Capstone Project - The Battle of Neighborhoods (Week 2)

In this week, you will continue working on your capstone project. Please remember by the end of this week, you will need to submit the following:

1. A full report consisting of all of the following components
2. A link to your Notebook on your Github repository pushed showing your code
3. Your choice of a presentation or blogpost

Import Libraries

In [1]:
```
import json, requests
import os
import geopandas
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors
from geopy.geocoders import Nominatim
from pandas.io.json import json_normalize
from sklearn.cluster import KMeans
import folium

pd.options.display.max_rows = None
pd.options.display.max_columns = None

%matplotlib inline
%autosave 60
```

Autosaving every 60 seconds

We proceed to import the Sandakan neighbourhood csv file which consists the places, names, location, latitude and longtitude.

Sandakan neighbourhood data description:

Number = Index number

Name of neighbourhood = Neighbourhood Names

Area = Area in acres

Residential units = Number of residential homes

Location = Location of neighbourhood

Latitude = Latitude coordinates

Longitude = Longitude coordinates

Load data

In [2]:
```
df = pd.read_csv('sandakan.csv', index_col="Number")
```

In [3]:
```
df.head()
```

Out[3]:

Number	Neighbourhood	Area	Residential Units	Location	Latitude	Longtitude
1	Airport	41.630	649	Batu 7, Jalan Lapangan Terbang	5.898035	118.061205
2	Anggerik	15.828	408	Jalan Lintas Sibuga	5.861322	118.037246
3	Astana Height	100.270	483	Batu 1, Jalan Lalang	5.853584	118.116925
4	Berhala Darat	23.200	192	Jalan Sim-Sim	5.850209	118.130763
5	Bukit Permai	270.890	4142	Batu 3 1/2, Jalan Utara	5.864637	118.084975

In [4]:
```
df.tail()
```

Out[4]:

Number	Neighbourhood	Area	Residential Units	Location	Latitude	Longtitude
69	Vista	10.700	172	Batu 7, Jalan Lintas Sibuga	5.858524	118.041216
70	Wijaya	6.03	196	Batu 7, Jalan Labuk	5.884665	118.045531
71	Wira	9.93	312	Jalan Sibuga	5.849188	118.042680
72	Yeng Seng	11.75	125	Batu 2 1/2, Jalan Utara	5.858835	118.098056
73	Yii Villa	1.30	100	Jalan Bulis Sim-Sim	5.854239	118.126795

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

Out[5]: (73, 6)

The dataset consists of 73 rows and 6 columns

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 73 entries, 1 to 73
Data columns (total 6 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Neighbourhood      73 non-null     object
 1   Area               72 non-null     float64
 2   Residential Units  73 non-null     int64
 3   Location           70 non-null     object
 4   Latitude           68 non-null     float64
 5   Longtitude         68 non-null     float64
dtypes: float64(3), int64(1), object(2)
memory usage: 4.0+ KB
```

In [7]:
```
df.isnull().sum()  #Count NaN values
```

Out[7]:
```
Neighbourhood        0
Area                 1
Residential Units    0
Location             3
Latitude             5
Longtitude           5
dtype: int64
```

Since we need to explore and plot neighbourhoods, I decided to drop NaNs for Latitude and Longtitude

In [8]:
```
df.dropna(inplace=True)
```

In [9]:
```
df.shape
```

Out[9]: (65, 6)

In [10]:
```
df.head()
```

Out[10]:

Number	Neighbourhood	Area	Residential Units	Location	Latitude	Longtitude
1	Airport	41.630	649	Batu 7, Jalan Lapangan Terbang	5.898035	118.061205
2	Anggerik	15.828	408	Jalan Lintas Sibuga	5.861322	118.037246
3	Astana Height	100.270	483	Batu 1, Jalan Lalang	5.853584	118.116925
4	Berhala Darat	23.200	192	Jalan Sim-Sim	5.850209	118.130763
5	Bukit Permai	270.890	4142	Batu 3 1/2, Jalan Utara	5.864637	118.084975

In [11]:
```
df.reset_index()
```

Out[11]:

	Number	Neighbourhood	Area	Residential Units	Location	Latitude	Longtitude
0	1	Airport	41.630	649	Batu 7, Jalan Lapangan Terbang	5.898035	118.061205
1	2	Anggerik	15.828	408	Jalan Lintas Sibuga	5.861322	118.037246
2	3	Astana Height	100.270	483	Batu 1, Jalan Lalang	5.853584	118.116925
3	4	Berhala Darat	23.200	192	Jalan Sim-Sim	5.850209	118.130763
4	5	Bukit Permai	270.890	4142	Batu 3 1/2, Jalan Utara	5.864637	118.084975
5	6	Bunga Matahari	11.880	172	Batu 4, Jalan Utara	5.865810	118.075874
6	7	Casa San Uno	38.890	307	Batu 4, Jalan Utara	5.865233	118.072556
7	8	Chrysanthemum	11.400	154	Batu 1 1/2, Jalan Utara	5.857480	118.105876
8	9	Damai & Sri Taman	21.670	123	Batu 4, Jalan Utara	5.858482	118.078921
9	10	Evergreen	23.990	48	Batu 6, Jalan Utara	5.873464	118.057834
10	11	Fajar	55.610	572	Batu 7, Jalan Lapangan Terbang	5.836976	118.098463
11	12	Fajar Perdana	15.030	185	Batu 7, Jalan Lapangan Terbang	5.884339	118.057220
12	13	Fortune	22.400	126	Batu 8, Jalan Labuk	5.885541	118.030970
13	14	Fulliwa	19.660	164	Batu 3 1/2, Jalan Utara	5.862592	118.085900
14	15	Garden Villa	25.760	82	Batu 6, Jalan Utara	5.864005	118.048945
15	16	Grandview	93.000	746	Batu 1 1/2, Jalan Buli Sim-Sim	5.862512	118.119377
16	17	Hap Seng Properties	16.350	74	Jalan Batu Sapi	5.833646	118.092576
17	19	Hing Lee	11.750	227	Batu 3 1/2, Jalan Utara	5.862033	118.090287
18	23	Indah	56.270	356	Batu 4, Jalan Utara	5.842067	118.066095
19	24	Indah Jaya	235.680	2752	Batu 4, Jalan Utara	5.843796	118.067200
20	25	Jade Garden	1154.000	8	Batu 1 1/2, Jalan Utara	5.862170	118.110320
21	26	Kam Jai Yen	8.510	250	Batu 1, Jalan Aman	5.849344	118.110234
22	27	Karamunting Baru	17.670	40	Jalan Karamunting	5.810315	118.072635
23	28	Karamunting Flat	20.870	513	Jalan Batu Sapi	5.813614	118.065127
24	29	Kenari	3.440	590	Batu 7, Jalan Lapangan Terbang	5.895072	118.043659
25	30	Khong Lok (Hillside)	24.809	168	Batu 7, Jalan Lapangan Terbang	5.878357	118.059916
26	31	LPPB	27.210	984	Batu 3 1/2, Jalan Utara	5.862075	118.084777
27	32	Lucky & Wemin	43.544	260	Batu 5, Jalan Utara	5.863112	118.062768
28	33	Mawar	175.386	2396	Jalan Sibuga	5.842216	118.032957
29	34	Megah	44.940	478	Batu 8, Jalan Utara	5.875798	118.042150
30	35	Melanta	9.040	143	Jalan Karamunting	5.810121	118.079529
31	36	Melrose	14.580	44	Batu 3 1/2, Jalan Utara	5.839134	118.115892
32	37	Merpati	154.200	494	Batu 8, Jalan Lapangan Terbang	5.889157	118.042522
33	38	Mesra	23.180	1000	Batu 4, Jalan Utara	5.861271	118.077664
34	40	Mutiara	64.790	836	Batu 3, Jalan Utara	5.854958	118.087570
35	41	Nuri (Oscaraya)	32.650	837	Batu 7, Jalan Lapangan Terbang	5.891094	118.040690
36	42	Pak Tak	29.886	148	Batu 7, Jalan Lapangan Terbang	5.880687	118.056512
37	43	Pecky Valley	36.610	166	Batu 2 1/2, Jalan Utara	5.857141	118.100194
38	44	Pertama	53.320	438	Batu 5, Jalan Utara	5.861339	118.069276
39	45	Po Hing	190.170	160	Jalan Batu Sapi	5.838391	118.047363
40	46	Rajawali	154.200	1004	Batu 8, Jalan Lapangan Terbang	5.890578	118.039349
41	47	Regent	3.449	46	Batu 1, Jalan Utara	5.839134	118.115892
42	48	Ria	3.406	44	Batu 7, Jalan Lapangan Terbang	5.838738	118.116348
43	49	Rimba	38.980	280	Batu 7, Jalan Utara	5.884839	118.048295
44	50	Samudera	20.010	586	Batu 3 1/2, Jalan Utara	5.855790	118.091154
45	51	Sanny Estate	34.800	154	Batu 1 1/2, Jalan Utara	5.855878	118.118020
46	52	Seaview	3.763	152	Batu 1 1/2, Jalan Utara	5.867749	118.115667
47	53	Sejati	22.890	1086	Batu 7, Jalan Lapangan Terbang	5.887911	118.056251
48	54	Sejati Ujana	99.640	1850	Batu 7, Jalan Lapangan Terbang	5.891526	118.049406
49	55	Seletar	65.000	50	Jalan Lintas Sibuga	5.838134	118.045310
50	56	Sentosa	49.300	928	Batu 2 1/2, Jalan Utara	5.856344	118.098159
51	57	Seri Labuk	15.000	178	Batu 8, Jalan Kampung Melayu	5.871849	118.026715
52	58	Sibuga	102.600	647	Jalan Sibuga	5.871911	118.048039
53	61	Sri Rimba	29.200	414	Batu 7, Jalan Labuk	5.885663	118.047306
54	63	Tai Fai Yin	27.600	160	Batu 8, Jalan Labuk	5.883818	118.037509
55	64	Tinosan	24.510	235	Batu 4, Jalan Utara	5.861032	118.074517
56	65	Tshun Ngen	70.260	304	Batu 5, Jalan Utara	5.858728	118.065804
57	66	Tyng	111.700	585	Batu 5, Jalan Utara	5.867669	118.059997
58	67	Utama	41.263	329	Batu 6, Jalan Utara	5.864601	118.058569
59	68	Villa Permai Jaya	20.000	968	Jalan Sibuga	5.856154	118.057149
60	69	Vista	10.700	172	Batu 7, Jalan Lintas Sibuga	5.858524	118.041216
61	70	Wijaya	6.030	196	Batu 7, Jalan Labuk	5.884665	118.045531
62	71	Wira	9.930	312	Jalan Sibuga	5.849188	118.042680
63	72	Yeng Seng	11.750	125	Batu 2 1/2, Jalan Utara	5.858835	118.098056
64	73	Yii Villa	1.300	100	Jalan Bulis Sim-Sim	5.854239	118.126795

Drop number and location columns from dataframe

In [12]:
```
df = df[['Neighbourhood','Area','Residential Units','Latitude','Longtitude']]
```

In [13]:
```
df.reset_index(drop="Number", inplace=True)
```

In [14]:
```
#Save a cleaned csv file for backup

