Capstone Project - A Reasonable Location for a New Bookstore in Greater Melbourne (Week1)

by Fendy Gao

date: 2020-Aug-15

Table of Contens

1 Problem Description

Melbourne is the capital and most populous city of the Australian state of Victoria, and the second most populous city in Australia, where I am living. The area of Greater Melbourne area is about 9,993 km2, with a metropolitan area with 31 municipalities. It has a population of about 5 million.

One of my favorite shopping places is bookstore. Due to the impact of on-line sales and electronic publications, it gets more and more difficult to access a regular bookstore now. However, a bookstore is still a very nice place for adults to have some readings and for children to have some learnings. Hence, in this study, I focus myself on bookstores.

To setup a bookstore would be determined by multiple factors, such as population, transportation, rental rate and etc. Also the existing numbers of bookstores in a certain region would be also imortant. I would like to focus on the factors in the population side. Hence, I'd like to collect data about the number of bookstores in each sub-region of Greater Melbourne Area, the bookstore number vs. population, the accessibility of bookstores to each sub-region, and etc. With all these data analyzed using machine learning technologies, I should be able to yield a recommendation for a location to setup a new bookstore.

In summary, in this project, a study was performed to determine a reasonable location (or locations) to setup a new bookstore (bookstores) in the Greater Melbourne area.

2 Data

2.1 Data of Greater Melbourne

The local government areas (LGAs) of Melbourne are collected from the follow wiki link. I use the LGAs as the basic regions for the coming analysis. The wiki table provided the LGA areas and population in 2018, which is acceptable for the current study.

Local government areas sorted by region [edit]

Greater Melbourne [edit]

| Local government area | Council seat | Region | Year est. | Land area[1] | | | Population | | Councillors (2012) |
				km²	sq mi	Density (2018)[1]	(2013)[2]	(2018)[2][1]	
City of Melbourne	Melbourne	Inner Melbourne	12 August 1842	37	14	4,550	118,357	169,961	11
City of Port Phillip	St Kilda	Inner Melbourne	22 June 1994	21	8	5,466	102,156	113,200	7
City of Stonnington	Malvern	Inner Melbourne	22 June 1994	26	10	4,530	103,487	116,207	9

Local government areas of Victoria (https://en.wikipedia.org/wiki/Local_government_areas_of_Victoria)

The wiki table of LGAs of Greater Melbourne did not provide any geographical coordinate information. Hence, the following Python package was used to extract geographical coordinate of each LGA.

Python Folim (https://github.com/python-visualization/folium)

2.2 Data of Bookstores

In order to find all bookstores in each LGA, Foursquare API was used to find all bookstores using a key word "book" and a reasonable large radius. Some bookstores might be missed from this search, but the current method should be acceptable.

Foursquare API (https://developer.foursquare.com/)

Then, the data was carefully processed. The first action was to remove those venues not in the category of bookstore. The second action was to remove those duplicated venues. The reason was that I used a large radius and certain venues might be found from different LGAs. The bookstore latitudes and longitudes were automatically retrieved via Foursquare API and saved into the dataframe.

	name	categories	address	lat	lng	distance	city
0	Federation Square Book Market	Bookstore	Federation Square	-37.817229	144.969340	638	Melbourne
1	The Book Grocer	Bookstore	455 Bourke Street	-37.815270	144.960540	258	Melbourne
2	The Book Grocer	Bookstore	287 Little Collins St	-37.814927	144.965139	191	Melbourne
3	The Book Grocer	Bookstore	165 Swanston Street	-37.813913	144.965694	225	Melbourne
4	Book Grocer	Bookstore	206 Bourke St	-37.812910	144.967020	369	Melbourne

3 Methodology

The following assumption was used to judge where to build a new book store, that a bookstore should be setup where it is needed. The word "needed" is interpreted as that the number of bookstores (or the average number of bookstores per certain population) is less than the average value.

In the study, the number of bookstore in each LGAs were studied, as well as the number of bookstores vs. population, the number of bookstore vs. area, and etc. Then based on the study, the results and recoomendations were presented. Serveral machine learning tools and data plots were used, including bar chart, linear regression, and k-means clustering.

In [ ]:

In [ ]: