IBM Data Science Capstone Project

Made by - Sayali

Description and business problem :

Singapore is an island country off the southern tip of the Malay Peninsula in Southeast Asia. It is separated from Malaysia by the Straits of Johor, and from Indonesia's Riau Islands by the Singapore Strait. Singapore has a strategic location for Southeast Asian sea routes. While it is small in size it is pretty advanced in terms of financial sector and various other developmental sectors, it is also a well known tourist attraction. Singapore has a wide variety of food.

I was wondering if I had to help someone set up a food business, like a cafe or a small restaurant, How would I go by my initial stages of my research?? How would I use foursquare data to help with this? What kind of food is popular in what area

Scraping data from the web

```
In [49]: df = df.drop(columns=['Malay', 'Chinese', 'Pinyin','Tamil'])

In [50]: df.shape
Out[50]: (55, 5)

In [51]: df.head()
```

Out[51]:

	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

Geographical co-ordinates of Singapore

In [54]:
```
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 [63]:
```
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[63]:

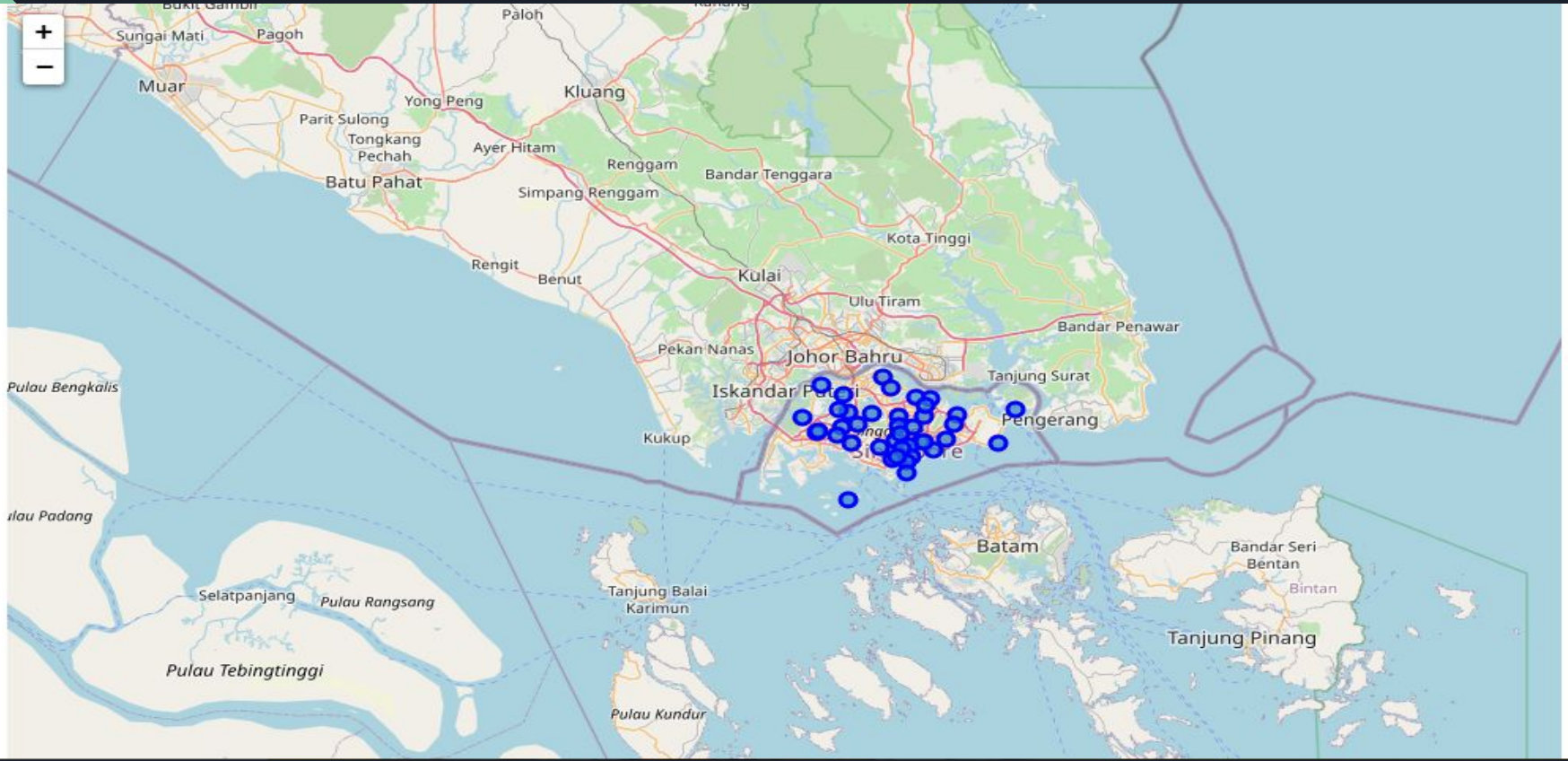
	Name (English)	Region	Area (km2)	Population[7]	Density (/km2)	Latitude	Longitude
0	Ang Mo Kio	North-East	13.94	163950	13400	1.370080	103.849523
1	Bedok	East	21.69	279380	13000	1.323976	103.930216
2	Bishan	Central	7.62	88010	12000	1.350986	103.848255
3	Boon Lay	West	8.23	30	3.6	1.338550	103.705812
4	Bukit 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

