Analyzing the best business destination - Dubai or Kuala Lumpur

Introduction

AmerCave Group is a business entity who has spread its wings across the US. It has a chain of shopping malls and restaurants in almost all major cities of different states in America. As part of expanding business, they are strategically stepping outside US. AmerCave have chosen East Asia and Middle East for deploying their first shopping mall and an American restaurant. The reason for selecting both are given below.

a. East Asia and Middle East are fast growing economies, both having outstanding and competitive regional hubs for starting a business. Middle East, especially Dubai, UAE is an extremely attractive business destination and a great tourism spot. Lot of investments pouring in for the oil sector makes Dubai an advantage for investors. Apart from that, it is the gateway for Eastern markets as it acts as a hub for imports and exports. The city attracts a lot of expat investors as well.

b. Kuala Lumpur, Malaysia in the East Asian region have an increasing consumer demand from various nearby countries like India, China etc. It is set to be the center of digital economy and have a multitude of business forms. A diversified workforce is another advantage for opening a business in Malaysia. Kuala Lumpur offers a competitive market as well.

Business Problem

AmerCave Group have a good strategy, but they need a clarity in the following.

a. Where to start restaurant business? – Dubai or Kuala Lumpur

b. Where to start the shopping mall? – Dubai or Kuala Lumpur

AmerCave don’t want their business to be opened in some suburb area anywhere. The location of the new business should be within city limits

Problems to be resolved

Problems to be resolved

1. Provide a list of best locations in Dubai or in Kuala Lumpur for starting shopping mall

2. Provide a list of best locations in Dubai or in Kuala Lumpur for starting restaurant

3. Towards the end, it should be clear where to start shopping mall and where to start restaurant

4. The location should be within 2 Km from vicinity of both cities

Interested Audience

I believe the methodology and strategy for this project would cater to the needs of a business entity who is moving to either of two cities. The approach used can be used to explore more business opportunities and not confined to restaurants or shopping malls.

Data Source & Data Requirements

Data Section

1. Dubai neighborhoods information which can be web scrapped from Wikipedia page -- https://en.wikipedia.org/wiki/List_of_communities_in_Dubai (https://en.wikipedia.org/wiki/List_of_communities_in_Dubai)

2. Kuala Lumpur neighborhood information which can be web scrapped from Wikipedia page -- https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur (https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur)

3. Geodata of Dubai and Kuala Lumpur with top venue information to be collected from Foursquare

Data Requirements

1. Kuala Lumpur and Dubai maps is to be created using Nominatim, Foursquare and Folium mapping

2. Neighborhood data for both Kuala Lumpur and Dubai have to clubbed with their geo data for mapping

Program set to start below

1. Start by installing all the required packages and importing them

In [15]:
```
!pip install geopy
!pip install folium
!pip install geocoder

import pandas as pd #for data analysis
import numpy as np #for handling data
import requests
import folium # map rendering library
import geocoder # to get coordinates
import matplotlib.pyplot as plt # Matplotlib and associated plotting modules
import matplotlib.cm as cm # Matplotlib and associated plotting modules
import matplotlib.colors as colors # Matplotlib and associated plotting module
s
import json # library to handle JSON files

from pandas.io.json import json_normalize # tranform JSON file into a pandas d
ataframe
from bs4 import BeautifulSoup
from geopy.geocoders import Nominatim  # convert an address into latitude and
 longitude values
from sklearn.cluster import KMeans # import k-means from clustering stage


def warn(*args, **kwargs):
    pass
import warnings
warnings.warn = warn

print("All required packages installed and imported")
```

```
Requirement already satisfied: geopy in /opt/conda/envs/Python36/lib/python3.
6/site-packages (1.18.1)
Requirement already satisfied: geographiclib<2,>=1.49 in /opt/conda/envs/Pyth
on36/lib/python3.6/site-packages (from geopy) (1.49)
Requirement already satisfied: folium in /opt/conda/envs/Python36/lib/python
3.6/site-packages (0.10.0)
Requirement already satisfied: numpy in /opt/conda/envs/Python36/lib/python3.
6/site-packages (from folium) (1.15.4)
Requirement already satisfied: requests in /opt/conda/envs/Python36/lib/pytho
n3.6/site-packages (from folium) (2.21.0)
Requirement already satisfied: branca>=0.3.0 in /opt/conda/envs/Python36/lib/
python3.6/site-packages (from folium) (0.3.1)
Requirement already satisfied: jinja2>=2.9 in /opt/conda/envs/Python36/lib/py
thon3.6/site-packages (from folium) (2.10)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/envs/Python3
6/lib/python3.6/site-packages (from requests->folium) (2019.6.16)
Requirement already satisfied: urllib3<1.25,>=1.21.1 in /opt/conda/envs/Pytho
n36/lib/python3.6/site-packages (from requests->folium) (1.24.1)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /opt/conda/envs/Pytho
n36/lib/python3.6/site-packages (from requests->folium) (3.0.4)
Requirement already satisfied: idna<2.9,>=2.5 in /opt/conda/envs/Python36/li
b/python3.6/site-packages (from requests->folium) (2.8)
Requirement already satisfied: six in /opt/conda/envs/Python36/lib/python3.6/
site-packages (from branca>=0.3.0->folium) (1.12.0)
Requirement already satisfied: MarkupSafe>=0.23 in /opt/conda/envs/Python36/l
ib/python3.6/site-packages (from jinja2>=2.9->folium) (1.1.0)
Requirement already satisfied: geocoder in /opt/conda/envs/Python36/lib/pytho
n3.6/site-packages (1.38.1)
Requirement already satisfied: requests in /opt/conda/envs/Python36/lib/pytho
n3.6/site-packages (from geocoder) (2.21.0)
Requirement already satisfied: click in /opt/conda/envs/Python36/lib/python3.
6/site-packages (from geocoder) (7.0)
Requirement already satisfied: future in /opt/conda/envs/Python36/lib/python
3.6/site-packages (from geocoder) (0.17.1)
Requirement already satisfied: ratelim in /opt/conda/envs/Python36/lib/python
3.6/site-packages (from geocoder) (0.1.6)
Requirement already satisfied: six in /opt/conda/envs/Python36/lib/python3.6/
site-packages (from geocoder) (1.12.0)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /opt/conda/envs/Pytho
n36/lib/python3.6/site-packages (from requests->geocoder) (3.0.4)
Requirement already satisfied: idna<2.9,>=2.5 in /opt/conda/envs/Python36/li
b/python3.6/site-packages (from requests->geocoder) (2.8)
Requirement already satisfied: urllib3<1.25,>=1.21.1 in /opt/conda/envs/Pytho
n36/lib/python3.6/site-packages (from requests->geocoder) (1.24.1)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/envs/Python3
6/lib/python3.6/site-packages (from requests->geocoder) (2019.6.16)
Requirement already satisfied: decorator in /opt/conda/envs/Python36/lib/pyth
on3.6/site-packages (from ratelim->geocoder) (4.3.2)
All required packages installed and imported
```

2. Setting the display options

In [16]:
```
pd.set_option("display.max_columns", None)
pd.set_option("display.max_rows", None)
```

3. Get the neighborhood data for Kuala Lumpur from Wiki page

All the data related to Kuala Lumpur can be identified by prefix 'kl' in this code

kl_df is the dataframe created for storing neighborhood information about Kuala Lumpur

In [17]:
```
#send the GET request
kl_data = requests.get("https://en.wikipedia.org/wiki/Category:Suburbs_in_Kual
a_Lumpur").text

#Parse the data from the html. The data will be available in the BeautifulSoup
object
kl_soup = BeautifulSoup(kl_data, 'html.parser')

#Create a list for storing neighborhood information
kl_neighborhoodList = []

#Read the parsed data from the tag named "mw-category". This tag contains the
 information pertaining neighborhoods of KL. This data needs to be appended on
e by one to the list
for row in kl_soup.find_all("div", class_="mw-category")[0].findAll("li"):
  kl_neighborhoodList.append(row.text)

#Add a column KL_Neighborhood in kl_df dataframe and add values from the list
 kl_neighborhoodList
kl_df = pd.DataFrame({"KL_Neighborhood": kl_neighborhoodList})

#Let's see the dimensions of new data frame for Kuala Lumpur (KL) - kl_df
print("The dimensions of kl_df dataframe is ", kl_df.shape)

#Let's have a visual of first 5 rows of kl_df
kl_df.head()
```

```
The dimensions of kl_df dataframe is  (71, 1)
```

Out[17]:

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

3. Get the neighborhood data for Dubai from Wiki page

All the data related to Dubai can be identified by prefix 'dxb' in this code

dxb_df is the dataframe created for storing neighborhood information about Dubai

In [18]:
```
#send the GET request
dxb_data = requests.get("https://en.wikipedia.org/wiki/List_of_communities_in_
Dubai").text

#Parse the data from the html. The data will be available in the BeautifulSoup
object
dxb_soup = BeautifulSoup(dxb_data, 'html.parser')

#Create a list for storing neighborhood information
dxb_neighborhoodList = []

#Read the parsed data from the tag named "mw-parser-output". This tag contains
the information pertaining neighborhoods of Dubai. This data needs to be appen
ded one by one to the list
#The tag is different from that of KL data.
for row in dxb_soup.find_all("div", class_="mw-parser-output")[0].findAll("li"
):
  dxb_neighborhoodList.append(row.text)

#Add a column DXB_Neighborhood in dxb_df dataframe and add values from the lis
t dxb_neighborhoodList
dxb_df = pd.DataFrame({"DXB_Neighborhood": dxb_neighborhoodList})

#Let's see the dimensions of new data frame for Dubai (DXB) - dxb_df
print("The dimensions of dxb_df dataframe is ", dxb_df.shape)

#Let's have a visual of first 5 rows of dxb_df
dxb_df.head()
```

```
The dimensions of dxb_df dataframe is  (83, 1)
```

Out[18]:

	DXB_Neighborhood
0	List of Industrial areas in Dubai
1	List of development projects in Dubai
2	Developments in Dubai
3	^ Dubai FAQs. "Communities in Dubai". dubaifaq...
4	Communities in Dubai

4. Cleaning the DXB data frame

Let's have a detailed closer look into the dxb_df, as from above first five rows, the DXB_Neighborhood entries doesn't look quite like name of places. If some unwanted information is present, it needs to be cleaned.

Start first by getting a full display of dxb_df

In [19]:
```
dxb_df
```

Out[19]:

	DXB_Neighborhood
0	List of Industrial areas in Dubai
1	List of development projects in Dubai
2	Developments in Dubai
3	^ Dubai FAQs. "Communities in Dubai". dubaifaq...
4	Communities in Dubai
5	Location of communities in Dubai
6	v
7	t
8	e
9	Abu Hail
10	Al Baraha
11	Al Buteen
12	Al Dhagaya
13	Al Garhoud
14	Al Hamriya Port
15	Al Karama
16	Al Khabisi
17	Al Mamzar
18	Al Mizhar
19	Al Muraqqabat
20	Al Murar
21	Al Muteena
22	Al Nahda
23	Al Qusais
24	Al Ras
25	Al Rashidiya
26	Al Rigga
27	Al Sabkha
28	Al Twar
29	Al Waheda
30	Al Warqaa
31	Ayal Nasir

	DXB_Neighborhood
32	Dubai International Airport
33	Hor Al Anz
34	Mirdif
35	Muhaisnah
36	Nad Al Hammar
37	Nad Shamma
38	Naif
39	Port Saeed
40	Rigga Al Buteen
41	Umm Ramool
42	Warisan
43	Al Amardhi
44	Al Bada
45	Al Barsha
46	Al Hamriya
47	Al Hudaiba
48	Al Jaddaf
49	Al Jafilia
50	Al Karama
51	Al Kefaf
52	Al Manara
53	Al Mankhool
54	Al Markada
55	Al Quoz
56	Al Rifa
57	Al Safa
58	Al Satwa
59	Al Shindagha
60	Al Souk Al Kabir
61	Al Sufouh
62	Al Wasl
63	Bu Kadra

	DXB_Neighborhood
64	Business Bay
65	Dubai Marina
66	Emirates Hills
67	Dubai International City
68	Jebel Ali
69	Jumeirah
70	Jumeirah Islands
71	Jumeirah Lake Towers
72	Nad Al Sheba
73	Oud Metha
74	Port Rashid
75	Ras Al Khor
76	Ras Al Khor Industrial Area
77	Trade Centre 1
78	Trade Centre 2
79	Umm Al Sheif
80	Umm Hurair
81	Umm Suqeim
82	Zabeel

A closer look reveals that the first 9 rows has no meaning as it is not place's name. This means the first 9 rows can be removed.

