FINDING THE BEST LOCATION FOR A TUITION CENTRE IN MELBOURNE, AUSTRALIA

DANIEL LEVY

February 4, 2021

Contents

1 INTRODUCTION

1.1 BACKGROUND

The primary and secondary school tuition services industry in Australia is valued at over $1 billion AUD *(Australian Financial Review, 2020)*. It is an extremely competitive market, comprising both private tutors who visit a student's home, as well as various commercial tuition centres dispersed geographically around the country.

1.2 PROBLEM AND INTEREST

Any tuition centre looking to either break into the market, or to open a new location for their existing entity, will want to do so in an area where such services are in demand, and especially to avoid opening in an area that is already saturated with tuition services.

This project aims to use data to determine the best locations in which to open a new tuition centre in the city of Melbourne, Australia. To do so will require the location data about schools and student enrolments (obtained externally), as well as existing education business venues (obtained from Foursquare).

2 DATA: SOURCES, CLEANING AND WRANGLING

2.1 LIST OF MELBOURNE POSTCODES AND THEIR CORRESPONDING LONGITUDE AND LATITUDES

The first port of call is to generate a map of Melbourne's postcodes. A .CSV file containing all Australian postcodes with corresponding latitude and longitude was sourced (Corra, 2015).

In preprocessing the latitude and longitude data, it was first necessary to drop the extraneous "state", "type" and "dc" columns. Next, all rows of data not inside Metropolitan Melbourne were dropped. Australia Post defines the postcodes of 3000-3207 and 8000-8873 as Metropolitan Melbourne (Adairs, 2018).

Finally, it was noticed that several postcodes were duplicated due to multiple suburbs belonging to the same postcode. All duplicate postcodes were combined into one row per postcode, joining the suburb strings together and geometrically averaging the latitude and longitude values for each postcode from their constituent suburbs.

This gave rise to the beginning of a master data frame, currently detailing for each postcode: suburb names, latitude and longitude.

2.2 LIST OF SCHOOLS, LOCATIONS AND THEIR ENROLMENT NUMBERS

It will be important to know student enrolment numbers per postcode, and so a comprehensive list of school enrolment and corresponding location data was sourced (Victorian Government, 2020). All columns in the enrolment table were subsequently dropped, except for the school's name and its student enrolment "grand total". Due to partial enrolments, this column was then cast to integer values.

A dataframe of school names and location data was subsequently read in, and all columns except school name and postcode dropped. This dataframe was then merged as an outer join along the postcode index with the table containing postcode and location data. Where no enrolment data existed for a postcode it was filled as 0 (which assumed that the data from the Victorian government was complete, and so if a postcode contained no enrolments, then there were 0 enrolments in that postcode).

One postcode contained NaN values but 58 enrolments. This is because that school had a postal address registered at a post office box with its own postcode, but no attached suburb. It was decided the best way to address this would be to simply drop the row as the number of enrolments was very small compared to the average number of enrolments per postcode.

The end result of this collation was that our master dataframe now had a column for total school enrolments by postcode.

2.3 SOCIO-ECONOMIC DATA

Next a dataframe with Socio-economic Index for Area (SEIFA) scores listed by postcode was read in. Duplicate values were checked and a large number of them were found. Only unique SEIFA scores for each postcode were kept after removing duplicates.

NaN values were then checked and a large number of 48 were found. When inspecting these rows, it was discovered that all bar one coincided with zero enrolment postcodes. With both no enrolment data and no SES information it was decided that those rows held no value and were discarded.

The one postcode, 3062, that had enrolments but no SEIFA value had a substantial number of enrolments, 1924. It was decided to look up the suburb corresponding to postcode 3062 (Somerton) on a map and take the average of all suburbs that bordered it. These were: Craigieburn, Roxburgh Park, Meadow Heights, Coolaroo, Campbellfield, Lalor, Epping and Wollert. The SEIFA scores for these suburbs were looked up, averaged, and then inserted in place for Somerton (postcode 3062). All other rows with NaN values were then dropped.

The overall dataframe was thus updated to include the SEIFA score for each postcode.

