The Battle of Neighborhoods — Coursera IBM Capstone Project

Introduction & Business Problem :

Problem Background:

Singapore is a multi-racial society and as of mid-2018, the estimated population of Singapore was 5,638,700 people. The Singaporean economy is known as one of the most innovative, most competitive, most dynamic and most business-friendly. The 2015 Index of Economic Freedom ranks Singapore as the second freest economy in the world and the Ease of doing business index has also ranked Singapore as the easiest place to do business for the past decade. Singapore attracts a large amount of foreign investment as a result of its location, skilled workforce, low tax rates, advanced infrastructure and zero-tolerance against corruption. (Source: https://en.wikipedia.org/wiki/Singapore)

However, Singapore was among the bottom 10 for work-life balance and ranked the second most overworked city in a study of 40 cities. (Source: channelnewsasia.com/news/singapore/singapore-bottom-ranks-work-life-balance-second-most-overworked-11789264).

This has lead to recommendations by the ministry to increase flexibility at work such that employees will better manage their time. Many employees find it hard to work from home due to the lack of discipline and for working parents, the lack of personal space and time. For many employers, they find it difficult to trust that their employees when they are not physically in the office. (Source: https://www.population.sg/articles/why-worklife-balance-may-not-work-in-singapore)

Problem Description:

An group of overseas investors wishes to set up a inclusive, integrated 24-hours working hub for professionals, who find it difficult to work from home, to tackle problems such as childcare and/or lack of space. The hub also has a time log system which employers can digitally track their employees' working hours. This is the first of its kind and if it is successful, many similar hubs will be set up in the residential areas such that employees are able to work within their neighbourhood. The first hub will be in an area where amenities are not readily available.

Factors to be considered are:

- Residential land price

- Categories of shops located in that area (so as to determine what type of shops are to be included within the hub)
- Overview of the number of professionals residing in that area

Target Audience:

The main target will be working professionals as well as companies who do not have physical workspaces and/or whose employees require work flexibility.

Data

Data used in this project will include:

Data 1: Postal Codes and the various neighbourhoods in Singapore This dataset exists in Wikipedia. Link to the dataset is : https://en.wikipedia.org/wiki/Postal_codes_in_Singapore

Data 2 : Getting coordinate for the neighbourhoods from Geopy Client

Data 3: Foursquare API to get the nearby venues and shop categories of various districts. Data for Hawker Centres (local eateries in Singapore) which is not available on Foursqaure and will be obtained from : https://data.gov.sg/dataset/hawker-centres

	name_of_centre	location_of_centre	type_of_centre	owner	no_of_stalls	no_of_cooked_food_stalls	no_of_mkt_produce_stalls
0	Adam Road Food Centre	2, Adam Road, S(289876)	HC	Government	32	32	0
1	Amoy Street Food Centre	National Development Building, Annex B, Telok ...	HC	Government	135	134	1
2	Bedok Food Centre	1, Bedok Road, S(469572)	HC	Government	32	32	0
3	Beo Crescent Market	38A, Beo Crescent, S(169982)	MHC	Government	94	32	62
4	Berseh Food Centre	166, Jalan Besar, S(208877)	HC	Government	66	66	0

Data 4: Number of working professionals sorted by districts will be obtained from : https://data.gov.sg/dataset/resident-working-persons-aged-15-years-and-over-by-planning-area-and-occupation-2015

	Jobs	location_of_centre	No of Population
0	Total	Ang Mo Kio , Singapore	101200.0
1	Total	Bedok , Singapore	150300.0
2	Total	Bishan , Singapore	49500.0
3	Total	Bukit Batok , Singapore	75600.0
4	Total	Bukit Merah , Singapore	82200.0
5	Total	Bukit Panjang , Singapore	80100.0
6	Total	Bukit Timah , Singapore	36800.0

Methodology:

Business Understanding:

Our main goal is to find an optimum location to build the integrated hub for the investor.

Analytic Approach:

Singapore has 74 districts and the first part of the analysis is done by clustering the areas using Exploratory data analysis.

Exploratory Data Analysis:

Data 1 – Wikipedia

1. Data was extracted from Wikipedia to obtain the different postal codes in Singapore.

2. Data was then converted to dataframe and filtered into columns Postal District, Postalcode and Neighbourhood.

	Postal_District	Postalcode	Neighborhood
0	01	01, 02, 03, 04, 05, 06	Raffles Place, Cecil, Marina, People's Park
1	02	07, 08	Anson, Tanjong Pagar
2	03	14, 15, 16	Bukit Merah, Queenstown, Tiong Bahru
3	04	09, 10	Telok Blangah, Harbourfront
4	05	11, 12, 13	Pasir Panjang, Hong Leong Garden, Clementi New...