In [20]:
```
#Cleaning DXB Data

#Removing first 9 entries (index 0 to index 8) OR selecting the remaining data
dxb_df = dxb_df.iloc[9:]

#The index value remains untouched, and to start afresh it is required to rese
t the index
dxb_df = dxb_df.reset_index()

#Also, a new column named 'index' will be induced by running above commands. L
et's drop that column and get the real dataframe
del dxb_df['index']

#The new dimensions of data frame are
print("The current dimension of dxb_df is ", dxb_df.shape)

#The polished dataframe looks like this
dxb_df.head()
```

```
The current dimension of dxb_df is  (74, 1)
```

Out[20]:

	DXB_Neighborhood
0	Abu Hail
1	Al Baraha
2	Al Buteen
3	Al Dhagaya
4	Al Garhoud

To summarize, the current neighborhood information of KL is available in kl_df and that of DXB is available in dxb_df

5. Getting co-ordinates of KL and DXB

In [21]:
```
#define function for getting KL coordinates
def get_latlng_kl(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

#Creating a list of co-ordinates using the list comprehension method
kl_coords = [ get_latlng_kl(neighborhood) for neighborhood in kl_df["KL_Neighb
orhood"].tolist() ]


# define a function to get Dubai coordinates
def get_latlng_dxb(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('{}, Dubai, UAE'.format(neighborhood))
        lat_lng_coords = g.latlng
    return lat_lng_coords

#Creating a list of co-ordinates using the list comprehension method
dxb_coords = [ get_latlng_dxb(neighborhood) for neighborhood in dxb_df["DXB_Ne
ighborhood"].tolist() ]
```

In [22]:
```
#Create new temporary data frames for storing coordinatate information
kl_coords = pd.DataFrame(kl_coords, columns=['Latitude', 'Longitude'])
dxb_coords = pd.DataFrame(dxb_coords, columns=['Latitude', 'Longitude'])

#Introduce two new columns for latitude and longitude and get the correspoding
informtion from dataframe stored with coordinate information(kl_coords and dxb
_coords)
kl_df['Latitude'] = kl_coords['Latitude']
kl_df['Longitude'] = kl_coords['Longitude']
dxb_df['Latitude'] = dxb_coords['Latitude']
dxb_df['Longitude'] = dxb_coords['Longitude']
```

In [23]:
```
#Lets have a look of how the two modified dataframes look like
kl_df.head()
```

Out[23]:

	KL_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

In [24]:
```
dxb_df.head()
```

Out[24]:

	DXB_Neighborhood	Latitude	Longitude
0	Abu Hail	25.28308	55.33435
1	Al Baraha	25.28280	55.31678
2	Al Buteen	25.26925	55.29944
3	Al Dhagaya	25.27217	55.30157
4	Al Garhoud	25.24337	55.35267

In [25]:
```
#Let's now check if there are any null values in our data frame. If there are
 null values, it is going to affect our map plotting which is coming next
kl_df.isnull().values.any()
```

Out[25]: `False`

In [26]:
```
dxb_df.isnull().values.any()
```

Out[26]: `False`

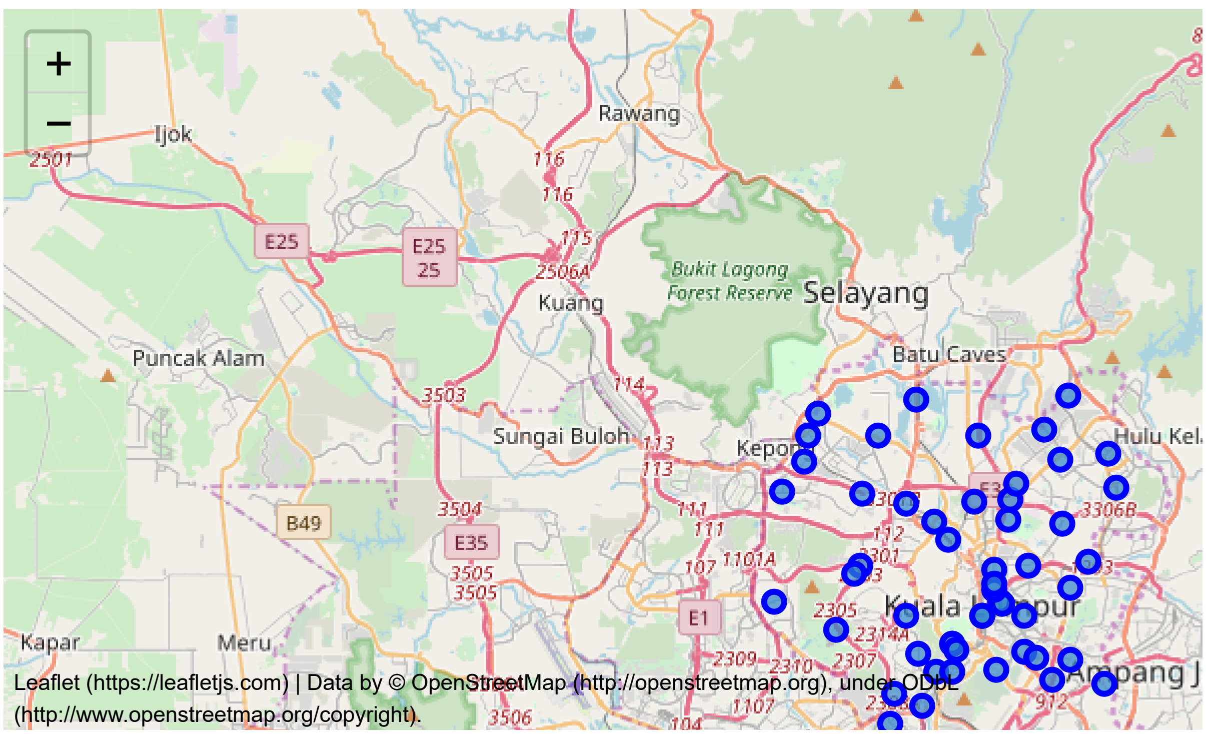
The above commands returned 'False', which means there are no null values and the commands for extracting coordinate information have worked fine.

6. Map of Kuala Lumpur (KL) with neighborhoods

In [27]:

```
#Get the co-ordinates of KL first into latitude and longitude
address = 'Kuala Lumpur, Malaysia'
geolocator = Nominatim(user_agent="my-application")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude


# create map of KL using latitude and longitude values
kl_map = 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
['KL_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(kl_map)

kl_map
```

Out[27]:

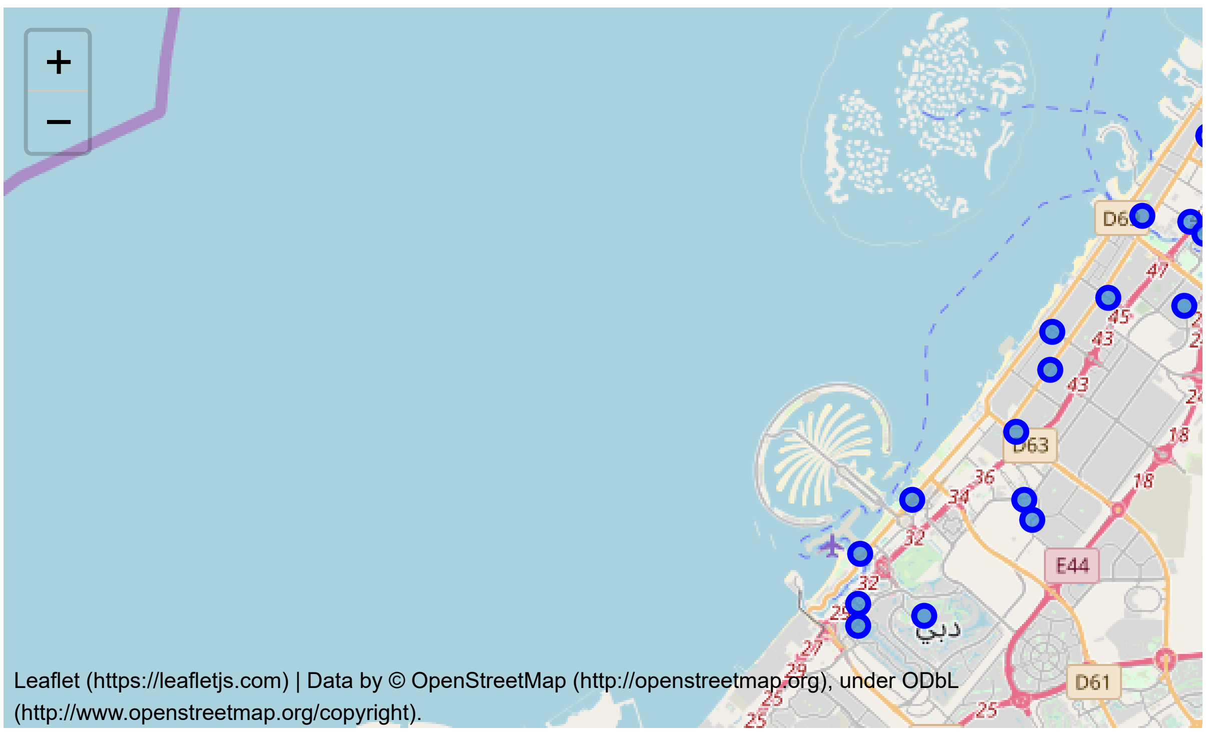


7. Map of Dubai (DXB) with neighborhoods

In [28]:

```
#Get the co-ordinates of DXB first into latitude and longitude
address = 'Dubai, UAE'
geolocator = Nominatim(user_agent="my-application")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude


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

# add markers to map
for lat, lng, neighborhood in zip(dxb_df['Latitude'], dxb_df['Longitude'], dxb
_df['DXB_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(dxb_map)

dxb_map
```

Out[28]:



8. Explore the KL Neighborhoods using the Foursquare API

Initiate the client ID and client secret ID capturing process

In [29]:
```
# define Foursquare Credentials and Version
CLIENT_ID = 'YV1J5SEQHSSMPE3VAJTKNIWYXJXTMVXKSS1OISQ2WNY21COG' # your Foursqua
re ID
CLIENT_SECRET = 'SI0O02XXDLUIVYBN3SNW3E3G04DQRGIN3CFJ0EIKFH1G4BFL' # your Four
square Secret
VERSION = '20180605' # Foursquare API version
```

Start the process of capturing the top 100 venues in Kuala Lumpur within a radius of 2000 meters

In [30]:
```
#Get top 100 venues within a radius of 2000 meters in Kuala Lumpur
radius = 2000
LIMIT = 100

venues = []

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

    # create the API request URL
    url = "https://api.foursquare.com/v2/venues/explore?client_id={}&client_se
cret={}&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']))
```

Define a new dataframe for capturing the neighborhood details and the venue details.

In [31]:
```
# convert the venues list into a new DataFrame
kl_venues_df = pd.DataFrame(venues)

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

kl_venues_df.head()
```

Out[31]:

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	Ven
0	Alam Damai	3.05769	101.74388	Pengedar Shaklee Kuala Lumpur	3.061235	101.740696	Sup Sho
1	Alam Damai	3.05769	101.74388	Machi Noodle 妈子面	3.057695	101.746635	Noo
2	Alam Damai	3.05769	101.74388	Restoran Ikbal	3.061134	101.750220	Res
3	Alam Damai	3.05769	101.74388	閒茶素食店 Leisure Tea Vegetarian	3.057673	101.747258	Veg Veg Res
4	Alam Damai	3.05769	101.74388	Ivy Sekinchan Seafood Noodle House 适耕莊特制魚丸海鲜面	3.065749	101.748718	Noo

Let's get a count of non-Null values in the dataframe based on values in 'Neighborhood' column

In [32]:
```
kl_venues_df.groupby(["Neighborhood"]).count()
```

Out[32]:

	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	Venue
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	100	100	100	100	100	100
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

	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	Venue
Neighborhood						
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	92	92	92	92	92	92
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

	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	Venue
Neighborhood						
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	91	91	91	91	91	91
Taman Sri Sinar	99	99	99	99	99	99
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

	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	Venue
Neighborhood						
Taman Wahyu	100	100	100	100	100	100
Titiwangsa	100	100	100	100	100	100
Wangsa Maju	100	100	100	100	100	100

Display the number of unique categories in KL

In [33]:
```
print('There are {} uniques categories in Kuala Lumpur.'.format(len(kl_venues_
df['VenueCategory'].unique())))
```

```
There are 300 uniques categories in Kuala Lumpur.
```

Get a list of 50 unique categories in the city of Kuala Lumpur

In [34]:
```
# List of categories in Kuala

kl_venues_df['VenueCategory'].unique()[:50]
```

Out[34]:
```
array(['Supplement Shop', 'Noodle House', 'Restaurant',
       'Vegetarian / Vegan Restaurant', 'Chinese Restaurant',
       'Breakfast Spot', 'Food Court', 'Asian Restaurant',
       'Other Great Outdoors', 'Park', 'Snack Place',
       'Dim Sum Restaurant', 'Food Truck', 'Japanese Restaurant',
       'Indian Restaurant', 'Bubble Tea Shop', 'Spa',
       'Seafood Restaurant', 'Chinese Breakfast Place',
       'Convenience Store', 'Dessert Shop', 'Cantonese Restaurant',
       'Farmers Market', 'Bakery', 'Pet Store', 'Café',
       'Malay Restaurant', 'Outlet Store', 'Gym / Fitness Center',
       'Hakka Restaurant', 'Steakhouse', 'Badminton Court',
       'Fast Food Restaurant', 'Athletics & Sports',
       'Middle Eastern Restaurant', 'Mamak Restaurant', 'Burger Joint',
       'Winery', 'College Bookstore', 'Grocery Store', 'Garden Center',
       'Basketball Court', 'Monument / Landmark', 'Hostel',
       'Latin American Restaurant', 'Hotel', 'Hotel Pool',
       'South Indian Restaurant', 'Soup Place', 'Dance Studio'],
      dtype=object)
```

Here is the most important part. Check whether there is a category for American restaurants and Shopping malls in the venues dataframe. If it returns 'True' we need a thorough analysis for a new shopping mall / restaurant location. If 'False' we can go ahead and get a best cluster for identifying the locations to start business.

