House to Rent in Selangor, Malaysia

Jayasantini Pandian

1.0 Introduction

In this section, we will discuss about the background and the problem to be solved

1.1 Background

Everyone needs a place to stay and some might have their own preferences. A good place to stay are where we have adequate amenities, the basic surroundings a person needs. This can be a clinic nearby, schools, hospitals, grocery shops, and bank. Entertainment places are also considered as important as others for some. Even if we were able to find a good place, a place we prefer, the rental or number of rooms will be a problem. For some, there will be parking trouble as there might be limited parking or free parking which depends on the time.

I am looking for a house to rent. I am doing a part-time art business such as graphite sketch, charcoal sketch and acrylic painting. Currently, due to lack of rooms, I had to combine my studio with my bedroom. This caused some imbalance in my room as I felt the space I need for my mini studio is expanding. Lack of space and having my important work papers around, it is easy for me to smudge the papers with my painting as well. Somehow, having a mini studio in my bedroom is disorganised in my view.

Currently I am living in a 3 bedroom house with my family members. Two rooms are occupied by my parents and brother, hence the studio was set up in my room. Since my business is growing and my art supplies are increasing, I would prefer to have a room allocated for my artspace. By that means, the mess will be in the studio room and will not extend to my bedroom. My current house is at Kajang, Selangor and I am looking for somewhere around Shah Alam.

1.2 Problem to be Solved

The challenge is to find a rental place in Shah Alam, Selangor which has some basic ammenities. The criterias I am looking for are as per below:

1. Rental price is not more than RM2,500.

2. A house consists of 4 bedrooms.

3. Basic ammenities such as clinic, grocery shop and good restaurants.

1.3 Interested Audience

This project would attract people who would like to move to their preferable city with certain criterias. The methodology applied in this project is generic, which can be applied for different criterias and places as well.

2.0 Data

In this section, we will discuss about data and resources

2.1 Required Data

Following data are used in this project:

1. List of Selangor's main addresses

2. Selangor state coordinates

3. List of rental details in Shah Alam, Selangor

4. Foursquare location data

List of Selangor's main address, Selangor state coordinates and list of rental details in Shah Alam, Selangor has been created and uploaded as CVS file in order to ease the file reading.

3.0 Methodology

In this section, we will discuss about how the data is used in this project.

Foursquare and geopy data will be used with the data mentioned above accordingly. List of Selangor's main addressess are merged with the Selangor coordinates and then merged with the rental price. To visualise the process, a few maps will be created by using foursquare and geopy which gives us better understanding on the area surrounding.

Below are the following key questions that will be answered in this project:

1. What is the rental price in the required area?

2. What are the houses that has 4 bedrooms and relevant rental?

3. What are the venues surrounding the preferred area?

4. What are the observations on overall data?

4.0 Findings

This section reflects on the findings step by step.

Install the necessary resources/items.

In [1]:
```
pip install geopy
```

```
Collecting geopy
  Downloading https://files.pythonhosted.org/packages/07/e1/9c72de674d5c2b8fcb0738a5ceeb5424941fefa080bfe4e240d0bacb5a38/geopy-2.0.0-py3-none-any.whl (111kB)
     |████████████████████████████████| 112kB 20.6MB/s eta 0:00:01
Collecting geographiclib<2,>=1.49 (from geopy)
  Downloading https://files.pythonhosted.org/packages/8b/62/26ec95a98ba64299163199e95ad1b0e34ad3f4e176e221c40245f211e425/geographiclib-1.50-py3-none-any.whl
Installing collected packages: geographiclib, geopy
Successfully installed geographiclib-1.50 geopy-2.0.0
Note: you may need to restart the kernel to use updated packages.
```

In [2]:
```
import pandas as pd
import requests
import numpy as np

from geopy.geocoders import Nominatim
from pandas.io.json import json_normalize

import folium

from sklearn.cluster import KMeans

import matplotlib.cm as cm
import matplotlib.colors as colors
```

Open Selangor list of addresses.

In [3]:
```
Selangor = open('selangor.csv', 'r').read()

print ('File downloaded and saved')
```

```
File downloaded and saved
```

In [4]:
```
data = pd.read_csv ('selangor.csv')

df = pd.DataFrame(data, columns= ['POSTCODE', 'CITY','ADDRESS'])

print(df)
```

```
     POSTCODE        CITY                                    ADDRESS
0       40000   Shah Alam                       Bukit Raja Selatan
1       40100   Shah Alam                          Persiaran Damai
2       40100   Shah Alam                        Persiaran Masjid
3       40100   Shah Alam                Persiaran Perbandaran
4       40100   Shah Alam           Shah Alam - Seksyen 9 - 13
..        ...         ...                                     ...
345     63300   Cyberjaya                Taman Pinggiran Cyber
346     64000        KLIA  Kuala Lumpur International Airport (KLIA)
347     68000      Ampang                             Zoo Negara
348     68100  Batu Caves                 Bandar Baru Selayang
349     68100  Batu Caves                             Batu Caves

[350 rows x 3 columns]
```

In [5]:
```
df.shape
```

Out[5]:
```
(350, 3)
```

In [6]:
```
df1 = df.groupby ('POSTCODE', sort=False).agg(','.join)

df1.reset_index(inplace=True)

df1.CITY=df1.CITY.str.split(',').apply(set).str.join(',')

df1
```

Out[6]:

	POSTCODE	CITY	ADDRESS
0	40000	Shah Alam	Bukit Raja Selatan
1	40100	Shah Alam	Persiaran Damai,Persiaran Masjid,Persiaran Per...
2	40150	Shah Alam	Batu Tiga,Bukit Cerakah,Bukit Jelutong,Desa Su...
3	40160	Shah Alam	Bandar Baru Sungai Buloh
4	40170	Shah Alam	Alam Budiman Seksyen U10,Setia Eco Park
...	...	...	...
266	63200	Cyberjaya	Majlis Perbandaran Sepang
267	63300	Cyberjaya	Jalan SP 2, Taman Sains Selangor 2,Taman Pingg...
268	64000	KLIA	Kuala Lumpur International Airport (KLIA)
269	68000	Ampang	Zoo Negara
270	68100	Batu Caves	Bandar Baru Selayang,Batu Caves

271 rows × 3 columns

In [7]:
```
df1.shape
```

Out[7]:
```
(271, 3)
```

Let's move on to find Latitude and Longitude of Selangor

In [8]:
```
lat_lon = pd.read_csv('Selangor_coordinates.csv')

lat_lon.head()
```

Out[8]:

	POSTCODE	LATITUDE	LONGITUDE
0	40000	3.0797	101.5186
1	40100	3.0799	101.5334
2	40150	3.1100	101.5538
3	40160	3.1643	101.5584
4	40170	3.0733	101.5054

