IBM Applied Data Science Capstone Course by Coursera

Week 5 Final Report

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 [1]:

```
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 an
d 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 panda
s 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 [2]:

```
# send the GET request
```

```
data = requests.get("https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur").text
```

In [3]:

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

In [4]:

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

In [5]:

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

In [6]:

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

kl_df.head()
```

Out[6]:

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

In [7]:

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

Out[7]:

```
(70, 1)
```

3. Get the geographical coordinates

In [10]:

```
# 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(neighborhoo
d))
        lat_lng_coords = g.latlng
    return lat_lng_coords
```

In [12]:

```
# 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 [13]:

```
coords
```

Out[13]:

```
[[3.0576900000000364, 101.74388000000005],
 [3.158092785454847, 101.74870260727259],
 [3.1903500000000236, 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],
 [3.09898000000004, 101.73499000000004],
 [3.12916000000007, 101.68406000000004],
 [3.147770000000037, 101.70855000000006],
 [3.057800000000043, 101.68965000000009],
 [3.1434800000000678, 101.64433000000008],
 [3.1520079194122164, 101.70102829894834],
 [3.129290000000026, 101.69892000000004],
 [3.17381000000006, 101.68276000000009],
 [3.061870000000056, 101.74675000000008],
 [3.163590000000056, 101.69811000000004],
 [3.147980000000075, 101.66798000000006],
 [3.1413156714557724, 101.62607997785557],
 [3.1413156714557724, 101.62607997785557],
 [3.155238081462032, 101.70152004297655],
 [3.08331000000004, 101.70438000000007],
 [3.1363700000000563, 101.68564000000003],
 [3.201630000000023, 101.72107000000005],
 [3.132961658243644, 101.72470311855207],
 [3.179388043398946, 101.67745432110556],
 [3.209500000000048, 101.65874000000008],
```

```
 [3.165460000000053, 101.71028000000007],
 [3.166400000000067, 101.73046000000005],
 [3.2175000000000296, 101.63763000000006],
 [3.117110000000025, 101.67349000000007],
 [3.0907400000000393, 101.67733000000004],
 [3.121216228040549, 101.66389741147111],
 [3.147890000000075, 101.69405000000006],
 [3.1592600000000743, 101.69834000000009],
 [3.147890000000075, 101.69405000000006],
 [3.165320000000065, 101.65243000000004],
 [3.0951600000000212, 101.66812000000004],
 [3.133540000000039, 101.71307000000007],
 [3.111867031417114, 101.71326779181086],
 [3.0810200000000236, 101.69724000000008],
 [3.186390000000074, 101.66810000000004],
 [3.1803929473666943, 101.7234139398852],
 [3.175374881541043, 101.69303426939257],
 [3.188160000000039, 101.70415000000008],
 [3.1784341541269487, 101.73701342752202],
 [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],
 [3.2100374598740182, 101.63449832484254],
 [3.1936000000000604, 101.70598000000007],
 [3.1900700000000484, 101.65293000000008],
 [3.0870700000000397, 101.73681000000005],
 [3.1528300000000513, 101.62271000000004],
 [3.1576500000000465, 101.72502000000009],
 [3.22240000000005, 101.67173000000008],
 [3.1806700000000205, 101.70322000000004],
 [3.2039100000000644, 101.73719000000006]]
```

In [14]:

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

In [15]:

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

```
kl_df['Longitude'] = df_coords['Longitude']
```

In [16]:

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

