Data Analysis of Candidate Areas of New Preschools in Singapore

For IBM Data Science Capstone Project

Author: Huang yuzhu

Date: 11 Aug. 2019

1. Introduction

Living in a city country like Singapore, looking for an accessible and quality preschool is in the top list of almost all mothers. Many have experienced long waiting list and difficulties on transport during the preschool hunting process. Research done by Economist Intelligence Unit (EIU) in year 2012 also ranked Singapore 29th out of 45 countries in early childhood education, which showed that there were still many opportunities for more quality preschools to be set up in Singapore. Companies in the early childhood education sector may be interested in investing on new centres in Singapore. To start a new centre, as any other business, companies will always look for a good location.

In this project I applied the data science techniques taught in the previous courses to help the companies find a good location for a new preschool.

Considering the nature of the preschool, I defined a good location for a new preschool as some place close to one MRT (Mass Rapid Transit) station fulfilling below criteria:

- close to as many residences as possible,
- close to as many bus stations as possible
- not already crowded with preschools

2. Data

Based on the definition of the stated problem, factors that will influence the analysis are:

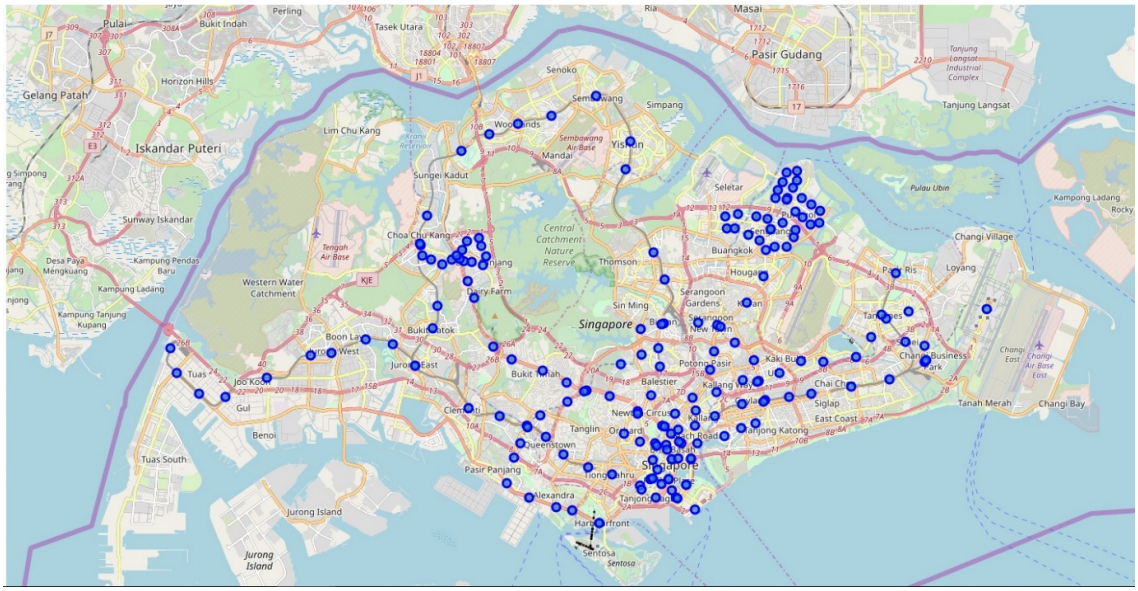
- number of existing preschools in the neighbourhood of a particular MRT station
- number of residences in the neighbourhood of a particular MRT station
- number of bus stops in the neighbourhood of a particular MRT station

Following data sources will be needed to extract/generate the required information:

- list of Singapore MRT stations with latitude and longitude info will be scraped from a data file shared in github
- number of preschools, bus stations or residences in the neighbourhood of every MRT station will be obtained using Foursquare API

3. Methodology

To start the analysis, I needed the list of Singapore MRT stations currently in operation. To do so, I first scraped the location info of the MRT stations in Singapore from a data file shared in github, then removed the MRT stations planned but not in operation yet. I used python folium library to visualize the distribution of the valid MRT stations in Singapore:



With the MRT station data ready, I could then compare and narrow down the interest areas to the areas of the MRT stations with the least number of existing preschools and the greatest number of residences and/or bus stations.

I utilized the Foursquare API to explore the vicinity of each valid MRT station and computed the number of preschools, residences and bus stations nearby. I limited the radius to 500 metres around each MRT station to make the preschool accessible by walking. Here's the processed data set:

	Neighbourhood	No. of Preschool	No. of Bus Station	No. of Residence
0	HARBOURFRONT MRT STATION	0.0	82.0	10.0
1	PROMENADE MRT STATION	0.0	54.0	2.0
2	MACPHERSON MRT STATION	0.0	35.0	44.0
3	NICOLL HIGHWAY MRT STATION	0.0	32.0	16.0
4	OUTRAM PARK MRT STATION	0.0	27.0	33.0
5	CHINATOWN MRT STATION	0.0	27.0	28.0
6	CLARKE QUAY MRT STATION	0.0	24.0	24.0
7	BAYFRONT MRT STATION	0.0	24.0	16.0
8	UBI MRT STATION	0.0	23.0	22.0
9	TAMPINES MRT STATION	0.0	22.0	17.0
10	BOON LAY MRT STATION	0.0	21.0	26.0
11	BUKIT BATOK MRT STATION	0.0	21.0	24.0
12	MARINA BAY MRT STATION	0.0	21.0	8.0
13	PHOENIX LRT STATION	0.0	20.0	24.0
14	CHOA CHU KANG LRT STATION	0.0	19.0	22.0
15	DHOBY GHAUT MRT STATION	0.0	18.0	24.0
16	JOO KOON MRT STATION	0.0	18.0	0.0
17	PIONEER MRT STATION	0.0	16.0	32.0
18	CHOA CHU KANG MRT STATION	0.0	16.0	25.0
19	TELOK BLANGAH MRT STATION	0.0	16.0	24.0
20	SERANGOON MRT STATION	0.0	16.0	13.0
21	UPPER CHANGI MRT STATION	0.0	15.0	27.0
22	SENGKANG LRT STATION	0.0	15.0	16.0
23	SENGKANG MRT STATION	0.0	15.0	16.0
24	LAKESIDE MRT STATION	0.0	14.0	36.0
25	QUEENSTOWN MRT STATION	0.0	14.0	29.0
26	KAKI BUKIT MRT STATION	0.0	14.0	28.0
27	TANAH MERAH MRT STATION	0.0	14.0	28.0
28	MATTAR MRT STATION	0.0	14.0	24.0
29	HAW PAR VILLA MRT STATION	0.0	14.0	15.0
...	...	...	...	...
132	LITTLE INDIA MRT STATION	2.0	28.0	32.0
133	BENDEMEER MRT STATION	2.0	21.0	23.0
134	WOODLANDS MRT STATION	2.0	19.0	24.0
135	NOVENA MRT STATION	2.0	16.0	29.0
136	BARTLEY MRT STATION	2.0	15.0	18.0
137	MOUNTBATTEN MRT STATION	2.0	14.0	26.0
138	CASHEW MRT STATION	2.0	13.0	12.0
139	NEWTON MRT STATION	2.0	12.0	20.0
140	LORONG CHUAN MRT STATION	2.0	11.0	32.0
141	BEDOK NORTH MRT STATION	2.0	11.0	31.0
142	SIXTH AVENUE MRT STATION	2.0	11.0	7.0
143	DAKOTA MRT STATION	2.0	10.0	32.0
144	HOLLAND VILLAGE MRT STATION	2.0	8.0	36.0
145	SUMANG LRT STATION	2.0	6.0	19.0
146	PUNGGOL MRT STATION	2.0	4.0	10.0
147	OASIS LRT STATION	2.0	1.0	8.0
148	SOO TECK LRT STATION	2.0	1.0	4.0
149	PAYA LEBAR MRT STATION	3.0	26.0	28.0
150	BOTANIC GARDENS MRT STATION	3.0	25.0	9.0
151	ONE-NORTH MRT STATION	3.0	22.0	4.0
152	KEMBANGAN MRT STATION	3.0	15.0	30.0
153	KOVAN MRT STATION	3.0	14.0	25.0
154	DAMAI LRT STATION	3.0	10.0	26.0
155	COMPASSVALE LRT STATION	3.0	9.0	20.0
156	SAM KEE LRT STATION	3.0	5.0	15.0
157	JURONG EAST MRT STATION	4.0	68.0	22.0
158	BEAUTY WORLD MRT STATION	4.0	12.0	22.0
159	GEYLANG BAHRU MRT STATION	4.0	12.0	22.0
160	TAMPINES WEST MRT STATION	4.0	8.0	30.0
161	POTONG PASIR MRT STATION	5.0	18.0	23.0

162 rows × 4 columns

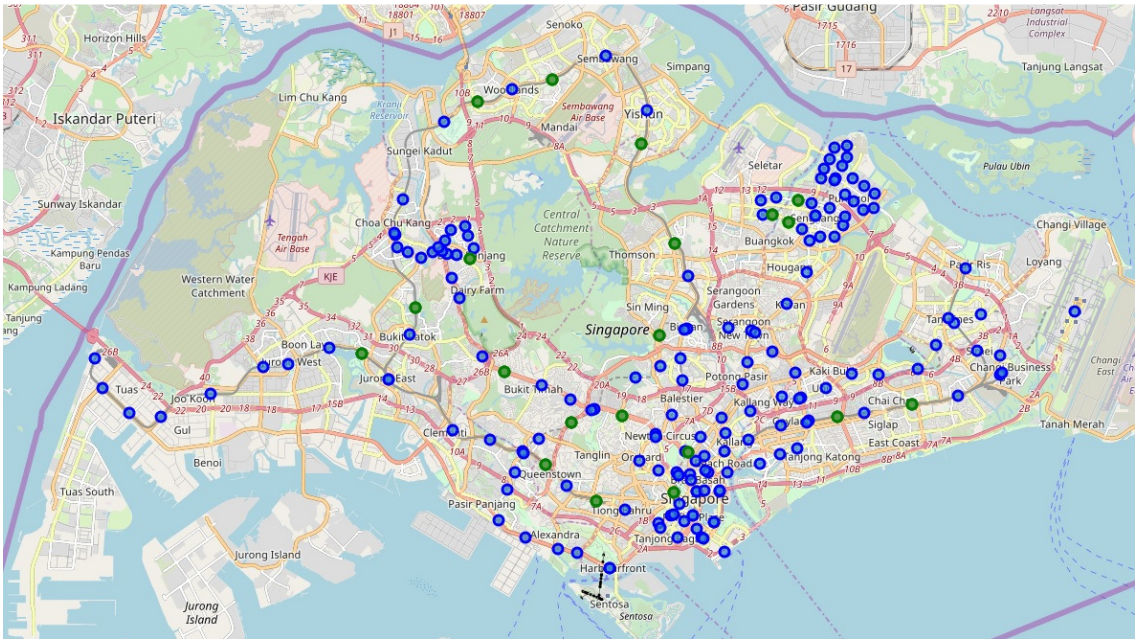
When I looked at the processed data set, I found many MRT stations without any preschools nearby. Zooming in at the MRT stations, I realized that it could be due to the business nature of the districts the MRT stations locate at. These areas were not considered interest areas as there were not enough residences to ensure sufficient enrolment. Therefore, I focused on the areas with only one existing preschool instead, which indicated that the area was suitable for business while less crowded. Here's the list of the filtered areas:

	Neighbourhood	No. of Preschool	No. of Bus Station	No. of Residence
0	CHANGI AIRPORT MRT STATION	1.0	29.0	5.0
1	KENT RIDGE MRT STATION	1.0	29.0	0.0
2	TELOK AYER MRT STATION	1.0	24.0	20.0
3	BUGIS MRT STATION	1.0	21.0	9.0
4	DOWNTOWN MRT STATION	1.0	20.0	17.0
5	FORT CANNING MRT STATION	1.0	19.0	26.0
6	ROCHOR MRT STATION	1.0	18.0	27.0
7	MARYMOUNT MRT STATION	1.0	17.0	26.0
8	YIO CHU KANG MRT STATION	1.0	17.0	26.0
9	BEDOK MRT STATION	1.0	17.0	23.0
10	TAI SENG MRT STATION	1.0	16.0	16.0
11	STEVENS MRT STATION	1.0	15.0	32.0
12	CHINESE GARDEN MRT STATION	1.0	15.0	30.0
13	REDHILL MRT STATION	1.0	14.0	28.0
14	EUNOS MRT STATION	1.0	14.0	25.0
15	CALDECOTT MRT STATION	1.0	14.0	18.0
16	CLEMENTI MRT STATION	1.0	14.0	13.0
17	LABRADOR PARK MRT STATION	1.0	14.0	3.0
18	PENDING LRT STATION	1.0	13.0	30.0
19	COMMONWEALTH MRT STATION	1.0	12.0	28.0
20	FARMWAY LRT STATION	1.0	12.0	23.0
21	MARSILING MRT STATION	1.0	11.0	32.0
22	BUKIT GOMBAK MRT STATION	1.0	11.0	31.0
23	KHATIB MRT STATION	1.0	11.0	31.0
24	ORCHARD MRT STATION	1.0	11.0	8.0
25	KING ALBERT PARK MRT STATION	1.0	10.0	25.0
26	TONGKANG LRT STATION	1.0	10.0	24.0
27	ADMIRALTY MRT STATION	1.0	9.0	38.0
28	FARRER ROAD MRT STATION	1.0	7.0	21.0
29	BENCOOLEN MRT STATION	1.0	7.0	9.0
30	THANGGAM LRT STATION	1.0	6.0	17.0
31	TANJONG PAGAR MRT STATION	1.0	5.0	9.0
32	LAYAR LRT STATION	1.0	4.0	25.0
33	RUMBIA LRT STATION	1.0	4.0	16.0
34	FERNVALE LRT STATION	1.0	4.0	10.0
35	LAVENDER MRT STATION	1.0	4.0	5.0
36	PUNGGOL LRT STATION	1.0	3.0	8.0
37	KADALOOR LRT STATION	1.0	2.0	9.0
38	TECK LEE LRT STATION	1.0	0.0	2.0

The list was sorted by the number of residences and bus stations in descending order respectively. To further narrow down to the most promising areas, I limited the areas to those having more than 20 residences. The final list looked like this:

	Neighbourhood	No. of Preschool	No. of Bus Station	No. of Residence
0	FORT CANNING MRT STATION	1.0	19.0	26.0
1	ROCHOR MRT STATION	1.0	18.0	27.0
2	MARYMOUNT MRT STATION	1.0	17.0	26.0
3	YIO CHU KANG MRT STATION	1.0	17.0	26.0
4	BEDOK MRT STATION	1.0	17.0	23.0
5	STEVENS MRT STATION	1.0	15.0	32.0
6	CHINESE GARDEN MRT STATION	1.0	15.0	30.0
7	REDHILL MRT STATION	1.0	14.0	28.0
8	EUNOS MRT STATION	1.0	14.0	25.0
9	PENDING LRT STATION	1.0	13.0	30.0
10	COMMONWEALTH MRT STATION	1.0	12.0	28.0
11	FARMWAY LRT STATION	1.0	12.0	23.0
12	MARSILING MRT STATION	1.0	11.0	32.0
13	BUKIT GOMBAK MRT STATION	1.0	11.0	31.0
14	KHATIB MRT STATION	1.0	11.0	31.0
15	KING ALBERT PARK MRT STATION	1.0	10.0	25.0
16	TONGKANG LRT STATION	1.0	10.0	24.0
17	ADMIRALTY MRT STATION	1.0	9.0	38.0
18	FARRER ROAD MRT STATION	1.0	7.0	21.0
19	LAYAR LRT STATION	1.0	4.0	25.0

I used the folium library again to visualize the distribution of the final candidate areas in the previous map of Singapore MRT stations, resembled by the green dots:



4. Results

Finally, I manually labelled the candidate areas with the geographical region it locates at, namely North, East, West or Central.

STN_NAME	No. of Preschool	No. of Residence	No. of Bus Station	Region
STEVENS MRT STATION	1	32	15	Central
REDHILL MRT STATION	1	28	14	Central
COMMONWEALTH MRT STATION	1	28	12	Central
ROCHOR MRT STATION	1	27	18	Central
FORT CANNING MRT STATION	1	26	19	Central
MARYMOUNT MRT STATION	1	26	17	Central
KING ALBERT PARK MRT STATION	1	25	10	Central
FARRER ROAD MRT STATION	1	21	7	Central
EUNOS MRT STATION	1	25	14	East
BEDOK MRT STATION	1	23	17	East
ADMIRALTY MRT STATION	1	38	9	North
MARSILING MRT STATION	1	32	11	North
KHATIB MRT STATION	1	31	11	North
YIO CHU KANG MRT STATION	1	26	17	North
LAYAR LRT STATION	1	25	4	North
TONGKANG LRT STATION	1	24	10	North
FARMWAY LRT STATION	1	23	12	North
BUKIT GOMBAK MRT STATION	1	31	11	West
CHINESE GARDEN MRT STATION	1	30	15	West
PENDING LRT STATION	1	30	13	West

5. Discussion

As the previous analysis showed, the potential good locations for operating a new preschool in Singapore mainly located in the Central and North region, with a few in the West and East.

Combining this observation with the population density of Singapore, I could conclude that **North is the most suitable region to set up a new preschool and the potential investors could focus on the 7 areas in the North region identified in the previous process**. Reasons being:

- Population is the least dense in the east of Singapore, and there were just a few candidate areas, so it could be safely removed from the list.
- Population is at appropriate density in the west of Singapore, but the options were limited with just a few candidate areas, so it could be removed as well.
- Population is at appropriate density in the central Singapore and there were adequate options there. However, considering the highest cost of living in the central Singapore, the profit margin may not be ideal. Hence central region was removed too.
- Population is at appropriate density and the cost of living is the lowest in the north of Singapore. With adequate candidate areas identified, it could be considered a good candidate region to set up a new preschool.

6. Conclusion

In this study, I explored the venue data around each MRT station in Singapore using the data science techniques learnt in previous courses. I computed the number of preschools, residences and bus stations and visualized the candidate areas to identify a good location for setting up a new preschool. The identified candidate areas can ensure the new preschool has:

- sufficient enrolment
- adequate transport options
- least number of competitors
- possible highest profit margin

This study can be further refined to recommend the type of the preschool, e.g. infant care, child care or kindergarten, if incorporating more relevant data for analysis.

References:

https://eiuperspectives.economist.com/talent-education/starting-well

https://github.com/hxchua/datadoubleconfirm/blob/master/datasets/mrtsg.csv

https://mothership.sg/2016/10/nifty-map-shows-which-parts-of-spore-has-densest-population-guess-which-is-no-1/