IBM Data Science Capstone Project

Abstract

In this report, we will explore restaurants in Singapore by location and category, for a businessman to decide where he can open his next restaurant in Singapore.

Angelina

Contents

A description of the problem and a discussion of the background

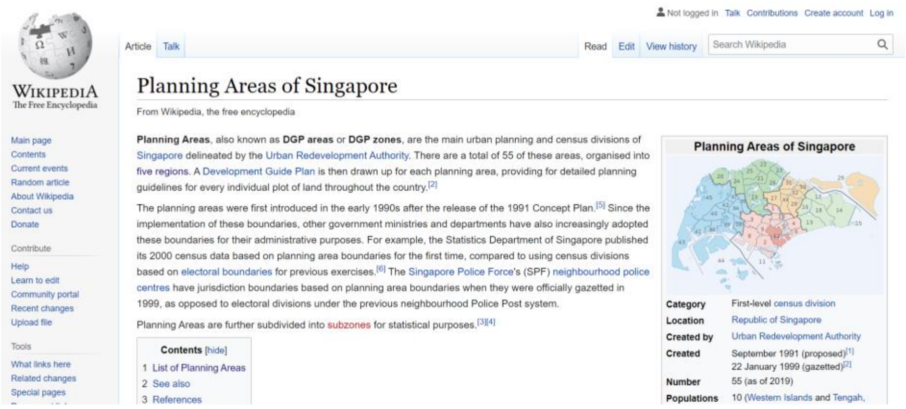
Singapore, whilst small, is known for many things. It is a regional financial centre, a popular tourist destination and a shipping hub, home to over a population of 5 million. On top of that, Singapore is also known for its gastronomical delights. If you were to ask locals, or anyone who has lived in Singapore before, what they would miss the most of about Singapore if they were to move abroad, the common answer would be: food.

There are hundreds of eating places in Singapore, ranging from hawker centres to restaurants and snack places dotted across the country. Singapore is geographically small, and one can easily travel across Singapore by car in less than an hour.

If a businessman is thinking of opening a restaurant in Singapore, how does he start his research? How can we leverage Foursquare location data and machine learning to help the businessman decide he can open his restaurant? What type of restaurants are popular in Singapore?

A description of the data and how it will be used to solve the problem

Foursquare data about restaurants in Singapore will be used to extract information such as its location, name and category of food. Also, the overview of planning areas in Singapore will be extracted from Wikipedia: https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore



Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages

Planning Areas of Singapore

From Wikipedia, the free encyclopedia

Planning Areas, also known as **DGP areas** or **DGP zones**, are the main urban planning and census divisions of Singapore delineated by the Urban Redevelopment Authority. There are a total of 55 of these areas, organised into five regions. A Development Guide Plan is then drawn up for each planning area, providing for detailed planning guidelines for every individual plot of land throughout the country.[2]

The planning areas were first introduced in the early 1990s after the release of the 1991 Concept Plan.[5] Since the implementation of these boundaries, other government ministries and departments have also increasingly adopted these boundaries for their administrative purposes. For example, the Statistics Department of Singapore published its 2000 census data based on planning area boundaries for the first time, compared to using census divisions based on electoral boundaries for previous exercises.[6] The Singapore Police Force's (SPF) neighbourhood police centres have jurisdiction boundaries based on planning area boundaries when they were officially gazetted in 1999, as opposed to electoral divisions under the previous neighbourhood Police Post system.

Planning Areas are further subdivided into subzones for statistical purposes.[3][4]

Contents [hide]
1 List of Planning Areas
2 See also
3 References

Planning Areas of Singapore

Category	First-level census division
Location	Republic of Singapore
Created by	Urban Redevelopment Authority
Created	September 1991 (proposed)[1] 22 January 1999 (gazetted)[2]
Number	55 (as of 2019)
Populations	10 (Western Islands and Tengah,

Methodology

Scrapping of Data into a Dataframe

We will begin by scrapping the list of planning areas in Singapore from Wikipedia and put it in a dataframe. Irrelevant columns will be dropped to get a nice dataframe with relevant columns.

```
In [5]: df = df.drop(columns=['Malay', 'Chinese', 'Pinyin','Tamil'])

In [6]: df.shape
Out[6]: (55, 5)

In [7]: df.head()
Out[7]:
```

	Name (English)	Region	Area (km2)	Population[7]	Density (/km2)
0	Ang Mo Kio	North-East	13.94	163950	13400
1	Bedok	East	21.69	279380	13000
2	Bishan	Central	7.62	88010	12000
3	Boon Lay	West	8.23	30	3.6
4	Bukit Batok	West	11.13	153740	14000

Obtain Coordinates of Planning Areas: Geopy Client

I used the nominatim function to add geospatial data to the data frame and added the Latitude and Longitude each planning area to the data frame.