In [9]:
```
df2 = pd.merge(df1,lat_lon,on='POSTCODE')
df2.head()
```

Out[9]:

	POSTCODE	CITY	ADDRESS	LATITUDE	LONGITUDE
0	40000	Shah Alam	Bukit Raja Selatan	3.0797	101.5186
1	40100	Shah Alam	Persiaran Damai,Persiaran Masjid,Persiaran Per...	3.0799	101.5334
2	40150	Shah Alam	Batu Tiga,Bukit Cerakah,Bukit Jelutong,Desa Su...	3.1100	101.5538
3	40160	Shah Alam	Bandar Baru Sungai Buloh	3.1643	101.5584
4	40170	Shah Alam	Alam Budiman Seksyen U10,Setia Eco Park	3.0733	101.5054

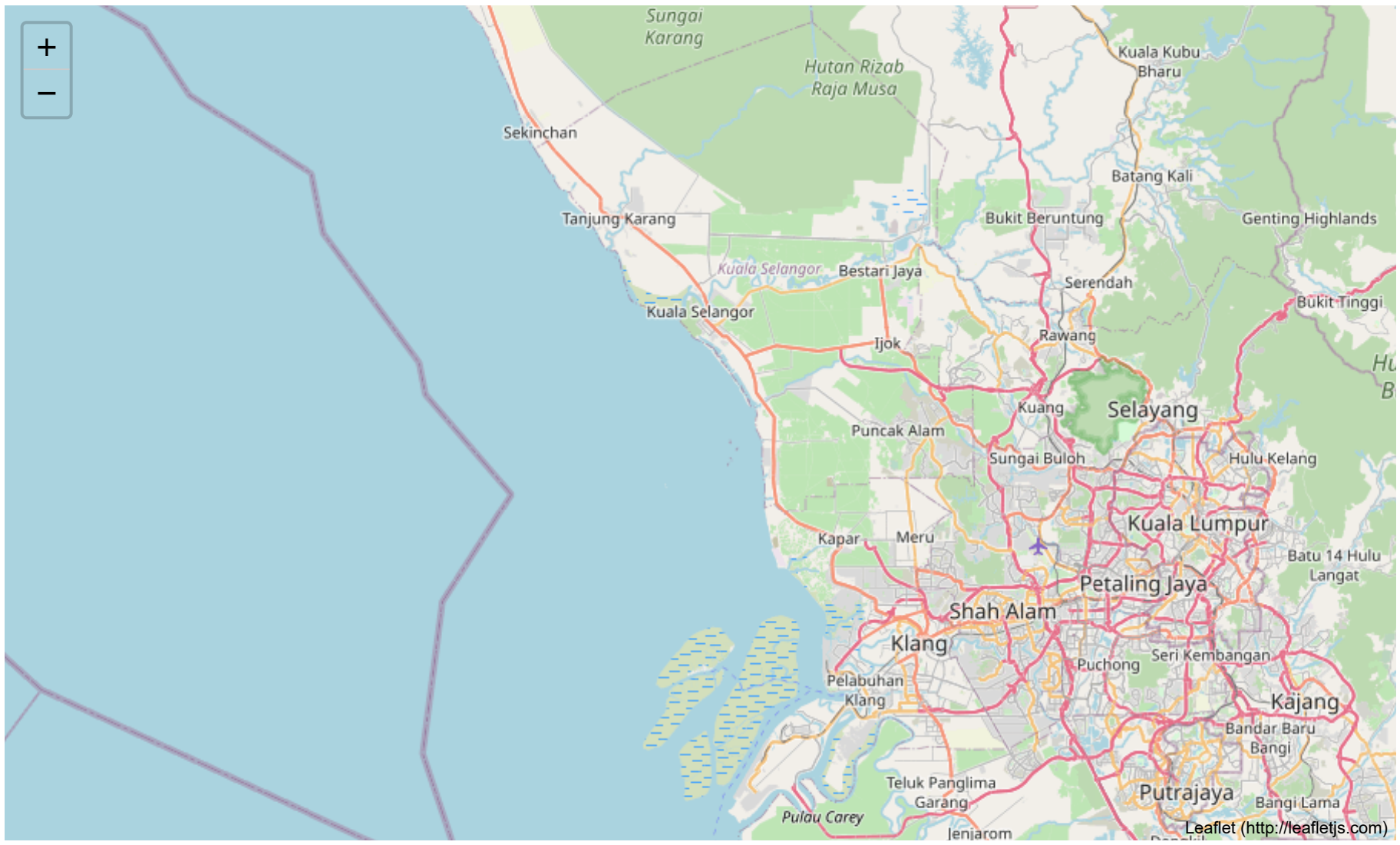
Now let's explore Selangor, Malaysia!

In [10]:
```
address = "Selangor"

geolocator = Nominatim(user_agent="selangor_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Selangor state is {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of Selangor state is 3.2083304, 101.304146.
```