In [35]:
```
#Unique categories in KL
"American Restaurant" in kl_venues_df['VenueCategory'].unique()
```

Out[35]:
```
True
```

In [36]:
```
#Unique categories in KL
"Shopping Mall" in kl_venues_df['VenueCategory'].unique()
```

Out[36]: True

The above checking returned true, indicating we need a thorough analysis of location. This means the restaurants & shopping mall the company, AmarCave, is intending to open is already existing in the city of Kuala Lumpur. So, we need to identify a better location for a new business.

9. Explore the DXB Neighborhoods using the Foursquare API

Before drilling down into more details of KL city, lets go and analyze the neighborhoods of DXB first.

Let's now get into top 100 venues ion Dubai city, within a span of 2Kms. We don't need the Foursquare Client ID and it's secret ID as it is already in there from the previous analysis for KL.

In [37]:
```
#get top 100 venues within 2000 meters in Dubai
radius = 2000
LIMIT = 100

venues = []

for lat, long, neighborhood in zip(dxb_df['Latitude'], dxb_df['Longitude'], dx
b_df['DXB_Neighborhood']):

    # create the API request URL
    url = "https://api.foursquare.com/v2/venues/explore?client_id={}&client_se
cret={}&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']))
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
<ipython-input-37-1a0a5c40ee1e> in <module>
     18 
     19     # make the GET request
---> 20     results = requests.get(url).json()["response"]['groups'][0]['item
s']
     21 
     22     # return only relevant information for each nearby venue

KeyError: 'groups'
```

This returns a list of values of top venues and let's now capture it to dataframe

In [38]:
```
# convert the venues list into a new DataFrame
dxb_venues_df = pd.DataFrame(venues)

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

dxb_venues_df.head()
```

Out[38]:

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	Ver
0	Abu Hail	25.28308	55.33435	Habib Bakery	25.281124	55.332774	Bak
1	Abu Hail	25.28308	55.33435	Gold's Gym	25.282698	55.341019	Gyı
2	Abu Hail	25.28308	55.33435	Al Douri Roastery	25.277057	55.328223	Bak
3	Abu Hail	25.28308	55.33435	Union Co-Operative Society	25.282769	55.340896	Dep Sto
4	Abu Hail	25.28308	55.33435	Fitness Time (وقت اللياقة)	25.289077	55.347913	Gyı

In [39]:
```
dxb_venues_df.groupby(["Neighborhood"]).count()
```

Out[39]:

	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	Venue
Neighborhood						
Abu Hail	100	100	100	100	100	100
Al Amardhi	5	5	5	5	5	5
Al Bada	100	100	100	100	100	100
Al Baraha	100	100	100	100	100	100
Al Barsha	100	100	100	100	100	100
Al Buteen	100	100	100	100	100	100
Al Dhagaya	100	100	100	100	100	100
Al Garhoud	100	100	100	100	100	100
Al Hamriya	100	100	100	100	100	100
Al Hamriya Port	52	52	52	52	52	52
Al Hudaiba	100	100	100	100	100	100
Al Jaddaf	75	75	75	75	75	75
Al Jafilia	100	100	100	100	100	100
Al Karama	200	200	200	200	200	200
Al Kefaf	100	100	100	100	100	100
Al Khabisi	100	100	100	100	100	100
Al Mamzar	38	38	38	38	38	38
Al Manara	100	100	100	100	100	100
Al Mankhool	100	100	100	100	100	100
Al Markada	100	100	100	100	100	100
Al Mizhar	47	47	47	47	47	47
Al Muraqqabat	100	100	100	100	100	100
Al Murar	100	100	100	100	100	100
Al Muteena	100	100	100	100	100	100
Al Nahda	100	100	100	100	100	100
Al Quoz	61	61	61	61	61	61
Al Qusais	74	74	74	74	74	74
Al Ras	100	100	100	100	100	100
Al Rashidiya	55	55	55	55	55	55

	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	Venue
Neighborhood						
Al Rifa	100	100	100	100	100	100
Al Rigga	100	100	100	100	100	100
Al Sabkha	100	100	100	100	100	100
Al Safa	84	84	84	84	84	84
Al Satwa	100	100	100	100	100	100
Al Shindagha	100	100	100	100	100	100
Al Souk Al Kabir	100	100	100	100	100	100
Al Sufouh	100	100	100	100	100	100
Al Twar	79	79	79	79	79	79
Al Waheda	79	79	79	79	79	79
Al Warqaa	26	26	26	26	26	26
Al Wasl	100	100	100	100	100	100
Ayal Nasir	100	100	100	100	100	100
Bu Kadra	7	7	7	7	7	7
Business Bay	100	100	100	100	100	100
Dubai International Airport	100	100	100	100	100	100
Dubai International City	100	100	100	100	100	100
Dubai Marina	100	100	100	100	100	100
Emirates Hills	36	36	36	36	36	36
Hor Al Anz	100	100	100	100	100	100
Jebel Ali	42	42	42	42	42	42
Jumeirah	100	100	100	100	100	100
Jumeirah Islands	100	100	100	100	100	100
Jumeirah Lake Towers	100	100	100	100	100	100
Mirdif	96	96	96	96	96	96
Muhaisnah	27	27	27	27	27	27

	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	Venue
Neighborhood						
Nad Al Hammar	42	42	42	42	42	42
Nad Al Sheba	4	4	4	4	4	4
Nad Shamma	49	49	49	49	49	49
Naif	100	100	100	100	100	100
Oud Metha	100	100	100	100	100	100
Port Rashid	100	100	100	100	100	100
Port Saeed	100	100	100	100	100	100
Ras Al Khor	7	7	7	7	7	7
Ras Al Khor Industrial Area	18	18	18	18	18	18
Rigga Al Buteen	100	100	100	100	100	100
Umm Ramool	72	72	72	72	72	72
Warisan	97	97	97	97	97	97

In [40]:
```
print('There are {} uniques categories in Dubai.'.format(len(dxb_venues_df['Ve
nueCategory'].unique())))
```

```
There are 267 uniques categories in Dubai.
```

In [41]:
```
# List of categories in DXB
dxb_venues_df['VenueCategory'].unique()[:50]
```

Out[41]:
```
array(['Bakery', 'Gym', 'Department Store', 'Market',
       'Seafood Restaurant', 'Indian Restaurant', 'Performing Arts Venue',
       'Middle Eastern Restaurant', 'Hotel', 'Fast Food Restaurant',
       'Restaurant', 'Mediterranean Restaurant', 'Fried Chicken Joint',
       'Ice Cream Shop', 'Lounge', 'BBQ Joint', 'Asian Restaurant',
       'Café', 'Dessert Shop', 'Burger Joint', 'Nightclub',
       'Bowling Alley', 'American Restaurant', 'Thai Restaurant',
       'Pool Hall', 'Sports Bar', 'Comedy Club', 'Italian Restaurant',
       'Convenience Store', 'Fishing Store', 'Pizza Place',
       'Burrito Place', 'Bavarian Restaurant', 'Korean Restaurant',
       'Gym / Fitness Center', 'Tea Room', 'Chinese Restaurant', 'Buffet',
       'Soccer Field', 'Hotel Bar', 'Smoke Shop', 'Food Court',
       'Hookah Bar', 'Arepa Restaurant', 'Coffee Shop',
       'Moroccan Restaurant', 'Filipino Restaurant', 'Bar',
       'Scenic Lookout', 'Accessories Store'], dtype=object)
```

In [42]:
```
#Unique categories in DXB
"American Restaurant" in dxb_venues_df['VenueCategory'].unique()
```

Out[42]: True

In [43]:
```
#Unique categories in DXB
"Shopping Mall" in dxb_venues_df['VenueCategory'].unique()
```

Out[43]: True

The above steps are similar for the one we did for KL, and finally it returned a 'True' for both restaurants and shopping malls in Dubai. Same is the case with Kuala Lumpur as well. This means, the business that we are intending to start are already present in both cities and so, for a newer one we need to go in and do a detailed analysis of the same.

10. Analyze each neighborhood for KL

As mentioned above, a detailed analysis of both cities is required. Let's first analyze the one for Kuala Lumpur and then for Dubai later

In [44]:
```
#Analyse each neighboorhood for KL

# one hot encoding
kl_onehot = pd.get_dummies(kl_venues_df[['VenueCategory']], prefix="", prefix_
sep="")

# add neighborhood column back to dataframe
kl_onehot['Neighborhoods'] = kl_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()
```

```
(7082, 301)
```

Out[44]:

	Neighborhoods	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Arcade	Art Gallery	Mus
0	Alam Damai	0	0	0	0	0	0	0
1	Alam Damai	0	0	0	0	0	0	0
2	Alam Damai	0	0	0	0	0	0	0
3	Alam Damai	0	0	0	0	0	0	0
4	Alam Damai	0	0	0	0	0	0	0

In [45]:
```
kl_grouped = kl_onehot.groupby(["Neighborhoods"]).mean().reset_index()

print(kl_grouped.shape)
kl_grouped
```

```
(71, 301)
```

Out[45]:

	Neighborhoods	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Arcade	Art Gallery	M
0	Alam Damai	0.00	0.00	0.00	0.00	0.000000	0.00	0
1	Ampang, Kuala Lumpur	0.00	0.00	0.00	0.00	0.000000	0.01	0
2	Bandar Menjalara	0.00	0.00	0.01	0.00	0.000000	0.00	0
3	Bandar Sri Permaisuri	0.01	0.00	0.00	0.00	0.000000	0.01	0
4	Bandar Tasik Selatan	0.01	0.00	0.00	0.00	0.010000	0.00	0
5	Bandar Tun Razak	0.01	0.00	0.00	0.00	0.000000	0.00	0
6	Bangsar	0.00	0.00	0.00	0.00	0.000000	0.00	0
7	Bangsar Park	0.00	0.00	0.00	0.00	0.000000	0.00	0
8	Bangsar South	0.00	0.00	0.00	0.00	0.000000	0.00	0
9	Batu 11 Cheras	0.00	0.00	0.00	0.00	0.000000	0.00	0
10	Batu, Kuala Lumpur	0.00	0.00	0.00	0.00	0.000000	0.00	0
11	Brickfields	0.00	0.00	0.01	0.00	0.000000	0.01	0
12	Bukit Bintang	0.00	0.00	0.00	0.00	0.000000	0.00	0
13	Bukit Jalil	0.00	0.00	0.00	0.00	0.000000	0.00	0
14	Bukit Kiara	0.00	0.00	0.00	0.00	0.000000	0.00	0
15	Bukit Nanas	0.00	0.00	0.00	0.00	0.000000	0.00	0
16	Bukit Petaling	0.01	0.00	0.01	0.00	0.000000	0.01	0
17	Bukit Tunku	0.00	0.00	0.00	0.00	0.010000	0.02	0
18	Cheras, Kuala Lumpur	0.00	0.00	0.00	0.00	0.000000	0.00	0
19	Chow Kit	0.00	0.01	0.00	0.00	0.000000	0.01	0
20	Damansara Heights	0.00	0.00	0.00	0.00	0.000000	0.00	0
21	Damansara Town Centre	0.00	0.00	0.01	0.00	0.000000	0.02	0
22	Damansara, Kuala Lumpur	0.00	0.00	0.01	0.00	0.000000	0.02	0

	Neighborhoods	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Arcade	Art Gallery	M
23	Dang Wangi	0.00	0.00	0.00	0.00	0.000000	0.01	0
24	Desa Petaling	0.01	0.00	0.00	0.00	0.000000	0.01	0
25	Federal Hill, Kuala Lumpur	0.00	0.00	0.01	0.00	0.000000	0.01	0
26	Happy Garden	0.00	0.00	0.00	0.00	0.000000	0.00	0
27	Jalan Cochrane, Kuala Lumpur	0.00	0.00	0.00	0.00	0.000000	0.00	0
28	Jalan Duta	0.00	0.00	0.00	0.00	0.010000	0.02	0
29	Jinjang	0.01	0.00	0.00	0.00	0.000000	0.00	0
30	KL Eco City	0.02	0.00	0.00	0.00	0.000000	0.00	0
31	Kampung Baru, Kuala Lumpur	0.01	0.00	0.00	0.00	0.000000	0.01	0
32	Kampung Datuk Keramat	0.00	0.00	0.00	0.01	0.000000	0.03	0
33	Kampung Padang Balang	0.00	0.00	0.00	0.00	0.010870	0.00	0
34	Kepong	0.01	0.00	0.00	0.01	0.000000	0.00	0
35	Kuchai Lama	0.00	0.00	0.00	0.00	0.000000	0.00	0
36	Lembah Pantai	0.00	0.00	0.00	0.00	0.000000	0.00	0
37	Maluri	0.00	0.01	0.00	0.00	0.000000	0.01	0
38	Medan Tuanku	0.00	0.00	0.00	0.00	0.000000	0.01	0
39	Miharja	0.00	0.01	0.00	0.00	0.000000	0.01	0
40	Mont Kiara	0.00	0.00	0.00	0.00	0.000000	0.01	0
41	Pantai Dalam	0.00	0.00	0.00	0.00	0.000000	0.00	0
42	Pudu, Kuala Lumpur	0.00	0.00	0.00	0.00	0.000000	0.00	0
43	Putrajaya	0.00	0.00	0.00	0.00	0.000000	0.00	0
44	Salak South	0.00	0.00	0.00	0.00	0.000000	0.00	0
45	Segambut	0.00	0.00	0.00	0.00	0.000000	0.02	0
46	Semarak	0.00	0.00	0.00	0.00	0.010000	0.03	0
47	Sentul Raya	0.00	0.00	0.00	0.00	0.010000	0.00	0
48	Setapak	0.00	0.00	0.00	0.00	0.010000	0.01	0

	Neighborhoods	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Arcade	Art Gallery	M
49	Setiawangsa	0.00	0.00	0.01	0.00	0.000000	0.00	0
50	Shamelin	0.00	0.00	0.00	0.00	0.000000	0.00	0
51	Sri Hartamas	0.00	0.00	0.00	0.00	0.000000	0.00	0
52	Sri Petaling	0.00	0.00	0.00	0.00	0.000000	0.00	0
53	Sungai Besi	0.00	0.00	0.00	0.00	0.000000	0.00	0
54	Taman Bukit Maluri	0.00	0.00	0.00	0.00	0.000000	0.00	0
55	Taman Cheras Hartamas	0.00	0.00	0.00	0.00	0.000000	0.00	0
56	Taman Connaught	0.00	0.00	0.00	0.00	0.000000	0.00	0
57	Taman Desa	0.00	0.00	0.00	0.00	0.000000	0.00	0
58	Taman Ibukota	0.00	0.00	0.00	0.01	0.000000	0.00	0
59	Taman Len Seng	0.00	0.00	0.00	0.00	0.000000	0.00	0
60	Taman Melati	0.00	0.00	0.00	0.01	0.000000	0.00	0
61	Taman Midah	0.00	0.00	0.00	0.00	0.000000	0.01	0
62	Taman OUG	0.01	0.00	0.00	0.00	0.000000	0.00	0
63	Taman P. Ramlee	0.00	0.00	0.00	0.00	0.010989	0.00	0
64	Taman Sri Sinar	0.00	0.00	0.00	0.00	0.000000	0.00	0
65	Taman Taynton View	0.00	0.00	0.00	0.00	0.000000	0.00	0
66	Taman Tun Dr Ismail	0.01	0.00	0.00	0.00	0.000000	0.00	0
67	Taman U-Thant	0.01	0.00	0.00	0.00	0.000000	0.01	0
68	Taman Wahyu	0.00	0.00	0.00	0.00	0.000000	0.00	0
69	Titiwangsa	0.00	0.00	0.00	0.00	0.010000	0.01	0
70	Wangsa Maju	0.00	0.00	0.00	0.00	0.000000	0.00	0

Let's go in a get a list of "Shopping Mall" in KL City. This will be followed by a similar analysis of "American Restaurants" later.

In [46]:
```
kl_shopping_malls = kl_grouped[["Neighborhoods","Shopping Mall"]]
```

In [47]:
```
kl_shopping_malls.head()
```

Out[47]:

	Neighborhoods	Shopping Mall
0	Alam Damai	0.00
1	Ampang, Kuala Lumpur	0.03
2	Bandar Menjalara	0.01
3	Bandar Sri Permaisuri	0.00
4	Bandar Tasik Selatan	0.02

Use KMeans Clustering

In [48]:
```
# set number of clusters
kclusters = 3

kl_clustering = kl_shopping_malls.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[48]: array([2, 1, 0, 2, 0, 2, 1, 1, 1, 2], dtype=int32)

In [49]:
```
# create a new dataframe that includes the cluster as well as the top 10 venue
s for each neighborhood.
kl_merged = kl_shopping_malls.copy()

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

In [50]:
```
kl_merged.rename(columns={"Neighborhoods": "Neighborhood"}, inplace=True)
kl_merged.head()
```

Out[50]:

	Neighborhood	Shopping Mall	Cluster Labels
0	Alam Damai	0.00	2
1	Ampang, Kuala Lumpur	0.03	1
2	Bandar Menjalara	0.01	0
3	Bandar Sri Permaisuri	0.00	2
4	Bandar Tasik Selatan	0.02	0

In [51]:
```
kl_merged = kl_merged.join(kl_df.set_index("KL_Neighborhood"), on="Neighborhoo
d")

print(kl_merged.shape)
kl_merged.head()
```

```
(71, 5)
```

Out[51]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
0	Alam Damai	0.00	2	3.057690	101.743880
1	Ampang, Kuala Lumpur	0.03	1	3.153153	101.700413
2	Bandar Menjalara	0.01	0	3.190350	101.625450
3	Bandar Sri Permaisuri	0.00	2	3.103910	101.712260
4	Bandar Tasik Selatan	0.02	0	3.072620	101.714710

In [52]:
```
print(kl_merged.shape)
kl_merged.sort_values(["Cluster Labels"], inplace=True)
kl_merged
```

```
(71, 5)
```

Out[52]:

	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
50	Shamelin	0.020000	0	3.124580	101.735970
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
41	Pantai Dalam	0.010000	0	3.094760	101.667470
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
49	Setiawangsa	0.010000	0	3.191803	101.740070
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
16	Bukit Petaling	0.010000	0	3.129290	101.698920
64	Taman Sri Sinar	0.010101	0	3.190070	101.652930
2	Bandar Menjalara	0.010000	0	3.190350	101.625450
4	Bandar Tasik Selatan	0.020000	0	3.072620	101.714710
12	Bukit Bintang	0.020000	0	3.147770	101.708550
10	Batu, Kuala Lumpur	0.020000	0	3.135760	101.708370
1	Ampang, Kuala Lumpur	0.030000	1	3.153153	101.700413
42	Pudu, Kuala Lumpur	0.040000	1	3.133540	101.713070
67	Taman U-Thant	0.040000	1	3.157700	101.724520
40	Mont Kiara	0.030000	1	3.165320	101.652430
6	Bangsar	0.050000	1	3.129200	101.678440
7	Bangsar Park	0.050000	1	3.134780	101.672620

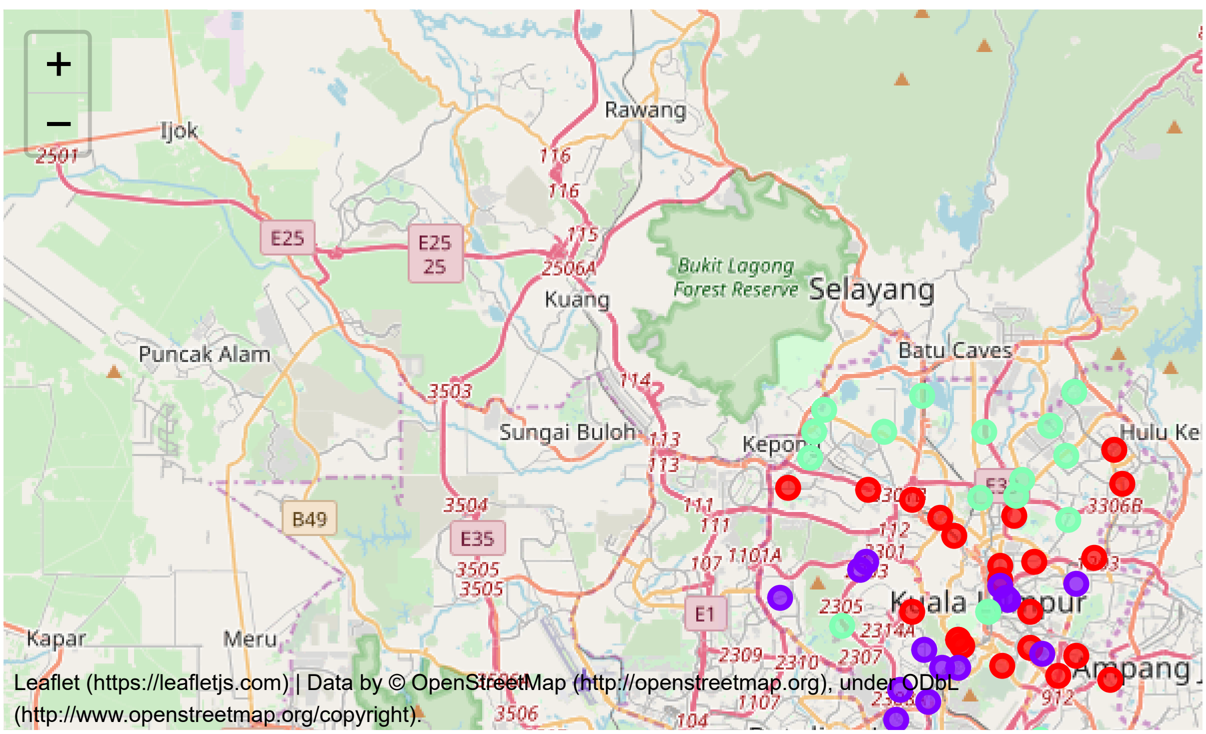
	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
36	Lembah Pantai	0.050000	1	3.121202	101.663899
8	Bangsar South	0.030000	1	3.111020	101.662830
51	Sri Hartamas	0.030000	1	3.162200	101.650360
15	Bukit Nanas	0.030000	1	3.152017	101.701028
66	Taman Tun Dr Ismail	0.030000	1	3.152830	101.622710
30	KL Eco City	0.040000	1	3.117140	101.673890
23	Dang Wangi	0.030000	1	3.156685	101.698076
11	Brickfields	0.030000	1	3.129160	101.684060
52	Sri Petaling	0.000000	2	3.072600	101.682520
54	Taman Bukit Maluri	0.000000	2	3.200660	101.633370
68	Taman Wahyu	0.000000	2	3.222400	101.671730
55	Taman Cheras Hartamas	0.000000	2	3.082630	101.746710
58	Taman Ibukota	0.000000	2	3.212160	101.715400
59	Taman Len Seng	0.000000	2	3.069080	101.742870
61	Taman Midah	0.000000	2	3.093590	101.728370
56	Taman Connaught	0.000000	2	3.082690	101.736890
60	Taman Melati	0.000000	2	3.223570	101.723990
63	Taman P. Ramlee	0.000000	2	3.193600	101.705980
62	Taman OUG	0.000000	2	3.210050	101.634508
65	Taman Taynton View	0.000000	2	3.087070	101.736810
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
26	Happy Garden	0.000000	2	3.201630	101.721070
29	Jinjang	0.000000	2	3.209500	101.658740
33	Kampung Padang Balang	0.000000	2	3.209430	101.693180

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
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
44	Salak South	0.000000	2	3.081020	101.697240
46	Semarak	0.000000	2	3.179927	101.721442
48	Setapak	0.000000	2	3.188160	101.704150
35	Kuchai Lama	0.000000	2	3.090740	101.677330

Let's now mark the "Shopping Malls" that we found in KL city

In [53]:
```
address = 'Kuala Lumpur, Malaysia'

geolocator = Nominatim(user_agent="my-application")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude

kl_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=T
rue)
    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(kl_map_clusters)

kl_map_clusters
```

Out[53]:



11. Analyze each cluster in KL

Analyze Cluster 0

In [54]:
```
#Cluster 0
kl_merged.loc[kl_merged['Cluster Labels'] == 0]
```

Out[54]:

	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
50	Shamelin	0.020000	0	3.124580	101.735970
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
41	Pantai Dalam	0.010000	0	3.094760	101.667470
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
49	Setiawangsa	0.010000	0	3.191803	101.740070
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
16	Bukit Petaling	0.010000	0	3.129290	101.698920
64	Taman Sri Sinar	0.010101	0	3.190070	101.652930
2	Bandar Menjalara	0.010000	0	3.190350	101.625450
4	Bandar Tasik Selatan	0.020000	0	3.072620	101.714710
12	Bukit Bintang	0.020000	0	3.147770	101.708550
10	Batu, Kuala Lumpur	0.020000	0	3.135760	101.708370

Analyze Cluster 1

In [55]:
```
#Cluster 1

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

Out[55]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
1	Ampang, Kuala Lumpur	0.03	1	3.153153	101.700413
42	Pudu, Kuala Lumpur	0.04	1	3.133540	101.713070
67	Taman U-Thant	0.04	1	3.157700	101.724520
40	Mont Kiara	0.03	1	3.165320	101.652430
6	Bangsar	0.05	1	3.129200	101.678440
7	Bangsar Park	0.05	1	3.134780	101.672620
36	Lembah Pantai	0.05	1	3.121202	101.663899
8	Bangsar South	0.03	1	3.111020	101.662830
51	Sri Hartamas	0.03	1	3.162200	101.650360
15	Bukit Nanas	0.03	1	3.152017	101.701028
66	Taman Tun Dr Ismail	0.03	1	3.152830	101.622710
30	KL Eco City	0.04	1	3.117140	101.673890
23	Dang Wangi	0.03	1	3.156685	101.698076
11	Brickfields	0.03	1	3.129160	101.684060

Analyze Cluster 2

In [56]:
```
#Cluster 2
kl_merged.loc[kl_merged['Cluster Labels'] == 2]
```

Out[56]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
52	Sri Petaling	0.0	2	3.072600	101.682520
54	Taman Bukit Maluri	0.0	2	3.200660	101.633370
68	Taman Wahyu	0.0	2	3.222400	101.671730
55	Taman Cheras Hartamas	0.0	2	3.082630	101.746710
58	Taman Ibukota	0.0	2	3.212160	101.715400
59	Taman Len Seng	0.0	2	3.069080	101.742870
61	Taman Midah	0.0	2	3.093590	101.728370
56	Taman Connaught	0.0	2	3.082690	101.736890
60	Taman Melati	0.0	2	3.223570	101.723990
63	Taman P. Ramlee	0.0	2	3.193600	101.705980
62	Taman OUG	0.0	2	3.210050	101.634508
65	Taman Taynton View	0.0	2	3.087070	101.736810
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
26	Happy Garden	0.0	2	3.201630	101.721070
29	Jinjang	0.0	2	3.209500	101.658740
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
44	Salak South	0.0	2	3.081020	101.697240
46	Semarak	0.0	2	3.179927	101.721442
48	Setapak	0.0	2	3.188160	101.704150
35	Kuchai Lama	0.0	2	3.090740	101.677330

Now, we have a detailed information of three clusters pertaining to shopping malls in a kl_merged dataframe. We will use this as reference for later analysis and conclusion. Before that, we can have a similar analysis for American restaurants in the same clusters.

As, kl_merged holds the shopping malls information in KL city, kl_rest_merged would be the dataframe storing the cluster information regarding American restaturants in KL city

In [57]:

```
# Repeating the above step for American Restaurants in KL
kl_restaurants = kl_grouped[["Neighborhoods","American Restaurant"]]
kl_restaurants.head()
```

Out[57]:

	Neighborhoods	American Restaurant
0	Alam Damai	0.0
1	Ampang, Kuala Lumpur	0.0
2	Bandar Menjalara	0.0
3	Bandar Sri Permaisuri	0.0
4	Bandar Tasik Selatan	0.0

In [58]:
```
kl_restaurants = kl_grouped[["Neighborhoods","American Restaurant"]]
kl_restaurants.head()
kclusters = 3
kl_restaurants_clustering = kl_restaurants.drop(["Neighborhoods"], 1)
# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(kl_restaurants_clust
ering)
kmeans.labels_[0:10]
# create a new dataframe that includes the cluster as well as the top 10 venue
s for each neighborhood.
kl_rest_merged = kl_restaurants.copy()
# add clustering labels
kl_rest_merged["Cluster Labels"] = kmeans.labels_
kl_rest_merged.rename(columns={"Neighborhoods": "Neighborhood"}, inplace=True)

kl_rest_merged = kl_rest_merged.join(kl_df.set_index("KL_Neighborhood"), on="N
eighborhood")
kl_rest_merged.sort_values(["Cluster Labels"], inplace=True)
kl_rest_merged
```

Out[58]:

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
0	Alam Damai	0.00	0	3.057690	101.743880
37	Maluri	0.00	0	3.147890	101.694050
38	Medan Tuanku	0.00	0	3.159260	101.698340
39	Miharja	0.00	0	3.147890	101.694050
40	Mont Kiara	0.00	0	3.165320	101.652430
41	Pantai Dalam	0.00	0	3.094760	101.667470
42	Pudu, Kuala Lumpur	0.00	0	3.133540	101.713070
43	Putrajaya	0.00	0	3.125862	101.718624
44	Salak South	0.00	0	3.081020	101.697240
45	Segambut	0.00	0	3.186390	101.668100
46	Semarak	0.00	0	3.179927	101.721442
47	Sentul Raya	0.00	0	3.187431	101.691453
48	Setapak	0.00	0	3.188160	101.704150
49	Setiawangsa	0.00	0	3.191803	101.740070
50	Shamelin	0.00	0	3.124580	101.735970
36	Lembah Pantai	0.00	0	3.121202	101.663899
51	Sri Hartamas	0.00	0	3.162200	101.650360
53	Sungai Besi	0.00	0	3.050640	101.706130
54	Taman Bukit Maluri	0.00	0	3.200660	101.633370
55	Taman Cheras Hartamas	0.00	0	3.082630	101.746710
56	Taman Connaught	0.00	0	3.082690	101.736890
57	Taman Desa	0.00	0	3.102970	101.684710
59	Taman Len Seng	0.00	0	3.069080	101.742870
61	Taman Midah	0.00	0	3.093590	101.728370
62	Taman OUG	0.00	0	3.210050	101.634508
63	Taman P. Ramlee	0.00	0	3.193600	101.705980
64	Taman Sri Sinar	0.00	0	3.190070	101.652930
65	Taman Taynton View	0.00	0	3.087070	101.736810
66	Taman Tun Dr Ismail	0.00	0	3.152830	101.622710
67	Taman U-Thant	0.00	0	3.157700	101.724520
68	Taman Wahyu	0.00	0	3.222400	101.671730

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
52	Sri Petaling	0.00	0	3.072600	101.682520
69	Titiwangsa	0.00	0	3.180670	101.703220
35	Kuchai Lama	0.00	0	3.090740	101.677330
15	Bukit Nanas	0.00	0	3.152017	101.701028
14	Bukit Kiara	0.00	0	3.143480	101.644330
13	Bukit Jalil	0.00	0	3.057800	101.689650
12	Bukit Bintang	0.00	0	3.147770	101.708550
11	Brickfields	0.00	0	3.129160	101.684060
10	Batu, Kuala Lumpur	0.00	0	3.135760	101.708370
9	Batu 11 Cheras	0.00	0	3.098980	101.734990
33	Kampung Padang Balang	0.00	0	3.209430	101.693180
8	Bangsar South	0.00	0	3.111020	101.662830
6	Bangsar	0.00	0	3.129200	101.678440
5	Bandar Tun Razak	0.00	0	3.082800	101.722810
4	Bandar Tasik Selatan	0.00	0	3.072620	101.714710
3	Bandar Sri Permaisuri	0.00	0	3.103910	101.712260
2	Bandar Menjalara	0.00	0	3.190350	101.625450
1	Ampang, Kuala Lumpur	0.00	0	3.153153	101.700413
7	Bangsar Park	0.00	0	3.134780	101.672620
16	Bukit Petaling	0.00	0	3.129290	101.698920
70	Wangsa Maju	0.00	0	3.203910	101.737190
18	Cheras, Kuala Lumpur	0.00	0	3.061870	101.746750
31	Kampung Baru, Kuala Lumpur	0.00	0	3.165460	101.710280
30	KL Eco City	0.00	0	3.117140	101.673890
29	Jinjang	0.00	0	3.209500	101.658740
28	Jalan Duta	0.00	0	3.180163	101.677880
27	Jalan Cochrane, Kuala Lumpur	0.00	0	3.132977	101.724669
26	Happy Garden	0.00	0	3.201630	101.721070
17	Bukit Tunku	0.00	0	3.173810	101.682760
25	Federal Hill, Kuala Lumpur	0.00	0	3.136370	101.685640

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
24	Desa Petaling	0.00	0	3.083310	101.704380
23	Dang Wangi	0.00	0	3.156685	101.698076
22	Damansara, Kuala Lumpur	0.00	0	3.138759	101.684046
21	Damansara Town Centre	0.00	0	3.138759	101.684046
20	Damansara Heights	0.00	0	3.147980	101.667980
19	Chow Kit	0.00	0	3.163590	101.698110
58	Taman Ibukota	0.01	1	3.212160	101.715400
32	Kampung Datuk Keramat	0.01	1	3.166400	101.730460
60	Taman Melati	0.01	1	3.223570	101.723990
34	Kepong	0.01	1	3.217500	101.637630

Analyse Cluster 0

In [59]:
```
kl_rest_merged.loc[kl_rest_merged['Cluster Labels'] == 0]
```

Out[59]:

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
0	Alam Damai	0.0	0	3.057690	101.743880
37	Maluri	0.0	0	3.147890	101.694050
38	Medan Tuanku	0.0	0	3.159260	101.698340
39	Miharja	0.0	0	3.147890	101.694050
40	Mont Kiara	0.0	0	3.165320	101.652430
41	Pantai Dalam	0.0	0	3.094760	101.667470
42	Pudu, Kuala Lumpur	0.0	0	3.133540	101.713070
43	Putrajaya	0.0	0	3.125862	101.718624
44	Salak South	0.0	0	3.081020	101.697240
45	Segambut	0.0	0	3.186390	101.668100
46	Semarak	0.0	0	3.179927	101.721442
47	Sentul Raya	0.0	0	3.187431	101.691453
48	Setapak	0.0	0	3.188160	101.704150
49	Setiawangsa	0.0	0	3.191803	101.740070
50	Shamelin	0.0	0	3.124580	101.735970
36	Lembah Pantai	0.0	0	3.121202	101.663899
51	Sri Hartamas	0.0	0	3.162200	101.650360
53	Sungai Besi	0.0	0	3.050640	101.706130
54	Taman Bukit Maluri	0.0	0	3.200660	101.633370
55	Taman Cheras Hartamas	0.0	0	3.082630	101.746710
56	Taman Connaught	0.0	0	3.082690	101.736890
57	Taman Desa	0.0	0	3.102970	101.684710
59	Taman Len Seng	0.0	0	3.069080	101.742870
61	Taman Midah	0.0	0	3.093590	101.728370
62	Taman OUG	0.0	0	3.210050	101.634508
63	Taman P. Ramlee	0.0	0	3.193600	101.705980
64	Taman Sri Sinar	0.0	0	3.190070	101.652930
65	Taman Taynton View	0.0	0	3.087070	101.736810
66	Taman Tun Dr Ismail	0.0	0	3.152830	101.622710
67	Taman U-Thant	0.0	0	3.157700	101.724520
68	Taman Wahyu	0.0	0	3.222400	101.671730

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
52	Sri Petaling	0.0	0	3.072600	101.682520
69	Titiwangsa	0.0	0	3.180670	101.703220
35	Kuchai Lama	0.0	0	3.090740	101.677330
15	Bukit Nanas	0.0	0	3.152017	101.701028
14	Bukit Kiara	0.0	0	3.143480	101.644330
13	Bukit Jalil	0.0	0	3.057800	101.689650
12	Bukit Bintang	0.0	0	3.147770	101.708550
11	Brickfields	0.0	0	3.129160	101.684060
10	Batu, Kuala Lumpur	0.0	0	3.135760	101.708370
9	Batu 11 Cheras	0.0	0	3.098980	101.734990
33	Kampung Padang Balang	0.0	0	3.209430	101.693180
8	Bangsar South	0.0	0	3.111020	101.662830
6	Bangsar	0.0	0	3.129200	101.678440
5	Bandar Tun Razak	0.0	0	3.082800	101.722810
4	Bandar Tasik Selatan	0.0	0	3.072620	101.714710
3	Bandar Sri Permaisuri	0.0	0	3.103910	101.712260
2	Bandar Menjalara	0.0	0	3.190350	101.625450
1	Ampang, Kuala Lumpur	0.0	0	3.153153	101.700413
7	Bangsar Park	0.0	0	3.134780	101.672620
16	Bukit Petaling	0.0	0	3.129290	101.698920
70	Wangsa Maju	0.0	0	3.203910	101.737190
18	Cheras, Kuala Lumpur	0.0	0	3.061870	101.746750
31	Kampung Baru, Kuala Lumpur	0.0	0	3.165460	101.710280
30	KL Eco City	0.0	0	3.117140	101.673890
29	Jinjang	0.0	0	3.209500	101.658740
28	Jalan Duta	0.0	0	3.180163	101.677880
27	Jalan Cochrane, Kuala Lumpur	0.0	0	3.132977	101.724669
26	Happy Garden	0.0	0	3.201630	101.721070
17	Bukit Tunku	0.0	0	3.173810	101.682760
25	Federal Hill, Kuala Lumpur	0.0	0	3.136370	101.685640

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
24	Desa Petaling	0.0	0	3.083310	101.704380
23	Dang Wangi	0.0	0	3.156685	101.698076
22	Damansara, Kuala Lumpur	0.0	0	3.138759	101.684046
21	Damansara Town Centre	0.0	0	3.138759	101.684046
20	Damansara Heights	0.0	0	3.147980	101.667980
19	Chow Kit	0.0	0	3.163590	101.698110

Analyse Cluster 1

In [60]:
```
kl_rest_merged.loc[kl_rest_merged['Cluster Labels'] == 1]
```

Out[60]:

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
58	Taman Ibukota	0.01	1	3.21216	101.71540
32	Kampung Datuk Keramat	0.01	1	3.16640	101.73046
60	Taman Melati	0.01	1	3.22357	101.72399
34	Kepong	0.01	1	3.21750	101.63763

Analyse Cluster 2

In [61]:
```
kl_rest_merged.loc[kl_rest_merged['Cluster Labels'] == 2]
```

Out[61]:

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude

Finally, we have information pertaining to shopping malls in kl_merged and American restaurants in kl_rest_merged. This will be used with the equivalent ones of Dubai city for further analysis later.

12. Analyse each neighborhood in Dubai

Every analysis that we did for KL city will be performed for DXB as well

In [62]:
```
# one hot encoding
dxb_onehot = pd.get_dummies(dxb_venues_df[['VenueCategory']], prefix="", prefi
x_sep="")