Data 2 – Coordinates for neighbourhoods

1. Based on the Neighbourhood generated, longitude and latitude of these areas were obtained via Geopy Client.

	location_of_centre	location	point	Latitude	Longitude
37	Katong, Singapore	112 Katong, East Coast Road, Marine Parade, So...	(1.3052327, 103.90505238485225, 0.0)	1.305233	103.905052
53	Ang Mo Kio, Singapore	2450, Ang Mo Kio Avenue 8, Ang Mo Kio, Singapo...	(1.3700733, 103.8495157, 0.0)	1.370073	103.849516
39	Amber Road, Singapore	Amber Road, Marine Parade, Southeast, 489886, ...	(1.303121, 103.900556, 0.0)	1.303121	103.900556
4	Anson, Singapore	Anson Road, Radin Mas, Bukit Merah, Singapore,...	(1.2718055, 103.843321, 0.0)	1.271805	103.843321
30	Balestier, Singapore	Balestier, Singapore, Central, 329842, Singapore	(1.326226, 103.8473149, 0.0)	1.326226	103.847315

2. Nearby venues, categories and top 10 most common venues were then obtained as well from Foursquare API.

Neighborhood	1st Most Common Venue	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 Cor
112 Katong, East Coast Road, Marine Parade, So...	Chinese Restaurant	Bakery	Hotel	Noodle House	Indian Restaurant	Japanese Restaurant	Coffee Shop	Asian Restaurant	Ma
2450, Ang Mo Kio Avenue 8, Ang Mo Kio, Singapo...	Coffee Shop	Food Court	Dessert Shop	Supermarket	Sandwich Place	Bubble Tea Shop	Fast Food Restaurant	Japanese Restaurant	C
Amber Road, Marine Parade, Southeast, 489886, ...	Hotel	Café	Japanese Restaurant	Indian Restaurant	Bar	Chinese Restaurant	Flower Shop	Ice Cream Shop	
Anson Road, Radin Mas, Bukit Merah, Singapore,...	Japanese Restaurant	Coffee Shop	Café	Hotel	Food Court	Soup Place	Soba Restaurant	Bakery	Rest

Data 3 – Obtaining nearby eateries (Hawker Centre) that are not available in Foursquare

1. Dataframe was downloaded from website and analysed.

	name_of_centre	location_of_centre	type_of_centre	owner	no_of_stalls	no_of_cooked_food_stalls	no_of_mk
0	Adam Road Food Centre	2, Adam Road, S(289876)	HC	Government	32	32	
1	Amoy Street Food Centre	National Development Building, Annex B, Telok ...	HC	Government	135	134	
2	Bedok Food Centre	1, Bedok Road, S(469572)	HC	Government	32	32	
3	Beo Crescent Market	38A, Beo Crescent, S(169982)	MHC	Government	94	32	
4	Berseh Food	166, Jalan Besar,	HC	Government	66	66	

2. After obtaining the longitudes and latitudes from Geopy Client, a folium map was generated, integrating the clustering map from Data 1 with the locations of the hawker centres.

Data 4 – Number of Professionals above age 15 based on districts

1. Dataframe was downloaded from website and analysed. From the data, the following was obtained:
   - Maximum number of professionals in one area: 158,700
   - District where maximum number professionals was situated: Jurong West, Singapore
   - Minimum number of professionals in one area: 45,900
   - District where minimum number of professionals was situated: Clementi, Singapore