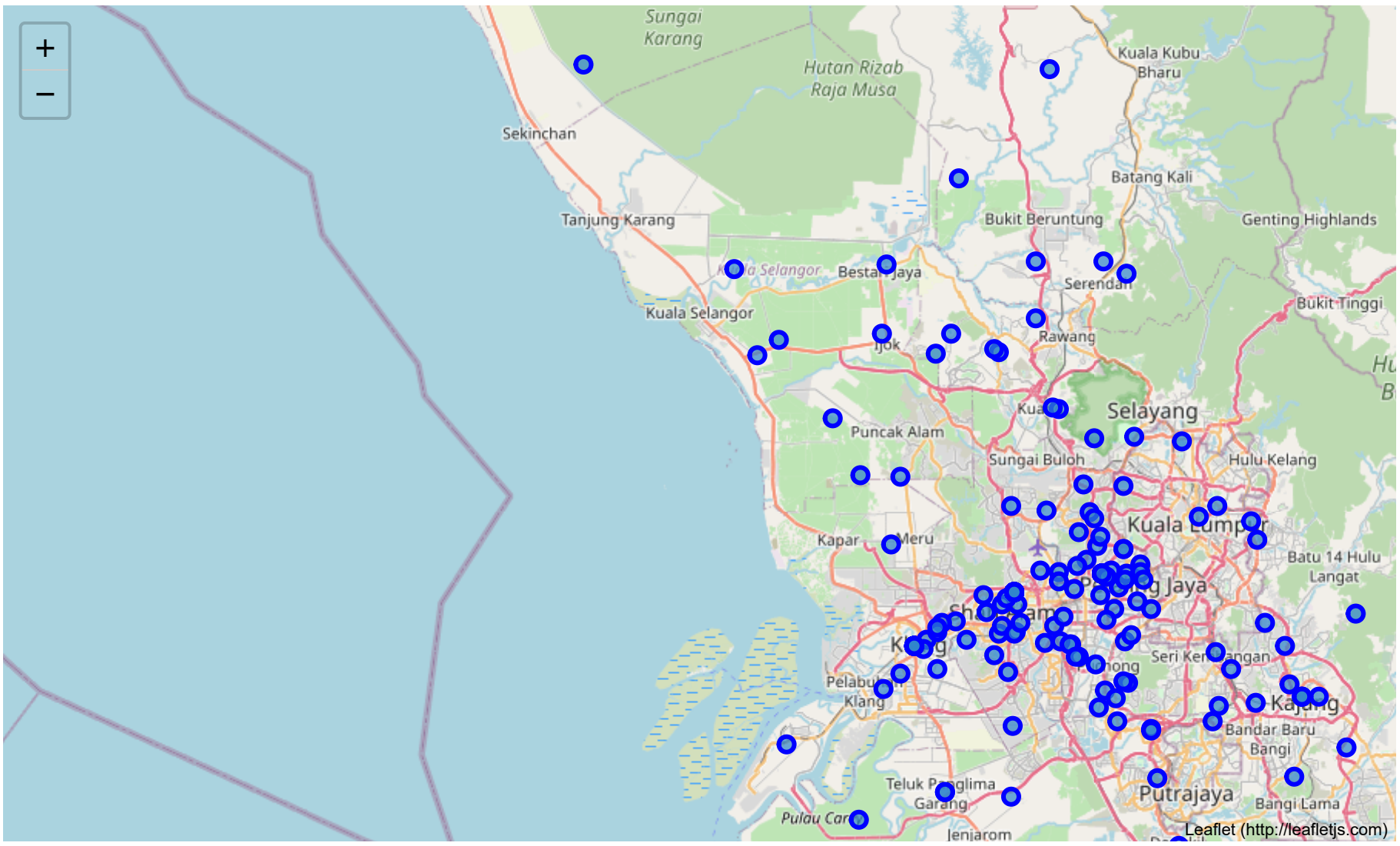
In [11]:
```
map_selangor = folium.Map(location=[latitude, longitude], zoom_start=10)
map_selangor
```

Out[11]:



In [12]:
```
for lat, lng, address, city in zip(
        df2['LATITUDE'],
        df2['LONGITUDE'],
        df2['CITY'],
        df2['ADDRESS']):
    label = '{}, {}'.format(address, city)
    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,
        parse_html=False).add_to(map_selangor)

map_selangor
```

Out[12]:



Now, let's narrow down to a specific part of Selangor!

In [13]:
```
df2_denc = df2[df2['CITY'].str.contains("Shah Alam")].reset_index(drop=True)
df2_denc.head()
```

Out[13]:

	POSTCODE	CITY	ADDRESS	LATITUDE	LONGITUDE
0	40000	Shah Alam	Bukit Raja Selatan	3.0797	101.5186
1	40100	Shah Alam	Persiaran Damai,Persiaran Masjid,Persiaran Per...	3.0799	101.5334
2	40150	Shah Alam	Batu Tiga,Bukit Cerakah,Bukit Jelutong,Desa Su...	3.1100	101.5538
3	40160	Shah Alam	Bandar Baru Sungai Buloh	3.1643	101.5584
4	40170	Shah Alam	Alam Budiman Seksyen U10,Setia Eco Park	3.0733	101.5054

It's time to define foursquare credential and get started with the quest!

In [14]:
```
CLIENT_ID = 'NMUQ4NSJQ24B1QS2CUQL1G5BWDE335MFSOCRA3EFEV13PH1X'
CLIENT_SECRET = 'LYOUHODAJH5SOLR3BI0WO2J3WSVBUBWYHLKKD2AVRNHXSBXZ'

VERSION = '20180604'
```

In [15]:
```
address_name = df2_denc.loc[0, 'ADDRESS']
print(f"The first address's name is '{address_name}'.")
```

```
The first address's name is 'Bukit Raja Selatan'.
```

In [16]:
```
address_latitude = df2_denc.loc[0, 'LATITUDE']
address_longitude = df2_denc.loc[0, 'LONGITUDE']

print( 'latitude and longitude values of {} are {}, {}.'.format(address_name,
                                                               address_latitude,
                                                               address_longitude))
```

```
latitude and longitude values of Bukit Raja Selatan are 3.0797, 101.5186.
```

In [17]:
```
LIMIT = 100
radius = 500
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.f
ormat(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    address_latitude,
    address_longitude,
    radius,
    LIMIT)

results = requests.get(url).json()

def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']

venues = results['response']['groups'][0]['items']
nearby_venues = pd.json_normalize(venues)

filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues =nearby_venues.loc[:, filtered_columns]

nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues
```

Out[17]:

	name	categories	lat	lng
0	Kompleks Kesenian Islam Shah Alam	Art Gallery	3.081152	101.522493
1	Shah Alam Lake Park	Garden	3.076369	101.515868
2	Kompleks Taman Seni Islam Selangor	History Museum	3.081232	101.522557
3	D'Arab Cafe	Middle Eastern Restaurant	3.078170	101.519039
4	Aroma Kitchen	Cafeteria	3.078156	101.518929
5	A&A Cafe	Café	3.078293	101.519084
6	Kedai Kopi, Tmn Tasik Shah Alam	Asian Restaurant	3.076918	101.516711
7	Kedai Kopi	Café	3.076930	101.516764

In [18]:
```
def getNearbyVenues(names, latitudes, longitudes, radius=500):
    venues_list=[]

    for name, lat, lng in zip(names, latitudes, longitudes):

        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&lim
it={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        results = requests.get(url).json()["response"]['groups'][0]['items']

        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['ADDRESS',
                  'Address Latitude',
                  'Address Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)

selangor_denc_venues = getNearbyVenues(names=df2_denc['ADDRESS'],
                                       latitudes=df2_denc['LATITUDE'],
                                       longitudes=df2_denc['LONGITUDE']
                                      )

selangor_denc_venues.head()
```

Out[18]:

	ADDRESS	Address Latitude	Address Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bukit Raja Selatan	3.0797	101.5186	Kompleks Kesenian Islam Shah Alam	3.081152	101.522493	Art Gallery
1	Bukit Raja Selatan	3.0797	101.5186	Shah Alam Lake Park	3.076369	101.515868	Garden
2	Bukit Raja Selatan	3.0797	101.5186	Kompleks Taman Seni Islam Selangor	3.081232	101.522557	History Museum
3	Bukit Raja Selatan	3.0797	101.5186	D'Arab Cafe	3.078170	101.519039	Middle Eastern Restaurant
4	Bukit Raja Selatan	3.0797	101.5186	Aroma Kitchen	3.078156	101.518929	Cafeteria

