangsar	3.129200	101.678440
7	Bangsar Park	3.134780	101.672620
8	Bangsar South	3.111020	101.662830
9	Batu 11 Cheras	3.098980	101.734990
10	Batu, Kuala Lumpur	3.135760	101.708370
11	Brickfields	3.129160	101.684060
12	Bukit Bintang	3.147770	101.708550
13	Bukit Jalil	3.057800	101.689650
14	Bukit Kiara	3.143480	101.644330
15	Bukit Nanas	3.152017	101.701028
16	Bukit Petaling	3.129290	101.698920
17	Bukit Tunku	3.173810	101.682760
18	Cheras, Kuala Lumpur	3.061870	101.746750
19	Chow Kit	3.163590	101.698110
20	Damansara Heights	3.147980	101.667980
21	Damansara Town Centre	3.138759	101.684046
22	Damansara, Kuala Lumpur	3.138759	101.684046
23	Dang Wangi	3.156685	101.698076
24	Desa Petaling	3.083310	101.704380
25	Federal Hill, Kuala Lumpur	3.136370	101.685640
26	Happy Garden	3.201630	101.721070
27	Jalan Cochrane, Kuala Lumpur	3.132977	101.724669
28	Jalan Duta	3.180163	101.677880
29	Jinjang	3.209500	101.658740
30	Kampung Baru, Kuala Lumpur	3.165460	101.710280
31	Kampung Datuk Keramat	3.166400	101.730460
32	Kampung Padang Balang	3.209430	101.693180
33	Kepong	3.217500	101.637630
34	KL Eco City	3.117140	101.673890
35	Kuchai Lama	3.090740	101.677330
36	Lembah Pantai	3.121202	101.663899
37	Maluri	3.147890	101.694050
38	Medan Tuanku	3.159260	101.698340

39	Miharja	3.147890	101.694050
40	Mont Kiara	3.165320	101.652430
41	Pantai Dalam	3.094760	101.667470
42	Pudu, Kuala Lumpur	3.133540	101.713070
43	Putrajaya	3.125862	101.718624
44	Salak South	3.081020	101.697240
45	Segambut	3.186390	101.668100
46	Semarak	3.179927	101.721442
47	Sentul Raya	3.187431	101.691453
48	Setapak	3.188160	101.704150
49	Setiawangsa	3.191803	101.740070
50	Shamelin	3.124580	101.735970
51	Sri Hartamas	3.162200	101.650360
52	Sri Petaling	3.072600	101.682520
53	Sungai Besi	3.050640	101.706130
54	Taman Bukit Maluri	3.200660	101.633370
55	Taman Cheras Hartamas	3.082630	101.746710
56	Taman Connaught	3.082690	101.736890
57	Taman Desa	3.102970	101.684710
58	Taman Ibukota	3.212160	101.715400
59	Taman Len Seng	3.069080	101.742870
60	Taman Melati	3.223570	101.723990
61	Taman Midah	3.093590	101.728370
62	Taman OUG	3.210050	101.634508
63	Taman P. Ramlee	3.193600	101.705980
64	Taman Sri Sinar	3.190070	101.652930
65	Taman Taynton View	3.087070	101.736810
66	Taman Tun Dr Ismail	3.152830	101.622710
67	Taman U-Thant	3.157700	101.724520
68	Taman Wahyu	3.222400	101.671730
69	Titiwangsa	3.180670	101.703220
70	Wangsa Maju	3.203910	101.737190

In [21]:

```
# save the DataFrame as CSV file
kl_df.to_csv("kl_df.csv", index=False)
```

4. Create a map of Kuala Lumpur with neighborhoods superimposed on top

In [22]:

```
# get the coordinates of Kuala Lumpur
address = 'Kuala Lumpur, Malaysia'

geolocator = Nominatim(user_agent="my-application")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geographical coordinate of Kuala Lumpur, Malaysiae {}, {}.'.format(latitude, longitude))
```

```
The geographical coordinate of Kuala Lumpur, Malaysiae 3.1516636, 101.6943028.
```

In [23]:

```
# create map of Toronto using latitude and longitude values
map_kl = folium.Map(location=[latitude, longitude], zoom_start=11)

# add markers to map
for lat, lng, neighborhood in zip(kl_df['Latitude'], kl_df['Longitude'], kl_df['Neighborhood']):
    label = '{}'.format(neighborhood)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7).add_to(map_kl)

map_kl
```

Out[23]:

In [24]:

```
# save the map as HTML file
map_kl.save('map_kl.html')
```

5. Use the Foursquare API to explore the neighborhoods

In [29]:

```
# define Foursquare Credentials and Version
CLIENT_ID = 'ID'  # your Foursquare ID
CLIENT_SECRET = 'SECRET' # your Foursquare Secret
VERSION = '20190930' # Foursquare API version

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: KKT0O13GOVTTFGZBZUPUGKPP4CSGPUTWAKSILLKMOMOUKTWY
CLIENT_SECRET:IOTG1WJ2S0JFZCCFCTL5AEUP1GIDWCTVLSZ50CBMO1FX0S4F
```

Now, let's get the top 100 venues that are within a radius of 2000 meters.

In [30]:

```
radius = 2000
LIMIT = 100

venues = []

for lat, long, neighborhood in zip(kl_df['Latitude'], kl_df['Longitude'], kl_df['Neighborhood']):

    # create the API request URL
    url = "https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&v={}&ll={},{}&radius={
}&limit={}".format(
        CLIENT_ID,
        CLIENT_SECRET,
        VERSION,
        lat,
        long,
        radius,
        LIMIT)

    # make the GET request
    results = requests.get(url).json()["response"]['groups'][0]['items']

    # return only relevant information for each nearby venue
    for venue in results:
        venues.append((
            neighborhood,
            lat,
            long,
            venue['venue']['name'],
            venue['venue']['location']['lat'],
            venue['venue']['location']['lng'],
            venue['venue']['categories'][0]['name']))
```

In [31]:

```
# convert the venues list into a new DataFrame
venues_df = pd.DataFrame(venues)

# define the column names
venues_df.columns = ['Neighborhood', 'Latitude', 'Longitude', 'VenueName', 'VenueLatitude', 'VenueLongitude'
, 'VenueCategory']

print(venues_df.shape)
venues_df.head()
```

(7083, 7)

Out[31]:

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Alam Damai	3.05769	101.74388	Pengedar Shaklee Kuala Lumpur	3.061235	101.740696	Supplement Shop
1	Alam Damai	3.05769	101.74388	Machi Noodle 妈子面	3.057695	101.746635	Noodle House
2	Alam Damai	3.05769	101.74388	628火焰鑫茶室	3.058442	101.747947	Chinese Restaurant
3	Alam Damai	3.05769	101.74388	Restoran Ikbal	3.061134	101.750220	Restaurant
4	Alam Damai	3.05769	101.74388	沙巴生肉面	3.057715	101.749096	Chinese Restaurant

Let's check how many venues were returned for each neighorhood

In [33]:

```
venues_df.groupby(["Neighborhood"]).count()
```

Out[33]:

	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
Neighborhood						
Alam Damai	100	100	100	100	100	100
Ampang, Kuala Lumpur	100	100	100	100	100	100
Bandar Menjalara	100	100	100	100	100	100

Bandar Sri Permaisuri	100	100	100	100	100	100
Bandar Tasik Selatan	99	99	99	99	99	99
Bandar Tun Razak	100	100	100	100	100	100
Bangsar	100	100	100	100	100	100
Bangsar Park	100	100	100	100	100	100
Bangsar South	100	100	100	100	100	100
Batu 11 Cheras	100	100	100	100	100	100
Batu, Kuala Lumpur	100	100	100	100	100	100
Brickfields	100	100	100	100	100	100
Bukit Bintang	100	100	100	100	100	100
Bukit Jalil	100	100	100	100	100	100
Bukit Kiara	100	100	100	100	100	100
Bukit Nanas	100	100	100	100	100	100
Bukit Petaling	100	100	100	100	100	100
Bukit Tunku	100	100	100	100	100	100
Cheras, Kuala Lumpur	100	100	100	100	100	100
Chow Kit	100	100	100	100	100	100
Damansara Heights	100	100	100	100	100	100
Damansara Town Centre	100	100	100	100	100	100
Damansara, Kuala Lumpur	100	100	100	100	100	100
Dang Wangi	100	100	100	100	100	100
Desa Petaling	100	100	100	100	100	100
Federal Hill, Kuala Lumpur	100	100	100	100	100	100
Happy Garden	100	100	100	100	100	100
Jalan Cochrane, Kuala Lumpur	100	100	100	100	100	100
Jalan Duta	100	100	100	100	100	100
Jinjang	100	100	100	100	100	100
KL Eco City	100	100	100	100	100	100
Kampung Baru, Kuala Lumpur	100	100	100	100	100	100
Kampung Datuk Keramat	100	100	100	100	100	100
Kampung Padang Balang	94	94	94	94	94	94
Kepong	100	100	100	100	100	100
Kuchai Lama	100	100	100	100	100	100
Lembah Pantai	100	100	100	100	100	100
Maluri	100	100	100	100	100	100
Medan Tuanku	100	100	100	100	100	100
Miharja	100	100	100	100	100	100
Mont Kiara	100	100	100	100	100	100
Pantai Dalam	100	100	100	100	100	100
Pudu, Kuala Lumpur	100	100	100	100	100	100
Putrajaya	100	100	100	100	100	100
Salak South	100	100	100	100	100	100
Segambut	100	100	100	100	100	100
Semarak	100	100	100	100	100	100
Sentul Raya	100	100	100	100	100	100
Setapak	100	100	100	100	100	100
Setiawangsa	100	100	100	100	100	100
Shamelin	100	100	100	100	100	100
Sri Hartamas	100	100	100	100	100	100
Sri Petaling	100	100	100	100	100	100

