Applied Data Science Capstone

IBM Data Science Professional Certificate

Neighbourhood Exploration for Expatriates Moving to Singapore

By:

Wenshuo Cui

June 2021

1. Introduction

1.1. Background

Expatriates, more commonly known as expats, are professionals who reside in a country other than their home countries. Migration to a new country marks the start of the journey for expats, and this process could be extremely dauting since it entails marching into the unknown. Amongst all the challenges, choosing a place to live is perhaps the most difficult of all; expats have to invest significant amount of time on researching the neighbourhoods to gain a brief understanding. This manual process could be extremely time-consuming, which further exacerbates the anxiety of the expats planning the move.

Situated in Southeast Asia, Singapore hosts 110,000 expats and 7,000 multinational companies[1]. Singapore is also one of the most popular expat destinations; Singapore ranked the best place overall for expats in 2018[2]. As a cultural melting pot, Singapore a highly multiracial and multicultural country. For example, Singapore has 4 official languages: Malay, English, Chinese and Tamil. The direct consequence of this is the increased difficulties for expats to research the different neighbourhoods due to the language and cultural barrier. As an example, it would be challenging to know the exact purpose of a store or restaurant by simply looking at the names. Therefore, the cultural diversity of Singapore further exacerbates the problem of researching a place to live for expats.

1.2. Business Problem

Currently, there is not any elegant solution for expats to quickly gain a bird's-eye-view of each neighbourhood and further restrict search area for a house in Singapore. Expats can choose to use software such as Google Maps or Foursquare to explore each neighbourhood separately. However, this is highly manual and does not provide an overview of the neighbourhoods. Alternatively, expats could read up on blogposts by other expats who have been living in Singapore. However, these blogposts could be biased and skewed since it is unlikely that anyone could know every neighbourhood well.

---

[1] Ministry of Manpower Singapore, "Living in Singapore", https://www.mom.gov.sg/working-in-singapore/living-in-singapore#:~:text=Businesses%20prize%20the%20Singapore%20workforce,forging%20a%20career%20in%20Singapore.

[2] HSBC, "Expat Explorer Survey 2018", https://www.expatexplorer.hsbc.com/survey/files/pdfs/overall-reports/2018/HSBC_Singapore_Report.pdf

Therefore, this project aims to employ the power of data science to provide a high-level overview of every neighbourhood in Singapore. This could save expats significant amount of time by providing information to help them to further restrict the neighbourhoods for the house search, based on their personal preferences of diet and lifestyle.

2. Data Acquisition

In order to achieve the goal of this project, a few datasets are required:

- List of neighbourhoods in Singapore
- Coordinates of each neighbourhood
- Information of the venues in each neighbourhood

2.1. List of Neighbourhoods in Singapore

There are many ways of categorising the neighbourhoods in Singapore: different sources may categorise differently. After much research, data from the Urban Redevelopment Authority (URA) of Singapore will be used. This is because URA is the government planning authority of Singapore and it's data should be the most reliable. On URA's website (https://www.ura.gov.sg/realEstateIIWeb/resources/misc/list_of_postal_districts.htm), a table of postal codes with "General Location" is found. This table could be extracted using webscraping techniques and the column of "General Location" could be split by delimiter to obtain the list of unique neighbourhoods in Singapore (see Figure 1).

Postal District	Postal Sector (1st 2 digits of 6-digit postal codes)	General Location
01	01, 02, 03, 04, 05, 06	Raffles Place, Cecil, Marina, People's Park
02	07, 08	Anson, Tanjong Pagar
03	14,15, 16	Queenstown, Tiong Bahru
04	09, 10	Telok Blangah, Harbourfront
05	11, 12, 13	Pasir Panjang, Hong Leong Garden, Clementi New Town
06	17	High Street, Beach Road (part)
07	18, 19	Middle Road, Golden Mile
08	20, 21	Little India
09	22, 23	Orchard, Cairnhill, River Valley
10	24, 25, 26, 27	Ardmore, Bukit Timah, Holland Road, Tanglin
11	28, 29, 30	Watten Estate, Novena, Thomson
12	31, 32, 33	Balestier, Toa Payoh, Serangoon
13	34, 35, 36, 37	Macpherson, Braddell
14	38, 39, 40, 41	Geylang, Eunos
15	42, 43, 44, 45	Katong, Joo Chiat, Amber Road
16	46, 47, 48	Bedok, Upper East Coast, Eastwood, Kew Drive
17	49, 50, 81	Loyang, Changi
18	51, 52	Tampines, Pasir Ris
19	53, 54, 55, 82	Serangoon Garden, Hougang, Punggol
20	56, 57	Bishan, Ang Mo Kio
21	58, 59	Upper Bukit Timah, Clementi Park, Ulu Pandan
22	60, 61, 62, 63, 64	Jurong
23	65, 66, 67, 68	Hillview, Dairy Farm, Bukit Panjang, Choa Chu Kang
24	69, 70, 71	Lim Chu Kang, Tengah
25	72, 73	Kranji, Woodgrove
26	77, 78	Upper Thomson, Springleaf
27	75, 76	Yishun, Sembawang
28	79, 80	Seletar

Figure 1: List of Postal Districts from URA Website

2.2. Coordinates of Neighbourhoods

After obtaining the list of neighbourhoods, the coordinates are required for subsequent operations. Since Singapore has a comprehensive postal code system where each building has a unique postal code, retrieving coordinates using postal codes would not be possible because there would be too many unique postal codes. Therefore, the coordinates will be retrieved based on location search using the geocoder.arcgis function from Python Geocoder package. The result of this would be a list of the neighbourhoods, each with a general coordinate based on the search.

2.3. Information of Venues in each Neighbourhood through Foursqure API

With the coordinates of each neighbourhood in Singapore, Foursqure API could be used to further extract huge amount of data for each neighbourhood. More specifically, the "explore " endpoint is used to return the venue recommendations in each neighbourhood. A radius of 3000 metres is set for the search because 3km roughly translates into a 10 minutes travel time on bicycles[3], which is considered nearby and very likely to be frequently visited by residents in the area.

These data will then be wrangled to output the top 10 venues in the neighbourhood, which would be the foundation of providing an overview of the neighbourhoods for expats. These data would also be used to run clustering algorithm to discover hidden structures in different regions of Singapore to gain more insights.

[3] Agnes Mazur, "Here's what bike commuting looks like in 12 major cities", *Vox,* https://www.vox.com/2015/10/8/9480951/bike-commute-data-strava