In [19]:
```
selangor_denc_venues.groupby('ADDRESS').count()
```

Out[19]:

ADDRESS	Address Latitude	Address Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Akauntan Negeri	6	6	6	6	6	6
Alam Budiman Seksyen U10,Setia Eco Park	25	25	25	25	25	25
Bahagian Pelesenan - Kementerian Penerangan	6	6	6	6	6	6
Bandar Baru Sungai Buloh	7	7	7	7	7	7
Batu Tiga,Bukit Cerakah,Bukit Jelutong,Desa Subang Permai	13	13	13	13	13	13
...	...	...	...	...	...	...
Telekom Malaysia Berhad	6	6	6	6	6	6
Tenaga Nasional Berhad	6	6	6	6	6	6
Unit Penyelidikan Sosio Ekonomi	6	6	6	6	6	6
Universiti Teknologi Mara (UiTM) Shah Alam	3	3	3	3	3	3
YAB Menteri Besar Selangor	13	13	13	13	13	13

80 rows × 6 columns

In [20]:
```
selangor_denc_onehot = pd.get_dummies(selangor_denc_venues[['Venue Category']], prefix="", prefix_sep="")

selangor_denc_onehot['ADDRESS'] = selangor_denc_venues['ADDRESS']

fixed_columns = [selangor_denc_onehot.columns[-1]] + list(selangor_denc_onehot.columns[:-1])
selangor_denc_onehot = selangor_denc_onehot[fixed_columns]

selangor_denc_onehot.head()
```

Out[20]:

	ADDRESS	American Restaurant	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	Auto Garage	Auto Workshop	Bakery	Bank	...	Sporting Goods Shop	Steakhouse	Street Art	Tea Room	Resta
0	Bukit Raja Selatan	0	0	1	0	0	0	0	0	0	...	0	0	0	0	
1	Bukit Raja Selatan	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
2	Bukit Raja Selatan	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
3	Bukit Raja Selatan	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
4	Bukit Raja Selatan	0	0	0	0	0	0	0	0	0	...	0	0	0	0	

5 rows × 114 columns

In [21]:
```
selangor_denc_grouped = selangor_denc_onehot.groupby('ADDRESS').mean().reset_index()
selangor_denc_grouped.head()
```

Out[21]:

	ADDRESS	American Restaurant	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	Auto Garage	Auto Workshop	Bakery	Bank	...	Sporting Goods Shop	Steakhouse	Street Art	Tea Room
0	Akauntan Negeri	0.0	0.0	0.0	0.00	0.000000	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.00	0.000000
1	Alam Budiman Seksyen U10,Setia Eco Park	0.0	0.0	0.0	0.08	0.080000	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.04	0.040000
2	Bahagian Pelesenan - Kementerian Penerangan	0.0	0.0	0.0	0.00	0.000000	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.00	0.000000
3	Bandar Baru Sungai Buloh	0.0	0.0	0.0	0.00	0.000000	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.00	0.142857
4	Batu Tiga,Bukit Cerakah,Bukit Jelutong,Desa Su...	0.0	0.0	0.0	0.00	0.076923	0.0	0.0	0.076923	0.0	...	0.0	0.0	0.00	0.000000

5 rows × 114 columns

In [22]:
```
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)
    return row_categories_sorted.index.values[0:num_top_venues]

num_top_venues = 10

indicators = ['st', 'nd', 'rd']

columns = ['ADDRESS']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))
```

In [23]:
```
address_venues_sorted = pd.DataFrame(columns=columns)
address_venues_sorted['ADDRESS'] = selangor_denc_grouped['ADDRESS']

for ind in np.arange(selangor_denc_grouped.shape[0]):
    address_venues_sorted.iloc[ind, 1:] = return_most_common_venues(selangor_denc_grouped.iloc[ind, :], num_top_venues
)

address_venues_sorted.head()
```

Out[23]:

	ADDRESS	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Akauntan Negeri	Golf Course	Gym	Café	Library	Dessert Shop	Women's Store	French Restaurant	Dive Bar	Donut Shop	Electronics Store
1	Alam Budiman Seksyen U10,Setia Eco Park	Burger Joint	Ice Cream Shop	Arts & Crafts Store	Asian Restaurant	Fast Food Restaurant	Soup Place	High School	Business Service	Print Shop	Food Court
2	Bahagian Pelesenan - Kementerian Penerangan	Golf Course	Gym	Café	Library	Dessert Shop	Women's Store	French Restaurant	Dive Bar	Donut Shop	Electronics Store
3	Bandar Baru Sungai Buloh	Malay Restaurant	Lounge	Tea Room	Coffee Shop	Field	Burger Joint	Women's Store	Donut Shop	Electronics Store	Farmers Market
4	Batu Tiga,Bukit Cerakah,Bukit Jelutong,Desa Su...	Music Venue	Bakery	Sandwich Place	Multiplex	Karaoke Bar	Bowling Alley	Pool Hall	Soccer Stadium	Fast Food Restaurant	Comfort Food Restaurant

Let's see how many clusters are there!

In [24]:
```
kclusters = 5

selangor_denc_grouped_clustering = selangor_denc_grouped.drop('ADDRESS', 1)

kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(selangor_denc_grouped_clustering)

kmeans.labels_[0:10]
```

Out[24]: array([0, 1, 0, 1, 1, 0, 1, 0, 0, 0], dtype=int32)