Let's get the geographical coordinates of Singapore

In [12]:
```
address = 'Singapore'

geolocator = Nominatim(user_agent="Singapore_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Singapore are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of Singapore are 1.357107, 103.8194992.
```

In [13]:
```
from geopy.geocoders import Nominatim # module to convert an address into latitude and longitude values
geolocator = Nominatim(user_agent="Singapore_explorer")

df['Major_Dist_Coord']= df['Name (English)'].apply(geolocator.geocode).apply(lambda x: (x.latitude, x.longitude))
df[['Latitude', 'Longitude']] = df['Major_Dist_Coord'].apply(pd.Series)

df.drop(['Major_Dist_Coord'], axis=1, inplace=True)
df
```

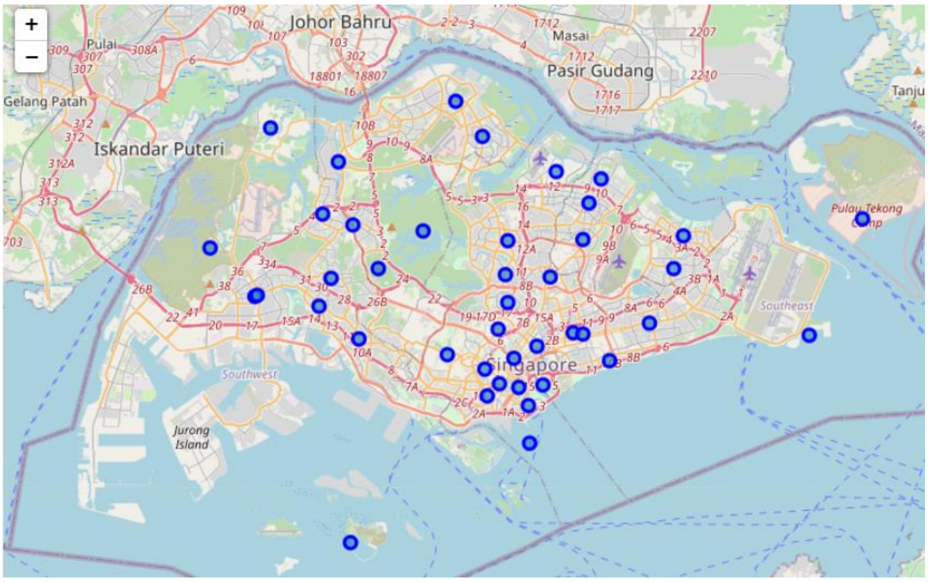
Out[13]:

	Name (English)	Region	Area (km2)	Population[7]	Density (/km2)	Latitude	Longitude
0	Ang Mo Kio	North-East	13.94	163950	13400	1.370080	103.849523
1	Bedok	East	21.69	279380	13000	1.323976	103.930216
2	Bishan	Central	7.62	88010	12000	1.350986	103.848255
3	Boon Lay	West	8.23	30	3.6	1.338550	103.705812
4	Bukit Batok	West	11.13	153740	14000	1.349057	103.749591
5	Bukit Merah	Central	14.34	151980	11000	4.561694	101.024037
6	Bukit Panjang	West	8.99	139280	15000	1.379149	103.761413
7	Bukit Timah	Central	17.53	77430	4400	1.354690	103.776372
8	Central Water Catchment	North	37.15	*	*	1.375708	103.801743
9	Changi	East	40.61	1830	80.62	43.880078	126.564903
10	Changi Bay	East	1.70	*	*	1.316850	104.020649
11	Choa Chu Kang	West	6.11	190890	30000	1.385317	103.744325
12	Clementi	West	9.49	92420	9800	1.315100	103.765231
13	Downtown Core	Central	4.34	2720	680	1.287475	103.856033
14	Geylang	Central	9.64	110200	11400	1.318186	103.887056
15	Hougang	North-East	13.93	226240	16000	1.370801	103.892544
16	Jurong East	West	17.83	79240	4400	1.333108	103.742294
17	Jurong West	West	14.69	264860	18000	1.339636	103.707339
18	Kallang	Central	9.17	101520	11000	1.310759	103.866262

Create Map with Folium

Using the folium package with the dataframe, I created a map with the planning areas plotted on it.

