The Battle of Neighbourhoods-Final Report

(Opening a Chinese Restaurant in Singapore)

QI ZHAO

IBM Data Science Professional Certificate

Contents

1. Introduction

1.1 Background and Target Audiences

Singapore is a sunny tropical island in Southeast Asia, which has a combination of world-class infrastructure, complete transportation systems throughout the island, vibrant living spaces, full of vigor and vitality of the business environment. The city-state has a high population density and is home to about 5.7 million people from four ethnic groups: the Majority Chinese, Malays, Indians and Eurasians.

The multi-culture has contributed to the diverse culinary scene. Although an eclectic mix of restaurants from all over the world can be easily spotted on the streets of Singapore, it is still highly profitable to open a restaurant considering the high population density and the increasing number of tourists and migrants. Since Chinese migrants take up the majority of the population, in this capstone project we will explore the neighbourhoods of Singapore to help potential stakeholders select optimal locations to open a Chinese restaurant.

The target audiences of this project include:

- **Group 1 : Potential stakeholders who want to invest in Singapore by opening a Chinese restaurant** （this group is the main target audiences of this project）**.**

  This project would be a useful starter guide for them to narrow down their location options.

- **Group 2: Singapore citizens who want to have a clearer idea of their neighbourhoods** (Even though they are already familiar with their own neighbourhood, it does not mean they are familiar with all the neighbourhoods.). This project is helpful to them when they want to find the top popular venues around a specific neighbourhood.

- **Group 3: Tourists and Migrants who are new to Singapore**. This project provides them with a scope with most common venues of every neighbourhood and the distribution of restaurants, and would be helpful when they want to find a restaurant for meals but have no idea about where to go.

1.2 Business Problem

We assume that a stakeholder wants to open a Chinese restaurant in Singapore, and has not decided the location yet. He is not very familiar with the city and asked us to recommend neighbourhoods or locations where he should open his restaurant.

Certain considerations must be taken when choosing the ideal location:

- When exploring all the neighborhoods, choose one where the Chinese restaurants are not among the 10 most common businesses in the neighborhood.
- It would be ideal if the neighborhood has hotels and entertainment areas which would indicate that there are many tourists around and people traveling for work, etc.
- The best 3 candidates who meet these requirements will be recommended.

2. Data Collection

Singapore is divided into regions, planning areas and subzones. The Planning Regions (total 5) are divided into smaller Planning Areas (total 55). Each Planning Area is further divided into smaller subzones (total more than 300). In this project we will explore the neighbourhoods in the level of Planning Area, i.e. we get the venue data around each planning area and select 3 planning area as the 'best' locations to open a Chinese restaurant.

The data in this project consists of two parts.

2.1 A List of Singapore Planning Areas and the Corresponding Latitudes & Longitudes

The list of planning areas defines the scope of this project which is confined to the country Singapore. The latitudes and longitudes of the planning areas are required to plot the map and get the venue data.

Data of Singapore Planning Area boundaries is available in the page https://data.gov.sg/dataset/master-plan-2019-planning-area-boundary-no-sea [1]. I downloaded the original .kml file, transformed it into a .csv file, and read the .csv data into a Pandas DataFrame *df_source*, the features of which contains information about regions, planning areas, coordinates, etc. Please refer to Figure 1 for more details.

```
df_source= pd.read_csv('planning-boundary-area.csv')
df_source.head()
```

	X	Y	gid	Name	description	PLN_AREA_N	PLN_AREA_C	CA_IND	REGION_N	REGION_C	INC_CRC	FMEL_UPD_D
0	103.793357	1.328117	3	kml_3	NaN	BUKIT TIMAH	BT	N	CENTRAL REGION	CR	6CCDADD1F85173E9	20191206144714
1	103.801664	1.376076	4	kml_4	NaN	CENTRAL WATER CATCHMENT	CC	N	NORTH REGION	NR	9F30125764C74984	20191206144714
2	103.748492	1.387486	6	kml_6	NaN	CHOA CHU KANG	CK	N	WEST REGION	WR	5224CD5C7960361F	20191206144714
3	104.049107	1.387936	14	kml_14	NaN	NORTH-EASTERN ISLANDS	NE	N	NORTH-EAST REGION	NER	E75708EADCFF04A6	20191206144714
4	103.725202	1.362108	34	kml_34	NaN	TENGAH	TH	N	WEST REGION	WR	0D2FF9150EC36DFE	20191206144714

```
df_source.columns
```

```
Index(['X', 'Y', 'gid', 'Name', 'description', 'PLN_AREA_N', 'PLN_AREA_C',
       'CA_IND', 'REGION_N', 'REGION_C', 'INC_CRC', 'FMEL_UPD_D'],
      dtype='object')
```