Creating map using Folium

In [64]:
```
# 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
```

Map of Singapore

Out[64]:



EXPLORATORY DATA ANALYSIS :

Out[80]:

	Venue_Category	Frequency
0	Chinese Restaurant	220
1	Japanese Restaurant	108
2	Fast Food Restaurant	86
3	Asian Restaurant	82
4	Indian Restaurant	71
5	Italian Restaurant	62
6	Thai Restaurant	60
7	Seafood Restaurant	49
8	Vegetarian / Vegan Restaurant	48
9	Restaurant	47

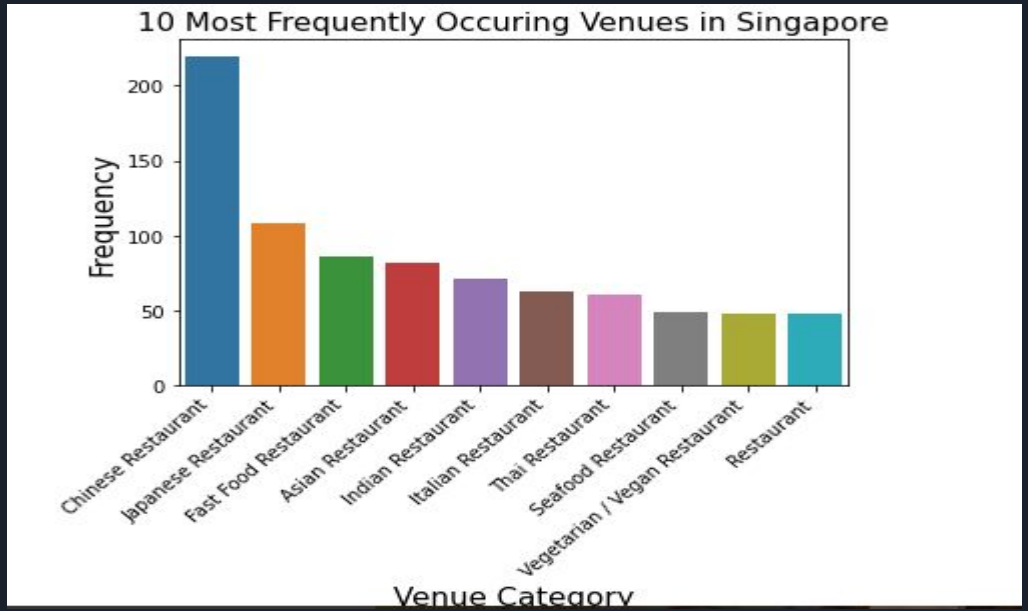
In [73]:
```
print ('{} unique categories in Ang Mo Kio.'.format(nearby_venues['categories'].value_counts().shape[0]))
```

```
13 unique categories in Ang Mo Kio.
```

In [74]:
```
print (nearby_venues['categories'].value_counts()[0:15])
```

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

TOP 10 VENUES IN SINGAPORE :



ONE HOT ENCODING ⅓ :

In [86]:
```
# 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()
```

Out[86]:

	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	Fren Restaur
1	0	1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0