In [14]:
```
# create map of Singapore  using latitude and longitude values
map_singapore = folium.Map(location=[latitude, longitude], zoom_start=11)

 # add markers to map
for lat, lng, label in zip(df['Latitude'], df['Longitude'], df['Name (English)']):
    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_singapore)

map_singapore
```



Exploratory Data Analysis

This will be done to gather insights and useful information to the reader, and businessman who may be thinking of how and where to open a restaurant in Singapore.

Gather Category of Unique Venues in a Planning Area

We first start by looking at the venues within a Planning Area – Ang Mo Kio.

```
In [27]: print ('{} unique categories in Ang Mo Kio.'.format(nearby_venues['categories'].value_counts().shape[0]))

         12 unique categories in Ang Mo Kio.
```

```
In [28]: print (nearby_venues['categories'].value_counts()[0:15])

         Coffee Shop              3
         Supermarket              2
         Japanese Restaurant      1
         Snack Place              1
         Bubble Tea Shop          1
         Burger Joint             1
         Noodle House             1
         Asian Restaurant         1
         Miscellaneous Shop       1
         Gym / Fitness Center     1
         Chinese Restaurant       1
         Fast Food Restaurant     1
         Name: categories, dtype: int64
```

We can see that it includes a variety of categories for venues such as supermarkets, gyms, snack places and restaurants.

Create a Dataframe for Restaurants Only

We will narrow the list to create a dataframe for restaurants only.

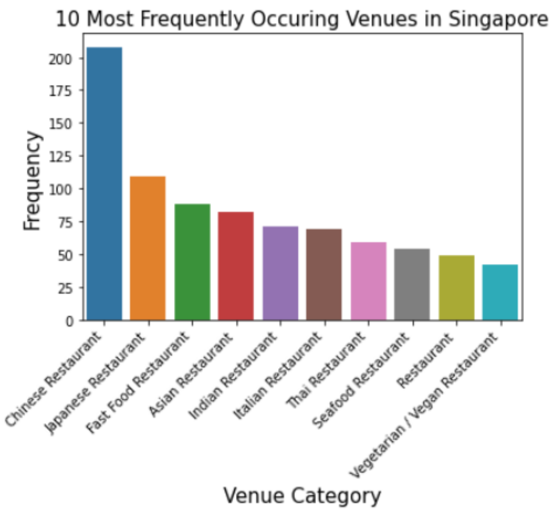
```
In [36]: # Create a Data-Frame out of it to Concentrate Only on Restaurants

         Singapore_Venues_only_restaurant = Singapore_venues[Singapore_venues['Venue Category'].str.contains('Restaurant')].reset_index(drop=True)
         Singapore_Venues_only_restaurant.index = np.arange(1, len(Singapore_Venues_only_restaurant )+1)
```

```
In [37]: print (Singapore_Venues_only_restaurant['Venue Category'].value_counts())
```

```
In [37]: print (Singapore_Venues_only_restaurant['Venue Category'].value_counts())
```

```
Chinese Restaurant                  208
Japanese Restaurant                 109
Fast Food Restaurant                 88
Asian Restaurant                     82
Indian Restaurant                    71
Italian Restaurant                   69
Thai Restaurant                      59
Seafood Restaurant                   54
Restaurant                           49
Vegetarian / Vegan Restaurant        42
Sushi Restaurant                     32
French Restaurant                    20
Hotpot Restaurant                    20
American Restaurant                  20
Korean Restaurant                    19
Dim Sum Restaurant                   18
Dumpling Restaurant                  14
Vietnamese Restaurant                14
Indonesian Restaurant                12
Malay Restaurant                     12
Spanish Restaurant                   11
Hainan Restaurant                    10
Modern European Restaurant           10
Cantonese Restaurant                  9
Ramen Restaurant                      8
Halal Restaurant                      7
```

Using seaborn, we reflected the top 10 frequently occurring venues in Singapore into a graph:



It seems that Singaporeans love Chinese food. We will now analyse each neighbourhood to know about the top 5 venues of each one, by creating a dataframe with pandas one hot encoding for the venue categories.

```
In [59]: # one hot encoding
Singapore_onehot = pd.get_dummies(Singapore_Venues_only_restaurant[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
Singapore_onehot['Neighborhood'] = Singapore_Venues_only_restaurant['Neighborhood']

Singapore_onehot.head()
```

```
In [60]: # move neighborhood column to the front
fixed_columns = [Singapore_onehot.columns[-1]] + list(Singapore_onehot.columns[:-1])
Singapore_onehot = Singapore_onehot[fixed_columns]

Singapore_onehot.head()
```

Out[60]:

	Neighborhood	American Restaurant	Asian Restaurant	Australian Restaurant	Cantonese Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	English Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	Fujian Restaurant
1	Ang Mo Kio	0	1	0	0	0	0	0	0	0	0	0	0	0
2	Ang Mo Kio	0	0	0	0	1	0	0	0	0	0	0	0	0
3	Ang Mo Kio	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Ang Mo Kio	0	0	0	0	0	0	0	0	0	0	0	0	0
5	Ang Mo Kio	0	0	0	0	1	0	0	0	0	0	0	0	0

Now, we will look at the mean of frequency occurring for each category of restaurants.

Out[62]:

	Neighborhood	American Restaurant	Asian Restaurant	Australian Restaurant	Cantonese Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	English Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	Fujian Restaurant
0	Ang Mo Kio	0.000000	0.081081	0.000000	0.000000	0.351351	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Bedok	0.046512	0.139535	0.000000	0.000000	0.302326	0.023256	0.000000	0.000000	0.000000	0.069767	0.000000	0.000000	0.000000
2	Bishan	0.000000	0.100000	0.000000	0.000000	0.400000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Boon Lay	0.000000	0.120000	0.000000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.280000	0.000000	0.000000	0.000000
4	Bukit Batok	0.025000	0.050000	0.000000	0.000000	0.275000	0.000000	0.050000	0.000000	0.000000	0.050000	0.000000	0.000000	0.000000
5	Bukit Merah	0.000000	0.266667	0.000000	0.000000	0.666667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Bukit Panjang	0.037037	0.074074	0.000000	0.037037	0.111111	0.000000	0.000000	0.000000	0.000000	0.185185	0.000000	0.000000	0.000000
7	Bukit Timah	0.000000	0.034483	0.000000	0.000000	0.172414	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	Central Water Catchment	0.000000	0.000000	0.000000	0.000000	0.250000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	Changi	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	Choa Chu Kang	0.000000	0.142857	0.000000	0.000000	0.142857	0.000000	0.000000	0.000000	0.000000	0.214286	0.000000	0.000000	0.000000
11	Clementi	0.029412	0.058824	0.000000	0.000000	0.176471	0.000000	0.029412	0.000000	0.000000	0.029412	0.000000	0.058824	0.000000
12	Downtown Core	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.125000	0.000000	0.000000	0.000000	0.000000	0.125000	0.000000

Now, we can update the table to view the most common venue in each neighbourhood.

Out[67]:

	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
0	Ang Mo Kio	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Italian Restaurant	Sushi Restaurant	Restaurant	Shaanxi Restaurant	Indian Restaurant
1	Bedok	Chinese Restaurant	Asian Restaurant	Seafood Restaurant	Fast Food Restaurant	Indian Restaurant	American Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant	Malay Restaurant
2	Bishan	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Italian Restaurant	Indian Restaurant	Szechuan Restaurant	Sushi Restaurant
3	Boon Lay	Fast Food Restaurant	Chinese Restaurant	Asian Restaurant	Japanese Restaurant	Indian Restaurant	Japanese Curry Restaurant	Sushi Restaurant	Restaurant	Thai Restaurant
4	Bukit Batok	Chinese Restaurant	Indian Restaurant	Korean Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Thai Restaurant	Japanese Restaurant	Hainan Restaurant
5	Bukit Merah	Chinese Restaurant	Asian Restaurant	Malay Restaurant	Vietnamese Restaurant	Hainan Restaurant	Kebab Restaurant	Japanese Restaurant	Japanese Curry Restaurant	Italian Restaurant
6	Bukit Panjang	Fast Food Restaurant	Chinese Restaurant	Italian Restaurant	Asian Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant	Indian Restaurant	American Restaurant	Japanese Restaurant
7	Bukit Timah	Italian Restaurant	Chinese Restaurant	Korean Restaurant	Thai Restaurant	Japanese Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Paella Restaurant

Here, we can see that in Ang Mo Kio, the most common venue is Chinese Restaurant whereas for Bukit Panjang, it is Fast Food Restaurant.