#df.to_csv('skanclean.csv', index=False)
```

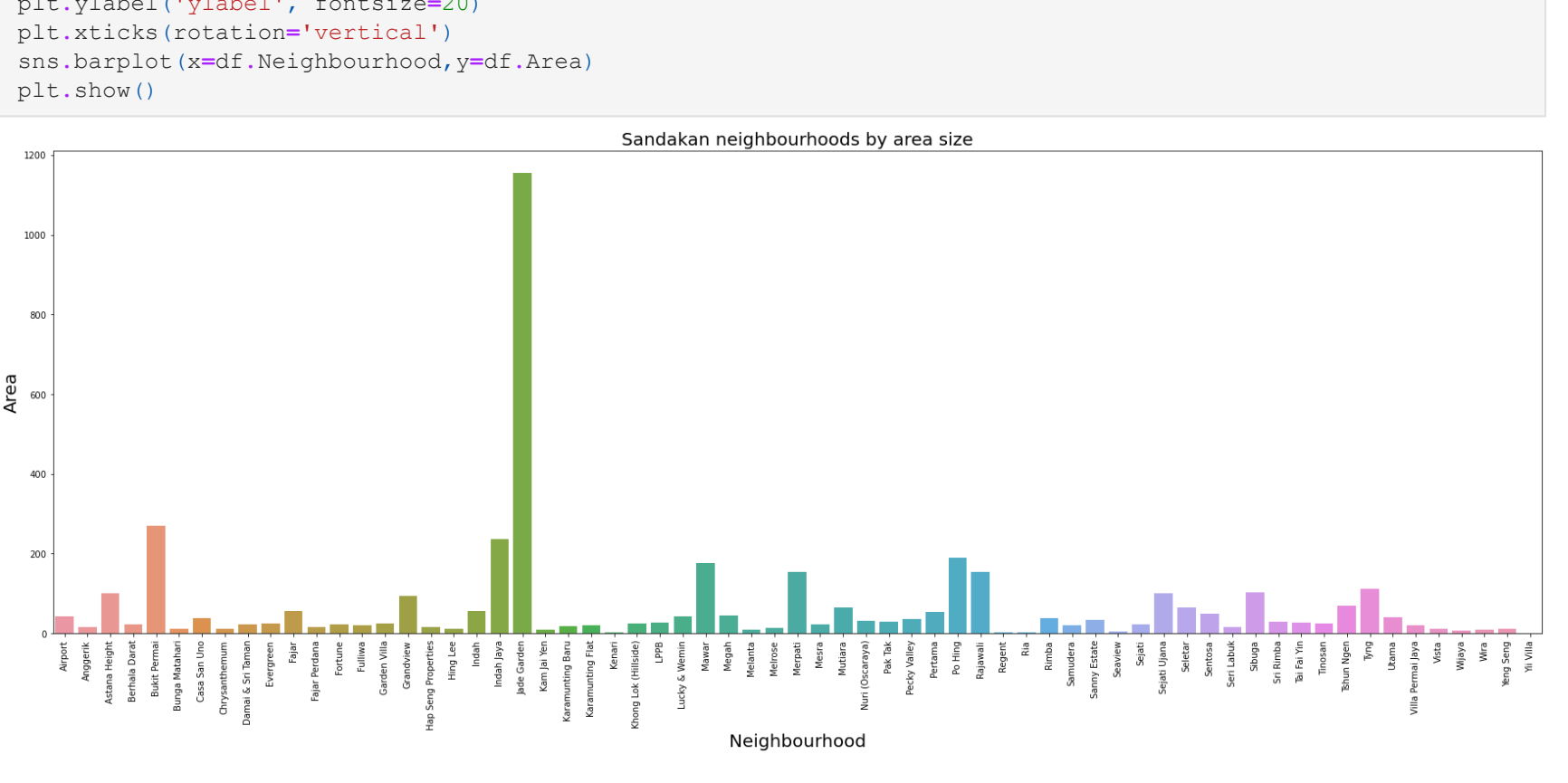
Data Visualization

In [15]:
```
df.head()
```

Out[15]:

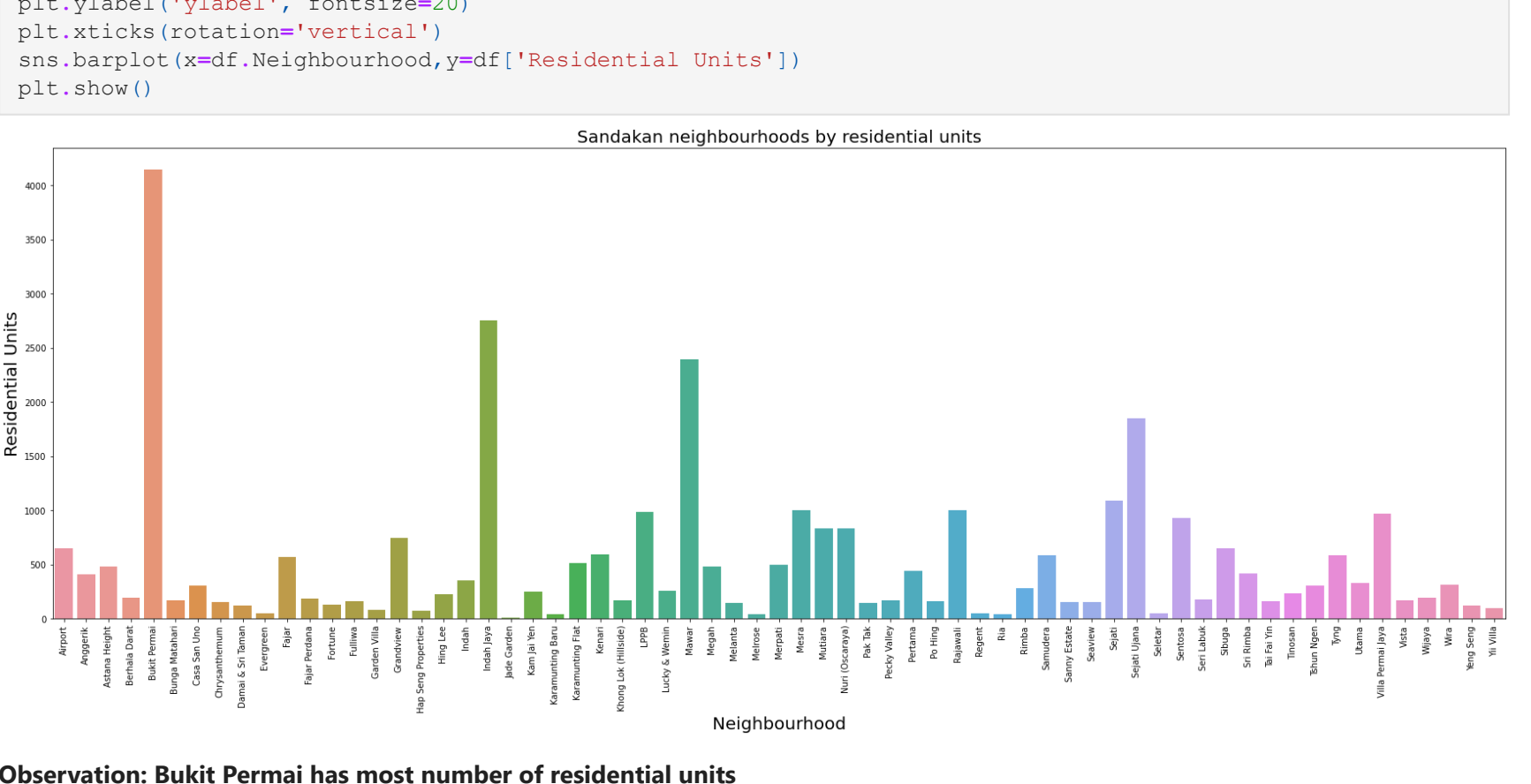
	Neighbourhood	Area	Residential Units	Latitude	Longtitude
0	Airport	41.630	649	5.898035	118.061205
1	Anggerik	15.828	408	5.861322	118.037246
2	Astana Height	100.270	483	5.853584	118.116925
3	Berhala Darat	23.200	192	5.850209	118.130763
4	Bukit Permai	270.890	4142	5.864637	118.084975

In [16]:
```
plt.figure(figsize=(30,10))
plt.title('Sandakan neighbourhoods by area size', fontsize=20)
plt.xlabel('xlabel', fontsize=20)
plt.ylabel('ylabel', fontsize=20)
plt.xticks(rotation='vertical')
sns.barplot(x=df.Neighbourhood,y=df.Area)
plt.show()
```



Observation: Jade Garden has largest area

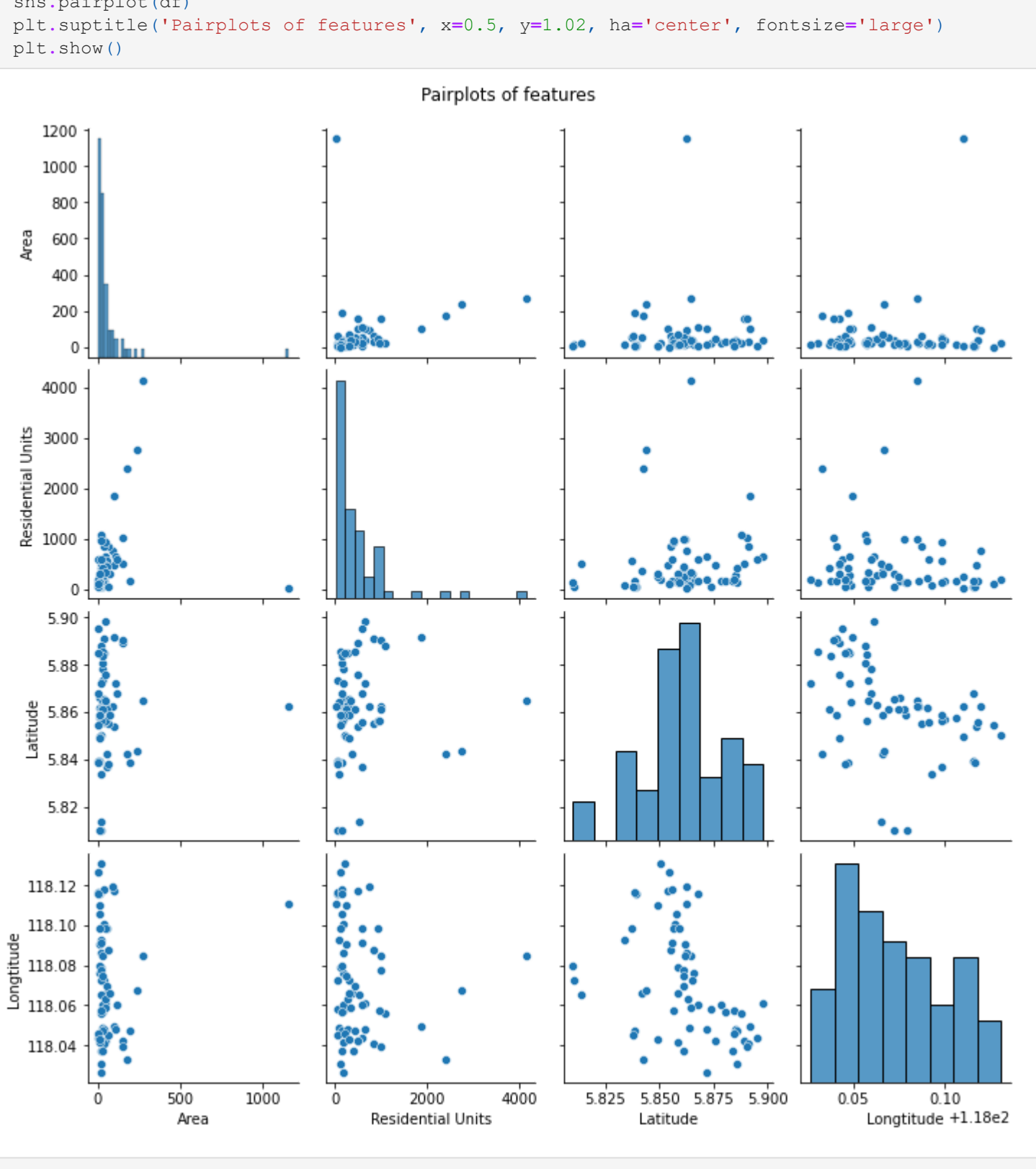
In [17]:
```
plt.figure(figsize=(30,10))
plt.title('Sandakan neighbourhoods by residential units', fontsize=20)
plt.xlabel('xlabel', fontsize=20)
plt.ylabel('ylabel', fontsize=20)
plt.xticks(rotation='vertical')
sns.barplot(x=df.Neighbourhood,y=df['Residential Units'])
plt.show()
```



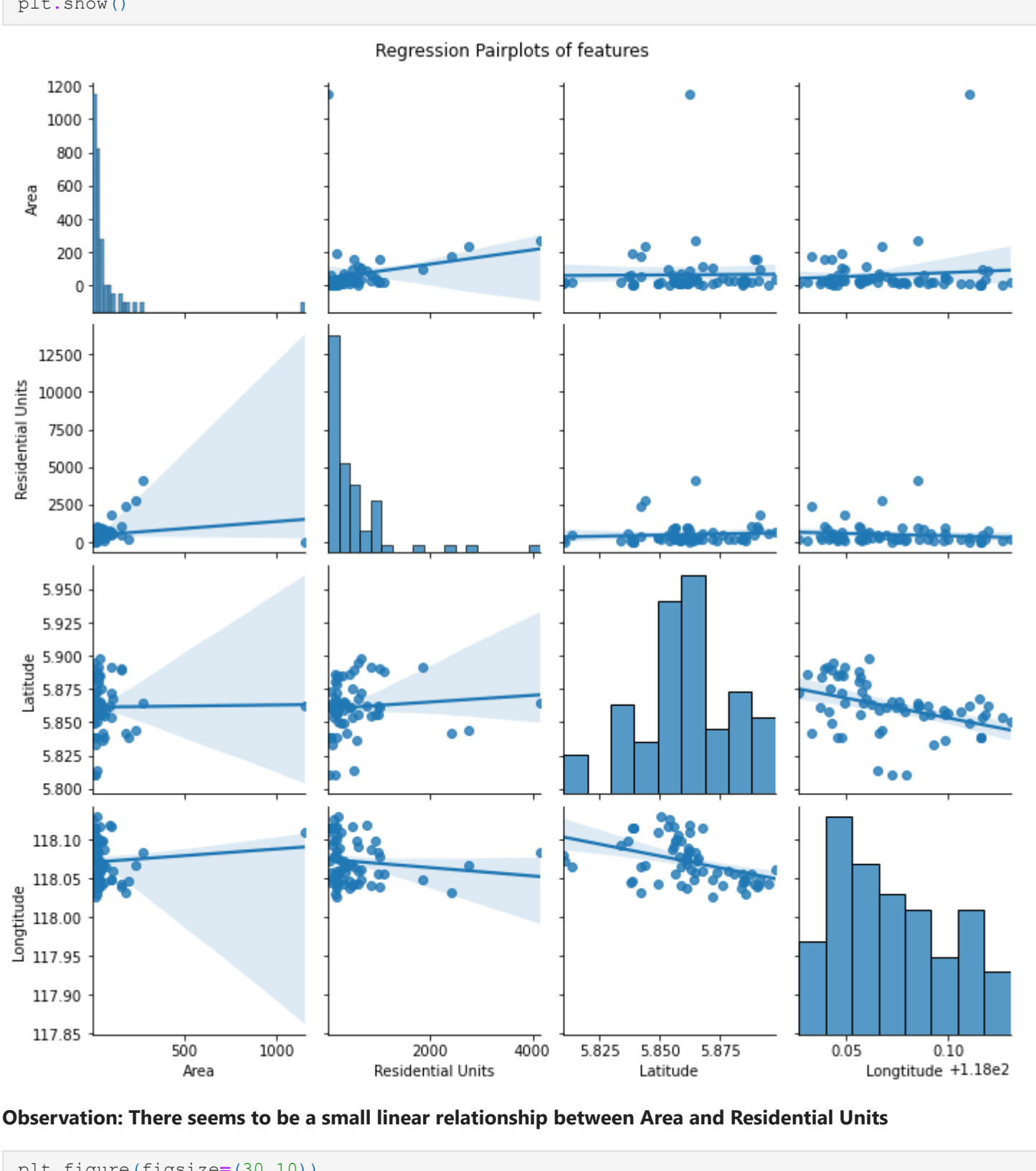
Observation: Bukit Permai has most number of residential units

Plotting pairplots to check for any correlation

In [18]:
```
sns.pairplot(df)
plt.suptitle('Pairplots of features', x=0.5, y=1.02, ha='center', fontsize='large')
plt.show()
```

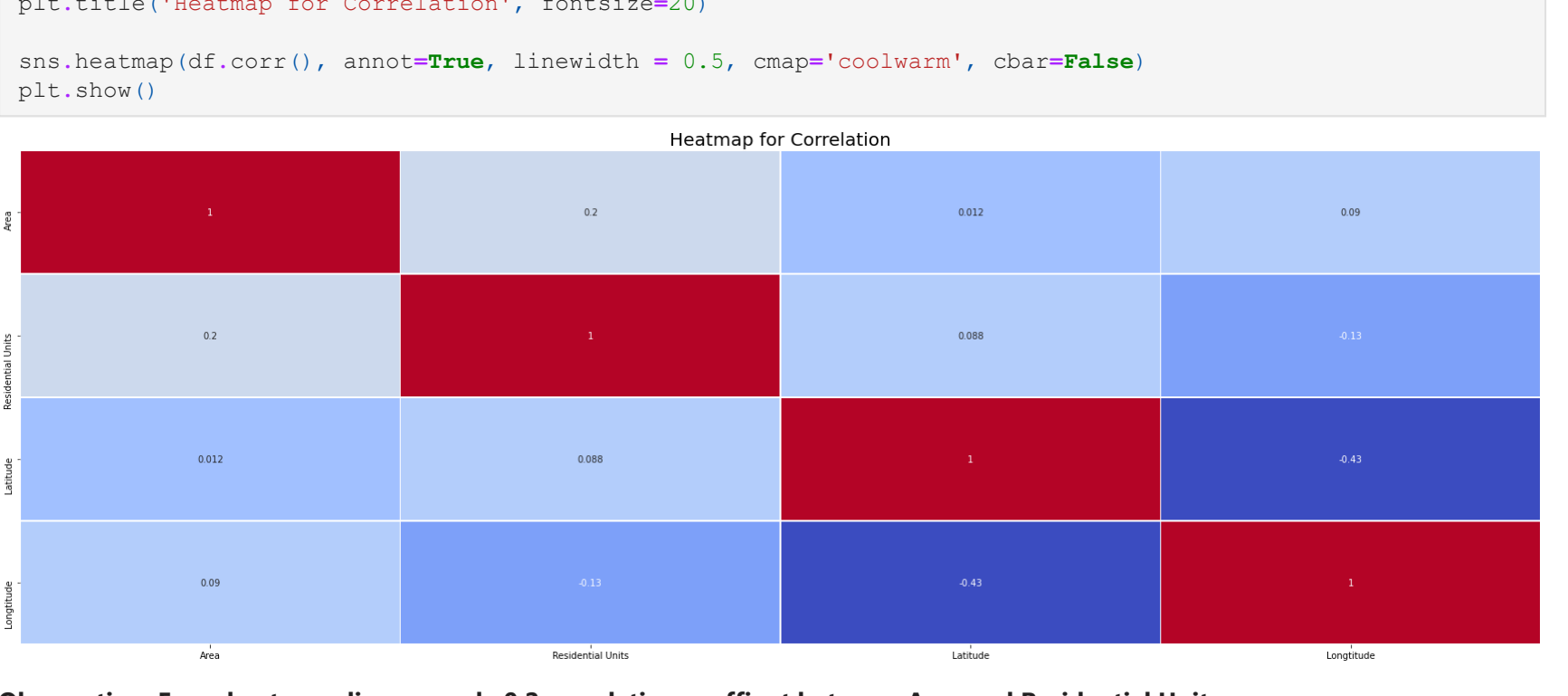


In [19]:
```
sns.pairplot(df, kind='reg')
plt.suptitle('Regression Pairplots of features', x=0.5, y=1.02, ha='center', fontsize='large')
plt.show()
```



Observation: There seems to be a small linear relationship between Area and Residential Units

In [20]:
```
plt.figure(figsize=(30,10))
plt.title('Heatmap for Correlation', fontsize=20)