Figure 1: DataFrame df_source

The original DataFrame *df_source* contains 12 columns. We only captured data in the columns *'X', 'Y', 'REGION_N' and 'PLN_AREA_N'*, then populated the data into a new DataFrame named *df_coord* for our analysis. Please refer to Figure 2 for more details.

```
df_coord = pd.DataFrame(columns = ['planning_area', 'region','latitude', 'longitude'])
```

```
df_coord['planning_area'] = df_source['PLN_AREA_N']
df_coord['region'] = df_source['REGION_N']
df_coord['latitude'] = df_source['Y']
df_coord['longitude'] = df_source['X']
df_coord.head()
```

	planning_area	region	latitude	longitude
0	BUKIT TIMAH	CENTRAL REGION	1.328117	103.793357
1	CENTRAL WATER CATCHMENT	NORTH REGION	1.376076	103.801664
2	CHOA CHU KANG	WEST REGION	1.387486	103.748492
3	NORTH-EASTERN ISLANDS	NORTH-EAST REGION	1.387936	104.049107
4	TENGAH	WEST REGION	1.362108	103.725202

```
df_coord.shape
```

```
(55, 4)
```

Figure 2: DataFrame df_coord

Next, we created a map using Folium packages with planning areas superimposed on top. Please refer to Figure 3.

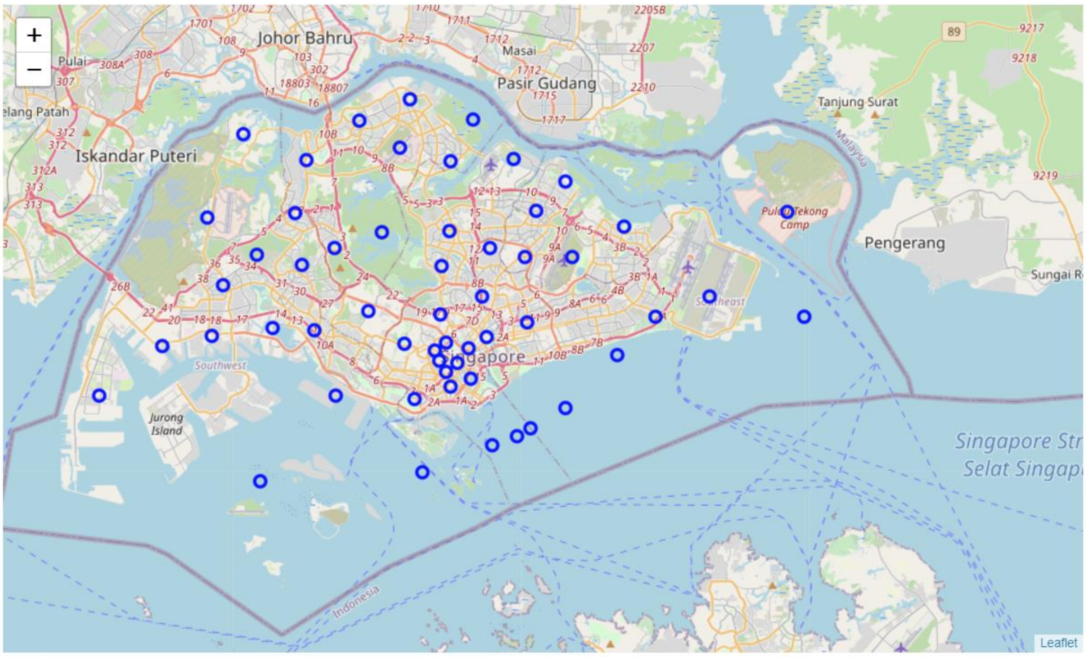

Figure 3: Map of Singapore with planning areas superimposed on top

2.2 Venue Data around Each Planning Area

Venue data is used to perform clustering on the planning areas.

I used Foursquare's "explore" API call [2].to get the information of venues around each planning area. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, latitudes, longitudes, venue categories, etc.