# add neighborhood column back to dataframe
dxb_onehot['Neighborhoods'] = dxb_venues_df['Neighborhood']

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

print(dxb_onehot.shape)
dxb_onehot.head()
```

```
(5572, 268)
```

Out[62]:

	Neighborhoods	Accessories Store	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Lounge	Airpo Servic
0	Abu Hail	0	0	0	0	0	0	0
1	Abu Hail	0	0	0	0	0	0	0
2	Abu Hail	0	0	0	0	0	0	0
3	Abu Hail	0	0	0	0	0	0	0
4	Abu Hail	0	0	0	0	0	0	0

In [63]:
```
dxb_grouped = dxb_onehot.groupby(["Neighborhoods"]).mean().reset_index()

print(dxb_grouped.shape)
dxb_grouped
```

```
(67, 268)
```

Out[63]:

	Neighborhoods	Accessories Store	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Lounge	A S
0	Abu Hail	0.010000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
1	Al Amardhi	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
2	Al Bada	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
3	Al Baraha	0.000000	0.010000	0.000000	0.00000	0.00	0.000000	0.0
4	Al Barsha	0.010000	0.010000	0.020000	0.00000	0.00	0.000000	0.0
5	Al Buteen	0.000000	0.010000	0.000000	0.00000	0.00	0.000000	0.0
6	Al Dhagaya	0.000000	0.010000	0.000000	0.00000	0.00	0.000000	0.0
7	Al Garhoud	0.010000	0.000000	0.000000	0.01000	0.00	0.100000	0.0
8	Al Hamriya	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
9	Al Hamriya Port	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
10	Al Hudaiba	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
11	Al Jaddaf	0.000000	0.000000	0.013333	0.00000	0.00	0.000000	0.0
12	Al Jafilia	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
13	Al Karama	0.000000	0.000000	0.010000	0.00000	0.00	0.000000	0.0
14	Al Kefaf	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
15	Al Khabisi	0.000000	0.000000	0.000000	0.00000	0.00	0.050000	0.0
16	Al Mamzar	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
17	Al Manara	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
18	Al Mankhool	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
19	Al Markada	0.000000	0.000000	0.010000	0.00000	0.00	0.000000	0.0
20	Al Mizhar	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
21	Al Muraqqabat	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
22	Al Murar	0.000000	0.010000	0.000000	0.00000	0.00	0.000000	0.0
23	Al Muteena	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
24	Al Nahda	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
25	Al Quoz	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
26	Al Qusais	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
27	Al Ras	0.000000	0.010000	0.000000	0.00000	0.00	0.000000	0.0
28	Al Rashidiya	0.000000	0.000000	0.000000	0.00000	0.00	0.018182	0.0
29	Al Rifa	0.000000	0.010000	0.000000	0.00000	0.00	0.000000	0.0

	Neighborhoods	Accessories Store	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Lounge	A S
30	Al Rigga	0.000000	0.010000	0.000000	0.00000	0.00	0.000000	0.0
31	Al Sabkha	0.000000	0.010000	0.000000	0.00000	0.00	0.000000	0.0
32	Al Safa	0.000000	0.011905	0.000000	0.00000	0.00	0.000000	0.0
33	Al Satwa	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
34	Al Shindagha	0.000000	0.010000	0.000000	0.00000	0.00	0.000000	0.0
35	Al Souk Al Kabir	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
36	Al Sufouh	0.000000	0.000000	0.010000	0.00000	0.00	0.000000	0.0
37	Al Twar	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
38	Al Waheda	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
39	Al Warqaa	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
40	Al Wasl	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
41	Ayal Nasir	0.000000	0.010000	0.000000	0.00000	0.00	0.000000	0.0
42	Bu Kadra	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
43	Business Bay	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
44	Dubai International Airport	0.010000	0.000000	0.000000	0.01000	0.01	0.070000	0.0
45	Dubai International City	0.000000	0.010000	0.000000	0.00000	0.00	0.000000	0.0
46	Dubai Marina	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
47	Emirates Hills	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
48	Hor Al Anz	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
49	Jebel Ali	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
50	Jumeirah	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
51	Jumeirah Islands	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
52	Jumeirah Lake Towers	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
53	Mirdif	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
54	Muhaisnah	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
55	Nad Al Hammar	0.023810	0.000000	0.000000	0.02381	0.00	0.000000	0.0

	Neighborhoods	Accessories Store	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Lounge	A S
56	Nad Al Sheba	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
57	Nad Shamma	0.000000	0.000000	0.000000	0.00000	0.00	0.040816	0.0
58	Naif	0.000000	0.010000	0.000000	0.00000	0.00	0.000000	0.0
59	Oud Metha	0.000000	0.000000	0.010000	0.00000	0.00	0.000000	0.0
60	Port Rashid	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
61	Port Saeed	0.000000	0.000000	0.010000	0.00000	0.00	0.010000	0.0
62	Ras Al Khor	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
63	Ras Al Khor Industrial Area	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
64	Rigga Al Buteen	0.000000	0.000000	0.000000	0.00000	0.00	0.000000	0.0
65	Umm Ramool	0.013889	0.000000	0.000000	0.00000	0.00	0.083333	0.0
66	Warisan	0.000000	0.010309	0.000000	0.00000	0.00	0.000000	0.0

In [64]:
```
dxb_shopping_malls = dxb_grouped[["Neighborhoods","Shopping Mall"]]
dxb_shopping_malls.head()
```

Out[64]:

	Neighborhoods	Shopping Mall
0	Abu Hail	0.00
1	Al Amardhi	0.00
2	Al Bada	0.02
3	Al Baraha	0.01
4	Al Barsha	0.01

Use KMeans Clustering

In [65]:
```
# set number of clusters
kclusters = 3
dxb_clustering = dxb_shopping_malls.drop(["Neighborhoods"], 1)
# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(dxb_clustering)

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

Out[65]:
```
array([0, 0, 1, 0, 0, 0, 0, 0, 0, 1], dtype=int32)
```

In [66]:
```
# create a new dataframe that includes the cluster as well as the top 10 venue
s for each neighborhood.
dxb_merged = dxb_shopping_malls.copy()
# add clustering labels
dxb_merged["Cluster Labels"] = kmeans.labels_
```

In [67]:
```
dxb_merged.rename(columns={"Neighborhoods": "Neighborhood"}, inplace=True)
dxb_merged.head()
```

Out[67]:

	Neighborhood	Shopping Mall	Cluster Labels
0	Abu Hail	0.00	0
1	Al Amardhi	0.00	0
2	Al Bada	0.02	1
3	Al Baraha	0.01	0
4	Al Barsha	0.01	0

In [68]:
```
dxb_merged = dxb_merged.join(dxb_df.set_index("DXB_Neighborhood"), on="Neighbo
rhood")
```

In [69]:
```
print(dxb_merged.shape)
dxb_merged.head()
```

```
(68, 5)
```

Out[69]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
0	Abu Hail	0.00	0	25.283080	55.334350
1	Al Amardhi	0.00	0	25.198412	55.521244
2	Al Bada	0.02	1	25.218610	55.264060
3	Al Baraha	0.01	0	25.282800	55.316780
4	Al Barsha	0.01	0	25.105640	55.200570

In [70]:
```
print(dxb_merged.shape)
dxb_merged.sort_values(["Cluster Labels"], inplace=True)
dxb_merged
```

```
(68, 5)
```

Out[70]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
0	Abu Hail	0.000000	0	25.283080	55.334350
30	Al Rigga	0.010000	0	25.267060	55.308900
31	Al Sabkha	0.010000	0	25.268950	55.302570
33	Al Satwa	0.010000	0	25.226130	55.280370
34	Al Shindagha	0.010000	0	25.269900	55.289840
35	Al Souk Al Kabir	0.010000	0	25.259800	55.293960
37	Al Twar	0.000000	0	25.262530	55.382250
38	Al Waheda	0.000000	0	25.291730	55.338220
41	Ayal Nasir	0.010000	0	25.272550	55.303130
42	Bu Kadra	0.000000	0	25.181420	55.328890
43	Business Bay	0.010000	0	25.187970	55.262730
44	Dubai International Airport	0.000000	0	25.258970	55.370800
46	Dubai Marina	0.000000	0	25.088910	55.144360
48	Hor Al Anz	0.000000	0	25.277430	55.337460
50	Jumeirah	0.010000	0	25.193590	55.241370
51	Jumeirah Islands	0.010000	0	25.066430	55.144100
52	Jumeirah Lake Towers	0.010000	0	25.073060	55.143670
54	Muhaisnah	0.000000	0	25.257750	55.416900
55	Nad Al Hammar	0.000000	0	25.200990	55.376400
56	Nad Al Sheba	0.000000	0	25.144360	55.350400
58	Naif	0.010000	0	25.271650	55.304620
59	Oud Metha	0.000000	0	25.235160	55.314500
60	Port Rashid	0.010000	0	25.255540	55.283780
61	Port Saeed	0.010000	0	25.256210	55.330700
62	Ras Al Khor	0.000000	0	25.185190	55.330390
64	Rigga Al Buteen	0.010000	0	25.262050	55.316490
29	Al Rifa	0.010000	0	25.099290	55.203850
27	Al Ras	0.010000	0	25.267580	55.294590
32	Al Safa	0.011905	0	25.168340	55.229780
7	Al Garhoud	0.000000	0	25.243370	55.352670
12	Al Jafilia	0.010000	0	25.233420	55.290010
13	Al Karama	0.000000	0	25.245290	55.303640