K Means Clustering

Using K Means clustering, we will put the planning areas into 5 clusters.

In [69]:
```
# set number of clusters
kclusters = 5

Singapore_grouped_clustering = Singapore_grouped.drop('Neighborhood', 1)

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```
Out[69]: `array([0, 0, 0, 1, 0, 4, 1, 0, 0, 4], dtype=int32)`

Out[95]:

	Cluster Labels	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
0	0	Ang Mo Kio	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Italian Restaurant	Sushi Restaurant	Restaurant	Shaanxi Restaurant
1	0	Bedok	Chinese Restaurant	Asian Restaurant	Seafood Restaurant	Fast Food Restaurant	Indian Restaurant	American Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant
2	0	Bishan	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Italian Restaurant	Indian Restaurant	Szechuan Restaurant
3	1	Boon Lay	Fast Food Restaurant	Chinese Restaurant	Asian Restaurant	Japanese Restaurant	Indian Restaurant	Japanese Curry Restaurant	Sushi Restaurant	Restaurant
4	0	Bukit Batok	Chinese Restaurant	Indian Restaurant	Korean Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Thai Restaurant	Japanese Restaurant
5	4	Bukit Merah	Chinese Restaurant	Asian Restaurant	Malay Restaurant	Vietnamese Restaurant	Hainan Restaurant	Kebab Restaurant	Japanese Restaurant	Japanese Curry Restaurant
6	1	Bukit Panjang	Fast Food Restaurant	Chinese Restaurant	Italian Restaurant	Asian Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant	Indian Restaurant	American Restaurant
7	0	Bukit Timah	Italian Restaurant	Chinese Restaurant	Korean Restaurant	Thai Restaurant	Japanese Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant
8	0	Central Water Catchment	Japanese Restaurant	Chinese Restaurant	Indian Restaurant	Seafood Restaurant	Vietnamese Restaurant	Greek Restaurant	Kebab Restaurant	Japanese Curry Restaurant

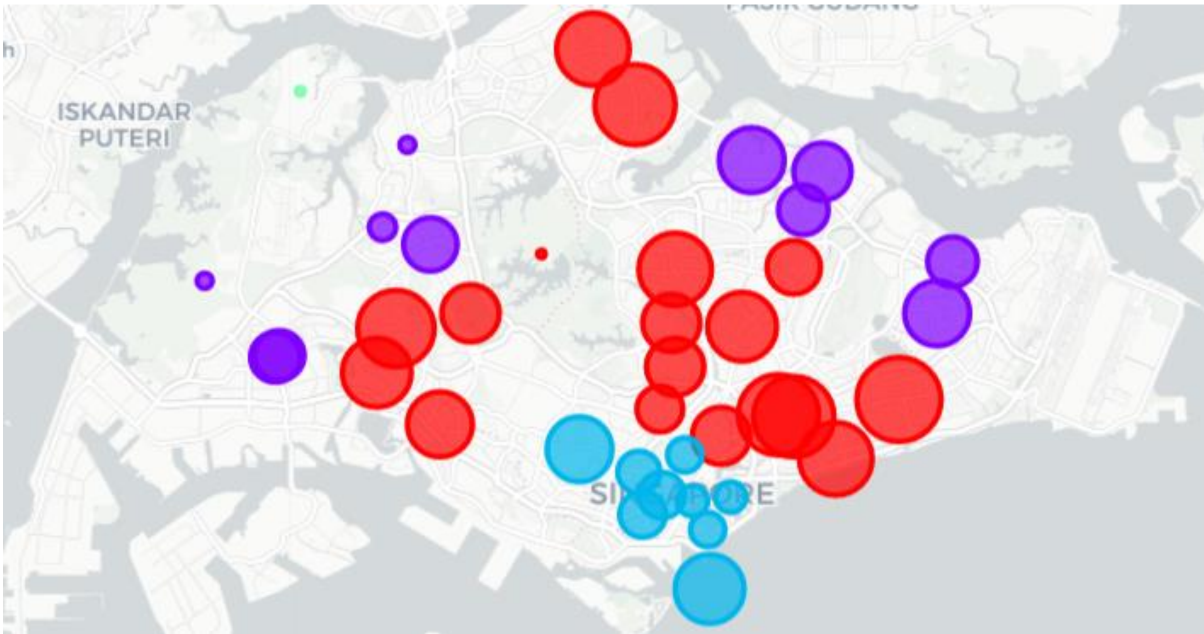
I then added the Longitude and Latitude to this table.

In [100]:
```
Singapore_merged.rename(columns={'Name (English)':'Neighborhood'}, inplace=True)
Singapore_merged = Singapore_merged.merge(neighborhoods_venues_sorted_w_clusters.set_index('Neighborhood'), on='Neighborhood')
Singapore_merged.head()
```
Out[100]:

	Neighborhood	Region	Area (km2)	Population[7]	Density (/km2)	Latitude	Longitude	Cluster Labels_x	1st Most Common Venue_x	2nd Most Common Venue_x	3rd Most Common Venue_x	4th Most Common Venue_x	5th Most Common Venue_x	6th Most Common Venue_x	7th Most Common Venue_x
0	Ang Mo Kio	North-East	13.94	163950	13400	1.370080	103.849523	0.0	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Italian Restaurant	Sushi Restaurant	Restaurant
1	Bedok	East	21.69	279380	13000	1.323976	103.930216	0.0	Chinese Restaurant	Asian Restaurant	Seafood Restaurant	Fast Food Restaurant	Indian Restaurant	American Restaurant	Thai Restaurant
2	Bishan	Central	7.62	88010	12000	1.350986	103.848255	0.0	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Italian Restaurant	Indian Restaurant
3	Boon Lay	West	8.23	30	3.6	1.338550	103.705812	1.0	Fast Food Restaurant	Chinese Restaurant	Asian Restaurant	Japanese Restaurant	Indian Restaurant	Japanese Curry Restaurant	Sushi Restaurant
4	Bukit Batok	West	11.13	153740	14000	1.349057	103.749591	0.0	Chinese Restaurant	Indian Restaurant	Korean Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Thai Restaurant

Results

Now, I used the cluster labels to show the areas with a cluster-specific color on a map, again using folium:



We now have 5 clusters of restaurant type concentrations in Singapore, where we can view the breakdown of clusters below.

Cluster 1

We can see that Cluster 1 consists mostly of Chinese Restaurants.

In [106]:
```
# Cluster 1
Singapore_merged.loc[Singapore_merged['Cluster Labels_x'] == 0, Singapore_merged.columns[[1] + list(range(5, Singapore_merged.shape[1]))]]
```

Out[106]:

	Region	Latitude	Longitude	Cluster Labels_x	1st Most Common Venue_x	2nd Most Common Venue_x	3rd Most Common Venue_x	4th Most Common Venue_x	5th Most Common Venue_x	6th Most Common Venue_x	7th Most Common Venue_x	8th Most Common Venue_x	9th Most Common Venue_x	10th Most Common Venue_x
0	North-East	1.370080	103.849523	0.0	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Italian Restaurant	Sushi Restaurant	Restaurant	Shaanxi Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant
1	East	1.323976	103.930216	0.0	Chinese Restaurant	Asian Restaurant	Seafood Restaurant	Fast Food Restaurant	Indian Restaurant	American Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant	Malay Restaurant	Italian Restaurant
2	Central	1.350986	103.848255	0.0	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Italian Restaurant	Indian Restaurant	Szechuan Restaurant	Sushi Restaurant	Hakka Restaurant
4	West	1.349057	103.749591	0.0	Chinese Restaurant	Indian Restaurant	Korean Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Thai Restaurant	Japanese Restaurant	Hainan Restaurant	Dim Sum Restaurant
7	Central	1.354690	103.776372	0.0	Italian Restaurant	Chinese Restaurant	Korean Restaurant	Thai Restaurant	Japanese Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Paella Restaurant	Halal Restaurant
8	North	1.375708	103.801743	0.0	Japanese Restaurant	Chinese Restaurant	Indian Restaurant	Seafood Restaurant	Vietnamese Restaurant	Greek Restaurant	Kebab Restaurant	Japanese Curry Restaurant	Italian Restaurant	Indonesian Restaurant
11	West	1.315100	103.765231	0.0	Chinese Restaurant	Japanese Restaurant	Seafood Restaurant	Thai Restaurant	Italian Restaurant	French Restaurant	Asian Restaurant	Indian Restaurant	Korean Restaurant	Malay Restaurant
13	Central	1.318186	103.887056	0.0	Chinese Restaurant	Asian Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Restaurant	Italian Restaurant	Hotpot Restaurant	Japanese Restaurant	Cantonese Restaurant	Dumpling Restaurant

Cluster 2

We can see that Cluster 2 consists mostly of Fast Food Restaurants

In [107]:
```
# Cluster 2
Singapore_merged.loc[Singapore_merged['Cluster Labels_x'] == 1, Singapore_merged.columns[[1] + list(range(5, Singapore_merged.shape[1]))]]
```

Out[107]:

	Region	Latitude	Longitude	Cluster Labels_x	1st Most Common Venue_x	2nd Most Common Venue_x	3rd Most Common Venue_x	4th Most Common Venue_x	5th Most Common Venue_x	6th Most Common Venue_x	7th Most Common Venue_x	8th Most Common Venue_x	9th Most Common Venue_x	10th Most Common Venue_x
3	West	1.338550	103.705812	1.0	Fast Food Restaurant	Chinese Restaurant	Asian Restaurant	Japanese Restaurant	Indian Restaurant	Japanese Curry Restaurant	Sushi Restaurant	Restaurant	Thai Restaurant	Halal Restaurant
6	West	1.379149	103.761413	1.0	Fast Food Restaurant	Chinese Restaurant	Italian Restaurant	Asian Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant	Indian Restaurant	American Restaurant	Japanese Restaurant	Malay Restaurant