For each planning area, we have chosen the limit to be 100, and the radius to be 2000 meters. Figure 4 below shows the process to get the nearby venue data.

```
venues=[]
for planning_area, lat, lng in zip(df_coord['planning_area'],df_coord['latitude'], df_coord['longitude']):
    print(planning_area)

    # create the API request URL
    url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
        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
    for venue in results:
        venues.append((
            planning_area,
            lat,
            lng,
            venue['venue']['name'],
            venue['venue']['location']['lat'],
            venue['venue']['location']['lng'],
            venue['venue']['categories'][0]['name']))
```

Figure 4: Process to get the nearby venue data

We only captured the venue data useful for us and populated it into a DataFrame named *venues_df*, refer to Figure 5. There are totally **298 unique categories** curated from all the returned venues.

```
# convert the venues list into a new DataFrame
venues_df = pd.DataFrame(venues)

# define the column names
venues_df.columns = ['planning_area', 'latitude', 'longitude', 'venue_name', 'venue_atitude', 'venue_longitude', 'venue_category']

print(venues_df.shape)
venues_df.head()
```

(3421, 7)

	planning_area	latitude	longitude	venue_name	venue_atitude	venue_longitude	venue_category
0	BUKIT TIMAH	1.328117	103.793357	Plank Sourdough Pizza By Baker & Cook	1.323890	103.796797	Pizza Place
1	BUKIT TIMAH	1.328117	103.793357	Brazil Churrasco	1.330798	103.795201	Churrascaria
2	BUKIT TIMAH	1.328117	103.793357	Ristorante Da Valentino	1.336949	103.794060	Italian Restaurant
3	BUKIT TIMAH	1.328117	103.793357	Sunny Heights	1.334700	103.794795	Dog Run
4	BUKIT TIMAH	1.328117	103.793357	Simply Bread	1.330535	103.795658	Bakery

```
print('There are {} uniques categories.'.format(len(venues_df['venue_category'].unique())))
```

There are 298 uniques categories.

Figure 5: Detailed Data of the DataFrame venues_df

Obtained venue data were used for the exploration, analysis and clustering the planning areas of Singapore.

3. Methodology

3.1 Analyze Each Planning Area

When analyzing each planning area, the objective is to prepare the data used for clustering and get the top 10 most common venues.

Firstly, One-hot Encoding technique was applied to the venue category data.

```
# one hot encoding
singapore_onehot = pd.get_dummies(venues_df[['venue_category']], prefix="", prefix_sep="")

# add planning_area column back to dataframe
singapore_onehot['planning_area'] = venues_df['planning_area']

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

print(singapore_onehot.shape)
singapore_onehot.head()
```

```
(3429, 299)
```

	planning_area	ATM	Airport	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Aquarium	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	G
0	BUKIT TIMAH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	BUKIT TIMAH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	BUKIT TIMAH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	BUKIT TIMAH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	BUKIT TIMAH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Figure 6: One-hot Encoding Technique

Then, the one-hot result was grouped by the mean frequency of occurrence of each category for each planning area.

Next, let's group rows by planning_area and by taking the mean of the frequency of occurrence of each category

```
singapore_grouped = singapore_onehot.groupby(["planning_area"]).mean().reset_index()

print(singapore_grouped.shape)
singapore_grouped
```

(55, 299)

	planning_area	ATM	Airport	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Aquarium	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Aust Resta
0	ANG MO KIO	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.030000	0.000000	0.0
1	BEDOK	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.085714	0.000000	0.0
2	BISHAN	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.020000	0.000000	0.0
3	BOON LAY	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.0
4	BUKIT BATOK	0.00	0.000000	0.000	0.000000	0.000000	0.010417	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.0
5	BUKIT MERAH	0.00	0.000000	0.000	0.000000	0.000000	0.010000	0.010000	0.00	0.000000	0.01	0.000000	0.010000	0.000000	0.0
6	BUKIT PANJANG	0.00	0.000000	0.000	0.000000	0.000000	0.021739	0.000000	0.00	0.000000	0.00	0.010870	0.021739	0.010870	0.0
7	BUKIT TIMAH	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.010000	0.000000	0.0

Figure 7: Grouped data by the mean frequency of occurrence of each category for each planning area for each planning area

