IBM Data Science Professional Certificate (Coursera)

- Applied Data Science Capstone

Project Title: Average 2 Bedrooms Units' Weekly Rental Price and Surroundings near The University of Melbourne Parkville Campus by Postcode

Project Proposal

By: Chai Hoon Lim

April 2018

Contents

I. Introduction

This is the proposal report for the capstone project of IBM Data Science Professional Certificate on the Coursera platform. In this project, I have to come up with an idea to leverage the Foursquare location data to explore or compare neighborhoods or cities of my choice or to come up with a problem that I can use the Foursquare location data to solve.

My idea is to explore areas near to the University of Melbourne Parkville campus, segment and cluster the areas and provide information on the average 2 bedrooms units' weekly rents index for the shortlisted postcode area. This idea arises from the housing need of new students moving to Melbourne to attend courses at the University of Melbourne at the beginning of each semester. For some students choose to rent the whole apartment near to the campus, they need to survey on areas which are near to campus before they can decide which area to check out for rental apartments. With this project, I hope that to assist them in deciding on which areas to look out for rental apartments and that they can find a suitable apartment and settle in before semester commences. And 2 bedrooms units are chosen in the analysis as 2 bedrooms units are the most popular housing options among students.

Even though the idea starts with housing needs for new students moving to Melbourne to attend courses at the University of Melbourne; the targeted place can be updated from University of Melbourne Parkville campus to any places at Melbourne that interested the audience.

II. Dataset

In order to achieve the objective of this project, the following datasets are acquired and cleansed accordingly:

Dataset 1: The coordinate of the University of Melbourne Parkville campus (the targeted place) is used to plot the targeted place on the map.

Dataset 2: List of postcodes for suburbs near to the campus are included to retrieve weekly rents index as well as for suburbs comparison.

Dataset 3: Average 2 bedrooms units' weekly rents index is scraped from the sqmresearch.com.au using the following query for each postcode in the **Dataset 2**:

https://sqmresearch.com.au/weekly-rents.php?postcode={}&t=1

After that, the result page is parsed to get the "2br Units" rents index from the table similar to the following:

SQM Research Weekly Rents Index							
Week ending 4 Apr 2019			Change on prev week	Rolling month % change	Rolling quarter % change	12 month % change	3 year % change
Melbourne	All Houses	543.2	1.8 ▲	-0.2% ▼	0.8% ▲	1.3% ▲	13.6% ▲
	3 br Houses	531.1	0.9 ▲	-0.4% ▼	1.7% ▲	2.3% ▲	9.9% ▲
	All Units	421.8	0.2 ▲	0.6% ▲	3.1% ▲	3.8% ▲	12.7% ▲
	2 br Units	424.5	0.5 ▲	0.5% ▲	1.4% ▲	2.8% ▲	10.9% ▲

Dataset 4: List of postcodes with locality and longitude and latitude coordinates from the following website:

https://www.matthewproctor.com/Content/postcodes/australian_postcodes.csv

The locality column in the file is translated to suburb name in the program.

Dataset 5: Geo Dataset for the Melbourne area is obtained from the following website that contains each postcode area's coordinate:

https://raw.githubusercontent.com/codeforamerica/click_that_hood/master/public/data/melbourne.geojson

Dataset 6: Foursquare API is used to retrieve venues near to each postcode coordinates.

1. Data Cleansing & Merging

Checking the datasets discover that **Dataset 4** has unmatched suburbs name compare to **Dataset 5** and the suburb name is corrected according to Dataset 5. 3 postcodes coordinates in **Dataset 4** are missing and are updated with the coordinates getting from http://www.corra.com.au/downloads/Australian_Post_Codes_Lat_Lon.zip

Dataset 2, **3** and **4** are merged to get a list of postcode, suburb name, latitude, longitude coordinates, and 2 Br units weekly rents index to create a "Melbourne 2Br Units' Weekly Rents Index" record for each postcode.

2. How the data is used

Dataset 1 and the merged dataset "Melbourne 2Br Units' Weekly Rents Index" are used to visualize postcode areas with 2Br Units Weekly Rent Index Superimpose on top the Melbourne map.

The neighboring venues data obtains from Dataset 6 are used to cluster postcode areas into 3 clusters using k-means clustering algorithm.