sns.heatmap(df.corr(), annot=True, linewidth = 0.5, cmap='coolwarm', cbar=False)
```



Observation: From heatmap diagram only 0.2 correlation coeffient between Area and Residential Units

Create maps to look at all neighbourhoods

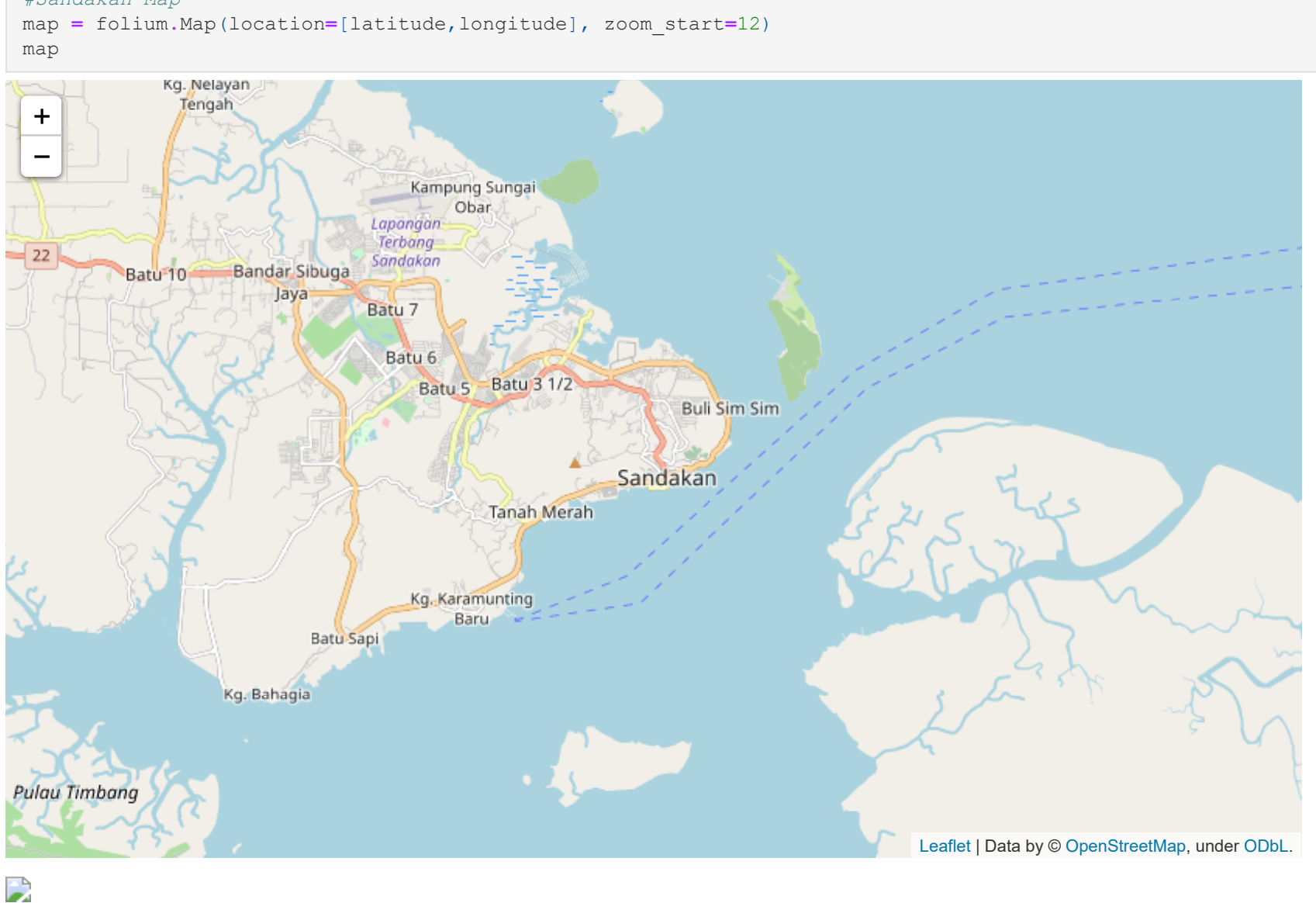
In [21]:
```
#Load the cleaned csv file

df = pd.read_csv("skanclean.csv")
```

In [22]:
```
#Get the lat and long coordinates for Sandakan
address = 'Sandakan'

geolocator = Nominatim(user_agent="foursquare_agent")

location = geolocator.geocode(address)

latitude = location.latitude
longitude = location.longitude
print(latitude, longitude)
```

5.8391337 118.1158919

In [23]:
```
#Sandakan Map
map = folium.Map(location=[latitude,longitude], zoom_start=12)
map
```

Out[23]:



In [24]:
```
#Segment suburbs coordinates