Sungai Besi	100	100	100	100	100	100
Taman Bukit Maluri	100	100	100	100	100	100
Taman Cheras Hartamas	100	100	100	100	100	100
Taman Connaught	100	100	100	100	100	100
Taman Desa	100	100	100	100	100	100
Taman Ibukota	100	100	100	100	100	100
Taman Len Seng	100	100	100	100	100	100
Taman Melati	100	100	100	100	100	100
Taman Midah	100	100	100	100	100	100
Taman OUG	100	100	100	100	100	100
Taman P. Ramlee	93	93	93	93	93	93
Taman Sri Sinar	97	97	97	97	97	97
Taman Taynton View	100	100	100	100	100	100
Taman Tun Dr Ismail	100	100	100	100	100	100
Taman U-Thant	100	100	100	100	100	100
Taman Wahyu	100	100	100	100	100	100
Titiwangsa	100	100	100	100	100	100
Wangsa Maju	100	100	100	100	100	100

Let's find out how many unique categories can be curated from all the returned venues

In [34]:

```
print('There are {} uniques categories.'.format(len(venues_df['VenueCategory'].unique())))
```

```
There are 307 uniques categories.
```

In [35]:

```
# print out the list of categories
venues_df['VenueCategory'].unique()[:50]
```

Out[35]:

```
array(['Supplement Shop', 'Noodle House', 'Chinese Restaurant',
       'Restaurant', 'Vegetarian / Vegan Restaurant', 'Breakfast Spot',
       'Food Court', 'Snack Place', 'Asian Restaurant', 'Park',
       'Other Great Outdoors', 'Indian Restaurant', 'Food Truck',
       'Dim Sum Restaurant', 'Japanese Restaurant', 'Spa',
       'Seafood Restaurant', 'Chinese Breakfast Place', 'Bubble Tea Shop',
       'Convenience Store', 'Dessert Shop', 'Pet Store', 'Bakery',
       'Cantonese Restaurant', 'Outlet Store', 'Malay Restaurant',
       'Farmers Market', 'Café', 'Gym / Fitness Center', 'Steakhouse',
       'Fast Food Restaurant', 'Badminton Court', 'Athletics & Sports',
       'Middle Eastern Restaurant', 'Hakka Restaurant',
       'Padangnese Restaurant', 'Burger Joint', 'Winery',
       'College Bookstore', 'Grocery Store', 'Halal Restaurant',
       'Bookstore', 'Monument / Landmark', 'Hostel',
       'Latin American Restaurant', 'Hotel', 'Hotel Pool',
       'South Indian Restaurant', 'Dance Studio', 'Soup Place'],
      dtype=object)
```

In [36]:

```
# check if the results contain "Shopping Mall"
"Neighborhood" in venues_df['VenueCategory'].unique()
```

Out[36]:

```
True
```

6. Analyze Each Neighborhood

In [37]:

```python
# one hot encoding
kl_onehot = pd.get_dummies(venues_df[['VenueCategory']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
kl_onehot['Neighborhoods'] = venues_df['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [kl_onehot.columns[-1]] + list(kl_onehot.columns[:-1])
kl_onehot = kl_onehot[fixed_columns]

print(kl_onehot.shape)
kl_onehot.head()
```