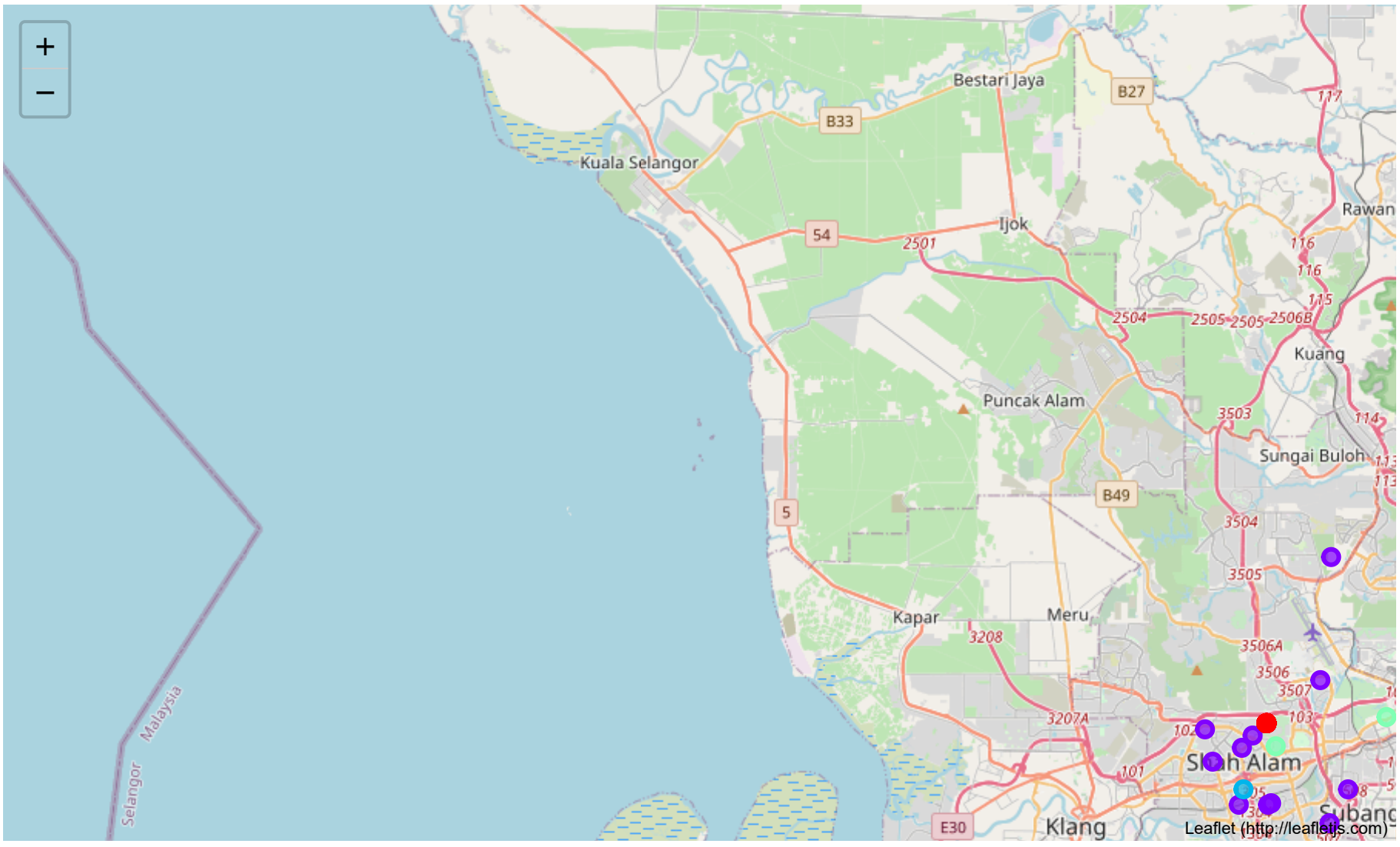
In [25]:
```
address_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

selangor_denc_merged = df2_denc

selangor_denc_merged = selangor_denc_merged.join(address_venues_sorted.set_index('ADDRESS'), on ='ADDRESS')

selangor_denc_merged.head()
```

Out[25]:

	POSTCODE	CITY	ADDRESS	LATITUDE	LONGITUDE	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Mo Commc Venu
0	40000	Shah Alam	Bukit Raja Selatan	3.0797	101.5186	1	Café	History Museum	Art Gallery	Asian Restaurant	Cafeteria	Garden	Midd Easte Restaura
1	40100	Shah Alam	Persiaran Damai,Persiaran Masjid,Persiaran Per...	3.0799	101.5334	3	Food Truck	Gym / Fitness Center	Motel	Dessert Shop	Women's Store	Fried Chicken Joint	Dive B
2	40150	Shah Alam	Batu Tiga,Bukit Cerakah,Bukit Jelutong,Desa Su...	3.1100	101.5538	1	Music Venue	Bakery	Sandwich Place	Multiplex	Karaoke Bar	Bowling Alley	Pool Ha
3	40160	Shah Alam	Bandar Baru Sungai Buloh	3.1643	101.5584	1	Malay Restaurant	Lounge	Tea Room	Coffee Shop	Field	Burger Joint	Women Sto
4	40170	Shah Alam	Alam Budiman Seksyen U10,Setia Eco Park	3.0733	101.5054	1	Burger Joint	Ice Cream Shop	Arts & Crafts Store	Asian Restaurant	Fast Food Restaurant	Soup Place	Hig Scho