df_suburbs = df[['Latitude','Longtitude']]
```

In [25]:
```
df_suburbs.head()
```

Out[25]:

	Latitude	Longtitude
0	5.898035	118.061205
1	5.861322	118.037246
2	5.853584	118.116925
3	5.850209	118.130763
4	5.864637	118.084975

In [26]:
```
df_suburbs.shape
```

Out[26]: (65, 2)

In [27]:
```
suburbs_list = df_suburbs.values.tolist()
```

In [28]:
```
suburbs_list_size = len(suburbs_list)
```

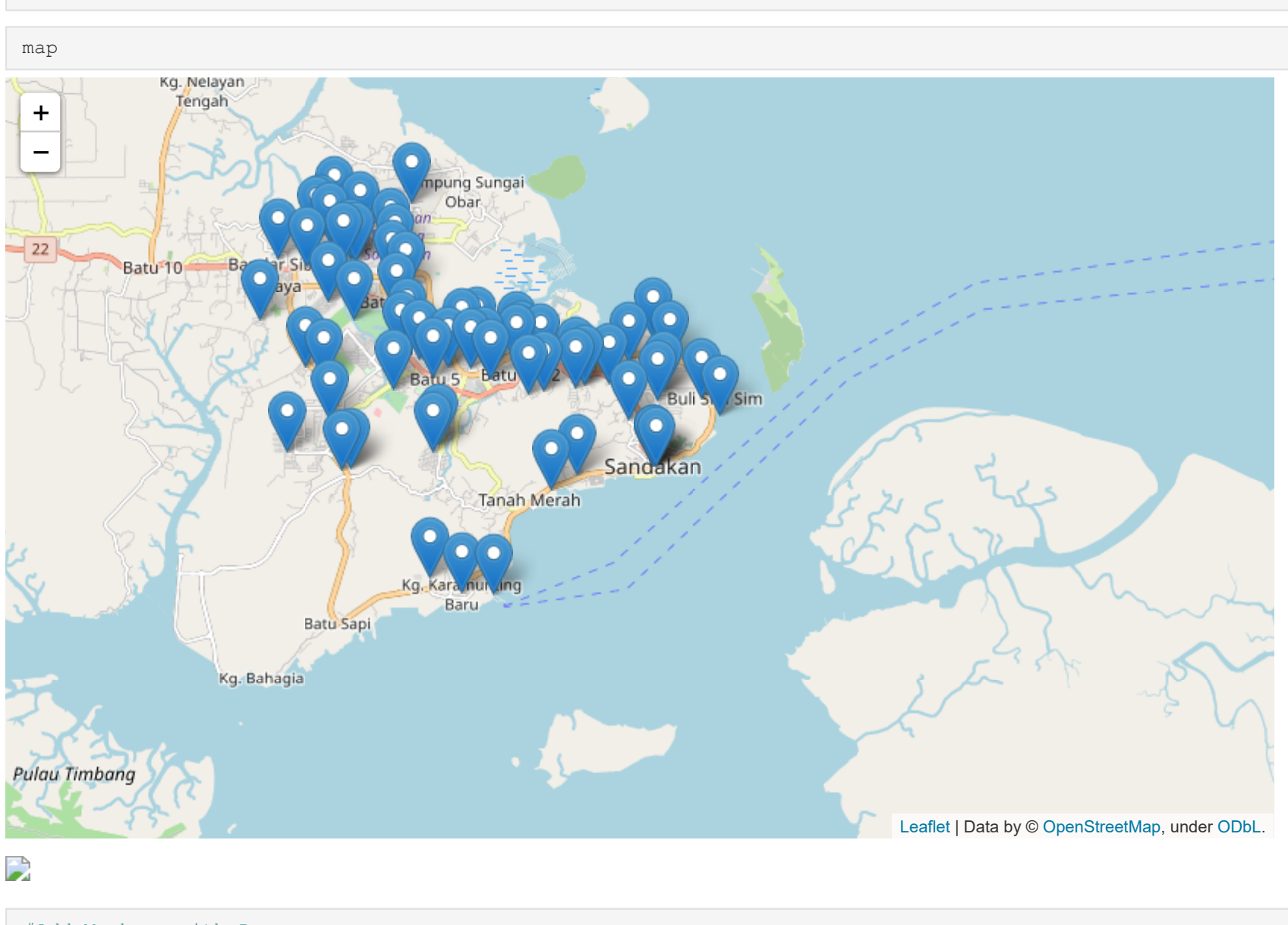
In [29]:
```
suburbs_list_size
```

Out[29]: 65

In [30]:
```
#Add Markers
for point in range(0,suburbs_list_size):
    folium.Marker(suburbs_list[point]).add_to(map)
```

In [31]:
```
map
```

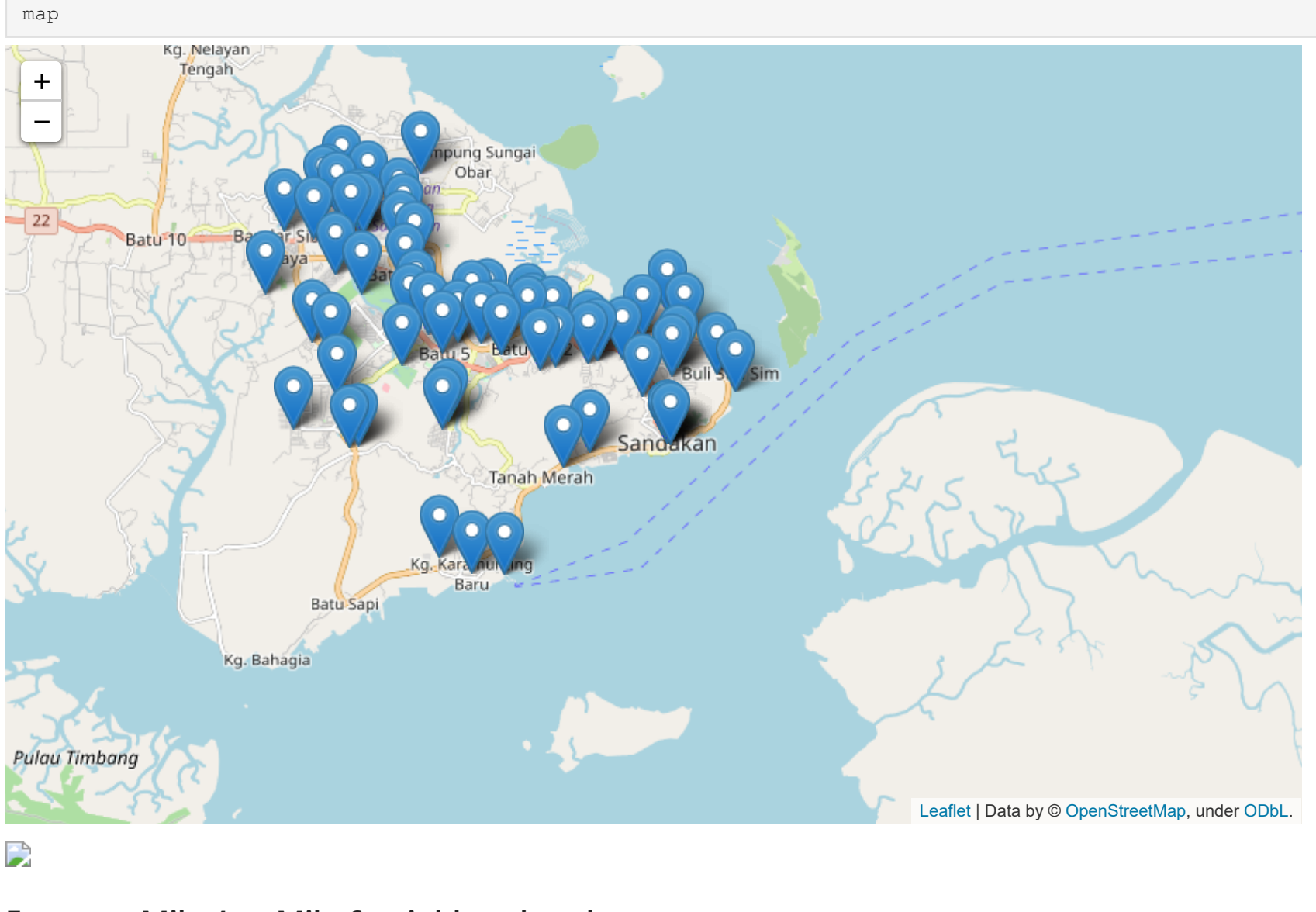
Out[31]:



In [32]:
```
#Add Markers with Popup
for point in range(0,suburbs_list_size):
    folium.Marker(suburbs_list[point], popup=df['Neighbourhood'][point]).add_to(map)
```

In [33]:
```
map
```

Out[33]:



Focus on Mile 4 to Mile 6 neighbourhoods

In [34]:
```
df1 = pd.read_csv("segment.csv")
```

In [35]:
```
df1
```

Out[35]:

	Neighbourhood	Area	Residential Units	Latitude	Longtitude
0	Bunga Matahari	11.880	172	5.865810	118.075874
1	Casa San Uno	38.890	307	5.865233	118.072556
2	Damai & Sri Taman	21.670	123	5.858482	118.078921
3	Evergreen	23.990	48	5.873464	118.057834
4	Garden Villa	25.760	82	5.863280	118.048945
5	Indah	56.270	356	5.842067	118.066095
6	Indah Jaya	235.680	2752	5.843796	118.067200
7	Lucky & Wemin	43.544	260	5.863112	118.062768
8	Mesra	23.180	1000	5.861271	118.077664
9	Pertama	53.320	438	5.861339	118.069276
10	Tinosan	24.510	235	5.861032	118.074517
11	Tshun Ngen	70.260	304	5.858728	118.065804
12	Tyng	111.700	585	5.867669	118.059997
13	Utama	41.263	329	5.864601	118.058569

In [36]:
```
df1.shape
```

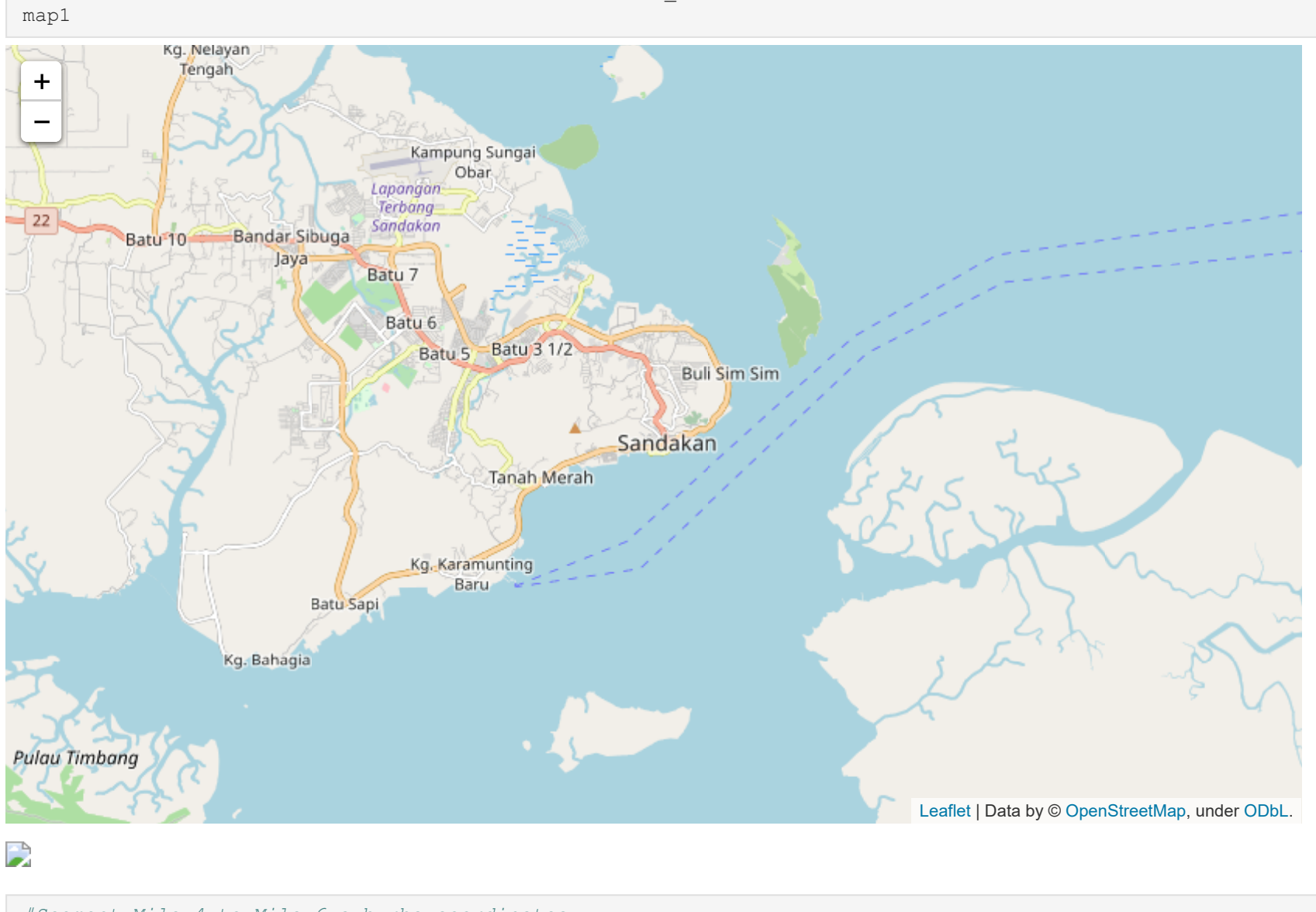
Out[36]: (14, 5)

In [37]:
```
address = 'Sandakan'

geolocator = Nominatim(user_agent="foursquare_agent")

location = geolocator.geocode(address)

latitude = location.latitude
longitude = location.longitude
print(latitude, longitude)
```

5.8391337 118.1158919

In [38]:
```
#Sandakan Map
map1 = folium.Map(location=[latitude,longitude], zoom_start=12)
map1
```

Out[38]:



In [39]:
```
#Segment Mile 4 to Mile 6 suburbs coordinates

df1_suburbs = df1[['Latitude','Longtitude']]
```

In [40]:
```
df1_suburbs
```

Out[40]:

	Latitude	Longtitude
0	5.865810	118.075874
1	5.865233	118.072556
2	5.858482	118.078921
3	5.873464	118.057834
4	5.863280	118.048945
5	5.842067	118.066095
6	5.843796	118.067200
7	5.863112	118.062768
8	5.861271	118.077664
9	5.861339	118.069276
10	5.861032	118.074517
11	5.858728	118.065804
12	5.867669	118.059997
13	5.864601	118.058569

In [41]:
```
df1_suburbs.shape
```

Out[41]: (14, 2)

In [42]:
```
suburbs1_list = df1_suburbs.values.tolist()
```

In [43]:
```
suburbs1_list
```

Out[43]:
```
[[5.8658095, 118.0758739],
 [5.8652331, 118.0725562],
 [5.8584821, 118.0789209],
 [5.8734638, 118.0578338],
 [5.8632801, 118.0489446],
 [5.8420670999999995, 118.06609499999999],
 [5.843796, 118.06720049999999],
 [5.8631125, 118.0627675],
 [5.8612709, 118.07766389999999],
 [5.8613392, 118.06927649999999],
 [5.8610315, 118.0745175],
 [5.85872755, 118.0658041],
 [5.8676695, 118.0599967],
 [5.864600599999999, 118.05856909999999]]
```

In [44]:
```
suburbs1_list_size = len(suburbs1_list)
```

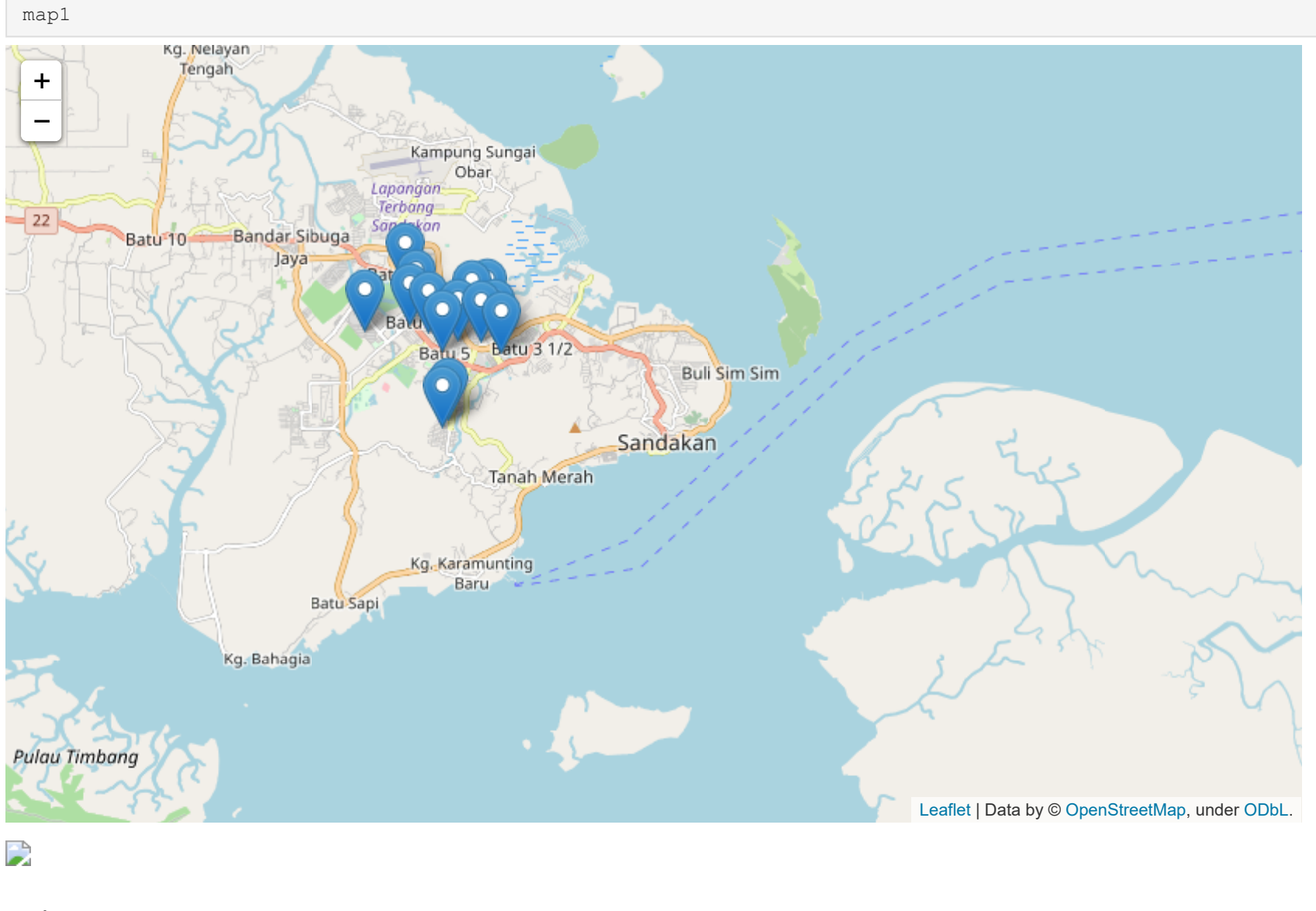
In [45]:
```
suburbs1_list_size
```

Out[45]: 14

In [46]:
```
#Add Markers with Popup
for point in range(0,suburbs1_list_size):
    folium.Marker(suburbs1_list[point], popup=df1['Neighbourhood'][point]).add_to(map1)
```

In [47]:
```
map1
```

Out[47]:



Using Foursquare API

Explore Neighborhoods with that focused segment

In [48]:
```
#define our Foursquare credentials and version
CLIENT_ID = 'G5ETWK2SC3RW4050IEELGL3COFWTLZWGQJ5WXKJFB5NZST3Z' # your Foursquare ID
CLIENT_SECRET = 'F4LEZEDAHXLETDPUPFQCTMII5LIQL4J34L2EZGGUYGVOKF4C' # your Foursquare Secret
VERSION = '20180604'
LIMIT = 15
```

In [49]:
```
neighborhoods_subset = df1[['Neighbourhood','Latitude','Longtitude']]
```

In [50]:
```
neighborhoods_subset
```

Out[50]:

	Neighbourhood	Latitude	Longtitude
0	Bunga Matahari	5.865810	118.075874
1	Casa San Uno	5.865233	118.072556
2	Damai & Sri Taman	5.858482	118.078921
3	Evergreen	5.873464	118.057834
4	Garden Villa	5.863280	118.048945
5	Indah	5.842067	118.066095
6	Indah Jaya	5.843796	118.067200
7	Lucky & Wemin	5.863112	118.062768
8	Mesra	5.861271	118.077664
9	Pertama	5.861339	118.069276
10	Tinosan	5.861032	118.074517
11	Tshun Ngen	5.858728	118.065804

12	Tyng	5.867669	118.059997
13	Utama	5.864601	118.058569

In [51]:
```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

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

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

        # return only relevant information for each nearby venue
        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 = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

In [52]:
```
target_venues = getNearbyVenues(names=neighborhoods_subset['Neighbourhood'],
                                   latitudes=neighborhoods_subset['Latitude'],
                                   longitudes=neighborhoods_subset['Longtitude']
                                  )
```

```
Bunga Matahari
Casa San Uno
Damai & Sri Taman
Evergreen
Garden Villa
Indah
Indah Jaya
Lucky & Wemin
Mesra
Pertama
Tinosan
Tshun Ngen
Tyng
Utama
```

