Capstone Project –
Battle of the neighborhoods

Singapore Neighborhood Recommendation

1. Introduction

1.1 Background:

Singapore is a highly developed market economy, it is known as one of the freest, most innovative, most competitive, most dynamic and most business-friendly. It is a global hub for business and commerce, it is diverse and a financial capital. Singapore is gateway to Asia, many multinational companies has their office located in Singapore, be it their head office, regional office or branch office. Singapore is also popular among tourist and one of the most visited countries in World.

Singapore is also one of the world's most expensive city to live in. There are people of all nationality come here and settle for work. Finding a suitable place to live is a big challenge. Expats have a difficult time in selecting the place to live with recommendations flying from different sources like websites, relatives, friends, colleagues etc.

1.2 Problem:

The objective of the project is to locate and recommend a suitable place to live for an expat looking to settle in Singapore with preferred places of interest. While executing the project the rationale for selecting the location, following points will be taken into account.

- Collect and provide a data driven recommendation that is supplemented with statistical data analysis
- Give a comparison of Average rent of all the towns(neighborhood) in Singapore
- Compare the most common places of interest in all towns and rank top 10 common venue.
- Make 5 clusters with similar attributes on the basis of most common venue and then attach the average rent columns.

For this demonstration, this Project will make use of the following data:

- Singapore Median Rental Prices by town.
- Popular Food venues in the vicinity.

1.3 Interest:

Apparently, Expats looking to settle in Singapore would be very interested in finding the appropriate location to settle with preferred place of their interest nearby. Students and Tourists may also find the output of project very useful.

2. Data Acquisition and Cleaning:

2.1 Data sources

The Project has sourced data from Singapore open data sources and FourSquare API venue recommendations.

Singapore Towns and median residential rental prices

Data is retrieved from Singapore open dataset from median rent by town from https://data.gov.sg website.

The original data source contains median rental prices of Singapore HDB units from 2005 Q2 up to Q4 of 2019. The data has four columns as given below

	quarter	Town	flat_type	median_rent
0	2005-Q2	ANG MO KIO	1-RM	na
1	2005-Q2	ANG MO KIO	2-RM	na
2	2005-Q2	ANG MO KIO	3-RM	800
3	2005-Q2	ANG MO KIO	4-RM	950
4	2005-Q2	ANG MO KIO	5-RM	-

We will retrieve rental the most recent recorded rental prices from this data source (2019-Q4) being the most relevant price available at this time. For this demonstration, we will simplify the analysis by using the average rental prices of all available flat type.

Singapore Top Venue Recommendations from FourSquare API

(FourSquare website: www.foursquare.com)

We have used the FourSquare API to explore neighborhoods in selected towns in Singapore. The Foursquare explore function is used to get the most common venue categories in each neighborhood, and then we will use this feature to group the neighborhoods into clusters.

	Town	Town Latitude	Town Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ANG MO KIO	1.370073	103.849516	FairPrice Xtra	1.369279	103.848886	Supermarket
1	ANG MO KIO	1.370073	103.849516	Old Chang Kee	1.369094	103.848389	Snack Place
2	ANG MO KIO	1.370073	103.849516	Subway	1.369136	103.847612	Sandwich Place
3	ANG MO KIO	1.370073	103.849516	MOS Burger	1.369170	103.847831	Burger Joint
4	ANG MO KIO	1.370073	103.849516	NTUC FairPrice	1.371507	103.847082	Supermarket

2.2 Data cleaning

For the towns rental data we have found the average rental of each town for the period of 2019-Q4. We have also removed unwanted columns as shown below

	Town	median_rent
0	ANG MO KIO	2066.666667
1	BEDOK	2087.500000
2	BISHAN	2200.000000
3	BUKIT BATOK	1962.500000
4	BUKIT MERAH	2416.666667
5	BUKIT PANJANG	1770.000000
6	CENTRAL	2500.000000
7	CHOA CHU KANG	1790.000000

Using geopy geocoders nominatim library, which convert an address into latitude and longitude values, we obtained the coordinates of towns and attached with the above table.

	median_rent	Latitude	Longitude
Town			
ANG MO KIO	2066.666667	1.370073	103.849516
BEDOK	2087.500000	1.323976	103.930216
BISHAN	2200.000000	1.351452	103.848250
BUKIT BATOK	1962.500000	1.349057	103.749591
BUKIT MERAH	2416.666667	1.280628	103.830591
BUKIT PANJANG	1770.000000	1.378629	103.762136
CENTRAL	2500.000000	1.340863	103.830392

3. Methodology:

For achieving the objective of our project we will take a recourse to unsupervised approach for machine learning such as segmenting and k-means clustering. Unsupervised learning is a type of machine learning that looks for previously undetected patterns in a data set with no pre-existing labels and with a minimum of human supervision.

In the first step we have downloaded the median rental prices of various towns in Singapore. Further we have taken average of rental prices of all available flat type and attached the coordinates to Towns. Using Foursquare API we have downloaded nearby places data with their ratings.

In the second step we will explore the neighborhoods and analyze each Singapore Town nearby recommended venues from the data downloaded using Foursquare.

In the third and final step we will rank the top ten most common nearby place of interest. We will also run *k*-means to cluster the Towns into 5 clusters based on Top ten most common nearby venue and present our findings in a map.