(7083, 308)

Out[37]:

	Neighborhoods	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	Alam Damai	0	0	0	0	0	0	0	0	0
1	Alam Damai	0	0	0	0	0	0	0	0	0
2	Alam Damai	0	0	0	0	0	0	0	0	0
3	Alam Damai	0	0	0	0	0	0	0	0	0
4	Alam Damai	0	0	0	0	0	0	0	0	0

Next, let's group rows by neighborhood and by taking the mean of the frequency of occurrence of each category

In [39]:

```python
kl_grouped = kl_onehot.groupby(["Neighborhoods"]).mean().reset_index()

print(kl_grouped.shape)
kl_grouped
```

(71, 308)

Out[39]:

	Neighborhoods	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	Alam Damai	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.050000
1	Ampang, Kuala Lumpur	0.000000	0.00	0.00	0.00	0.000000	0.01	0.00	0.00	0.000000
2	Bandar Menjalara	0.000000	0.00	0.01	0.00	0.000000	0.00	0.00	0.00	0.030000
3	Bandar Sri Permaisuri	0.010000	0.00	0.00	0.00	0.000000	0.01	0.00	0.00	0.040000
4	Bandar Tasik Selatan	0.010101	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.090909
5	Bandar Tun Razak	0.010000	0.00	0.00	0.00	0.000000	0.00	0.00	0.01	0.090000
6	Bangsar	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.02	0.010000
7	Bangsar Park	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.02	0.010000
8	Bangsar South	0.010000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.000000
9	Batu 11 Cheras	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.060000
10	Batu, Kuala Lumpur	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.010000
11	Brickfields	0.000000	0.00	0.01	0.00	0.000000	0.00	0.00	0.01	0.020000
12	Bukit Bintang	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.010000
13	Bukit Jalil	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.030000
14	Bukit Kiara	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.030000

15	Bukit Nanas	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.000000
16	Bukit Petaling	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.020000
17	Bukit Tunku	0.000000	0.00	0.00	0.00	0.010000	0.02	0.00	0.00	0.040000
18	Cheras, Kuala Lumpur	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.050000
19	Chow Kit	0.000000	0.01	0.00	0.00	0.000000	0.00	0.00	0.00	0.020000
20	Damansara Heights	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.02	0.020000
21	Damansara Town Centre	0.000000	0.00	0.01	0.00	0.000000	0.00	0.00	0.01	0.010000
22	Damansara, Kuala Lumpur	0.000000	0.00	0.01	0.00	0.000000	0.00	0.00	0.01	0.010000
23	Dang Wangi	0.000000	0.00	0.00	0.00	0.000000	0.01	0.00	0.00	0.010000
24	Desa Petaling	0.010000	0.00	0.00	0.00	0.000000	0.01	0.00	0.00	0.060000
25	Federal Hill, Kuala Lumpur	0.000000	0.00	0.01	0.00	0.000000	0.00	0.00	0.01	0.010000
26	Happy Garden	0.000000	0.00	0.00	0.00	0.000000	0.00	0.01	0.00	0.060000
27	Jalan Cochrane, Kuala Lumpur	0.000000	0.00	0.00	0.01	0.000000	0.00	0.00	0.00	0.060000
28	Jalan Duta	0.000000	0.00	0.00	0.00	0.010000	0.02	0.00	0.00	0.070000
29	Jinjang	0.010000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.150000
30	KL Eco City	0.020000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.010000
31	Kampung Baru, Kuala Lumpur	0.010000	0.00	0.00	0.00	0.000000	0.01	0.00	0.00	0.010000
32	Kampung Datuk Keramat	0.000000	0.00	0.00	0.01	0.000000	0.03	0.00	0.00	0.020000
33	Kampung Padang Balang	0.000000	0.00	0.00	0.00	0.010638	0.00	0.00	0.00	0.106383
34	Kepong	0.010000	0.00	0.00	0.01	0.000000	0.00	0.00	0.00	0.090000
35	Kuchai Lama	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.050000
36	Lembah Pantai	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.02	0.010000
37	Maluri	0.000000	0.01	0.00	0.00	0.000000	0.01	0.00	0.00	0.010000
38	Medan Tuanku	0.000000	0.00	0.00	0.00	0.000000	0.01	0.00	0.00	0.020000
39	Miharja	0.000000	0.01	0.00	0.00	0.000000	0.01	0.00	0.00	0.010000
40	Mont Kiara	0.000000	0.00	0.00	0.00	0.000000	0.01	0.00	0.00	0.030000
41	Pantai Dalam	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.040000
42	Pudu, Kuala Lumpur	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.030000
43	Putrajaya	0.000000	0.00	0.00	0.01	0.000000	0.00	0.00	0.00	0.070000
44	Salak South	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.050000
45	Segambut	0.000000	0.00	0.00	0.00	0.000000	0.02	0.00	0.00	0.080000
46	Semarak	0.000000	0.00	0.00	0.00	0.010000	0.03	0.01	0.01	0.060000
47	Sentul Raya	0.000000	0.00	0.00	0.00	0.010000	0.00	0.00	0.00	0.110000
48	Setapak	0.000000	0.00	0.00	0.00	0.010000	0.01	0.00	0.00	0.060000
49	Setiawangsa	0.000000	0.00	0.01	0.00	0.000000	0.00	0.01	0.00	0.070000
50	Shamelin	0.000000	0.00	0.00	0.01	0.000000	0.00	0.00	0.00	0.060000
51	Sri Hartamas	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.030000
52	Sri Petaling	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.060000
53	Sungai Besi	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.040000
54	Taman Bukit Maluri	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.070000
55	Taman Cheras Hartamas	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.010000
56	Taman Connaught	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.030000
57	Taman Desa	0.010000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.030000