In [53]:
```
print(target_venues.shape)
target_venues
```

```
(130, 7)
```

Out[53]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bunga Matahari	5.865810	118.075874	Sunflower Mini Market Taman Bunga Matahari	5.865544	118.076034	Convenience Store
1	Bunga Matahari	5.865810	118.075874	pizza hut,giant sandakan	5.866178	118.074039	Fast Food Restaurant
2	Bunga Matahari	5.865810	118.075874	Lee Yuan Chinese Restaurant	5.864253	118.073965	Chinese Restaurant
3	Bunga Matahari	5.865810	118.075874	BB yusri Lim	5.866836	118.072678	BBQ Joint
4	Bunga Matahari	5.865810	118.075874	Taman Tinosan	5.862603	118.073679	Other Great Outdoors
5	Bunga Matahari	5.865810	118.075874	Lubuk	5.870108	118.075571	Outdoors & Recreation
6	Casa San Uno	5.865233	118.072556	BB yusri Lim	5.866836	118.072678	BBQ Joint
7	Damai & Sri Taman	5.858482	118.078921	Servay Supermarket	5.858409	118.078295	Grocery Store
8	Damai & Sri Taman	5.858482	118.078921	Bandar Kim Fung 金凤市	5.856492	118.078332	Town
9	Damai & Sri Taman	5.858482	118.078921	Novelty Cafe & Cake House	5.857198	118.079499	Bakery
10	Damai & Sri Taman	5.858482	118.078921	Livingstone Hotel	5.857344	118.081836	Hotel
11	Damai & Sri Taman	5.858482	118.078921	双日Kopitiam茶餐室	5.856717	118.077535	Deli / Bodega
12	Damai & Sri Taman	5.858482	118.078921	Kim Fung Market	5.857622	118.079159	Food Court
13	Damai & Sri Taman	5.858482	118.078921	Faces Nasi Kuning Ayam	5.858331	118.080795	Wings Joint
14	Damai & Sri Taman	5.858482	118.078921	Tien Kee Restaurant 田記港式燒腊	5.856493	118.079025	BBQ Joint
15	Damai & Sri Taman	5.858482	118.078921	Digital Wise Sdn Bhd	5.857668	118.080302	Electronics Store
16	Damai & Sri Taman	5.858482	118.078921	三點三茶餐厅 Kedai Kopi Kong Fei	5.857155	118.080185	Deli / Bodega
17	Damai & Sri Taman	5.858482	118.078921	7 Eleven	5.857464	118.079115	Convenience Store
18	Damai & Sri Taman	5.858482	118.078921	2020 Restaurant	5.856228	118.077870	Chinese Restaurant
19	Damai & Sri Taman	5.858482	118.078921	KFC 肯德基	5.857699	118.078737	Fast Food Restaurant
20	Damai & Sri Taman	5.858482	118.078921	Sun Tai Chung Supermarket (S) Sdn. Bhd.	5.857285	118.079529	Fruit & Vegetable Store
21	Damai & Sri Taman	5.858482	118.078921	Tealive	5.856693	118.074871	Bubble Tea Shop
22	Evergreen	5.873464	118.057834	Sandakan Golf & Country Club	5.872148	118.054561	Golf Course
23	Evergreen	5.873464	118.057834	SGCC Gym & Fitness Centre	5.872099	118.054609	Gym
24	Evergreen	5.873464	118.057834	Sandakan Golf And Country Club	5.877885	118.058050	Golf Course
25	Garden Villa	5.863280	118.048945	Fresh & Tasty	5.864483	118.052092	Restaurant
26	Garden Villa	5.863280	118.048945	New Ocean King Seafood, IJM, Batu 5	5.864866	118.051724	Seafood Restaurant
27	Garden Villa	5.863280	118.048945	Crowd 99 Cafe	5.864673	118.052402	Café
28	Garden Villa	5.863280	118.048945	Up Case Café	5.865119	118.051937	Café
29	Garden Villa	5.863280	118.048945	KKI (Kari Kepala Ikan)	5.862239	118.052539	Bistro
30	Garden Villa	5.863280	118.048945	Nasi Kandar anak mami	5.863824	118.052159	Indian Restaurant
31	Garden Villa	5.863280	118.048945	Baker Field	5.864836	118.053030	Café
32	Garden Villa	5.863280	118.048945	Oppa Korean Restaurant	5.863989	118.051933	Korean Restaurant
33	Garden Villa	5.863280	118.048945	The Sunset Beer Garden	5.860161	118.051310	Beer Garden
34	Garden Villa	5.863280	118.048945	New Happy Family Restaurant 新家乐茶餐室	5.864302	118.052055	Breakfast Spot
35	Garden Villa	5.863280	118.048945	Framelicious Studio	5.862110	118.052486	Photography Studio
36	Garden Villa	5.863280	118.048945	Mell's Fresco Bites Cafe WK	5.864673	118.052495	Lounge
37	Garden Villa	5.863280	118.048945	家乡牛杂	5.864679	118.052592	Hakka Restaurant
38	Garden Villa	5.863280	118.048945	Papparato	5.865549	118.052506	Coffee Shop
39	Garden Villa	5.863280	118.048945	MingCha 茗茶	5.865332	118.052660	Bubble Tea Shop
40	Indah	5.842067	118.066095	Indah Jaya Recreational Club	5.842595	118.068523	Recreation Center
41	Indah	5.842067	118.066095	Fresh Market - 日日鲜蔬果鱼市场	5.842627	118.067305	Grocery Store
42	Indah	5.842067	118.066095	Restoran Syn Kee Siang 新奇香茶餐室	5.843448	118.067123	Chinese Restaurant
43	Indah	5.842067	118.066095	Indah Park	5.843113	118.066798	Baseball Stadium
44	Indah	5.842067	118.066095	Kedai Kopi Indah	5.842939	118.067024	Coffee Shop
45	Indah	5.842067	118.066095	Anjung Sukau Baru	5.840170	118.066819	Food Truck
46	Indah	5.842067	118.066095	Al Falah Curry House	5.843534	118.067780	Indian Restaurant
47	Indah	5.842067	118.066095	Happy Mart	5.842860	118.067219	Department Store
48	Indah	5.842067	118.066095	Indah Jaya Recreational Club Gymnasium	5.842736	118.068631	Athletics & Sports
49	Indah	5.842067	118.066095	Tomato Steamboat & Grill	5.842560	118.068717	Asian Restaurant
50	Indah	5.842067	118.066095	Sandakan Paintball	5.843511	118.068872	Athletics & Sports
51	Indah Jaya	5.843796	118.067200	Indah Jaya Recreational Club	5.842595	118.068523	Recreation Center
52	Indah Jaya	5.843796	118.067200	Indah Jaya Recreational Club Gymnasium	5.842736	118.068631	Athletics & Sports
53	Indah Jaya	5.843796	118.067200	Al Falah Curry House	5.843534	118.067780	Indian Restaurant
54	Indah Jaya	5.843796	118.067200	Indah Park	5.843113	118.066798	Baseball Stadium
55	Indah Jaya	5.843796	118.067200	Restoran Makanan Laut Riang Ria 新佳乐海鲜楼	5.843363	118.066555	Chinese Restaurant
56	Indah Jaya	5.843796	118.067200	Happy Mart	5.842860	118.067219	Department Store
57	Lucky & Wemin	5.863112	118.062768	restoran ikan bakar sederhana 2	5.864050	118.060908	Halal Restaurant
58	Mesra	5.861271	118.077664	Servay Supermarket	5.858409	118.078295	Grocery Store
59	Mesra	5.861271	118.077664	Novelty Cafe & Cake House	5.857198	118.079499	Bakery
60	Mesra	5.861271	118.077664	Kim Fung Market	5.857622	118.079159	Food Court
61	Mesra	5.861271	118.077664	Faces Nasi Kuning Ayam	5.858331	118.080795	Wings Joint
62	Mesra	5.861271	118.077664	Digital Wise Sdn Bhd	5.857668	118.080302	Electronics Store
63	Mesra	5.861271	118.077664	7 Eleven	5.857464	118.079115	Convenience Store
64	Mesra	5.861271	118.077664	KFC 肯德基	5.857699	118.078737	Fast Food Restaurant
65	Mesra	5.861271	118.077664	Servay Mile 4	5.859961	118.079498	Department Store
66	Mesra	5.861271	118.077664	Sun Tai Chung Supermarket (S) Sdn. Bhd.	5.857285	118.079529	Fruit & Vegetable Store
67	Mesra	5.861271	118.077664	Whimsy, Bandar Kim Fung	5.857785	118.078515	Music Venue
68	Mesra	5.861271	118.077664	别不同	5.858310	118.079790	Chinese Restaurant
69	Mesra	5.861271	118.077664	Santino's Pizza	5.858846	118.073881	Pizza Place
70	Mesra	5.861271	118.077664	ON ON	5.858700	118.074696	Electronics Store
71	Mesra	5.861271	118.077664	Taman Tinosan	5.862603	118.073679	Other Great Outdoors
72	Mesra	5.861271	118.077664	Bubble Tea House	5.858601	118.074406	Café
73	Pertama	5.861339	118.069276	Tai Lou Restaurant	5.858214	118.067683	Chinese Restaurant
74	Pertama	5.861339	118.069276	MyMee - 我的面	5.858351	118.066022	Noodle House
75	Pertama	5.861339	118.069276	Wong Yau Furnitures	5.859627	118.065442	Furniture / Home Store
76	Tinosan	5.861032	118.074517	Brew Cottage	5.857391	118.074090	Café
77	Tinosan	5.861032	118.074517	Hotel D' North Star	5.856912	118.072872	Bed & Breakfast
78	Tinosan	5.861032	118.074517	Restoran Tom Yam	5.857198	118.073182	Austrian Restaurant
79	Tinosan	5.861032	118.074517	Tealive	5.856693	118.074871	Bubble Tea Shop
80	Tinosan	5.861032	118.074517	7 Eleven	5.857466	118.073342	Convenience Store
81	Tinosan	5.861032	118.074517	Nurshinie Curry House	5.857267	118.075027	Indian Restaurant
82	Tinosan	5.861032	118.074517	Santino's Pizza	5.858846	118.073881	Pizza Place
83	Tinosan	5.861032	118.074517	Cielo Blu Tasty Recipe	5.856815	118.074085	Diner
84	Tinosan	5.861032	118.074517	Dragon Restaurant 良记鱼旦	5.856901	118.073410	Noodle House
85	Tinosan	5.861032	118.074517	Red Wagon Warung Kopi (WARKOP)	5.858697	118.074277	Café
86	Tinosan	5.861032	118.074517	Apple Pub & Bar	5.858367	118.073475	Pub
87	Tinosan	5.861032	118.074517	Stephen Restaurant 肥佬饭店	5.858307	118.073628	Chinese Restaurant
88	Tinosan	5.861032	118.074517	满香楼	5.858098	118.074042	Chinese Restaurant
89	Tinosan	5.861032	118.074517	Tac de Cafe 亚弟牛屎塔	5.857569	118.073700	Bakery
90	Tinosan	5.861032	118.074517	Restoran Modern Family	5.858700	118.074464	Malay Restaurant
91	Tshun Ngen	5.858728	118.065804	Tai Lou Restaurant	5.858214	118.067683	Chinese Restaurant
92	Tshun Ngen	5.858728	118.065804	MyMee - 我的面	5.858351	118.066022	Noodle House
93	Tshun Ngen	5.858728	118.065804	Cafeteria Guan Huat	5.858504	118.066122	Vegetarian / Vegan Restaurant
94	Tshun Ngen	5.858728	118.065804	Kedai Makanan Kim Boo MyMee 我的面	5.858421	118.065962	Chinese Restaurant
95	Tshun Ngen	5.858728	118.065804	源发茶餐室	5.858579	118.066385	Asian Restaurant
96	Tshun Ngen	5.858728	118.065804	Foodmart	5.860066	118.066120	Fruit & Vegetable Store
97	Tshun Ngen	5.858728	118.065804	Annie's Cooking Studio	5.858568	118.066010	Grocery Store
98	Tshun Ngen	5.858728	118.065804	Ikan Bakar Esso	5.858763	118.066376	Fish & Chips Shop
99	Tshun Ngen	5.858728	118.065804	Ikan Bakar Samudera	5.861129	118.064573	Food Truck
100	Tyng	5.867669	118.059997	Ice City Drink House 冰城冰果屋	5.867956	118.059335	Cupcake Shop
101	Tyng	5.867669	118.059997	Starbucks	5.865550	118.057200	Coffee Shop
102	Tyng	5.867669	118.059997	Moon n Star 生肉面	5.867019	118.058930	Noodle House
103	Tyng	5.867669	118.059997	Edo Ichi 江户一日本料理	5.867432	118.058584	Sushi Restaurant
104	Tyng	5.867669	118.059997	Tyng Garden Hotel	5.867625	118.060904	Hotel
105	Tyng	5.867669	118.059997	Fresh & Frozen Market 鲜美坊	5.866783	118.059605	Market
106	Tyng	5.867669	118.059997	The Fresh Meat Shop (non halal)	5.866112	118.060222	Butcher
107	Tyng	5.867669	118.059997	叶贵凉粉	5.867614	118.060908	Asian Restaurant
108	Tyng	5.867669	118.059997	Restaurant Yum Yum 老婆茶餐室	5.867546	118.059085	Chinese Restaurant
109	Tyng	5.867669	118.059997	Labuk Tyng Restaurant	5.867792	118.060783	Asian Restaurant
110	Tyng	5.867669	118.059997	Kedai Kopi Kim Lok - 金乐小食店	5.867607	118.060890	Noodle House
111	Tyng	5.867669	118.059997	After Taste Restaurant	5.867637	118.060870	Chinese Restaurant
112	Tyng	5.867669	118.059997	Mr Bun 好運食店	5.867682	118.060835	Bakery
113	Tyng	5.867669	118.059997	Kedai Kopi Kim Fah	5.868088	118.059140	Chinese Restaurant
114	Tyng	5.867669	118.059997	Yusirina Cafeteria	5.867620	118.060875	Asian Restaurant
115	Utama	5.864601	118.058569	Starbucks	5.865550	118.057200	Coffee Shop
116	Utama	5.864601	118.058569	Ice City Drink House 冰城冰果屋	5.867956	118.059335	Cupcake Shop
117	Utama	5.864601	118.058569	Mother's Cook & Catering	5.865463	118.055916	Indonesian Restaurant
118	Utama	5.864601	118.058569	Moon n Star 生肉面	5.867019	118.058930	Noodle House
119	Utama	5.864601	118.058569	McDonald's & McCafé	5.863777	118.056093	Fast Food Restaurant
120	Utama	5.864601	118.058569	Edo Ichi 江户一日本料理	5.867432	118.058584	Sushi Restaurant
121	Utama	5.864601	118.058569	Dania Kopitiam	5.865076	118.057538	Malay Restaurant
122	Utama	5.864601	118.058569	Old City	5.865347	118.057907	Café
123	Utama	5.864601	118.058569	Ginza Food & Drink 金沙美食軒	5.864797	118.057367	Asian Restaurant
124	Utama	5.864601	118.058569	Florence Boutique®	5.865016	118.056767	Women's Store
125	Utama	5.864601	118.058569	Moghul Tijani	5.865658	118.056883	Indian Restaurant
126	Utama	5.864601	118.058569	Giant Hypermarket	5.866648	118.057490	Department Store
127	Utama	5.864601	118.058569	Restaurant Sri Mersing	5.866143	118.056420	Restaurant
128	Utama	5.864601	118.058569	Iqram Corner	5.864508	118.055739	Malay Restaurant
129	Utama	5.864601	118.058569	Good Taste Restaurant 美口福飯店	5.864633	118.055695	Chinese Restaurant

In [54]:
```
#Save a copy of csv
#target_venues.to_csv("foursq.csv",index=False)
```

In [55]:
```
target_venues.groupby('Neighborhood').count()
```

Out[55]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bunga Matahari	6	6	6	6	6	6
Casa San Uno	1	1	1	1	1	1
Damai & Sri Taman	15	15	15	15	15	15
Evergreen	3	3	3	3	3	3
Garden Villa	15	15	15	15	15	15
Indah	11	11	11	11	11	11
Indah Jaya	6	6	6	6	6	6
Lucky & Wemin	1	1	1	1	1	1
Mesra	15	15	15	15	15	15
Pertama	3	3	3	3	3	3
Tinosan	15	15	15	15	15	15
Tshun Ngen	9	9	9	9	9	9
Tyng	15	15	15	15	15	15
Utama	15	15	15	15	15	15

In [56]:
```
print('There are {} uniques categories.'.format(len(target_venues['Venue Category'].unique())))
```

```
There are 54 uniques categories.
```

Analyze Each Neighborhood

In [57]:
```
# one hot encoding
target_onehot = pd.get_dummies(target_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
target_onehot['Neighborhood'] = target_venues['Neighborhood']

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

target_onehot.head()
```

Out[57]:

	Neighborhood	Asian Restaurant	Athletics & Sports	Austrian Restaurant	BBQ Joint	Bakery	Baseball Stadium	Bed & Breakfast	Beer Garden	Bistro	Breakfast Spot	Bubble Tea Shop	Butcher	Café
0	Bunga Matahari	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Bunga Matahari	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Bunga Matahari	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Bunga Matahari	0	0	0	1	0	0	0	0	0	0	0	0	0
4	Bunga Matahari	0	0	0	0	0	0	0	0	0	0	0	0	0

In [58]:
```
target_onehot.shape
```

Out[58]:
```
(130, 55)
```

In [59]:
```
target_grouped = target_onehot.groupby('Neighborhood').mean().reset_index()
target_grouped
```

Out[59]:

	Neighborhood	Asian Restaurant	Athletics & Sports	Austrian Restaurant	BBQ Joint	Bakery	Baseball Stadium	Bed & Breakfast	Beer Garden	Bistro	Breakfast Spot	Bubble Tea Shop	Butc
0	Bunga Matahari	0.000000	0.000000	0.000000	0.166667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
1	Casa San Uno	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
2	Damai & Sri Taman	0.000000	0.000000	0.000000	0.066667	0.066667	0.000000	0.000000	0.000000	0.000000	0.000000	0.066667	0.000
3	Evergreen	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
4	Garden Villa	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.066667	0.066667	0.066667	0.066667	0.000
5	Indah	0.090909	0.181818	0.000000	0.000000	0.000000	0.090909	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
6	Indah Jaya	0.000000	0.166667	0.000000	0.000000	0.000000	0.166667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
7	Lucky & Wemin	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
8	Mesra	0.000000	0.000000	0.000000	0.000000	0.066667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
9	Pertama	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
10	Tinosan	0.000000	0.000000	0.066667	0.000000	0.066667	0.000000	0.066667	0.000000	0.000000	0.000000	0.066667	0.000
11	Tshun Ngen	0.111111	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
12	Tyng	0.200000	0.000000	0.000000	0.000000	0.066667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.066
13	Utama	0.066667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000

In [60]:
```
target_grouped.shape
```

Out[60]:
```
(14, 55)
```

In [61]:
```
# Print out top 5 venues for each neighbourhood
num_top_venues = 5

for hood in target_grouped['Neighborhood']:
    print("----"+hood+"----")
    temp = target_grouped[target_grouped['Neighborhood'] == hood].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

```
----Bunga Matahari----
                   venue  freq
0  Outdoors & Recreation  0.17
1   Other Great Outdoors  0.17
2              BBQ Joint  0.17
3   Fast Food Restaurant  0.17
4     Chinese Restaurant  0.17


----Casa San Uno----
                  venue  freq
0             BBQ Joint   1.0
1      Asian Restaurant   0.0
2  Other Great Outdoors   0.0
3      Hakka Restaurant   0.0
4      Halal Restaurant   0.0


----Damai & Sri Taman----
               venue  freq
0      Deli / Bodega  0.13
1              Hotel  0.07
2      Grocery Store  0.07
3         Food Court  0.07
4  Convenience Store  0.07


----Evergreen----
              venue  freq
0       Golf Course  0.67
1               Gym  0.33
2      Noodle House  0.00
3  Hakka Restaurant  0.00
4  Halal Restaurant  0.00


----Garden Villa----
                venue  freq
0                Café  0.20
1     Bubble Tea Shop  0.07
2    Hakka Restaurant  0.07
3   Indian Restaurant  0.07
4  Photography Studio  0.07


----Indah----
                venue  freq
0  Athletics & Sports  0.18
1    Asian Restaurant  0.09
2          Food Truck  0.09
3         Coffee Shop  0.09
4  Chinese Restaurant  0.09


----Indah Jaya----
                venue  freq
0  Athletics & Sports  0.17
1    Baseball Stadium  0.17
2   Indian Restaurant  0.17
3   Recreation Center  0.17
4  Chinese Restaurant  0.17


----Lucky & Wemin----
              venue  freq
0  Halal Restaurant   1.0
1  Asian Restaurant   0.0
2      Noodle House   0.0
3               Gym   0.0
4  Hakka Restaurant   0.0


----Mesra----
                  venue  freq
0     Electronics Store  0.13
1  Fast Food Restaurant  0.07
2            Food Court  0.07
3           Wings Joint  0.07
4      Department Store  0.07


----Pertama----
                    venue  freq
0            Noodle House  0.33
1  Furniture / Home Store  0.33
2      Chinese Restaurant  0.33
3        Asian Restaurant  0.00
4        Hakka Restaurant  0.00


----Tinosan----
                venue  freq
0  Chinese Restaurant  0.13
1                Café  0.13
2   Convenience Store  0.07
3    Malay Restaurant  0.07
4         Pizza Place  0.07


----Tshun Ngen----
                     venue  freq
0       Chinese Restaurant  0.22
1         Asian Restaurant  0.11
2  Fruit & Vegetable Store  0.11
3               Food Truck  0.11
4        Fish & Chips Shop  0.11


----Tyng----
                venue  freq
0    Asian Restaurant  0.20
1  Chinese Restaurant  0.20
2        Noodle House  0.13
3             Butcher  0.07
4               Hotel  0.07


----Utama----
                venue  freq
0    Malay Restaurant  0.13
1    Asian Restaurant  0.07
2  Chinese Restaurant  0.07
3   Sushi Restaurant  0.07
4          Restaurant  0.07
```

In [62]:
```
#Create a function to return common venues
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]
```

In [63]:
```
#Sort each neighbourhood with top 10 venues
num_top_venues = 10

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

# create columns according to number of top venues
columns = ['Neighborhood']
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))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = target_grouped['Neighborhood']

for ind in np.arange(target_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(target_grouped.iloc[ind, :], num_t

neighborhoods_venues_sorted
```

Out[63]:

	Neighborhood	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 M Comm Ver
0	Bunga Matahari	Chinese Restaurant	BBQ Joint	Fast Food Restaurant	Outdoors & Recreation	Other Great Outdoors	Convenience Store	Women's Store	Food Court	Fish & Chips Shop	Electro St
1	Casa San Uno	BBQ Joint	Women's Store	Chinese Restaurant	Food Truck	Food Court	Fish & Chips Shop	Fast Food Restaurant	Electronics Store	Diner	Departm St
2	Damai & Sri Taman	Deli / Bodega	Fruit & Vegetable Store	Fast Food Restaurant	Chinese Restaurant	Food Court	Bubble Tea Shop	Hotel	Electronics Store	Grocery Store	Bak
3	Evergreen	Golf Course	Gym	Women's Store	Fruit & Vegetable Store	Food Court	Fish & Chips Shop	Fast Food Restaurant	Electronics Store	Diner	Departm St
4	Garden Villa	Café	Photography Studio	Indian Restaurant	Korean Restaurant	Lounge	Bubble Tea Shop	Breakfast Spot	Bistro	Beer Garden	Co Sl
5	Indah	Athletics & Sports	Asian Restaurant	Recreation Center	Chinese Restaurant	Food Truck	Indian Restaurant	Department Store	Coffee Shop	Baseball Stadium	Groc St
6	Indah Jaya	Athletics & Sports	Baseball Stadium	Recreation Center	Department Store	Indian Restaurant	Chinese Restaurant	Women's Store	Coffee Shop	Food Court	Fis Chips Sl
7	Lucky & Wemin	Halal Restaurant	Women's Store	Chinese Restaurant	Food Court	Fish & Chips Shop	Fast Food Restaurant	Electronics Store	Diner	Department Store	De Bod
8	Mesra	Electronics Store	Pizza Place	Convenience Store	Café	Food Court	Music Venue	Other Great Outdoors	Department Store	Chinese Restaurant	Groc St
9	Pertama	Furniture / Home Store	Chinese Restaurant	Noodle House	Austrian Restaurant	BBQ Joint	Food Court	Fish & Chips Shop	Fast Food Restaurant	Electronics Store	Di
10	Tinosan	Chinese Restaurant	Café	Pub	Malay Restaurant	Indian Restaurant	Convenience Store	Noodle House	Bubble Tea Shop	Bed & Breakfast	Di
11	Tshun Ngen	Chinese Restaurant	Asian Restaurant	Noodle House	Fish & Chips Shop	Grocery Store	Food Truck	Fruit & Vegetable Store	Vegetarian / Vegan Restaurant	Baseball Stadium	Be Break
12	Tyng	Asian Restaurant	Chinese Restaurant	Noodle House	Hotel	Cupcake Shop	Coffee Shop	Market	Butcher	Sushi Restaurant	Bak
13	Utama	Malay Restaurant	Women's Store	Noodle House	Café	Chinese Restaurant	Coffee Shop	Cupcake Shop	Department Store	Fast Food Restaurant	Inc Restau

Clustering Neighborhoods - Using K-means method

In [64]:
```
# set number of clusters
kclusters = 5

target_grouped_clustering = target_grouped.drop('Neighborhood', 1)

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:15]
```

Out[64]:
```
array([1, 0, 1, 2, 1, 1, 1, 3, 1, 4, 1, 1, 1, 1])
```

In [65]:
```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

target_merged = neighborhoods_subset

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
target_merged = target_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighbourhood'

target_merged.head() # check the last columns!
```

Out[65]:

	Neighbourhood	Latitude	Longtitude	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 M Comm Ven
0	Bunga Matahari	5.865810	118.075874	1	Chinese Restaurant	BBQ Joint	Fast Food Restaurant	Outdoors & Recreation	Other Great Outdoors	Convenience Store	Women's Store	Fo Co
1	Casa San Uno	5.865233	118.072556	0	BBQ Joint	Women's Store	Chinese Restaurant	Food Truck	Food Court	Fish & Chips Shop	Fast Food Restaurant	Electron St
2	Damai & Sri Taman	5.858482	118.078921	1	Deli / Bodega	Fruit & Vegetable Store	Fast Food Restaurant	Chinese Restaurant	Food Court	Bubble Tea Shop	Hotel	Electron St
3	Evergreen	5.873464	118.057834	2	Golf Course	Gym	Women's Store	Fruit & Vegetable Store	Food Court	Fish & Chips Shop	Fast Food Restaurant	Electron St
4	Garden Villa	5.863280	118.048945	1	Café	Photography Studio	Indian Restaurant	Korean Restaurant	Lounge	Bubble Tea Shop	Breakfast Spot	Bis

In [66]:
```
target_merged #Do a check for all
```

Out[66]:

	Neighbourhood	Latitude	Longtitude	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
0	Bunga Matahari	5.865810	118.075874	1	Chinese Restaurant	BBQ Joint	Fast Food Restaurant	Outdoors & Recreation	Other Great Outdoors	Convenience Store	Women's Store
1	Casa San Uno	5.865233	118.072556	0	BBQ Joint	Women's Store	Chinese Restaurant	Food Truck	Food Court	Fish & Chips Shop	Fast Food Restaurant
2	Damai & Sri Taman	5.858482	118.078921	1	Deli / Bodega	Fruit & Vegetable Store	Fast Food Restaurant	Chinese Restaurant	Food Court	Bubble Tea Shop	Hotel
3	Evergreen	5.873464	118.057834	2	Golf Course	Gym	Women's Store	Fruit & Vegetable Store	Food Court	Fish & Chips Shop	Fast Food Restaurant
4	Garden Villa	5.863280	118.048945	1	Café	Photography Studio	Indian Restaurant	Korean Restaurant	Lounge	Bubble Tea Shop	Breakfast Spot
5	Indah	5.842067	118.066095	1	Athletics & Sports	Asian Restaurant	Recreation Center	Chinese Restaurant	Food Truck	Indian Restaurant	Department Store
6	Indah Jaya	5.843796	118.067200	1	Athletics & Sports	Baseball Stadium	Recreation Center	Department Store	Indian Restaurant	Chinese Restaurant	Women's Store
7	Lucky & Wemin	5.863112	118.062768	3	Halal Restaurant	Women's Store	Chinese Restaurant	Food Court	Fish & Chips Shop	Fast Food Restaurant	Electronics Store
8	Mesra	5.861271	118.077664	1	Electronics Store	Pizza Place	Convenience Store	Café	Food Court	Music Venue	Other Great Outdoors
9	Pertama	5.861339	118.069276	4	Furniture / Home Store	Chinese Restaurant	Noodle House	Austrian Restaurant	BBQ Joint	Food Court	Fish & Chips Shop
10	Tinosan	5.861032	118.074517	1	Chinese Restaurant	Café	Pub	Malay Restaurant	Indian Restaurant	Convenience Store	Noodle House
11	Tshun Ngen	5.858728	118.065804	1	Chinese Restaurant	Asian Restaurant	Noodle House	Fish & Chips Shop	Grocery Store	Food Truck	Fruit & Vegetable Store
12	Tyng	5.867669	118.059997	1	Asian Restaurant	Chinese Restaurant	Noodle House	Hotel	Cupcake Shop	Coffee Shop	Market
13	Utama	5.864601	118.058569	1	Malay Restaurant	Women's Store	Noodle House	Café	Chinese Restaurant	Coffee Shop	Cupcake Shop

In [67]:
```
#target_merged.drop(index=1, inplace=True) #Drop Casa San Uno as there are NaNs
```

In [68]:
```
#target_merged
```

In [69]:
```
target_merged.info()
```

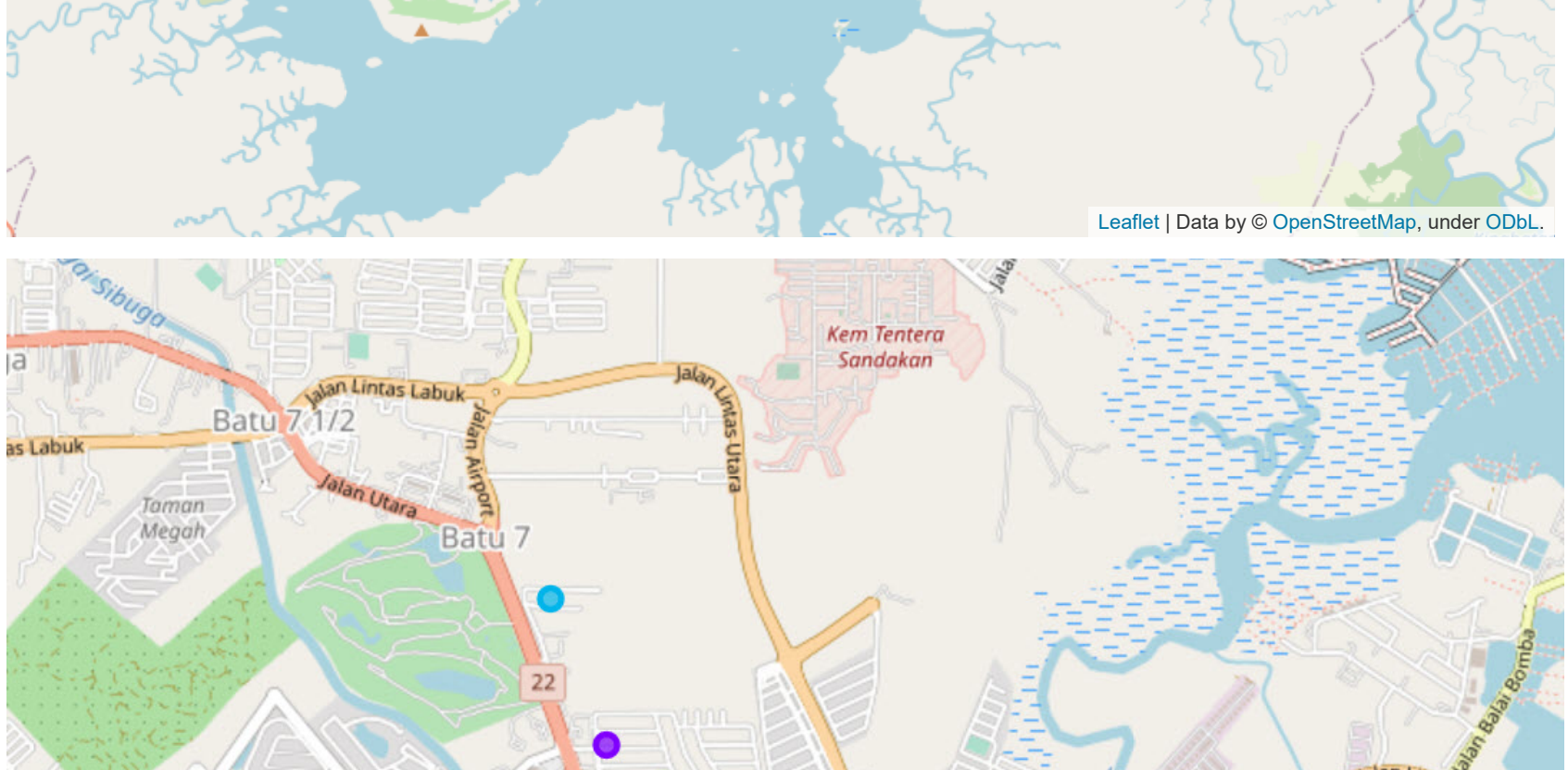
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14 entries, 0 to 13
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Neighbourhood           14 non-null     object
 1   Latitude                14 non-null     float64
 2   Longtitude              14 non-null     float64
 3   Cluster Labels          14 non-null     int32
 4   1st Most Common Venue   14 non-null     object
 5   2nd Most Common Venue   14 non-null     object
 6   3rd Most Common Venue   14 non-null     object
 7   4th Most Common Venue   14 non-null     object
 8   5th Most Common Venue   14 non-null     object
 9   6th Most Common Venue   14 non-null     object
 10  7th Most Common Venue   14 non-null     object
 11  8th Most Common Venue   14 non-null     object
 12  9th Most Common Venue   14 non-null     object
 13  10th Most Common Venue  14 non-null     object
dtypes: float64(2), int32(1), object(11)
memory usage: 1.6+ KB
```