10	West	1.385317	103.744325	1.0	Fast Food Restaurant	Chinese Restaurant	Asian Restaurant	Thai Restaurant	Japanese Restaurant	Restaurant	Indonesian Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Italian Restaurant
16	West	1.339636	103.707339	1.0	Chinese Restaurant	Fast Food Restaurant	Asian Restaurant	Japanese Restaurant	Indian Restaurant	Japanese Curry Restaurant	Sushi Restaurant	Restaurant	Thai Restaurant	Halal Restaurant
24	East	1.373031	103.949255	1.0	Fast Food Restaurant	Thai Restaurant	Asian Restaurant	Italian Restaurant	Mediterranean Restaurant	Seafood Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dumpling Restaurant	Hotpot Restaurant
26	North-East	1.405197	103.902350	1.0	Fast Food Restaurant	Chinese Restaurant	Seafood Restaurant	Asian Restaurant	Japanese Restaurant	Sushi Restaurant	Vietnamese Restaurant	Cantonese Restaurant	French Restaurant	Hotpot Restaurant
30	North-East	1.409849	103.877379	1.0	Fast Food Restaurant	Asian Restaurant	Chinese Restaurant	Seafood Restaurant	Thai Restaurant	Sushi Restaurant	Japanese Restaurant	Restaurant	Vietnamese Restaurant	English Restaurant

Cluster 3

We can see that Cluster 3 consists mostly of Japanese Restaurants.

In [109]:
```
# Cluster 3
Singapore_merged.loc[Singapore_merged['Cluster Labels_x'] == 2, Singapore_merged.columns[[1] + list(range(5, Singapore_merged.shape[1]))
```

Out[109]:

	Region	Latitude	Longitude	Cluster Labels_x	1st Most Common Venue_x	2nd Most Common Venue_x	3rd Most Common Venue_x	4th Most Common Venue_x	5th Most Common Venue_x	6th Most Common Venue_x	7th Most Common Venue_x	8th Most Common Venue_x	9th Most Common Venue_x	10th Most Common Venue_x
12	Central	1.287475	103.856033	2.0	Japanese Restaurant	Italian Restaurant	French Restaurant	Dim Sum Restaurant	Thai Restaurant	Modern European Restaurant	Spanish Restaurant	Restaurant	Seafood Restaurant	Indonesian Restaurant
19	Central	1.288624	103.869827	2.0	Italian Restaurant	Japanese Restaurant	Thai Restaurant	Dim Sum Restaurant	Modern European Restaurant	Indonesian Restaurant	Dumpling Restaurant	Restaurant	Vietnamese Restaurant	Hainan Restaurant
20	Central	1.276998	103.861500	2.0	Japanese Restaurant	Italian Restaurant	Dim Sum Restaurant	Restaurant	Seafood Restaurant	Modern European Restaurant	Indian Restaurant	Dumpling Restaurant	Peking Duck Restaurant	Japanese Curry Restaurant
23	Central	1.282870	103.837860	2.0	Japanese Restaurant	Spanish Restaurant	Restaurant	Seafood Restaurant	Tapas Restaurant	Australian Restaurant	Kebab Restaurant	Chinese Restaurant	Italian Restaurant	Korean Restaurant
27	Central	-45.032192	168.661000	2.0	Restaurant	Japanese Restaurant	Italian Restaurant	Seafood Restaurant	Tapas Restaurant	Indian Restaurant	Asian Restaurant	Spanish Restaurant	Japanese Curry Restaurant	Indonesian Restaurant
28	Central	1.297582	103.836514	2.0	Japanese Restaurant	French Restaurant	Sushi Restaurant	Hotpot Restaurant	Seafood Restaurant	Modern European Restaurant	Dumpling Restaurant	Persian Restaurant	Dim Sum Restaurant	Spanish Restaurant
29	Central	1.303918	103.852789	2.0	Japanese Restaurant	Italian Restaurant	French Restaurant	Indian Restaurant	Restaurant	Hotpot Restaurant	Dim Sum Restaurant	Spanish Restaurant	Modern European Restaurant	Middle Eastern Restaurant
35	Central	1.289178	103.845154	2.0	Japanese Restaurant	Seafood Restaurant	Spanish Restaurant	Restaurant	French Restaurant	Modern European Restaurant	Persian Restaurant	Greek Restaurant	Chinese Restaurant	Australian Restaurant