58	Taman Ibukota	0.000000	0.00	0.00	0.01	0.000000	0.00	0.01	0.00	0.020000
59	Taman Len Seng	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.020000
60	Taman Melati	0.000000	0.00	0.00	0.01	0.000000	0.00	0.00	0.00	0.050000
61	Taman Midah	0.000000	0.00	0.00	0.00	0.000000	0.01	0.00	0.00	0.070000
62	Taman OUG	0.010000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.080000
63	Taman P. Ramlee	0.000000	0.00	0.00	0.00	0.010753	0.00	0.00	0.00	0.043011
64	Taman Sri Sinar	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.072165
65	Taman Taynton View	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.050000
66	Taman Tun Dr Ismail	0.010000	0.00	0.00	0.00	0.000000	0.00	0.00	0.02	0.000000
67	Taman U-Thant	0.000000	0.00	0.00	0.00	0.000000	0.01	0.00	0.00	0.000000
68	Taman Wahyu	0.000000	0.00	0.00	0.00	0.000000	0.00	0.00	0.00	0.070000
69	Titiwangsa	0.000000	0.00	0.00	0.00	0.010000	0.01	0.00	0.00	0.060000
70	Wangsa Maju	0.000000	0.00	0.00	0.00	0.000000	0.00	0.01	0.00	0.070000

In [40]:

```
len(kl_grouped[kl_grouped["Shopping Mall"] > 0])
```

Out[40]:

39

Create a new DataFrame for Shopping Mall data only

In [41]:

```
kl_mall = kl_grouped[["Neighborhoods","Shopping Mall"]]
```

In [42]:

```
kl_mall.head()
```

Out[42]:

	Neighborhoods	Shopping Mall
0	Alam Damai	0.000000
1	Ampang, Kuala Lumpur	0.030000
2	Bandar Menjalara	0.010000
3	Bandar Sri Permaisuri	0.000000
4	Bandar Tasik Selatan	0.010101

7. Cluster Neighborhoods

Run k-means to cluster the neighborhoods in Kuala Lumpur into 3 clusters.

In [43]:

```
# set number of clusters
kclusters = 3

kl_clustering = kl_mall.drop(["Neighborhoods"], 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(kl_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

Out[43]:

```
array([2, 1, 0, 2, 0, 2, 1, 1, 0, 2], dtype=int32)
```

In [44]:

```
# create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.
kl_merged = kl_mall.copy()

# add clustering labels
kl_merged["Cluster Labels"] = kmeans.labels_
```

In [45]:

```
kl_merged.rename(columns={"Neighborhoods": "Neighborhood"}, inplace=True)
kl_merged.head()
```

Out[45]:

	Neighborhood	Shopping Mall	Cluster Labels
0	Alam Damai	0.000000	2
1	Ampang, Kuala Lumpur	0.030000	1
2	Bandar Menjalara	0.010000	0
3	Bandar Sri Permaisuri	0.000000	2
4	Bandar Tasik Selatan	0.010101	0

In [46]:

```
# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
kl_merged = kl_merged.join(kl_df.set_index("Neighborhood"), on="Neighborhood")

print(kl_merged.shape)
kl_merged.head() # check the last columns!
```

(71, 5)

Out[46]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
0	Alam Damai	0.000000	2	3.057690	101.743880
1	Ampang, Kuala Lumpur	0.030000	1	3.153153	101.700413
2	Bandar Menjalara	0.010000	0	3.190350	101.625450
3	Bandar Sri Permaisuri	0.000000	2	3.103910	101.712260
4	Bandar Tasik Selatan	0.010101	0	3.072620	101.714710

In [47]:

```
# sort the results by Cluster Labels
print(kl_merged.shape)
kl_merged.sort_values(["Cluster Labels"], inplace=True)
kl_merged
```

(71, 5)

Out[47]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
70	Wangsa Maju	0.010000	0	3.203910	101.737190
21	Damansara Town Centre	0.010000	0	3.138759	101.684046
22	Damansara, Kuala Lumpur	0.010000	0	3.138759	101.684046
23	Dang Wangi	0.020000	0	3.156685	101.698076
49	Setiawangsa	0.010000	0	3.191803	101.740070
27	Jalan Cochrane, Kuala Lumpur	0.020000	0	3.132977	101.724669
28	Jalan Duta	0.010000	0	3.180163	101.677880
31	Kampung Baru, Kuala Lumpur	0.020000	0	3.165460	101.710280
32	Kampung Datuk Keramat	0.010000	0	3.166400	101.730460
69	Titiwangsa	0.010000	0	3.180670	101.703220
57	Taman Desa	0.010000	0	3.102970	101.684710
38	Medan Tuanku	0.020000	0	3.159260	101.698340

43	Putrajaya	0.020000	0	3.125862	101.718624
45	Segambut	0.020000	0	3.186390	101.668100
53	Sungai Besi	0.010000	0	3.050640	101.706130
51	Sri Hartamas	0.020000	0	3.162200	101.650360
20	Damansara Heights	0.020000	0	3.147980	101.667980
19	Chow Kit	0.020000	0	3.163590	101.698110
25	Federal Hill, Kuala Lumpur	0.010000	0	3.136370	101.685640
17	Bukit Tunku	0.020000	0	3.173810	101.682760
2	Bandar Menjalara	0.010000	0	3.190350	101.625450
4	Bandar Tasik Selatan	0.010101	0	3.072620	101.714710
8	Bangsar South	0.020000	0	3.111020	101.662830
10	Batu, Kuala Lumpur	0.020000	0	3.135760	101.708370
12	Bukit Bintang	0.020000	0	3.147770	101.708550
50	Shamelin	0.020000	0	3.124580	101.735970
64	Taman Sri Sinar	0.010309	0	3.190070	101.652930
16	Bukit Petaling	0.010000	0	3.129290	101.698920
67	Taman U-Thant	0.030000	1	3.157700	101.724520
11	Brickfields	0.030000	1	3.129160	101.684060
1	Ampang, Kuala Lumpur	0.030000	1	3.153153	101.700413
30	KL Eco City	0.030000	1	3.117140	101.673890
66	Taman Tun Dr Ismail	0.030000	1	3.152830	101.622710
7	Bangsar Park	0.050000	1	3.134780	101.672620
36	Lembah Pantai	0.040000	1	3.121202	101.663899
6	Bangsar	0.050000	1	3.129200	101.678440
15	Bukit Nanas	0.030000	1	3.152017	101.701028
40	Mont Kiara	0.030000	1	3.165320	101.652430
42	Pudu, Kuala Lumpur	0.040000	1	3.133540	101.713070
52	Sri Petaling	0.000000	2	3.072600	101.682520
68	Taman Wahyu	0.000000	2	3.222400	101.671730
54	Taman Bukit Maluri	0.000000	2	3.200660	101.633370
61	Taman Midah	0.000000	2	3.093590	101.728370
56	Taman Connaught	0.000000	2	3.082690	101.736890
62	Taman OUG	0.000000	2	3.210050	101.634508
63	Taman P. Ramlee	0.000000	2	3.193600	101.705980
59	Taman Len Seng	0.000000	2	3.069080	101.742870
60	Taman Melati	0.000000	2	3.223570	101.723990
65	Taman Taynton View	0.000000	2	3.087070	101.736810
55	Taman Cheras Hartamas	0.000000	2	3.082630	101.746710
58	Taman Ibukota	0.000000	2	3.212160	101.715400
0	Alam Damai	0.000000	2	3.057690	101.743880
47	Sentul Raya	0.000000	2	3.187431	101.691453
3	Bandar Sri Permaisuri	0.000000	2	3.103910	101.712260
5	Bandar Tun Razak	0.000000	2	3.082800	101.722810
9	Batu 11 Cheras	0.000000	2	3.098980	101.734990
13	Bukit Jalil	0.000000	2	3.057800	101.689650
14	Bukit Kiara	0.000000	2	3.143480	101.644330
18	Cheras, Kuala Lumpur	0.000000	2	3.061870	101.746750
24	Desa Petaling	0.000000	2	3.083310	101.704380
48	Setapak	0.000000	2	3.188160	101.704150
26	Happy Garden	0.000000	2	3.201630	101.721070
33	Kampung Padang Balang	0.000000	2	3.209430	101.693180

34	Kepong	0.000000	2	3.217500	101.637630
37	Maluri	0.000000	2	3.147890	101.694050
39	Miharja	0.000000	2	3.147890	101.694050
41	Pantai Dalam	0.000000	2	3.094760	101.667470
44	Salak South	0.000000	2	3.081020	101.697240
46	Semarak	0.000000	2	3.179927	101.721442
29	Jinjang	0.000000	2	3.209500	101.658740
35	Kuchai Lama	0.000000	2	3.090740	101.677330

Finally, let's visualize the resulting clusters

In [48]:

```
# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i+x+(i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(kl_merged['Latitude'], kl_merged['Longitude'], kl_merged['Neighborhood'],
kl_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' - Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[48]:

In [49]:

```
# save the map as HTML file
map_clusters.save('map_clusters.html')
```

8. Examine Clusters

Cluster 0

In [50]:

```
kl_merged.loc[kl_merged['Cluster Labels'] == 0]
```

Out[50]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
70	Wangsa Maju	0.010000	0	3.203910	101.737190
21	Damansara Town Centre	0.010000	0	3.138759	101.684046
22	Damansara, Kuala Lumpur	0.010000	0	3.138759	101.684046
23	Dang Wangi	0.020000	0	3.156685	101.698076
49	Setiawangsa	0.010000	0	3.191803	101.740070
27	Jalan Cochrane, Kuala Lumpur	0.020000	0	3.132977	101.724669
28	Jalan Duta	0.010000	0	3.180163	101.677880
31	Kampung Baru, Kuala Lumpur	0.020000	0	3.165460	101.710280
32	Kampung Datuk Keramat	0.010000	0	3.166400	101.730460
69	Titiwangsa	0.010000	0	3.180670	101.703220
57	Taman Desa	0.010000	0	3.102970	101.684710
38	Medan Tuanku	0.020000	0	3.159260	101.698340
43	Putrajaya	0.020000	0	3.125862	101.718624
45	Segambut	0.020000	0	3.186390	101.668100
53	Sungai Besi	0.010000	0	3.050640	101.706130
51	Sri Hartamas	0.020000	0	3.162200	101.650360
20	Damansara Heights	0.020000	0	3.147980	101.667980
19	Chow Kit	0.020000	0	3.163590	101.698110
25	Federal Hill, Kuala Lumpur	0.010000	0	3.136370	101.685640
17	Bukit Tunku	0.020000	0	3.173810	101.682760
2	Bandar Menjalara	0.010000	0	3.190350	101.625450
4	Bandar Tasik Selatan	0.010101	0	3.072620	101.714710
8	Bangsar South	0.020000	0	3.111020	101.662830
10	Batu, Kuala Lumpur	0.020000	0	3.135760	101.708370
12	Bukit Bintang	0.020000	0	3.147770	101.708550
50	Shamelin	0.020000	0	3.124580	101.735970
64	Taman Sri Sinar	0.010309	0	3.190070	101.652930
16	Bukit Petaling	0.010000	0	3.129290	101.698920