Finally, the grouped data was sorted by descending order and the top 10 most venues were populated into a DataFrame.

```
# First, let's write a function to sort the venues in descending order.

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]
```

```
# Now let's create the new dataframe and display the top 10 venues for each planning area.
num_top_venues = 10

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

# create columns according to number of top venues
columns = ['planning_area']
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
planningarea_venues_sorted = pd.DataFrame(columns=columns)
planningarea_venues_sorted['planning_area'] = singapore_grouped['planning_area']

for ind in np.arange(singapore_grouped.shape[0]):
    planningarea_venues_sorted.iloc[ind, 1:] = return_most_common_venues(singapore_grouped.iloc[ind, :], num_top_venues)

planningarea_venues_sorted.head()
```

	planning_area	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 Most Common Venue
0	ANG MO KIO	Food Court	Chinese Restaurant	Coffee Shop	Park	Café	Noodle House	Fast Food Restaurant	Japanese Restaurant	Asian Restaurant	Snack Place
1	BEDOK	Chinese Restaurant	Seafood Restaurant	Beach	Asian Restaurant	Park	Skate Park	Harbor / Marina	Pier	Wings Joint	Bike Rental / Bike Share
2	BISHAN	Chinese Restaurant	Coffee Shop	Café	Supermarket	Japanese Restaurant	Thai Restaurant	Park	Food Court	Spa	Ice Cream Shop
3	BOON LAY	Exhibit	Zoo Exhibit	Fishing Spot	Café	Restaurant	Boat or Ferry	Bus Station	Bus Stop	Other Great Outdoors	Theater

Figure 8: DataFrame of planning areas with top 10 most common venues

3.2 Cluster Planning Areas based on Data of Chinese Restaurants

Since we want to know where to open a Chinese Restaurant, we filtered venue category "Chinese Restaurant" from the DataFrame *singapore_grouped* and created a new DataFrame *ChineseRes_grouped*. Please refer to Figure 9.

```
ChineseRes_grouped = singapore_grouped[["planning_area","Chinese Restaurant"]]
ChineseRes_grouped.head()
```

	planning_area	Chinese Restaurant
0	ANG MO KIO	0.090000
1	BEDOK	0.142857
2	BISHAN	0.120000
3	BOON LAY	0.000000
4	BUKIT BATOK	0.062500

Figure 9: DataFrame of planning areas with only category of Chinese Restaurant

We performed clustering based on DataFrame *ChineseRes_grouped* by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project.

We clustered the planning areas into "3" clusters based on their frequency of occurrence for "Chinese Restaurant". The results allow us to identify which planning areas have higher concentration of Chinese restaurants while which have fewer number of Chinese restaurants. Based on the occurrence of Chinese restaurants in different planning areas, it would help us to select the most suitable planning areas to open new Chinese restaurants.

```
# set number of clusters
kclusters= 3

ChineseRes_grouped_clustering = ChineseRes_grouped.drop('planning_area', 1)

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:30]
```

```
array([2, 1, 1, 0, 2, 2, 0, 2, 0, 1, 0, 0, 2, 0, 2, 2, 0, 2, 2, 0, 2, 0,
       0, 0, 0, 2, 0, 2, 0, 0], dtype=int32)
```

```
# add clustering labels
planningarea_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)


# merge     with  df_coord    to add latitude/longitude for each planning area
planningarea_venues_sorted = df_coord.join(planningarea_venues_sorted.set_index('planning_area'), on='planning_area')

planningarea_venues_sorted.head()
```

	planning_area	region	latitude	longitude	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
0	BUKIT TIMAH	CENTRAL REGION	1.328117	103.793357	2	Bakery	Café	Chinese Restaurant	Italian Restaurant	Japanese Restaurant	Pizza Place	Soccer Field	Supermarket
1	CENTRAL WATER CATCHMENT	NORTH REGION	1.376076	103.801664	0	Reservoir	Rest Area	Boat or Ferry	Park	Bridge	Coworking Space	Fast Food Restaurant	French Restaurant
2	CHOA CHU KANG	WEST REGION	1.387486	103.748492	0	Fast Food Restaurant	Coffee Shop	Asian Restaurant	Café	Food Court	Sandwich Place	Bus Station	Supermarket
3	NORTH-EASTERN ISLANDS	NORTH-EAST REGION	1.387936	104.049107	0	Gun Range	Campground	Asian Restaurant	Military Base	Zoo Exhibit	Flea Market	Filipino Restaurant	Fish & Chips Shop
4	TENGAH	WEST REGION	1.362108	103.725202	0	Food Court	Coffee Shop	Asian Restaurant	Flower Shop	Café	Gas Station	Fast Food Restaurant	Pool

Figure 10: Clustering process and DataFrame with cluster labels and top 10 most common venues

4. Results and Discussion

4.1 Examine Each Cluster

We visualized the three clusters of planning areas on a map (each circle marker represents a planning area) and examine each cluster.