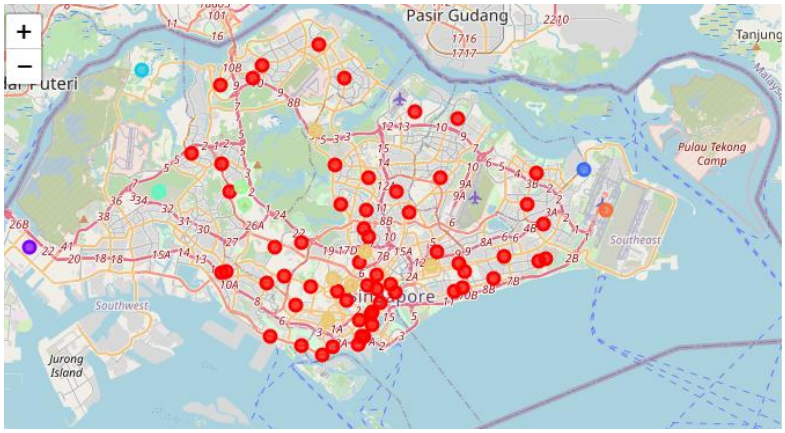
	Jobs	location_of_centre	No of Population	location	point
0	Total	Ang Mo Kio , Singapore	101200.0	(2450, Ang Mo Kio Avenue 8, Ang Mo Kio, Singap...	(1.3700733, 103.8495157, 0.0)
1	Total	Bedok , Singapore	150300.0	(Bedok, New Upper Changi Road, Simpang Bedok, ...	(1.3239765, 103.930216, 0.0)
2	Total	Bishan , Singapore	49500.0	(Bishan, Bishan Place, Bishan, Singapore, Cent...	(1.3509859, 103.84825507492937, 0.0)
3	Total	Bukit Batok , Singapore	75600.0	(Bukit Batok, Bukit Batok Central Link, Bukit ...	(1.3490572, 103.7495906, 0.0)
4	Total	Bukit Merah , Singapore	82200.0	(Bukit Merah, Singapore, Central, Singapore, (...	(1.2704395, 103.82831840176755, 0.0)

Results:

Neighbourhood K-means clustering based on mean occurrence of venue category:

To cluster the needihoods into 8 clusters, we used the K-Means clustering Algorithm. The cluster number was established after multiple samplings and iterations.

1. Clustering (total of 8) was done and visualised into a folium map.



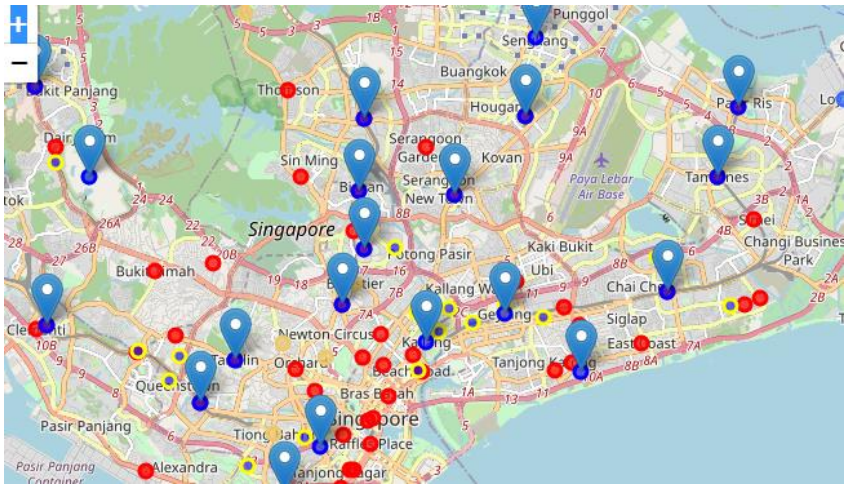
2. Clusters were then examined to see which cluster has the most and least amenities and facilities. The clusters in red or cluster = 0 are the areas have the most amenities while clusters 1,2,3,4 and 7 have very few amenities.

Examine cluster 1

```
new_merge.loc[new_merge['Cluster Labels'] == 1, new_merge.columns[[1] + list(range(5, new_merge.shape[1]))]]
```

	Neighborhood	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
58	Tuas, Southwest, Singapore	1	Bar	Coffee Shop	Food Court	Fishing Spot	Flea Market	Flower Shop	Food	Food & Drink Shop	Yoga Studio	Fish & Chips Shop

2. Using Geopy Client and Folium, a map integrating data 1,2 and 3 was generated.



3. With the given numbers, we are able to see that clusters 1, 2, 3 and 7 are not suitable areas. There were no recorded professionals living in those clusters 1 and 2 and cluster 7 is the airport.

Discussion:

- Based on the criteria given by the investor group and the cluster data, we focused on clusters 4 - Tengah.
- This cluster is surrounded by 3 areas where a lot of professionals lived.

- The 3 areas are Jurong West, Jurong East and Bukit Batok.
- In addition, despite the high number of people living there (Jurong West: 158,700, Jurong East: 46,100, Bukit Batok: 75,600), there were only a few amenities and 2 local eateries.
- This is an ideal location where an integrated hub will serve its goal and most likely to be profitable.

Conclusion:

The analysis is performed on limited data. The information afforded us may be dated due to relying on available data on websites which is usually not updated in real time. To add on, land price should be monitored to maximise profits. However, this information was not readily available. Overall, the model created can be easily replicated and monitored closely once recent numbers are available.