



# 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 .....	2
A description of the data and how it will be used to solve the problem .....	2
Methodology.....	2
Scrapping of Data into a Dataframe.....	2
Obtain Coordinates of Planning Areas: Geopy Client.....	2
Create Map with Folium .....	3

## A description of the problem and a discussion of the background

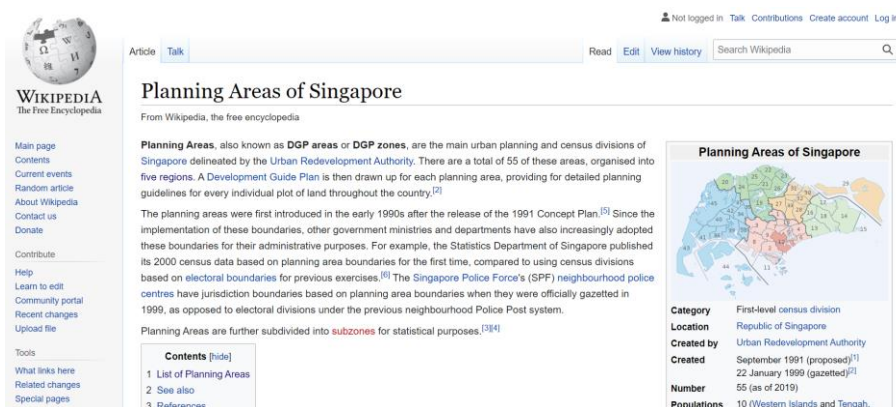
Singapore, whilst small, is known for many things. It is a regional financial centre, a popular tourist destination and a shipping hub, home to over a population of 5 million. On top of that, Singapore is also known for its gastronomical delights. If you were to ask locals, or anyone who has lived in 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](https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore)



**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.<sup>[2]</sup>

The planning areas were first introduced in the early 1990s after the release of the 1991 Concept Plan.<sup>[3]</sup> 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.<sup>[4]</sup> 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.<sup>[3][4]</sup>

Category	Location	Created by	Created	Number	Populations
First-level census division	Republic of Singapore	Urban Redevelopment Authority	September 1991 (proposed) <sup>[1]</sup> 22 January 1999 (gazetted) <sup>[2]</sup>	55 (as of 2019)	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
```