In [26]:
```
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

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]

markers_colors = []
for lat, lon, poi, cluster in zip(
        selangor_denc_merged['LATITUDE'],
        selangor_denc_merged['LONGITUDE'],
        selangor_denc_merged['ADDRESS'],
        selangor_denc_merged['Cluster Labels']):

    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[26]:



In [27]:
```
selangor_denc_merged.loc[selangor_denc_merged['Cluster Labels'] == 0, selangor_denc_merged.columns[[1] + list(range(5, selangor_denc_merged.shape[1]))]]
```

Out[27]:

	CITY	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
11	Shah Alam	0	Golf Course	Gym	Café	Library	Dessert Shop	Women's Store	French Restaurant	Dive Bar	Donut Shop	Electronics Store
13	Shah Alam	0	Golf Course	Gym	Café	Library	Dessert Shop	Women's Store	French Restaurant	Dive Bar	Donut Shop	Electronics Store
14	Shah Alam	0	Golf Course	Gym	Café	Library	Dessert Shop	Women's Store	French Restaurant	Dive Bar	Donut Shop	Electronics Store
15	Shah Alam	0	Golf Course	Gym	Café	Library	Dessert Shop	Women's Store	French Restaurant	Dive Bar	Donut Shop	Electronics Store
17	Shah Alam	0	Golf Course	Gym	Café	Library	Dessert Shop	Women's Store	French Restaurant	Dive Bar	Donut Shop	Electronics Store
...	...	...	...	...	...	...	...	...	...	...	...	...
75	Shah Alam	0	Golf Course	Gym	Café	Library	Dessert Shop	Women's Store	French Restaurant	Dive Bar	Donut Shop	Electronics Store
76	Shah Alam	0	Golf Course	Gym	Café	Library	Dessert Shop	Women's Store	French Restaurant	Dive Bar	Donut Shop	Electronics Store
77	Shah Alam	0	Golf Course	Gym	Café	Library	Dessert Shop	Women's Store	French Restaurant	Dive Bar	Donut Shop	Electronics Store
78	Shah Alam	0	Golf Course	Gym	Café	Library	Dessert Shop	Women's Store	French Restaurant	Dive Bar	Donut Shop	Electronics Store
79	Shah Alam	0	Golf Course	Gym	Café	Library	Dessert Shop	Women's Store	French Restaurant	Dive Bar	Donut Shop	Electronics Store

62 rows × 12 columns

In [28]:
```
selangor_denc_merged.loc[selangor_denc_merged['Cluster Labels'] == 1, selangor_denc_merged.columns[[1] + list(range(5, selangor_denc_merged.shape[1]))]]
```

Out[28]:

	CITY	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Shah Alam	1	Café	History Museum	Art Gallery	Asian Restaurant	Cafeteria	Garden	Middle Eastern Restaurant	Fried Chicken Joint	Electronics Store	Farmers Market
2	Shah Alam	1	Music Venue	Bakery	Sandwich Place	Multiplex	Karaoke Bar	Bowling Alley	Pool Hall	Soccer Stadium	Fast Food Restaurant	Comfort Food Restaurant
3	Shah Alam	1	Malay Restaurant	Lounge	Tea Room	Coffee Shop	Field	Burger Joint	Women's Store	Donut Shop	Electronics Store	Farmers Market
4	Shah Alam	1	Burger Joint	Ice Cream Shop	Arts & Crafts Store	Asian Restaurant	Fast Food Restaurant	Soup Place	High School	Business Service	Print Shop	Food Court
5	Shah Alam	1	Soup Place	Restaurant	Train Station	Department Store	Thai Restaurant	Asian Restaurant	Snack Place	Women's Store	Food Court	Flea Market
6	Shah Alam	1	Park	Food Truck	Noodle House	Grocery Store	Garden Center	Sake Bar	French Restaurant	Dim Sum Restaurant	Dive Bar	Donut Shop
7	Shah Alam	1	Pizza Place	Food Truck	Boutique	Asian Restaurant	Auto Garage	Auto Workshop	Food Court	Field	French Restaurant	Flea Market
9	Shah Alam	1	Sake Bar	Beach	Hookah Bar	Art Gallery	Malay Restaurant	Noodle House	Gaming Cafe	College Library	Food Truck	Speakeasy
10	Shah Alam	1	Convenience Store	Indian Restaurant	Garden Center	Garden	Food Truck	Women's Store	Fried Chicken Joint	Dive Bar	Donut Shop	Electronics Store
12	Shah Alam	1	Park	Basketball Court	Palace	Convenience Store	Recreation Center	Food Court	Malay Restaurant	Food Truck	Burger Joint	Breakfast Spot
16	Shah Alam	1	Malay Restaurant	Food Court	Coffee Shop	Museum	Sandwich Place	Cafeteria	Light Rail Station	IT Services	Dim Sum Restaurant	Café
22	Shah Alam	1	Malay Restaurant	Asian Restaurant	Café	Indian Restaurant	Japanese Restaurant	Bakery	Pharmacy	Boutique	Food Truck	Golf Course
39	Shah Alam	1	Malay Restaurant	Indian Restaurant	Asian Restaurant	Restaurant	Comfort Food Restaurant	Seafood Restaurant	Indonesian Restaurant	Food Court	Middle Eastern Restaurant	Moroccan Restaurant
47	Shah Alam	1	Malay Restaurant	Burger Joint	Asian Restaurant	Seafood Restaurant	Breakfast Spot	Café	Clothing Store	Bubble Tea Shop	Flea Market	Fast Food Restaurant

In [29]:
```
selangor_denc_merged.loc[selangor_denc_merged['Cluster Labels'] == 2, selangor_denc_merged.columns[[1] + list(range(5, selangor_denc_merged.shape[1]))]]
```

Out[29]:

	CITY	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
8	Shah Alam	2	Pet Store	Art Gallery	Bookstore	Furniture / Home Store	Dive Bar	Donut Shop	Electronics Store	Farmers Market	Fast Food Restaurant	Field

In [30]:
```
selangor_denc_merged.loc[selangor_denc_merged['Cluster Labels'] == 3, selangor_denc_merged.columns[[1] + list(range(5, selangor_denc_merged.shape[1]))]]
```

Out[30]:

	CITY	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Shah Alam	3	Food Truck	Gym / Fitness Center	Motel	Dessert Shop	Women's Store	Fried Chicken Joint	Dive Bar	Donut Shop	Electronics Store	Farmers Market
40	Shah Alam	3	Food Truck	Malay Restaurant	Dessert Shop	Japanese Restaurant	Cafeteria	Office	Gym	Halal Restaurant	Dive Bar	Donut Shop

In [31]:
```
selangor_denc_merged.loc[selangor_denc_merged['Cluster Labels'] == 4, selangor_denc_merged.columns[[1] + list(range(5, selangor_denc_merged.shape[1]))]]
```

Out[31]:

	CITY	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
37	Shah Alam	4	Golf Course	Café	Park	Malay Restaurant	Women's Store	French Restaurant	Dim Sum Restaurant	Dive Bar	Donut Shop	Electronics Store

We have explored the clusters. Now let's dive into the rental details.

In [32]:
```
shah_a_rent=pd.read_csv('Shah Alam Rental.csv')
shah_a_rent.head()
```

Out[32]:

	ID	BuiltUpSize	City	CreatedDateTime	CreatedUser	DataSourceID	Furnishing	Latitude	Longitude	NoOfBathroom	...	POSTCC
0	42102	820	Shah Alam	18/9/2018 13:00	SYSTEM_AUTOMATION	23	Fully Furnished	0	0	2	...	40
1	42103	820	Shah Alam	18/9/2018 13:00	SYSTEM_AUTOMATION	23	Fully Furnished	0	0	2	...	40
2	42104	839	Shah Alam	18/9/2018 13:00	SYSTEM_AUTOMATION	23	Fully Furnished	0	0	1	...	40
3	42105	839	Shah Alam	18/9/2018 13:00	SYSTEM_AUTOMATION	23	Fully Furnished	0	0	1	...	40
4	42108	1123	Shah Alam	18/9/2018 13:00	SYSTEM_AUTOMATION	23	Fully Furnished	0	0	2	...	40

5 rows × 22 columns

In [33]:
```
lat_lon1 = pd.read_csv('Selangor_coordinates.csv')