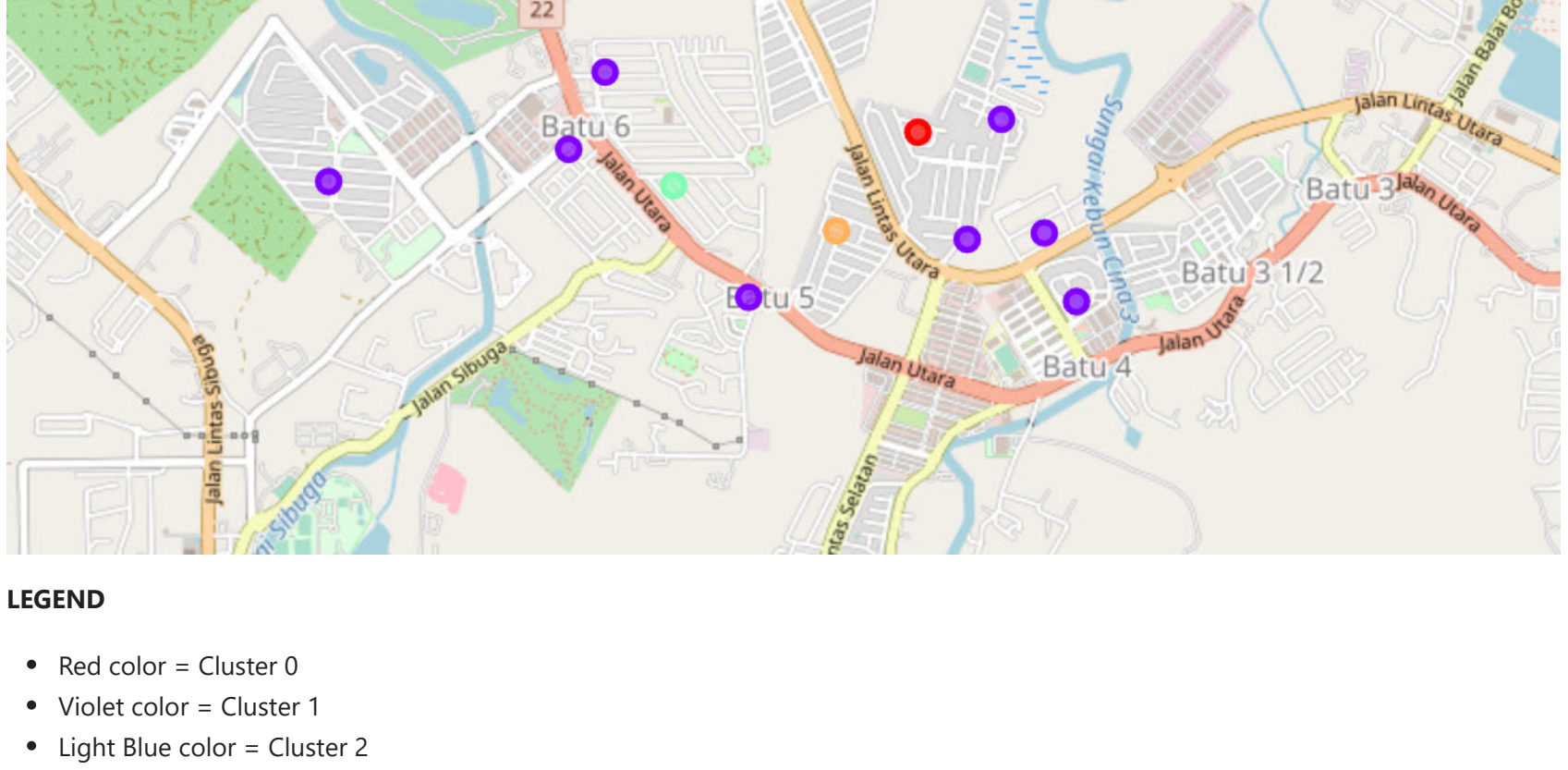
In [70]:
```
#Convert float to int for Cluster Labels
#target_merged['Cluster Labels'] = target_merged['Cluster Labels'].astype(int)
```

In [71]:
```
#target_merged
```

In [72]:
```
# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

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

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(target_merged['Latitude'], target_merged['Longtitude'], target_merged['Nei
    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[72]:





LEGEND

- Red color = Cluster 0
- Violet color = Cluster 1
- Light Blue color = Cluster 2
- Green color = Cluster 3
- Orange color = Cluster 4

Display each cluster

In [73]:
```
target_merged.loc[target_merged['Cluster Labels'] == 0]
```

Out[73]:

Neighbourhood	Latitude	Longtitude	Cluster	1st Most	2nd	3rd Most	4th Most	5th Most	6th Most	7th Most	8th Most	9t

				Labels	Common Venue	Most Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Co
1	Casa San Uno	5.865233	118.072556	0	BBQ Joint	Women's Store	Chinese Restaurant	Food Truck	Food Court	Fish & Chips Shop	Fast Food Restaurant	Electronics Store	

In [74]:

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

Out[74]:

	Neighbourhood	Latitude	Longtitude	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	
0	Bunga Matahari	5.865810	118.075874	1	Chinese Restaurant	BBQ Joint	Fast Food Restaurant	Outdoors & Recreation	Other Great Outdoors	Convenience Store	Women's Store	
2	Damai & Sri Taman	5.858482	118.078921	1	Deli / Bodega	Fruit & Vegetable Store	Fast Food Restaurant	Chinese Restaurant	Food Court	Bubble Tea Shop	Hotel	
4	Garden Villa	5.863280	118.048945	1	Café	Photography Studio	Indian Restaurant	Korean Restaurant	Lounge	Bubble Tea Shop	Breakfast Spot	
5	Indah	5.842067	118.066095	1	Athletics & Sports	Asian Restaurant	Recreation Center	Chinese Restaurant	Food Truck	Indian Restaurant	Department Store	
6	Indah Jaya	5.843796	118.067200	1	Athletics & Sports	Baseball Stadium	Recreation Center	Department Store	Indian Restaurant	Chinese Restaurant	Women's Store	
8	Mesra	5.861271	118.077664	1	Electronics Store	Pizza Place	Convenience Store	Café	Food Court	Music Venue	Other Great Outdoors	I
10	Tinosan	5.861032	118.074517	1	Chinese Restaurant	Café	Pub	Malay Restaurant	Indian Restaurant	Convenience Store	Noodle House	
11	Tshun Ngen	5.858728	118.065804	1	Chinese Restaurant	Asian Restaurant	Noodle House	Fish & Chips Shop	Grocery Store	Food Truck	Fruit & Vegetable Store	
12	Tyng	5.867669	118.059997	1	Asian Restaurant	Chinese Restaurant	Noodle House	Hotel	Cupcake Shop	Coffee Shop	Market	
13	Utama	5.864601	118.058569	1	Malay Restaurant	Women's Store	Noodle House	Café	Chinese Restaurant	Coffee Shop	Cupcake Shop	I

In [75]:

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

Out[75]:

	Neighbourhood	Latitude	Longtitude	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 Co
3	Evergreen	5.873464	118.057834	2	Golf Course	Gym	Women's Store	Fruit & Vegetable Store	Food Court	Fish & Chips Shop	Fast Food Restaurant	Electronics Store	

In [76]:

```
target_merged.loc[target_merged['Cluster Labels'] == 3]
```

Out[76]:

	Neighbourhood	Latitude	Longtitude	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	
7	Lucky & Wemin	5.863112	118.062768	3	Halal Restaurant	Women's Store	Chinese Restaurant	Food Court	Fish & Chips Shop	Fast Food Restaurant	Electronics Store	Diner	D

In [77]:

```
target_merged.loc[target_merged['Cluster Labels'] == 4]
```

Out[77]:

	Neighbourhood	Latitude	Longtitude	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	9 C
9	Pertama	5.861339	118.069276	4	Furniture / Home Store	Chinese Restaurant	Noodle House	Austrian Restaurant	BBQ Joint	Food Court	Fish & Chips Shop	Fast Food Restaurant	Ele

Results and Discussion

The clustering results gave the most number of neighbourhoods are **cluster 1**. Startup Businesses who are keen in setting up any businesses can refer to the clustering results and what sort of businesses are there.

Business people need to factor in costs like rental, utilities, land prices, transportation, labor etc before setting any businesses.

To recap, we collected data from relevant websites and merged them into a single csv file. Some data exploration were performed to look for any patterns amongst the features.

Then we decided to focus on Mile 4 to Mile 6 neighbourhood areas since majority of them are concentrated there.

We mapped these locations using Folium. We used Foursquare API to get the common venues visited by people who live there.

K-Means clustering is applied to cluster these neighbourhoods to five clusters and the result will give new business owners to analyze what sort of opportunities available.

Conclusion

The purpose of this project is to explore business opportunities in Sandakan neighbourhoods. Using clustering methods, we can identify suitable venues which can be considered by business people.