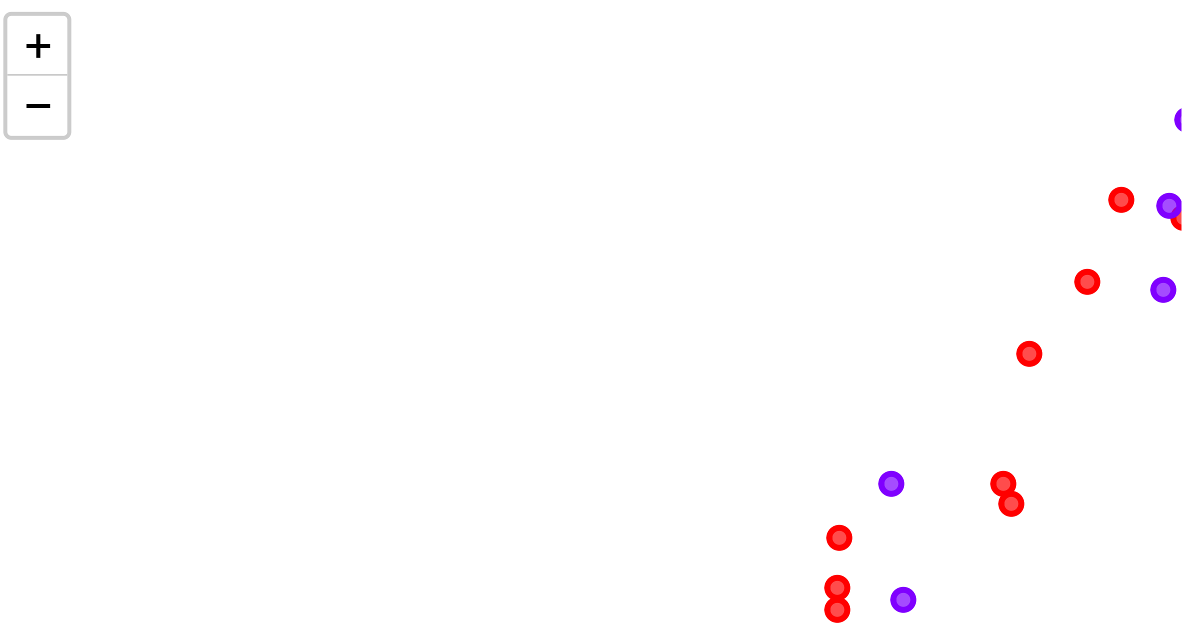
	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
26	Al Qusais	0.013514	0	25.270480	55.385070
13	Al Karama	0.000000	0	25.245290	55.303640
8	Al Hamriya	0.000000	0	25.256960	55.302460
14	Al Kefaf	0.000000	0	25.238090	55.297790
15	Al Khabisi	0.000000	0	25.271770	55.337620
16	Al Mamzar	0.000000	0	25.300940	55.340030
17	Al Manara	0.000000	0	25.145790	55.209580
11	Al Jaddaf	0.013333	0	25.220540	55.341660
10	Al Hudaiba	0.000000	0	25.237130	55.277070
18	Al Mankhool	0.000000	0	25.245160	55.293330
19	Al Markada	0.000000	0	25.241020	55.304880
6	Al Dhagaya	0.010000	0	25.272170	55.301570
5	Al Buteen	0.010000	0	25.269250	55.299440
4	Al Barsha	0.010000	0	25.105640	55.200570
22	Al Murar	0.010000	0	25.278350	55.308170
3	Al Baraha	0.010000	0	25.282800	55.316780
23	Al Muteena	0.010000	0	25.272280	55.322910
1	Al Amardhi	0.000000	0	25.198412	55.521244
2	Al Bada	0.020000	1	25.218610	55.264060
9	Al Hamriya Port	0.019231	1	25.298710	55.335460
57	Nad Shamma	0.040816	1	25.221280	55.379330
28	Al Rashidiya	0.036364	1	25.225430	55.390090
47	Emirates Hills	0.027778	1	25.069530	55.166200
49	Jebel Ali	0.023810	1	25.027770	55.126730
45	Dubai International City	0.030000	1	25.176790	55.410880
40	Al Wasl	0.020000	1	25.191810	55.257830
39	Al Warqaa	0.038462	1	25.191320	55.422330
36	Al Sufouh	0.020000	1	25.105620	55.162480
21	Al Muraqqabat	0.020000	1	25.268360	55.326340
65	Umm Ramool	0.027778	1	25.231600	55.377240
24	Al Nahda	0.020000	1	25.293090	55.379620
25	Al Quoz	0.016393	1	25.165900	55.255850

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
53	Mirdif	0.041667	1	25.226140	55.424210
66	Warisan	0.020619	1	25.167440	55.407080
20	Al Mizhar	0.085106	2	25.245060	55.439790
63	Ras Al Khor Industrial Area	0.055556	2	25.178100	55.368860

In [71]:

```
#Mapping clusters in map
address = 'Dubai, UAE'

geolocator = Nominatim(user_agent="my-application")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude


dxb_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(dxb_merged['Latitude'], dxb_merged['Longitud
e'], dxb_merged['Neighborhood'], dxb_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' - Cluster ' + str(cluster), parse_html=T
rue)
    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(dxb_map_clusters)

dxb_map_clusters
```

Out[71]:



Leaflet (https://leafletjs.com) | Data by © OpenStreetMap (http://openstreetmap.org), under ODbL (http://www.openstreetmap.org/copyright).

Similar to KL, we have Shopping malls information in dxb_merged dataframe for DXB city.

Analyze Cluster 0

In [72]:
```
dxb_merged.loc[dxb_merged['Cluster Labels'] == 0]
```

Out[72]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
0	Abu Hail	0.000000	0	25.283080	55.334350
30	Al Rigga	0.010000	0	25.267060	55.308900
31	Al Sabkha	0.010000	0	25.268950	55.302570
33	Al Satwa	0.010000	0	25.226130	55.280370
34	Al Shindagha	0.010000	0	25.269900	55.289840
35	Al Souk Al Kabir	0.010000	0	25.259800	55.293960
37	Al Twar	0.000000	0	25.262530	55.382250
38	Al Waheda	0.000000	0	25.291730	55.338220
41	Ayal Nasir	0.010000	0	25.272550	55.303130
42	Bu Kadra	0.000000	0	25.181420	55.328890
43	Business Bay	0.010000	0	25.187970	55.262730
44	Dubai International Airport	0.000000	0	25.258970	55.370800
46	Dubai Marina	0.000000	0	25.088910	55.144360
48	Hor Al Anz	0.000000	0	25.277430	55.337460
50	Jumeirah	0.010000	0	25.193590	55.241370
51	Jumeirah Islands	0.010000	0	25.066430	55.144100
52	Jumeirah Lake Towers	0.010000	0	25.073060	55.143670
54	Muhaisnah	0.000000	0	25.257750	55.416900
55	Nad Al Hammar	0.000000	0	25.200990	55.376400
56	Nad Al Sheba	0.000000	0	25.144360	55.350400
58	Naif	0.010000	0	25.271650	55.304620
59	Oud Metha	0.000000	0	25.235160	55.314500
60	Port Rashid	0.010000	0	25.255540	55.283780
61	Port Saeed	0.010000	0	25.256210	55.330700
62	Ras Al Khor	0.000000	0	25.185190	55.330390
64	Rigga Al Buteen	0.010000	0	25.262050	55.316490
29	Al Rifa	0.010000	0	25.099290	55.203850
27	Al Ras	0.010000	0	25.267580	55.294590
32	Al Safa	0.011905	0	25.168340	55.229780
7	Al Garhoud	0.000000	0	25.243370	55.352670
12	Al Jafilia	0.010000	0	25.233420	55.290010
13	Al Karama	0.000000	0	25.245290	55.303640

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
26	Al Qusais	0.013514	0	25.270480	55.385070
13	Al Karama	0.000000	0	25.245290	55.303640
8	Al Hamriya	0.000000	0	25.256960	55.302460
14	Al Kefaf	0.000000	0	25.238090	55.297790
15	Al Khabisi	0.000000	0	25.271770	55.337620
16	Al Mamzar	0.000000	0	25.300940	55.340030
17	Al Manara	0.000000	0	25.145790	55.209580
11	Al Jaddaf	0.013333	0	25.220540	55.341660
10	Al Hudaiba	0.000000	0	25.237130	55.277070
18	Al Mankhool	0.000000	0	25.245160	55.293330
19	Al Markada	0.000000	0	25.241020	55.304880
6	Al Dhagaya	0.010000	0	25.272170	55.301570
5	Al Buteen	0.010000	0	25.269250	55.299440
4	Al Barsha	0.010000	0	25.105640	55.200570
22	Al Murar	0.010000	0	25.278350	55.308170
3	Al Baraha	0.010000	0	25.282800	55.316780
23	Al Muteena	0.010000	0	25.272280	55.322910
1	Al Amardhi	0.000000	0	25.198412	55.521244

Analyze Cluster 1

In [73]:
```
dxb_merged.loc[dxb_merged['Cluster Labels'] == 1]
```

Out[73]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
2	Al Bada	0.020000	1	25.21861	55.26406
9	Al Hamriya Port	0.019231	1	25.29871	55.33546
57	Nad Shamma	0.040816	1	25.22128	55.37933
28	Al Rashidiya	0.036364	1	25.22543	55.39009
47	Emirates Hills	0.027778	1	25.06953	55.16620
49	Jebel Ali	0.023810	1	25.02777	55.12673
45	Dubai International City	0.030000	1	25.17679	55.41088
40	Al Wasl	0.020000	1	25.19181	55.25783
39	Al Warqaa	0.038462	1	25.19132	55.42233
36	Al Sufouh	0.020000	1	25.10562	55.16248
21	Al Muraqqabat	0.020000	1	25.26836	55.32634
65	Umm Ramool	0.027778	1	25.23160	55.37724
24	Al Nahda	0.020000	1	25.29309	55.37962
25	Al Quoz	0.016393	1	25.16590	55.25585
53	Mirdif	0.041667	1	25.22614	55.42421
66	Warisan	0.020619	1	25.16744	55.40708

Analyze Cluster 2

In [74]:
```
dxb_merged.loc[dxb_merged['Cluster Labels'] == 2]
```

Out[74]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
20	Al Mizhar	0.085106	2	25.24506	55.43979
63	Ras Al Khor Industrial Area	0.055556	2	25.17810	55.36886

And now we have shopping mall information in Dubai with its cluster information in dxb_merged dataframe. Let's now dig in and do a similar analysis for American restaurants there(This information can be stored in dxb_rest_merged dataframe)

In [75]:
```
#Repeating the above process for American Restaurants in Dubai
dxb_restaurants = dxb_grouped[["Neighborhoods","American Restaurant"]]
dxb_restaurants.head()
```

Out[75]:

	Neighborhoods	American Restaurant
0	Abu Hail	0.01
1	Al Amardhi	0.00
2	Al Bada	0.02
3	Al Baraha	0.02
4	Al Barsha	0.02

In [76]:
```
dxb_restaurants = dxb_grouped[["Neighborhoods","American Restaurant"]]
dxb_restaurants.head()

kclusters = 3

dxb_restaurants_clustering = dxb_restaurants.drop(["Neighborhoods"], 1)

dxb_restaurants = dxb_grouped[["Neighborhoods","American Restaurant"]]
dxb_restaurants.head()
kclusters = 3
dxb_restaurants_clustering = dxb_restaurants.drop(["Neighborhoods"], 1)
# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(dxb_restaurants_clus
tering)
kmeans.labels_[0:10]
# create a new dataframe that includes the cluster as well as the top 10 venue
s for each neighborhood.
dxb_rest_merged = dxb_restaurants.copy()
# add clustering labels
dxb_rest_merged["Cluster Labels"] = kmeans.labels_
dxb_rest_merged.rename(columns={"Neighborhoods": "Neighborhood"}, inplace=True
)

dxb_rest_merged = dxb_rest_merged.join(dxb_df.set_index("DXB_Neighborhood"), o
n="Neighborhood")
dxb_rest_merged.sort_values(["Cluster Labels"], inplace=True)
dxb_rest_merged
```

Out[76]:

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
32	Al Safa	0.000000	0	25.168340	55.229780
29	Al Rifa	0.000000	0	25.099290	55.203850
65	Umm Ramool	0.000000	0	25.231600	55.377240
33	Al Satwa	0.000000	0	25.226130	55.280370
34	Al Shindagha	0.000000	0	25.269900	55.289840
35	Al Souk Al Kabir	0.000000	0	25.259800	55.293960
36	Al Sufouh	0.000000	0	25.105620	55.162480
37	Al Twar	0.000000	0	25.262530	55.382250
38	Al Waheda	0.000000	0	25.291730	55.338220
39	Al Warqaa	0.000000	0	25.191320	55.422330
42	Bu Kadra	0.000000	0	25.181420	55.328890
44	Dubai International Airport	0.000000	0	25.258970	55.370800
45	Dubai International City	0.000000	0	25.176790	55.410880
47	Emirates Hills	0.000000	0	25.069530	55.166200
48	Hor Al Anz	0.000000	0	25.277430	55.337460
49	Jebel Ali	0.000000	0	25.027770	55.126730
50	Jumeirah	0.000000	0	25.193590	55.241370
54	Muhaisnah	0.000000	0	25.257750	55.416900
55	Nad Al Hammar	0.000000	0	25.200990	55.376400
56	Nad Al Sheba	0.000000	0	25.144360	55.350400
57	Nad Shamma	0.000000	0	25.221280	55.379330
59	Oud Metha	0.000000	0	25.235160	55.314500
62	Ras Al Khor	0.000000	0	25.185190	55.330390
63	Ras Al Khor Industrial Area	0.000000	0	25.178100	55.368860
27	Al Ras	0.000000	0	25.267580	55.294590
26	Al Qusais	0.000000	0	25.270480	55.385070
66	Warisan	0.000000	0	25.167440	55.407080
19	Al Markada	0.000000	0	25.241020	55.304880
13	Al Karama	0.000000	0	25.245290	55.303640
14	Al Kefaf	0.000000	0	25.238090	55.297790