lat_lon1.head()
```

Out[33]:

	POSTCODE	LATITUDE	LONGITUDE
0	40000	3.0797	101.5186
1	40100	3.0799	101.5334
2	40150	3.1100	101.5538
3	40160	3.1643	101.5584
4	40170	3.0733	101.5054

In [34]:
```
df3 = pd.merge(shah_a_rent ,lat_lon,on='POSTCODE')
df3.head()
```

Out[34]:

	ID	BuiltUpSize	City	CreatedDateTime	CreatedUser	DataSourceID	Furnishing	Latitude	Longitude	NoOfBathroom	...	Property
0	42102	820	Shah Alam	18/9/2018 13:00	SYSTEM_AUTOMATION	23	Fully Furnished	0	0	2	...	Jln Pega
1	42103	820	Shah Alam	18/9/2018 13:00	SYSTEM_AUTOMATION	23	Fully Furnished	0	0	2	...	Jln Pega
2	43034	2600	Shah Alam	18/9/2018 13:00	SYSTEM_AUTOMATION	23	Partly Furnished	0	0	5	...	
3	44443	850	Shah Alam	18/9/2018 13:00	SYSTEM_AUTOMATION	23	Partly Furnished	0	0	2	...	Ja Pe 40170 S
4	44444	850	Shah Alam	18/9/2018 13:00	SYSTEM_AUTOMATION	23	Partly Furnished	0	0	2	...	Ja Pe 40170 S

5 rows × 24 columns

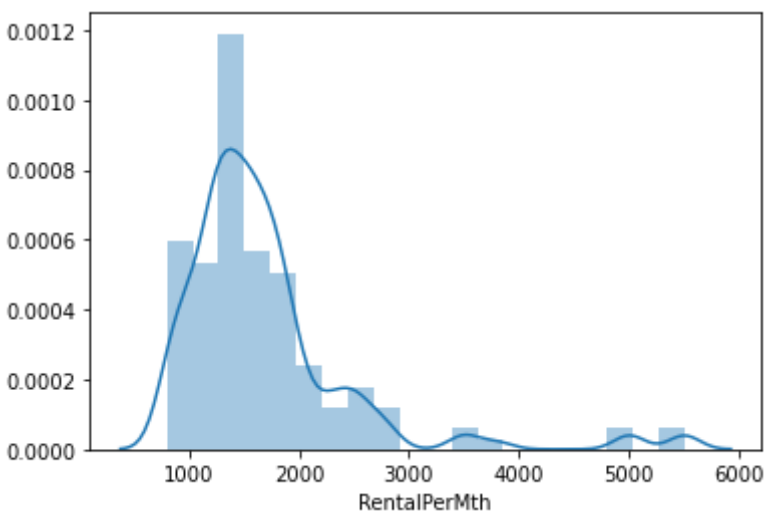
In [35]:
```
df4 = df3.drop(['CreatedDateTime', 'ID', 'CreatedUser', 'DataSourceID', 'Latitude', 'Longitude', 'SourceUrl', 'Updated
DateTime', 'UpdatedUser'], axis=1)

df4.head()
```

Out[35]:

	BuiltUpSize	City	Furnishing	NoOfBathroom	NoOfBedroom	NoOfParking	POSTCODE	PostedDate	PropertyAddress	PropertyName	Property
0	820	Shah Alam	Fully Furnished	2	3	2	40170	17/9/2018 0:00	Jln Pegaga U12/1	Seri Pegaga	Apart
1	820	Shah Alam	Fully Furnished	2	3	2	40170	17/9/2018 0:00	Jln Pegaga U12/1	Seri Pegaga	Apart
2	2600	Shah Alam	Partly Furnished	5	0	0	40170	16/9/2018 0:00	Jalan U13	Setia Eco Park	deta H
3	850	Shah Alam	Partly Furnished	2	3	2	40170	14/9/2018 0:00	Jalan Setia Permai U13 40170 Shah Alam	Seri Pinang	Apart
4	850	Shah Alam	Partly Furnished	2	3	2	40170	14/9/2018 0:00	Jalan Setia Permai U13 40170 Shah Alam	Seri Pinang	Apart

In [36]:
```
import seaborn as sns

sns.distplot(df4['RentalPerMth'],bins=20)
```

Out[36]:
```
<AxesSubplot:xlabel='RentalPerMth'>
```



In [37]:
```
df4.groupby(['LATITUDE', 'LONGITUDE']).groups.keys()
```

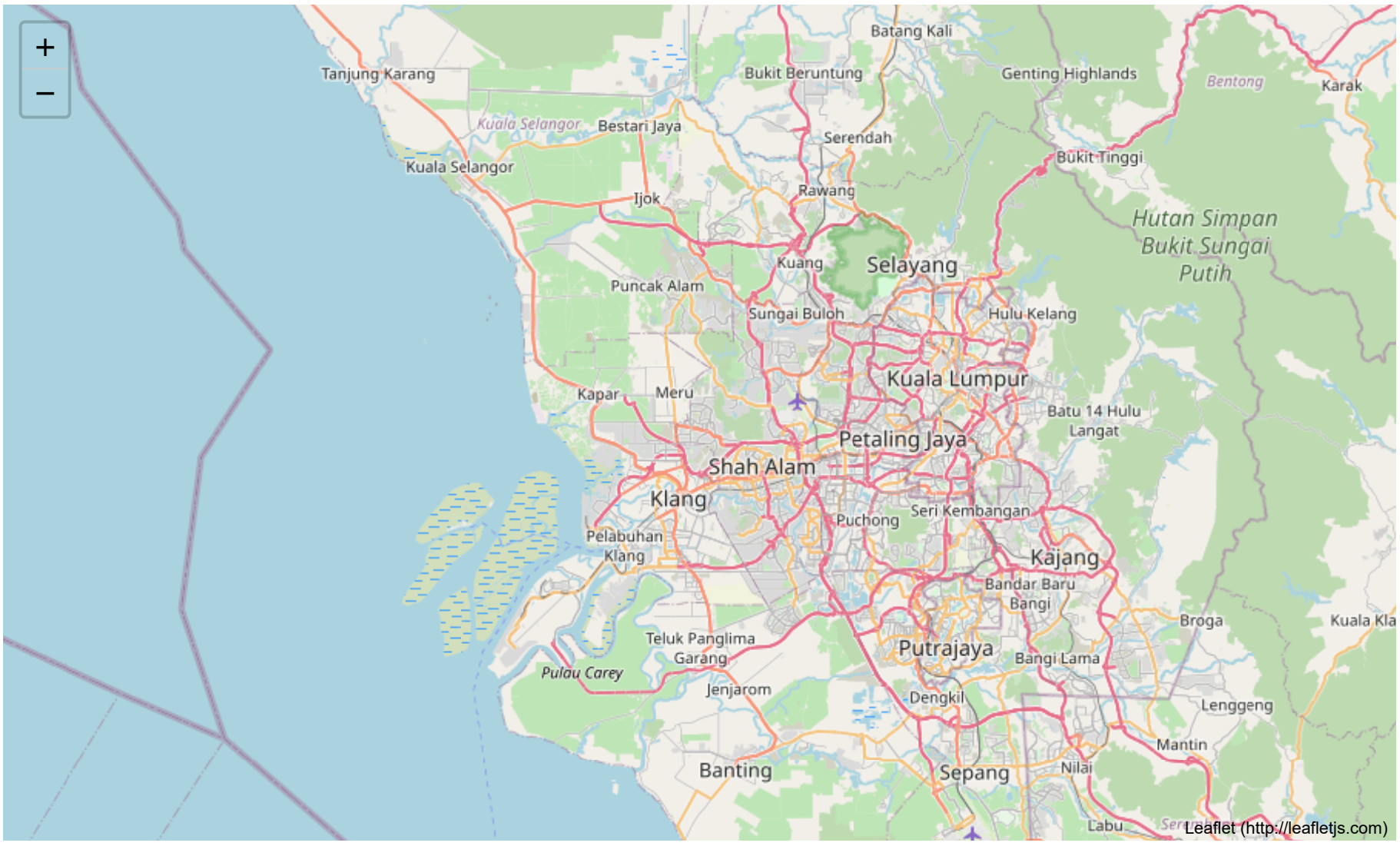
Out[37]:
```
dict_keys([(3.0457, 101.5573), (3.0538, 101.5304), (3.0544, 101.5169), (3.0549, 101.5314), (3.0733, 101.5054), (3.079
7, 101.5186), (3.0799, 101.5334), (3.11, 101.5538)])
```

In [38]:
```
print('The geograpical coordinate of preferable area is 3.0797, 101.5186.'.format(latitude, longitude))
```