```
(70, 3)
```

Out[16]:

	Neighborhood	Latitude	Longitude
0	Alam Damai	3.05769 0	101.74388 0
1	Ampang, Kuala Lumpur	3.15809 3	101.74870 3
2	Bandar Menjalara	3.19035 0	101.62545 0
3	Bandar Sri Permaisuri	3.1039	101.71226 0

	Neighborhood	**Latitude**	**Longitude**
		10	
4	Bandar Tasik Selatan	3.07262 0	101.71471 0
5	Bandar Tun Razak	3.08280 0	101.72281 0
6	Bangsar	3.12920 0	101.67844 0
7	Bangsar Park	3.13478 0	101.67262 0

	Neighborhood	**Latitude**	**Longitude**
8	Bangsar South	3.11102 0	101.662830
9	Batu 11 Cheras	3.09890 80	101.734990
10	Batu, Kuala Lumpur	3.09890 80	101.734990
11	Brickfields	3.12916 0	101.684060
12	Bukit Bintang	3.147	101.70855

	Neighborhood	**Latitude**	**Longitude**
		770	0
13	Bukit Jalil	3.05780 0	101.689650
14	Bukit Kiara	3.14348 0	101.644330
15	Bukit Nanas	3.15200 8	101.701028
16	Bukit Petaling	3.12929 0	101.698920

	Neighborhood	**Latitude**	**Longitude**
17	Bukit Tunku	3.17381 0	101.682760
18	Cheras, Kuala Lumpur	3.06187 0	101.746750
19	Chow Kit	3.16359 0	101.698110
20	Damansara Heights	3.14798 0	101.667980
21	Damansara Town	3.141	101.62608

	Neighborhood	**Latitude**	**Longitude**
	Centre	3 1 6	0
2 2	Damansara, Kuala Lumpur	3. 1 4 1 3 1 6	101 .62 608 0
2 3	Dang Wangi	3. 1 5 5 2 3 8	101 .70 152 0
2 4	Desa Petaling	3. 0 8 3 3 1 0	101 .70 438 0
2 5	Federal Hill, Kuala Lumpur	3. 1 3 6 3 7 0	101 .68 564 0

	Neighborhood	**Latitude**	**Longitude**
26	Happy Garden	3.20163 0	101.721070
27	Jalan Cochrane, Kuala Lumpur	3.132962	101.724703
28	Jalan Duta	3.17938 8	101.677454
29	Jinjang	3.20950 0	101.658740
30	Kampung Baru, Kuala	3.165	101.71028

	Neighborhood	**Latitude**	**Longitude**
	Lumpur	460	0
31	Kampung Datuk Keramat	3.16640 0	101.730460
32	Kepong	3.21750 0	101.637630
33	KL Eco City	3.11710	101.673490
34	Kuchai Lama	3.09074 0	101.677330

	Neighborhood	**Latitude**	**Longitude**
35	Lembah Pantai	3.12121 6	101.66389 7
36	Maluri	3.14789 0	101.69405 0
37	Medan Tuanku	3.15926 0	101.69834 0
38	Miharja	3.14789 0	101.69405 0
39	Mont Kiara	3.165	101.65243

	Neighborhood	**Latitude**	**Longitude**
		3 2 0	0
4 0	Pantai Dalam	3. 0 9 5 1 6 0	101 .66 812 0
4 1	Pudu, Kuala Lumpur	3. 1 3 3 5 4 0	101 .71 307 0
4 2	Putrajaya	3. 1 1 1 8 6 7	101 .71 326 8
4 3	Salak South	3. 0 8 1 0 2 0	101 .69 724 0

	Neighborhood	**Latitude**	**Longitude**
44	Segambut	3.186390	101.668100
45	Semarak	3.180393	101.723414
46	Sentul Raya	3.175375	101.693034
47	Setapak	3.188160	101.704150
48	Setiawangsa	3.178	101.73701

	Neighborhood	**Latitude**	**Longitude**
		434	3
49	Shamelin	3.12458 0	101.73597 0
50	Sri Hartamas	3.16220 0	101.65036 0
51	Sri Petaling	3.07260 0	101.68252 0
52	Sungai Besi	3.05064 0	101.70613 0

	Neighborhood	**Latitude**	**Longitude**
53	Taman Bukit Maluri	3.20066 0	101.633370
54	Taman Cheras Hartamas	3.08263 0	101.746710
55	Taman Connaught	3.08269 0	101.736890
56	Taman Desa	3.10297 0	101.684710
57	Taman Ibukot	3.212	101.71540

	Neighborhood	**Latitude**	**Longitude**
	a	160	0
58	Taman Len Seng	3.069080	101.742870
59	Taman Melati	3.223570	101.723990
60	Taman Midah	3.093590	101.728370
61	Taman OUG	3.210037	101.634498

	Neighborhood	Latitude	Longitude
62	Taman P. Ramlee	3.19360 0	101.705980
63	Taman Sri Sinar	3.19000 70	101.65293 0
64	Taman Taynton View	3.08707 0	101.73681 0
65	Taman Tun Dr Ismail	3.15283 0	101.62271 0
66	Taman U-Thant	3.157	101.72502

	Neighborhood	Latitude	Longitude
		650	0
67	Taman Wahyu	3.22240 0	101.671730
68	Titiwangsa	3.18067 0	101.703220
69	Wangsa Maju	3.20391 0	101.737190

In [17]:

```
# 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 [18]:

```
# 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 geograpical coordinate of Kuala Lumpur, Malaysiae {}, {}.'.forma
t(latitude, longitude))
```

```
The geograpical coordinate of Kuala Lumpur, Malaysiae 3.1516636, 101.694302
8.
```

In [19]:

```
# 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[19]:

In [77]:

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

5. Use the Foursquare API to explore the neighborhoods

In [1]:

```
# define Foursquare Credentials and Version
CLIENT_ID = 'your Foursquare ID' # your Foursquare ID
CLIENT_SECRET = 'your Foursquare Secret' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version

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

```
Your credentails:
CLIENT_ID: your Foursquare ID
CLIENT_SECRET:your Foursquare Secret
```

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

In [21]:

```
radius = 2000
```

```
LIMIT = 100

venues = []

for lat, long, neighborhood in zip(kl_df['Latitude'], kl_df['Longitude'], k
l_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 [22]:

```
# 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()
```

```
(6970, 7)
```

Out[22]:

	Neighborhood		Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
	Alam Damai		101.74388	Machi Noodle 如子面	3.057695	101.746635	Noodle House
	Alam Damai		101.74388	Pengedar Shakl	3.061235	101.740696	Supplement Sho

	Neighborhood		Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
				ee KL | COD Vivix Shaklee Kuala .			p

	Neighborhood		**Longitude**	**VenueName**	**VenueLatitude**	**VenueLongitude**	**VenueCategory**
				. .			
	Alam Damai		101.74388	628火焰鍋茶室	3.05844 2	101.747947	Chinese Restaurant
	Alam Dama		101.7438	Ivy Sekin	3.065749	101.748711	Noodle Ho

	Neighborhood		Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
	i		8	chan Seafood Noodle House 运利特制鱼		8	use

	Neighborhood		Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
				大海鮮面			
	Alam Damai		101.7438	Restoran Ikbal	3.061134	101.750220	Restaurant

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

In [23]:

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

Out[23]:

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Alam Damai	100	100	100	100	100	100
Ampang	100	100	100	100	100	100

	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
Neighborhood						
, Kuala Lumpur						
Ban	10	10	10	10	10	10

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
dar Menjalara	0	0	0	0	0	0
Band	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
ar Sri Permaisuri						

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Bandar Tasik Selat	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
an						
Bandar Tun Raza	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
k						
Bangsar	100	100	100	100	100	100
Bangsar	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Park						
Bangsar South	100	100	100	100	100	100

	Latitude	**Longitude**	**VenueName**	**VenueLatitude**	**VenueLongitude**	**VenueCategory**
Neighborhood						
h						
Batu 11 Cheras	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Batu, Kuala Lumpu	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
r						
Brickfields	100	100	100	100	100	100
Buk	10	10	10	10	10	10

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
it Bintang	0	0	0	0	0	0
Bukit J	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
alil						
Bukit Kiara	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Bukit Nanas	100	100	100	100	100	100
Bukit	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Petaling						
Bukit Tu	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
nku						
Cheras, Kuala	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Lumpur						
Chow Kit	100	100	100	100	100	100
D	1	1	1	1	1	1

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
amansara Heights	00	00	00	00	00	00
D	1	1	1	1	1	1

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
amansara Town Centr	0 0	0 0	0 0	0 0	0 0	0 0

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
e						
Damansara, Kuala	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Lumpur						
Dang Wang	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
i						
Desa Petaling	100	100	100	100	100	100
F	1	1	1	1	1	1

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
ederal Hill, Kuala	0 0	0 0	0 0	0 0	0 0	0 0

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Lumpur						
Happy Garde	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
n						
Jalan Cochrane,	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Kuala Lumpur						
Jalan	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Duta						
Jinjang	100	100	100	100	100	100
KL	10	10	10	10	10	10

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Eco City	0	0	0	0	0	0
Kampung B	100	100	100	100	100	100

	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
Neighborhood						
aru, Kuala Lumpur						

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Kampung Datuk Kera	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
mat						
Kepong	100	100	100	100	100	100
Kuchai	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Lama						
Lembah Pantai	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
i						
Maluri	100	100	100	100	100	100
Medan Tu	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
ankU						
Mihar ja	100	100	100	100	100	100
Mont	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Kiara						
Pantai Dalla	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
m						
Pudu, Kuala Lump	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
ur						
Putrajaya	78	78	78	78	78	78
Sala	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
k South						
Segambut	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Semarak	100	100	100	100	100	100
Sentul Ra	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
ya						
Setapak	100	100	100	100	100	100
Setiaw	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
angsa						
Shamelin	100	100	100	100	100	100
Sr	10	10	10	10	10	10

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
i Hartamas	0	0	0	0	0	0
Sri Pet	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
aling						
Sungai Besi	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Taman Bukit Malur	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
i						
Taman Cheras Har	100	100	100	100	100	100

	Latitude	**Longitude**	**VenueName**	**VenueLatitude**	**VenueLongitude**	**VenueCategory**
Neighborhood						
tamas						
Tamaan Connnau	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
ght						
Taman Desa	100	100	100	100	100	100
Ta	10	10	10	10	10	10

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
man Ibukota	0	0	0	0	0	0
Taman	99	99	99	99	99	99

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Len Seng						
Taman Mel	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
ati						
Taman Midah	100	100	100	100	100	100
T	1	1	1	1	1	1

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
aman OUG	00	00	00	00	00	00
Taman P.	93	93	93	93	93	93

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Ramlee						
Taman Sri S	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
inar						
Taman Taynton	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
View						
Taman Tun Dr	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Ismail						
Taman U-Th	100	100	100	100	100	100

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
ant						
Taman Wahyu	100	100	100	100	100	100
T	1	1	1	1	1	1

	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
itiwangsa	00	00	00	00	00	00
Wangsa M	100	100	100	100	100	100

	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
Neighborhood						
aju						

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

In [24]:

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