ONE HOT ENCODING : ½

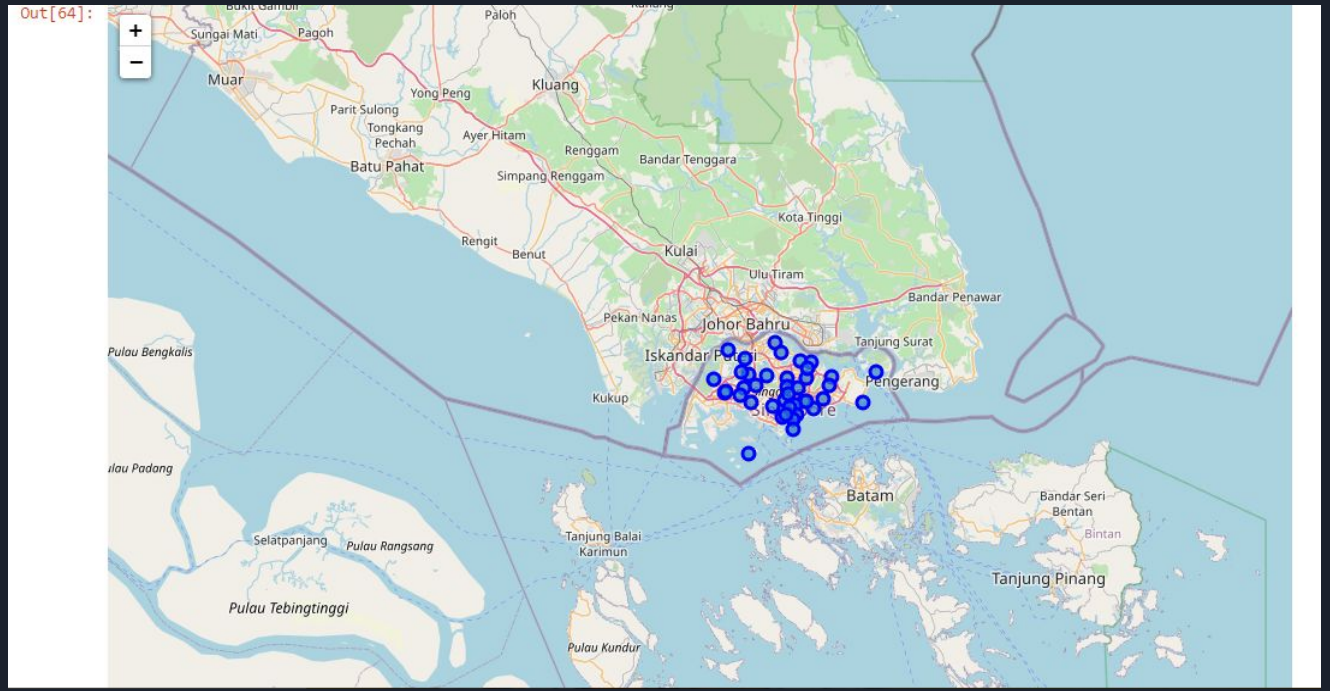
In [87]:
```
# move neighborhood column to the front
fixed_columns = [Singapore_onehot.columns[-1]] + list(Singapore_onehot.columns[:-1])
Singapore_onehot = Singapore_onehot[fixed_columns]

Singapore_onehot.head()
```

Out[87]:

	Neighborhood	American Restaurant	Asian Restaurant	Australian Restaurant	Cantonese Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	English Restaurant	Fast Food Restaurant	F Resta
1	Ang Mo Kio	0	1	0	0	0	0	0	0	0	0	0
2	Ang Mo Kio	0	0	0	0	1	0	0	0	0	0	0
3	Ang Mo Kio	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
5	Ang Mo Kio	0	0	0	0	0	0	0	0	0	0	0

CLUSTER MAP :

Out[64]:



DISCUSSION :

From our ground research we could derive certain conclusions and insights. Chinese restaurants and chinese food is one of the most common types of food in Singapore, Japanese food also being amongst famous food options.
Vegetarian and vegan options are very limited around Singapore.
Fast food options are also seen around a lot in the far away areas of certain regions.

CONCLUSIONS :

1) These insights are based on the data that we have gathered from Foursquare.
2) The data can be further improved by gaining insights from various other sources and leveraging data from there.
3) However these insights can be primary and still be used on a primary level to give recommendations to the tourists visiting Singapore and businessmen can use this data while they’re primarily thinking about setting up a restaurant.

THANK YOU