Cluster 4

We can see that Cluster 4 consists mostly of Japanese Restaurants.

In [110]:
```
# Cluster 4
Singapore_merged.loc[Singapore_merged['Cluster Labels_x'] == 3, Singapore_merged.columns[[1] + list(range(5, Singapore_merged.shape[1
```

Out[110]:

	Region	Latitude	Longitude	Cluster Labels_x	1st Most Common Venue_x	2nd Most Common Venue_x	3rd Most Common Venue_x	4th Most Common Venue_x	5th Most Common Venue_x	6th Most Common Venue_x	7th Most Common Venue_x	8th Most Common Venue_x	9th Most Common Venue_x	10th Most Common Venue_x
18	North	1.434217	103.714987	3.0	Vegetarian / Vegan Restaurant	Asian Restaurant	Restaurant	Greek Restaurant	Kebab Restaurant	Japanese Restaurant	Japanese Curry Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant

Cluster 5

We can see that Cluster 5 consists mostly of Japanese Restaurants.

In [111]:
```
# Cluster 5
Singapore_merged.loc[Singapore_merged['Cluster Labels_x'] == 4, Singapore_merged.columns[[1] + list(range(5, Singapore_merged.shape[1]))]]
```

Out[111]:

	Region	Latitude	Longitude	Cluster Labels_x	1st Most Common Venue_x	2nd Most Common Venue_x	3rd Most Common Venue_x	4th Most Common Venue_x	5th Most Common Venue_x	6th Most Common Venue_x	7th Most Common Venue_x	8th Most Common Venue_x	9th Most Common Venue_x	10th Most Common Venue_x	Cluster Labels_y
5	Central	4.561694	101.024037	4.0	Chinese Restaurant	Asian Restaurant	Malay Restaurant	Vietnamese Restaurant	Hainan Restaurant	Kebab Restaurant	Japanese Restaurant	Japanese Curry Restaurant	Italian Restaurant	Indonesian Restaurant	4
9	East	43.880078	126.564903	4.0	Chinese Restaurant	Vietnamese Restaurant	Greek Restaurant	Kebab Restaurant	Japanese Restaurant	Japanese Curry Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	4

Discussion

The above clusters give us a good overview about the current restaurant scene in Singapore. Here are some of the key findings:

- Chinese Restaurants top the list as the most common type of restaurants, and it is quite spaced out across Singapore
- Fast Food Restaurants are popular along the outskirts of Singapore's Central area
- Japanese Restaurants seem to concentrate in the Central area, particularly close to where Singapore's Business Districts are
- There do not seem to be many Vegetarian/Vegan Restaurants around

It is important to note that the above clustering data is solely based on information available via Foursquare, where we retrieved the information from.

It is important to note that there are other factors that determine the location of the restaurants, such as the distance from the venue to closest train stations, and type of housing or offices in each area. Hence, from the above, the businessman can decide on the type of restaurant he wishes to open and narrow down the geographical location he wishes to explore for his restaurant.

Conclusion

It is interesting how data can be pulled from various sources to gather insights about a problem that one wishes to solve. For example, Foursquare's data is most helpful in giving us an overview about restaurants in Singapore without having to do any legwork. It can give the businessman some good understanding about the types of restaurants and cuisines that he can introduce in Singapore.

From this project, the data is malleable can be used to solve other problems. For example, it can also be pivoted to produce food recommendations for tourists.