Cluster 1

In [51]:

```
kl_merged.loc[kl_merged['Cluster Labels'] == 1]
```

Out[51]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
67	Taman U-Thant	0.03	1	3.157700	101.724520
11	Brickfields	0.03	1	3.129160	101.684060
1	Ampang, Kuala Lumpur	0.03	1	3.153153	101.700413
30	KL Eco City	0.03	1	3.117140	101.673890
66	Taman Tun Dr Ismail	0.03	1	3.152830	101.622710
7	Bangsar Park	0.05	1	3.134780	101.672620
36	Lembah Pantai	0.04	1	3.121202	101.663899
6	Bangsar	0.05	1	3.129200	101.678440
15	Bukit Nanas	0.03	1	3.152017	101.701028
40	Mont Kiara	0.03	1	3.165320	101.652430
42	Pudu, Kuala Lumpur	0.04	1	3.133540	101.713070

Cluster 2

In [52]:

```
kl_merged.loc[kl_merged['Cluster Labels'] == 2]
```

Out[52]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
52	Sri Petaling	0.0	2	3.072600	101.682520
68	Taman Wahyu	0.0	2	3.222400	101.671730
54	Taman Bukit Maluri	0.0	2	3.200660	101.633370
61	Taman Midah	0.0	2	3.093590	101.728370
56	Taman Connaught	0.0	2	3.082690	101.736890
62	Taman OUG	0.0	2	3.210050	101.634508
63	Taman P. Ramlee	0.0	2	3.193600	101.705980
59	Taman Len Seng	0.0	2	3.069080	101.742870
60	Taman Melati	0.0	2	3.223570	101.723990
65	Taman Taynton View	0.0	2	3.087070	101.736810
55	Taman Cheras Hartamas	0.0	2	3.082630	101.746710
58	Taman Ibukota	0.0	2	3.212160	101.715400
0	Alam Damai	0.0	2	3.057690	101.743880
47	Sentul Raya	0.0	2	3.187431	101.691453
3	Bandar Sri Permaisuri	0.0	2	3.103910	101.712260
5	Bandar Tun Razak	0.0	2	3.082800	101.722810
9	Batu 11 Cheras	0.0	2	3.098980	101.734990
13	Bukit Jalil	0.0	2	3.057800	101.689650
14	Bukit Kiara	0.0	2	3.143480	101.644330
18	Cheras, Kuala Lumpur	0.0	2	3.061870	101.746750
24	Desa Petaling	0.0	2	3.083310	101.704380
48	Setapak	0.0	2	3.188160	101.704150
26	Happy Garden	0.0	2	3.201630	101.721070
33	Kampung Padang Balang	0.0	2	3.209430	101.693180
34	Kepong	0.0	2	3.217500	101.637630
37	Maluri	0.0	2	3.147890	101.694050
39	Miharja	0.0	2	3.147890	101.694050
41	Pantai Dalam	0.0	2	3.094760	101.667470
44	Salak South	0.0	2	3.081020	101.697240
46	Semarak	0.0	2	3.179927	101.721442
29	Jinjang	0.0	2	3.209500	101.658740
35	Kuchai Lama	0.0	2	3.090740	101.677330

Result

Categorized the neighbourhoods into 3 clusters:

- Cluster 0: Neighbourhoods with moderate number of shopping malls
- Cluster 1: Neighbourhoods with low number to no existence of shopping malls
- Cluster 2: Neighbourhoods with high concentration of shopping malls

Discussion

Most of the shopping malls are concentrated in the central area of Kuala Lumpur city, with the highest number in cluster 2 and moderate number in cluster 0. On the other hand, cluster 1 has very low number to totally no shopping mall in the neighborhoods. This represents a great opportunity and high potential areas to open new shopping malls as there is very little to no competition from existing malls. Meanwhile, shopping malls in cluster 2 are likely suffering from intense competition due to oversupply and high concentration of shopping malls. From another perspective, this also shows that the oversupply of shopping malls mostly happened in the central area of the city, with the suburb area still have very few shopping malls. Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighborhoods in cluster 1 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighborhoods in cluster 0 with moderate competition. Lastly, property developers are advised to avoid neighborhoods in cluster 2 which already have high concentration of shopping malls and suffering from intense competition.

Conclusion

- Answer to business question: The neighbourhoods in cluster 1 are the most preferred locations to open a new shopping mall
- Findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new shopping mall

In [ ]:

```

```