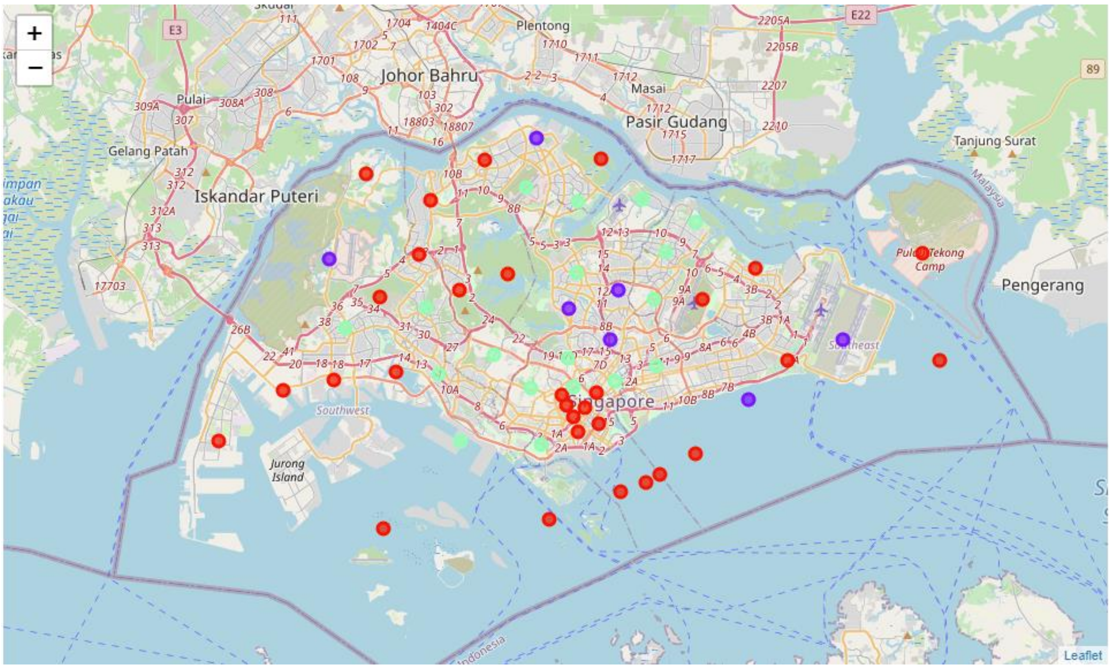

Figure 11: Map showing 3 clusters of all planning areas

• **Cluster 0** (red circle markers on the above map): This cluster contains 30 planning areas and has the lowest frequency of Chinese restaurants, which will be an ideal cluster to open a Chinese restaurant. Below Figure shows a fraction of planning areas in this cluster.

```
Chinese_merged.loc[Chinese_merged['Cluster Labels'] == 0]
```

	planning_area	Chinese Restaurant	Cluster Labels	region	latitude	longitude
19	LIM CHU KANG	0.000000	0	NORTH REGION	1.435699	103.716885
53	WOODLANDS	0.025000	0	NORTH REGION	1.443618	103.787925
26	NORTH-EASTERN ISLANDS	0.000000	0	NORTH-EAST REGION	1.387936	104.049107
43	SOUTHERN ISLANDS	0.000000	0	CENTRAL REGION	1.229460	103.826155
24	MUSEUM	0.000000	0	CENTRAL REGION	1.295972	103.847505
23	MARINE PARADE	0.000000	0	CENTRAL REGION	1.268724	103.913258
22	MARINA SOUTH	0.000000	0	CENTRAL REGION	1.251698	103.884283

Figure 12: A fraction of planning areas in cluster 0

• **Cluster 1** (purple circle markers on the above map): This cluster contains 7 planning areas and has the highest frequency of Chinese restaurants, so those areas are not good locations to open a Chinese restaurant. Below Figure shows all the planning areas in this cluster.

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

	planning_area	Chinese Restaurant	Cluster Labels	region	latitude	longitude
38	SEMBAWANG	0.129630	1	NORTH REGION	1.457080	103.818933
49	TOA PAYOH	0.160000	1	CENTRAL REGION	1.336457	103.862479
9	CHANGI	0.125000	1	EAST REGION	1.337079	104.001643
52	WESTERN WATER CATCHMENT	0.250000	1	WEST REGION	1.384956	103.694919
2	BISHAN	0.120000	1	CENTRAL REGION	1.355160	103.837734
1	BEDOK	0.142857	1	EAST REGION	1.301108	103.944936
40	SERANGOON	0.110000	1	NORTH-EAST REGION	1.366010	103.867606

Figure 13: A fraction of planning areas in cluster 1

• **Cluster 2** (green circle markers on the above map): This cluster contains 18 planning areas and has the medium frequency of Chinese Restaurants, which is also not an ideal cluster to open a Chinese Restaurant. Below Figure shows a fraction of planning areas in this cluster.