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
12	Al Jafilia	0.000000	0	25.233420	55.290010
16	Al Mamzar	0.000000	0	25.300940	55.340030
25	Al Quoz	0.000000	0	25.165900	55.255850
13	Al Karama	0.000000	0	25.245290	55.303640
1	Al Amardhi	0.000000	0	25.198412	55.521244
8	Al Hamriya	0.000000	0	25.256960	55.302460
9	Al Hamriya Port	0.000000	0	25.298710	55.335460
3	Al Baraha	0.020000	1	25.282800	55.316780
43	Business Bay	0.020000	1	25.187970	55.262730
11	Al Jaddaf	0.026667	1	25.220540	55.341660
58	Naif	0.020000	1	25.271650	55.304620
52	Jumeirah Lake Towers	0.020000	1	25.073060	55.143670
4	Al Barsha	0.020000	1	25.105640	55.200570
24	Al Nahda	0.020000	1	25.293090	55.379620
40	Al Wasl	0.020000	1	25.191810	55.257830
61	Port Saeed	0.020000	1	25.256210	55.330700
2	Al Bada	0.020000	1	25.218610	55.264060
20	Al Mizhar	0.021277	1	25.245060	55.439790
21	Al Muraqqabat	0.040000	1	25.268360	55.326340
22	Al Murar	0.020000	1	25.278350	55.308170
28	Al Rashidiya	0.018182	1	25.225430	55.390090
23	Al Muteena	0.020000	1	25.272280	55.322910
41	Ayal Nasir	0.020000	1	25.272550	55.303130
6	Al Dhagaya	0.020000	1	25.272170	55.301570
5	Al Buteen	0.010000	2	25.269250	55.299440
60	Port Rashid	0.010000	2	25.255540	55.283780
64	Rigga Al Buteen	0.010000	2	25.262050	55.316490
46	Dubai Marina	0.010000	2	25.088910	55.144360
53	Mirdif	0.010417	2	25.226140	55.424210
51	Jumeirah Islands	0.010000	2	25.066430	55.144100
10	Al Hudaiba	0.010000	2	25.237130	55.277070
15	Al Khabisi	0.010000	2	25.271770	55.337620

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
17	Al Manara	0.010000	2	25.145790	55.209580
18	Al Mankhool	0.010000	2	25.245160	55.293330
31	Al Sabkha	0.010000	2	25.268950	55.302570
30	Al Rigga	0.010000	2	25.267060	55.308900
7	Al Garhoud	0.010000	2	25.243370	55.352670
0	Abu Hail	0.010000	2	25.283080	55.334350

Analyse Cluster 0

In [77]:
```
dxb_rest_merged.loc[dxb_rest_merged['Cluster Labels'] == 0]
```

Out[77]:

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
32	Al Safa	0.0	0	25.168340	55.229780
29	Al Rifa	0.0	0	25.099290	55.203850
65	Umm Ramool	0.0	0	25.231600	55.377240
33	Al Satwa	0.0	0	25.226130	55.280370
34	Al Shindagha	0.0	0	25.269900	55.289840
35	Al Souk Al Kabir	0.0	0	25.259800	55.293960
36	Al Sufouh	0.0	0	25.105620	55.162480
37	Al Twar	0.0	0	25.262530	55.382250
38	Al Waheda	0.0	0	25.291730	55.338220
39	Al Warqaa	0.0	0	25.191320	55.422330
42	Bu Kadra	0.0	0	25.181420	55.328890
44	Dubai International Airport	0.0	0	25.258970	55.370800
45	Dubai International City	0.0	0	25.176790	55.410880
47	Emirates Hills	0.0	0	25.069530	55.166200
48	Hor Al Anz	0.0	0	25.277430	55.337460
49	Jebel Ali	0.0	0	25.027770	55.126730
50	Jumeirah	0.0	0	25.193590	55.241370
54	Muhaisnah	0.0	0	25.257750	55.416900
55	Nad Al Hammar	0.0	0	25.200990	55.376400
56	Nad Al Sheba	0.0	0	25.144360	55.350400
57	Nad Shamma	0.0	0	25.221280	55.379330
59	Oud Metha	0.0	0	25.235160	55.314500
62	Ras Al Khor	0.0	0	25.185190	55.330390
63	Ras Al Khor Industrial Area	0.0	0	25.178100	55.368860
27	Al Ras	0.0	0	25.267580	55.294590
26	Al Qusais	0.0	0	25.270480	55.385070
66	Warisan	0.0	0	25.167440	55.407080
19	Al Markada	0.0	0	25.241020	55.304880
13	Al Karama	0.0	0	25.245290	55.303640
14	Al Kefaf	0.0	0	25.238090	55.297790

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
12	Al Jafilia	0.0	0	25.233420	55.290010
16	Al Mamzar	0.0	0	25.300940	55.340030
25	Al Quoz	0.0	0	25.165900	55.255850
13	Al Karama	0.0	0	25.245290	55.303640
1	Al Amardhi	0.0	0	25.198412	55.521244
8	Al Hamriya	0.0	0	25.256960	55.302460
9	Al Hamriya Port	0.0	0	25.298710	55.335460

Analyse Cluster 1

In [78]:
```
dxb_rest_merged.loc[dxb_rest_merged['Cluster Labels'] == 1]
```

Out[78]:

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
3	Al Baraha	0.020000	1	25.28280	55.31678
43	Business Bay	0.020000	1	25.18797	55.26273
11	Al Jaddaf	0.026667	1	25.22054	55.34166
58	Naif	0.020000	1	25.27165	55.30462
52	Jumeirah Lake Towers	0.020000	1	25.07306	55.14367
4	Al Barsha	0.020000	1	25.10564	55.20057
24	Al Nahda	0.020000	1	25.29309	55.37962
40	Al Wasl	0.020000	1	25.19181	55.25783
61	Port Saeed	0.020000	1	25.25621	55.33070
2	Al Bada	0.020000	1	25.21861	55.26406
20	Al Mizhar	0.021277	1	25.24506	55.43979
21	Al Muraqqabat	0.040000	1	25.26836	55.32634
22	Al Murar	0.020000	1	25.27835	55.30817
28	Al Rashidiya	0.018182	1	25.22543	55.39009
23	Al Muteena	0.020000	1	25.27228	55.32291
41	Ayal Nasir	0.020000	1	25.27255	55.30313
6	Al Dhagaya	0.020000	1	25.27217	55.30157

Analyse Cluster 2

In [79]:
```
dxb_rest_merged.loc[dxb_rest_merged['Cluster Labels'] == 2]
```

Out[79]:

	Neighborhood	American Restaurant	Cluster Labels	Latitude	Longitude
5	Al Buteen	0.010000	2	25.26925	55.29944
60	Port Rashid	0.010000	2	25.25554	55.28378
64	Rigga Al Buteen	0.010000	2	25.26205	55.31649
46	Dubai Marina	0.010000	2	25.08891	55.14436
53	Mirdif	0.010417	2	25.22614	55.42421
51	Jumeirah Islands	0.010000	2	25.06643	55.14410
10	Al Hudaiba	0.010000	2	25.23713	55.27707
15	Al Khabisi	0.010000	2	25.27177	55.33762
17	Al Manara	0.010000	2	25.14579	55.20958
18	Al Mankhool	0.010000	2	25.24516	55.29333
31	Al Sabkha	0.010000	2	25.26895	55.30257
30	Al Rigga	0.010000	2	25.26706	55.30890
7	Al Garhoud	0.010000	2	25.24337	55.35267
0	Abu Hail	0.010000	2	25.28308	55.33435

12. Result Analysis

All the required information is now available in 4 data frames with each containing 3 clusters each.

Kuala Lumpur

Shopping Mall information in Kuala Lumpur is available in kl_merged

American Restaurant information in Kuala Lumpur is available in kl_rest_merged

Dubai

Shopping Mall information in Dubai is available in dxb_merged

American Restaurant information in Dubai is available in dxb_rest_merged

A detailed look of the above data frames with its cluster information reveals that, the more the values in the "Shopping Malls" column or the "American Restaurants" column, the higher the presence of such business in the locations in the cluster. It also depends on the number of neighborhoods in each cluster too. For example, take the restaurant information for Dubai. If looked closer the values of "American Restaurant" column in Cluster 2 is way less than that of Cluster 1. This means there is higher presence of American restaurants in the locations in Cluster 1 of dxb_rest_merged than that of Cluster 2.

So to draw conclusion, let's sum up the respective columns and corresponding number of neighborhoods

Let's begin by creating a data frame for summing up the results

In [80]:
```
row_names = {'KL_ShoppingMall', 'KL_AmericanRestaurant', 'DXB_ShoppingMall',
'DXB_AmericanRestaurant'}
col_names = {'Cluster 0', 'Cluster 1', 'Cluster 2'}
```

In [81]:
```
res_df = pd.DataFrame(index = row_names, columns = col_names)

#Sorting the indexes in alphabetical order
res_df = res_df.sort_index(ascending = 0)
res_df = res_df.sort_index(axis = 1)
```

In [82]:
```
#The resulting dataframe looks like this
res_df
```

Out[82]:

	Cluster 0	Cluster 1	Cluster 2
KL_ShoppingMall	NaN	NaN	NaN
KL_AmericanRestaurant	NaN	NaN	NaN
DXB_ShoppingMall	NaN	NaN	NaN
DXB_AmericanRestaurant	NaN	NaN	NaN

In [83]:
```
#Let's now begin to add data depending on different cluster values to datafram
e

row = 0
col = 0
cluster = {0,1,2}
lab = ('Shopping Mall', 'American Restaurant')

while col < 3:
    #print(col)
    for x in cluster:
        res_df.iloc[row,col] = round(kl_merged.loc[kl_merged['Cluster Labels']
== col][lab[0]].sum(),2)

    row = 2
    for x in cluster:
        res_df.iloc[row,col] = round(dxb_merged.loc[dxb_merged['Cluster Label
s'] == col][lab[0]].sum(),2)

    row = 1
    for x in cluster:
        res_df.iloc[row,col] = round(kl_rest_merged.loc[kl_rest_merged['Cluste
r Labels'] == col][lab[1]].sum(),2)

    row = 3
    for x in cluster:
        res_df.iloc[row,col] = round(dxb_rest_merged.loc[dxb_rest_merged['Clus
ter Labels'] == col][lab[1]].sum(),2)
    col = col + 1
    row = 0
```

In [84]:
```
#Once the corresponding cluster values are summed up, the resulting dataframe
 looks like this
res_df
```

Out[84]:

	Cluster 0	Cluster 1	Cluster 2
KL_ShoppingMall	0.38	0.51	0
KL_AmericanRestaurant	0	0.04	0
DXB_ShoppingMall	0.27	0.42	0.14
DXB_AmericanRestaurant	0	0.37	0.14

13. Conclusion

We need to compare the number of neighborhoods in each clusters of DXB and KL to draw insights from the above data.

From the resultant data frame it is pretty much clear that there is a high density of shopping malls in the locations in Cluster 0 and Cluster 1 of Kuala Lumpur. And there is very less density of shopping malls in locations of Cluster 2 in Kuala Lumpur.

Also analysing the values of shopping malls in Dubai, all the three clusters have sufficiently good amount of shopping malls. Therefore, it is not advised to start a new shopping mall in the vicinity of locations in either of the locations in Dubai clusters.

Therefore, we are advising AmerCave Group to start a new shopping mall in any of the locations of Cluster 2 in Kuala Lumpur, as there will be very less competition

A similar analysis can be done for American Restaurant. From the data frame it is pretty much clear that there is very less density of American restaurants in locations of various clusters in Kuala Lumpur. Also, from the values it is clear that Kuala Lumpur is not a potential market for American restaurants.

While on the other hand, there is sufficintly greater number of American Restaurants in locations of Cluster 1 and Cluster 2 of Dubai. And there is negligible amount of American restaurants in Dubai. The values of those in Cluster 1 & 2 reveals that Dubai is a potential market for American restaurants.

Therefore, we are advising AmerCave Group to start a new American restaurant in any of the locations of Cluster 0 in Dubai, as there will be very less competition and there is a potential for growth when comparing the values with other two clusters.

Thanks for going through this project

(All other requisites of report - methodology, discussion etc - are available in the report pdf in Github repository)

Arunjith M(arunjithece@gmail.com)