```
There are 304 uniques categories.
```

In [25]:

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

Out[25]:

```
array(['Noodle House', 'Supplement Shop', 'Chinese Restaurant',
       'Restaurant', 'Food Court', 'Vegetarian / Vegan Restaurant',
       'Asian Restaurant', 'Dim Sum Restaurant', 'Snack Place',
       'Other Great Outdoors', 'Seafood Restaurant', 'Spa', 'Food Truck',
       'Café', 'Park', 'Chinese Breakfast Place', 'Indian Restaurant',
```

```
       'Japanese Restaurant', 'Outlet Store', 'Convenience Store',
       'Bubble Tea Shop', 'Dessert Shop', 'Farmers Market',
       'Cantonese Restaurant', 'Malay Restaurant', 'Bakery',
       'Hakka Restaurant', 'Supermarket', 'Steakhouse', 'Pet Store',
       'Middle Eastern Restaurant', 'Badminton Court',
       'Athletics & Sports', 'Hookah Bar', 'Winery', 'Burger Joint',
       'Gym / Fitness Center', 'Bistro', 'Grocery Store',
       'Halal Restaurant', 'College Bookstore', 'Flea Market',
       'Vietnamese Restaurant', 'Italian Restaurant', 'Coffee Shop',
       'Juice Bar', 'Korean Restaurant', 'Pizza Place', 'Ice Cream Shop',
       'Massage Studio'], dtype=object)
```

In [42]:

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

Out[42]:

```
True
```

6. Analyze Each Neighborhood

In [43]:

```
# 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()
```

```
(6970, 305)
```

Out[43]:

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

In [47]:

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

print(kl_grouped.shape)
kl_grouped
```

```
(70, 305)
```

Out[47]:

In [48]:

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

Out[48]:

```
39
```

Create a new DataFrame for Shopping Mall data only

In [52]:

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

In [53]:

```
kl_mall.head()
```

Out[53]:

	Neighborhoods	Shopping Mall

	Neighborhoods	Shopping Mall
0	Alam Damai	0.00
1	Ampang, Kuala Lumpur	0.00
2	Bandar Menjalara	0.00
3	Bandar Sri Permaisuri	0.00
4	Bandar Tasik Selatan	0.01