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

	planning_area	Chinese Restaurant	Cluster Labels	region	latitude	longitude
47	TANGLIN	0.040000	2	CENTRAL REGION	1.307610	103.815114
0	ANG MO KIO	0.090000	2	NORTH-EAST REGION	1.376729	103.842565
27	NOVENA	0.070000	2	CENTRAL REGION	1.326091	103.837228
37	SELETAR	0.066667	2	NORTH-EAST REGION	1.420722	103.881743
34	QUEENSTOWN	0.073171	2	CENTRAL REGION	1.276400	103.773753
33	PUNGGOL	0.060000	2	NORTH-EAST REGION	1.406764	103.913796
25	NEWTON	0.040000	2	CENTRAL REGION	1.308506	103.840985
20	MANDAI	0.050847	2	NORTH REGION	1.427104	103.812815
18	KALLANG	0.060000	2	CENTRAL REGION	1.312327	103.865107

Figure 14: A fraction of planning areas in cluster 2

4.2 Select the Optimal Planning Areas

Next, we would examine which planning areas in cluster 0 have more hotels. More hotels indicate that there are many tourists around and people traveling for work, thus appeal more people to the Chinese restaurant. Below DataFrame shows the planning areas with the most hotels and least Chinese restaurants.

	Chinese Restaurant	planning_area	Hotel	region	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 Most Common Venue	10th Most Common Venue
13	0	DOWNTOWN CORE	16	CENTRAL REGION	0	Hotel	Waterfront	Event Space	Japanese Restaurant	Shopping Mall	Italian Restaurant	Plaza	Buffet	Restaurant	Cocktail Bar
24	0	MUSEUM	11	CENTRAL REGION	0	Hotel	Japanese Restaurant	Cocktail Bar	Wine Bar	Shopping Mall	Café	Arts & Crafts Store	Performing Arts Venue	Concert Hall	Coffee Shop
29	0	OUTRAM	11	CENTRAL REGION	0	Hotel	Japanese Restaurant	Gym / Fitness Center	Coffee Shop	Restaurant	Cocktail Bar	Korean Restaurant	Bar	Spanish Restaurant	Café
28	2	ORCHARD	10	CENTRAL REGION	0	Hotel	Japanese Restaurant	Shopping Mall	Bakery	Sushi Restaurant	Boutique	Clothing Store	Coffee Shop	Bubble Tea Shop	Cocktail Bar
36	1	ROCHOR	10	CENTRAL REGION	0	Hotel	Café	Coffee Shop	Indian Restaurant	Ice Cream Shop	Italian Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Bakery
42	0	SINGAPORE RIVER	6	CENTRAL REGION	0	Japanese Restaurant	Hotel	Wine Bar	Bookstore	Cocktail Bar	Spanish Restaurant	Hotpot Restaurant	French Restaurant	Bar	Bistro

Figure 15: DataFrame of planning areas with most hotels

All the planning areas in the above figure are in Central Region and have no Chinese restaurants in the top 10 most common venues (For ORCHARD and ROCHOR, Chinese restaurants are within the 2000 meters radius but not in the top 10 list). **DOWNTOWN CORE, MUSEUM, and OUTRAM**, these three planning areas have many hotels around , but no Chinese restaurants in both top 10 list and 2000 meters radius, so these three planning areas are the most optimal locations to open Chinese Restaurants.

5. Conclusion

In this project, we have gone through the process of introducing the background and identifying the business problem, collecting the data and preparing the data for analysis, performing machine learning clustering into “3” clusters based on data of Chinese restaurants, and lastly providing recommendations to potential stakeholders regarding the best planning areas to open a new Chinese restaurant.

Three planning areas DOWNTOWN CORE, MUSEUM, and OUTRAM in cluster 0 are most recommended.

There are also some **limitations** of this project. Firstly, the venue data extracted from Foursquare is confined to the radius 2000 meters, which may not represent the characteristics the whole planning area perfectly. In addition, we didn’t consider the competition of other types of restaurants in each planning area. And except hotels, other types of entertainment venues which also appeal a large number of tourists are not taken into account. All those aspects can be improved in the future work.

6. References

- [1] Data of Singapore Planning Area boundaries
- [2] Foursquare API