Relocation Recommendation in Singapore

Business problem

► This project will be useful for agencies and their customers or also direct individuals who are in search of a right accommodation for themselves.

► In the project, the city of Singapore is considered which is undoubtedly the most attracted for people across the world.

► Singapore is segmented into different neighborhoods using the geographical coordinates of different buildings and the geographical coordinates of top 100 venues around a particular building.

► Finally, the neighborhoods are grouped into clusters.

Data

- The main datasets used are 'Singapore_geo.csv' and 'sg_random_samles_clean' which provide information about all the postal codes.

- The latitude, longitude, postcode and building are extracted from the datasets.

- The dataset is cleaned by removing all the null values.

- Further, Foursquare API is used to extend the dataset and select random buildings for analysis.

Output screens

| | | 4 | 5 | 18925 | DBS Marina Bay MRT Station | 1.276427 | 103.854598 |

In [99]:
```python
#Randomly select only 500 neighborhoods
neighborhoods_subset = neighborhoods.sample(500
                                            )

neighborhoods_subset.head()
```

Out[99]:

	Unnamed: 0	Postcode	Neighborhood	Latitude	Longitude
34839	52820	419910	FRANKEL ESTATE	1.319396	103.914566
127685	194594	588905	MAYFAIR PARK	1.340750	103.779755
98572	156599	465607	TANAH MERAH GREEN	1.329652	103.943983
39790	65432	458411	OPERA ESTATE	1.319086	103.926208
95575	152847	456648	CASA NALLUR	1.311367	103.932649

In [100]:
```python
neighborhoods_subset=neighborhoods_subset.loc[:, ~neighborhoods_subset.columns.str.contains('^Unnamed')]
print(neighborhoods_subset.shape)
neighborhoods_subset.head()
```

(500, 4)

Out[100]:

	Postcode	Neighborhood	Latitude	Longitude
34839	419910	FRANKEL ESTATE	1.319396	103.914566
127685	588905	MAYFAIR PARK	1.340750	103.779755
98572	465607	TANAH MERAH GREEN	1.329652	103.943983
39790	458411	OPERA ESTATE	1.319086	103.926208
95575	456648	CASA NALLUR	1.311367	103.932649

```
In [109]: ▶ def subset_data_frame(input_df1):
              Nghbr = []
              Lat = []
              Long =[]

              for name in input_df1.Neighborhood :
                  Nghbr.append(name)
                  coordinates = neighborhoods_subset[neighborhoods_subset["Neighborhood"]==name][["Latitude","Longitude"]]
                  Lat.append(coordinates.iloc[0,0])
                  Long.append(coordinates.iloc[0,1])

              df_temp_coord = pd.DataFrame({"Neighborhood":Nghbr,"Latitude":Lat,"Longitude":Long})
              df_temp_coord=df_temp_coord[["Neighborhood","Latitude","Longitude"]]

              return df_temp_coord
```

```
In [110]: ▶ df_part1 = subset_data_frame(df_tmp)
```

```
In [111]: ▶ df_part2 = subset_data_frame(df_tmp_2)
            df_part2.head()
```

Out[111]:

	Neighborhood	Latitude	Longitude
0	24-HOUR WALK-IN CLINIC (MOUNT ALVERNIA HOSPITAL)	1.341499	103.837774
1	ADELPHI PARK ESTATE	1.355015	103.828192
2	AFFLUENCE COURT	1.350150	103.883062
3	ALEGRIA	1.324303	103.841018
4	ALIWAL PARK HOTEL	1.303419	103.860235

```
In [112]: ▶ df_part1=df_part1.append(df_part2)
            df_part1.shape
```

Out[112]: (345, 3)