```
The geograpical coordinate of preferable area is 3.0797, 101.5186.
```

In [39]:
```
map_location = folium.Map(location=[3.0797, 101.5186], zoom_start=10)
map_location
```

Out[39]:



In [40]:
```
df5 = pd.DataFrame(data, columns= ['City', 'PropertyAddress','PropertyName','NoOfBedroom','PropertyType','RentalPerMth','LATITUDE','LONGITUDE'])

df5 = df4.loc[df4[('NoOfBedroom')] == (4) , df4.columns[[1] + list(range(8, df4.shape[1]))]]

df5.head()
```

Out[40]:

	City	PropertyAddress	PropertyName	PropertyType	RentalPerMth	State	LATITUDE	LONGITUDE
7	Shah Alam	40170	Setia Impian 7	Semi-detached House	2500	selangor	3.0733	101.5054
8	Shah Alam	40170	Setia Impian 7	Semi-detached House	2500	selangor	3.0733	101.5054
13	Shah Alam	JALAN U16/1D	ELMINA VALLEY 3	2-sty Terrace/Link House	1500	selangor	3.0733	101.5054
14	Shah Alam	JALAN U16/1D	ELMINA VALLEY 3	2-sty Terrace/Link House	1500	selangor	3.0733	101.5054
22	Shah Alam	40170	Setia Impian 7	Semi-detached House	2500	selangor	3.0733	101.5054

In [41]:
```
kk = 1
selangor_denc_merged.loc[selangor_denc_merged['Cluster Labels'] == kk, selangor_denc_merged.columns[[1] + list(range(5
, selangor_denc_merged.shape[1]))]]
```

Out[41]:

	CITY	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Shah Alam	1	Café	History Museum	Art Gallery	Asian Restaurant	Cafeteria	Garden	Middle Eastern Restaurant	Fried Chicken Joint	Electronics Store	Farmers Market
2	Shah Alam	1	Music Venue	Bakery	Sandwich Place	Multiplex	Karaoke Bar	Bowling Alley	Pool Hall	Soccer Stadium	Fast Food Restaurant	Comfort Food Restaurant
3	Shah Alam	1	Malay Restaurant	Lounge	Tea Room	Coffee Shop	Field	Burger Joint	Women's Store	Donut Shop	Electronics Store	Farmers Market
4	Shah Alam	1	Burger Joint	Ice Cream Shop	Arts & Crafts Store	Asian Restaurant	Fast Food Restaurant	Soup Place	High School	Business Service	Print Shop	Food Court
5	Shah Alam	1	Soup Place	Restaurant	Train Station	Department Store	Thai Restaurant	Asian Restaurant	Snack Place	Women's Store	Food Court	Flea Market
6	Shah Alam	1	Park	Food Truck	Noodle House	Grocery Store	Garden Center	Sake Bar	French Restaurant	Dim Sum Restaurant	Dive Bar	Donut Shop
7	Shah Alam	1	Pizza Place	Food Truck	Boutique	Asian Restaurant	Auto Garage	Auto Workshop	Food Court	Field	French Restaurant	Flea Market
9	Shah Alam	1	Sake Bar	Beach	Hookah Bar	Art Gallery	Malay Restaurant	Noodle House	Gaming Cafe	College Library	Food Truck	Speakeasy
10	Shah Alam	1	Convenience Store	Indian Restaurant	Garden Center	Garden	Food Truck	Women's Store	Fried Chicken Joint	Dive Bar	Donut Shop	Electronics Store
12	Shah Alam	1	Park	Basketball Court	Palace	Convenience Store	Recreation Center	Food Court	Malay Restaurant	Food Truck	Burger Joint	Breakfast Spot
16	Shah Alam	1	Malay Restaurant	Food Court	Coffee Shop	Museum	Sandwich Place	Cafeteria	Light Rail Station	IT Services	Dim Sum Restaurant	Café
22	Shah Alam	1	Malay Restaurant	Asian Restaurant	Café	Indian Restaurant	Japanese Restaurant	Bakery	Pharmacy	Boutique	Food Truck	Golf Course
39	Shah Alam	1	Malay Restaurant	Indian Restaurant	Asian Restaurant	Restaurant	Comfort Food Restaurant	Seafood Restaurant	Indonesian Restaurant	Food Court	Middle Eastern Restaurant	Moroccan Restaurant
47	Shah Alam	1	Malay Restaurant	Burger Joint	Asian Restaurant	Seafood Restaurant	Breakfast Spot	Café	Clothing Store	Bubble Tea Shop	Flea Market	Fast Food Restaurant

5.0 Discussion

Based on the clusters we derived, Cluster 1 is a more big area in terms on ammenities aroundvwhichis located in 3.0797, 101.5186. Venues that attracted me are convenience store, indian restaurant, garden, park, coffee shop and pharmacy. One of my favourite spot is hookah bar and art & craft store.

When I explored the houses in cluster 1, the house rental is what I am looking for. At Setia Impian 7, a semi-detached house rental is RM 2,500 with the above mentioned venues. There is also another house for RM 1,500 rental price which is located in Jalan U16 1D, Elmina Valley, which is a 2-storey terrace/link house. On the whole, I would prefer the Elmina Valley house for RM 1,500 which is much more cheaper than my budget and contains 4 rooms.

6.0 Conclusion

The tools that I have used in this project including the data I uploaded has helped me to make this project workable. I was able to find the location and address that I prefer the most without too many google searches. The commons venues by using foursquare has definitely helped alot!

To those who are looking for rental around Shah Alam or Selangor on the whole, can apply this methodology to yours as it is generic!

The End