2.4 TUITION BUSINESS VENUE DATA FROM FOURSQUARE API

The final source of data was Foursquare's venue data for education businesses in Melbourne. A function was built to scrape the FourSquare API for all venues within a 2 km radius of each postcode that matched the search terms "VCE" (Victorian Certificate of Education), "tuition" and "tutoring". This was done after inspecting a number of search queries and determining which ones most accurately returned lists of school tuition centres (some search terms returned things

like Adult Education Centres, tutorial rooms within universities etc. which are not school tuition centres in the desired target market).

Due to limitations of the Foursquare API, these dataframes were saved to external CSVs within the project structure to be called again at will, rather than relying on the API call.

The next step was to combine the results of all those searches together and drop any duplicates. The number of unique tuition centre venues was 33 spread out across Melbourne. The latitude and longitude of each postcode in the dataframe was matched against the latitude and longitude of each venue to measure whether a given venue was in close proximity to that postcode. A value of 5 km was considered to be a "close" distance of the venue to the postcode. Each venue that satisifed this criterion was added to the postcode's total number of tuition/education facilities in close proximity.

Thus, the final column to complete the main dataframe was a count of the number of tuition centres in close proximity to each postcode.

2.5 COLLATED DATA

From four distinct sources of raw data, a dataframe was wrangled that featured a list of every postcode in Metropolitan Melbourne along with the corresponding: suburbs, latitude, longitude, student enrolments, socio-economic index score and number of school tuition venues in a 5 km radius.

A screenshot of the dataframe containing all the pre-processed data is as follows:

```
df_melbdata.head(50)
```

9]:

	postcode	suburb	lat	lon	enrolments	ses_score	tuition_venues
0	3000	MELBOURNE	-37.814563	144.970267	0.0	1030.0	11
2	3002	EAST MELBOURNE	-37.816640	144.987811	0.0	1126.0	11
3	3003	WEST MELBOURNE	-37.806255	144.941123	0.0	1088.0	10
4	3004	MELBOURNE	-37.837324	144.976335	1044.0	1116.0	11
5	3005	WORLD TRADE CENTRE	-37.822262	144.954856	0.0	1104.0	10
6	3006	SOUTHBANK	-37.823258	144.965926	861.0	1110.0	11
7	3008	DOCKLANDS	-37.814719	144.948039	0.0	1115.0	10
9	3011	FOOTSCRAY, SEDDON, SEDDON WEST	-37.801199	144.887090	501.0	980.0	2
10	3012	BROOKLYN, KINGSVILLE, KINGSVILLE WEST, MAIDSTO...	-37.800197	144.867860	1952.0	973.0	3
11	3013	YARRAVILLE, YARRAVILLE WEST	-37.817099	144.886678	2217.0	1054.0	2
12	3015	NEWPORT, SOUTH KINGSVILLE, SPOTSWOOD	-37.835258	144.879655	1413.0	1058.0	0
13	3016	WILLIAMSTOWN, WILLIAMSTOWN NORTH	-37.857292	144.892369	2872.0	1073.0	0
14	3018	ALTONA, SEAHOLME	-37.868634	144.836948	958.0	1009.0	0
15	3019	BRAYBROOK, BRAYBROOK NORTH, ROBINSON	-37.819588	144.930285	1844.0	839.0	10
16	3020	ALBION, GLENGALA, SUNSHINE NORTH, SUNSHINE WEST	-37.783259	144.819402	3355.0	888.0	2

Figure 1: The dataframe "df_melbdata", showing all the relevant data for each postcode.

References

[1] Theo Chapman, *Australian Financial Review*: Tutoring in Australia is a billion-dollar industry https://www.afr.com/policy/health-and-education/tutoring-in-australia-is-a-billion-dollar-industry-20201029-p569mp

[2] Australian Bureau of Statistics: SEIFA Data https://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa: :text=Socio%2DEconomic%20Indexes%20for%20Areas%20(SEIFA)

[3] https://www.corra.com.au/australian-postcode-location-data/

[4] https://www.adairs.com.au/globalassets/standarddelivery_christmas_cutoffs.pdf

[5] Victorian Government Education Data https://www.education.vic.gov.au/about/department/Pages/factsandfigures.aspx