7. Cluster Neighborhoods

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

In [79]:

```
# 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[79]:

```
array([1, 1, 1, 1, 1, 1, 2, 2, 0, 1])
```

In [80]:

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

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

In [81]:

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

```
kl_merged.head()
```

Out[81]:

	Neighborhood	Shopping Mall	Cluster Labels
0	Alam Damai	0.00	1
1	Ampang, Kuala Lumpur	0.00	1
2	Bandar Menjalara	0.00	1
3	Bandar Sri Permaisuri	0.00	1
4	Bandar Tasik Selatan	0.01	1

In [82]:

```
# 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")
```

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

```
(70, 5)
```

Out[82]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
0	Alam Damai	0.00	1	3.057690	101.743880
1	Ampang, Kuala Lu	0.00	1	3.158093	101.748703

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	mpur				
2	Bandar Menjalara	0.00	1	3.190350	101.625450
3	Bandar Sr	0.00	1	3.103910	101.71226

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	i Permaisuri				0
4	Bandar Tasik Selata	0.01	1	3.072620	101.714710

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	n				

In [83]:

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

```
(70, 5)
```

Out[83]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
19	Chow K	0.03	0	3.1635	101.69

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	it			90	8110
37	Medan Tuanku	0.02	0	3.159260	1011.698340
44	Segambut	0.02	0	3.186390	1011.668100

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
31	Kampung Baru, Kuala Lumpur	0.03	0	3.165460	101.710280
48	Seti	0.0	0	3.17	101.

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	awangsa	2		8434	737013
49	Shamelin	0.02	0	3.124580	1011.735970
28	Jalan Duta	0.02	0	3.179388	1011.677454

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
27	Jalan Cochrane, Kuala Lumpur	0.02	0	3.132962	101.724703
2	Da	0.	0	3.	10

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
3	ng Wangi	02		155238	1.701520
22	Damansara, Kuala Lumpu	0.03	0	3.141316	101.626080

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	r				
21	Damansara Town Centre	0.03	0	3.141316	101.626080
20	Damansa	0.02	0	3.14798	101.667

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	ra Heights			0	980
17	Bukit Tunku	0.02	0	3.173810	101.682760
15	Bukit Na	0.03	0	3.15200	101.7010

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	nas			8	28
39	Mont Kiara	0.02	0	3.165320	101.652430
56	Taman Desa	0.02	0	3.102970	101.684710
12	Buki	0.0	0	3.14	101.

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	t Bintang	3		7770	708550
65	Taman Tun Dr Ismail	0.03	0	3.15283 0	101.62271 0

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
8	Bangsar South	0.03	0	3.111020	101.662830
67	Taman Wahyu	0.00	1	3.222400	101.671730
43	Sala	0.0	1	3.08	101.

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	k South	0		10.20	69.7240
45	Semarak	0.00	1	3.180393	101.723414
46	Sentul Raya	0.01	1	3.175375	101.693034

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
47	Setapak	0.00	1	3.188160	101.704150
64	Taman Taynton View	0.00	1	3.087070	101.736810

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
63	Taman Sri Sinar	0.01	1	3.190070	101.652930
62	Taman P. Ramle	0.00	1	3.193600	101.705980

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	e				
61	Taman OUG	0.00	1	3.210037	101.634498
60	Taman Midah	0.00	1	3.093590	101.728370
50	Sri H	0.00	1	3.162	101.6

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	artamas			200	50360
51	Sri Petaling	0.00	1	3.07260 0	101.68252 0
52	Sungai Bes	0.01	1	3.05064 0	101.70613

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	i				0
59	Taman Melati	0.00	1	3.223570	101.723990
58	Taman Len Seng	0.00	1	3.069080	101.742870

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
57	Taman Ibukota	0.00	1	3.21216 0	101.71540 0
42	Putrajaya	0.00	1	3.11186 7	101.71326 8
53	Taman	0.00	1	3.200	101.6

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	Bukit Maluri			6.60	33.370
54	Taman Cheras Harta	0.00	1	3.08263	101.74671

	Neighborhood	**Shopping Mall**	**Cluster Labels**	**Latitude**	**Longitude**
	mas				
55	Taman Connaught	0.00	1	3.08269 0	101.736890
0	Alam Damai	0.00	1	3.05769 0	101.743880

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
34	Kuchai Lama	0.00	1	3.090740	101.677330
14	Bukit Kiara	0.00	1	3.143480	101.644330
1	Ampang	0.00	1	3.1580	101.74

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	, Kuala Lumpur			93	8703
2	Bandar Menjalara	0.00	1	3.190350	101.625450

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
3	Bandar Sri Permaisuri	0.00	1	3.103910	101.712260
4	Bandar Ta	0.01	1	3.072620	101.71471

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	sik Selatan				0
5	Bandar Tun Razak	0.00	1	3.08280	101.72810
9	Ba	0.	1	3.	10

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	tu 11 Cheras	00		0.98980	1.73490
10	Batu, Kuala Lumpur	0.00	1	3.09880	101.73490

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
13	Bukit Jalil	0.00	1	3.057800	101.689650
40	Pantai Dalam	0.01	1	3.095160	101.668120
16	Bukit	0.01	1	3.129	101.6

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	Petaling			290	98920
18	Cheras, Kuala Lumpur	0.00	1	3.061870	101.746750

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
69	Wangsa Maju	0.01	1	3.203910	101.737190
25	Federal Hill, Kuala	0.01	1	3.136370	101.685640

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	Lumpur				
26	Happy Garden	0.01	1	3.201630	101.721070
29	Jinjang	0.00	1	3.209500	101.65874

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
					0
32	Kampung Datuk Keramat	0.01	1	3.16640 0	101.730460
33	Kepong	0.00	1	3.21750	101.637

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
				0	630
68	Titiwangsa	0.01	1	3.180670	101.703220
36	Maluri	0.01	1	3.147890	101.694050
38	Miha	0.0	1	3.14	101.

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	rja	1		7890	694050
24	Desa Petaling	0.00	1	3.083310	101.704380
11	Brickfield	0.05	2	3.129160	101.68406

	Neighborhoods	Shopping Mall	Cluster Labels	Latitude	Longitude
	s				0
30	KL Eco City	0.05	2	3.117110	101.673490
7	Bangsar Park	0.04	2	3.134780	101.672620
6	Ban	0.0	2	3.1	101

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	gsar	5		29200	.678440
35	Lembah Pantai	0.05	2	3.12121 6	101.66389 7
66	Taman U-T	0.05	2	3.15765 0	101.72502

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	hant				0
41	Pudu, Kuala Lumpur	0.05	2	3.133540	101.713070

Finally, let's visualize the resulting clusters

In [91]:

```
# 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['Longitu
de'], kl_merged['Neighborhood'], kl_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' - Cluster ' + str(cluster), parse_htm
l=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[91]:

In [92]:

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

8. Examine Clusters

Cluster 0

In [93]:

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

Out[93]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
1	Ch	0.	0	3.	10

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
9	ow Kit	03		16359 0	1.69811 0
37	Medan Tuanku	0.02	0	3.15926 0	101.69834 0
44	Segambut	0.02	0	3.18639 0	101.66810

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
					0
31	Kampung Baru, Kuala Lumpur	0.03	0	3.16546 0	101.71028 0
4	Se	0.	0	3.	10

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
8	tiawangsa	02		178434	1.737013
49	Shamelin	0.02	0	3.124580	101.735970
28	Jalan Dut	0.02	0	3.179388	101.67745

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	a				4
27	Jalan Cochrane, Kuala Lumpur	0.02	0	3.132962	101.724703

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
23	Dang Wangi	0.02	0	3.155238	101.701520
22	Damansara, Kuala Lum	0.03	0	3.141316	101.626080

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	pur				
21	Damansara Town Centre	0.03	0	3.14131 6	101.626080
20	Daman	0.02	0	3.147	101.6

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	sara Heights			980	67980
17	Bukit Tunku	0.02	0	3.17381 0	1011.68276 0
15	Bukit	0.03	0	3.1520	1011.70

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	Nanas			08	1028
39	Mont Kiara	0.02	0	3.16532 0	101.65243 0
56	Taman Desa	0.02	0	3.10297 0	101.68471 0
1	Bu	0.	0	3.	10

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
2	kit Bintang	03		147770	1.708550
65	Taman Tun Dr Ismail	0.03	0	3.152830	101.622710

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
8	Bangsar South	0.03	0	3.111020	101.662830

Cluster 1

In [94]:

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

Out[94]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
67	Taman Wahyu	0.00	1	3.22400	101.67173 0
43	Salak South	0.00	1	3.08102 0	101.69724 0
45	Semara	0.00	1	3.1803	101.72

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	k			93	3414
46	Sentul Raya	0.01	1	3.175375	101.693034
47	Setapak	0.00	1	3.188160	101.704150
6	Ta	0.	1	3.	10

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
4	man Taynton View	00		0870700	1.736810
63	Taman Sri Sina	0.01	1	3.190070	101.652930

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	r				
62	Taman P. Ramlee	0.00	1	3.193600	101.705980
61	Taman OUG	0.00	1	3.210037	101.634498

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
60	Taman Midah	0.00	1	3.093590	101.728370
50	Sri Hartamas	0.00	1	3.162200	101.650360
51	Sri P	0.00	1	3.072	101.6

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	etaling			600	82520
52	Sungai Besi	0.01	1	3.05064 0	101.706130
59	Taman Melat	0.00	1	3.223570	101.723990

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	i				
58	Taman Len Seng	0.00	1	3.069080	101.742870
57	Taman Ibukota	0.00	1	3.212160	101.715400

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
42	Putrajaya	0.00	1	3.111867	101.713268
53	Taman Bukit Maluri	0.00	1	3.200660	101.633370

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
54	Taman Cheras Hartamas	0.00	1	3.08263 0	101.746710
55	Taman Con	0.00	1	3.08269 0	101.73689

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	naught				0
0	Alam Damai	0.00	1	3.057690	101.743880
34	Kuchai Lama	0.00	1	3.090740	101.677330

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
14	Bukit Kiara	0.00	1	3.143480	101.644330
1	Ampang, Kuala Lumpu	0.00	1	3.158093	101.748703

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	r				
2	Bandar Menjalara	0.00	1	3.190350	101.625450
3	Bandar Sri P	0.00	1	3.103910	101.712260

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	ermaisuri				
4	Bandar Tasik Selatan	0.01	1	3.072620	101.714710

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
5	Bandar Tun Razak	0.00	1	3.08280	101.72810
9	Batu 11 Chera	0.00	1	3.09890	101.73490

	Neighborhoods	Shopping Mall	Cluster Labels	Latitude	Longitude
10	Batu, Kuala Lumpur	0.00	1	3.09890	101.73490
13	Bukit Jalil	0.00	1	3.05780	101.68965

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	l				0
40	Pantai Dalam	0.01	1	3.095160	101.668120
16	Bukit Petaling	0.01	1	3.129290	101.698920

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
18	Cheras, Kuala Lumpur	0.00	1	3.061870	101.746750
69	Wangsa Maj	0.01	1	3.203910	101.737199

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	u				0
25	Federal Hill, Kuala Lumpur	0.01	1	3.136370	101.685640
2	Ha	0.	1	3.	10

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
6	ppy Garden	01		201630	1.721070
29	Jinjang	0.00	1	3.209500	101.658740
32	Kampung D	0.01	1	3.166400	101.730466

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	atuk Keramat				0
33	Kepong	0.00	1	3.21750	101.63763
68	Titiwan	0.01	1	3.18067	101.703

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	gsa			0	220
36	Maluri	0.01	1	3.147890	1011.694050
38	Miharja	0.01	1	3.147890	1011.694050
24	Desa	0.0	1	3.08	1011.

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	Petaling	0		3310	704380

Cluster 2

In [95]:

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

Out[95]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
1	Br	0.	2	3.	10

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
1	ickfields	05		129160	1.684060
30	KL Eco City	0.05	2	3.117110	101.673490
7	Bangsar P	0.04	2	3.134780	101.67262

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	ark				0
6	Bangsar	0.05	2	3.129200	101.678440
35	Lembah Pantai	0.05	2	3.121216	101.663897

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
66	Taman U-Thant	0.05	2	3.157650	101.725020
41	Pudu, Kuala Lumpu	0.05	2	3.133540	101.713070

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
	r				

Observations:

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.