Applied Data Science Capstone
Coursera IBM Specialisation

A Tale of Two Districts

Ivan Ong

23rd May 2020

CONTENTS

1. INTRODUCTION

1.1 Business Case Scenario

I recently got engaged with my lovely girlfriend of 3 years and our wedding is scheduled to be in the last months of the year. We often have debates on our post-marriage plans such as whose house should we reside in for the time being until our own apartment is ready. Both **Kembangan** and **Seng Kang** are fairly mature estates and each have their own strengths and weaknesses. To make our comparison more objective, I will measure the attractiveness of both districts based on 3 questions:

a. Which district has better **schools** in its vicinity? (Education planning)
b. What **amenities** are available in each district?
c. Which district has more favourable **popular spots**?

1.2 Key Considerations

a. Walking distance to public transports
b. Local attractions and venues available

1.3 Other Target Audience

a. Other couples comparing similar locations
b. Property investors targeting the district prospects
c. Inspiring data analysts to learn from this analysis

2. DATA

2.1 Singapore Dataset / Map

- FourSquare API Explore Call

2.2 Data Tools

- FourSquare API will be used to collect coordinates and locations
- The visual map will be created using Folium
- Pandas will be used to transform and load data onto data frames

3. METHODOLOGY

3.1 Foursquare API + Clustering

Using the FourSquare API, we will explore a 10km radius around the two districts where the locations will be grouped into clusters

3.2 Using Folium, the neighbourhood boundary data of the 2 locations would be plotted onto a choropleth map, graduated by crime rate statistics.

4. RESULTS
4.1 Dataframe Load / API Call

Schools near “Her Place”

	id	name	categories	referralId	hasPerk	location.address	location.lat	location.lng
0	4d5c98a26f6d6ea8c60567ec	North Vista Secondary School Lecture Theatre	[]	v-1590240511	False	North Vista Secondary School	1.381524	103.899081
1	4bcee055b6c49c746abe9791	North Spring Primary School	[{'id': '4bf58dd8d48988d13b941735', 'name': 'S...	v-1590240511	False	1 Rivervale St	1.387882	103.904102
2	4cda52bc5aeda1cd1e47b611	North Vista Secondary School	[{'id': '4bf58dd8d48988d13d941735', 'name': 'H...	v-1590240511	False	11 Rivervale Link	1.382298	103.898478
3	4cc9212dbfe1f04d56360c75	North Vista Primary School	[{'id': '4f4533804b9074f6e4fb0105', 'name': 'E...	v-1590240511	False	20 Compassvale Link	1.383070	103.895602
4	4e0d5a2e18a8bf9784c248fa	Serangoon Secondary School (Learning Hub)	[{'id': '4bf58dd8d48988d1a7941735', 'name': 'C...	v-1590240511	False	11 Upper Serangoon View	1.374992	103.901579

Schools near “My Place”

	id	name	categories	referralId	hasPerk	location.address	location.lat	location.lng
	click to scroll output; double click to hide							
0	4bc584f269369521d22a8488	St. Stephen's School	[{'id': '4bf58dd8d48988d13b941735', 'name': 'S...	v-1590240510	False	20 Siglap View	1.319152	103.917914
1	4cad0636a6e08cfadf70b094	Opera Estate Primary School	[{'id': '4f4533804b9074f6e4fb0105', 'name': 'E...	v-1590240510	False	48 Fidelio St	1.319669	103.924077
2	4d5507fd48ea6ea81e6bd0a3	PCF School	[{'id': '4f4533814b9074f6e4fb0107', 'name': 'N...	v-1590240510	False	59 Chai Chee Rd	1.324769	103.920616
3	4c75d9c5d8948cfaf31369da	MINDS Towner Gardens School	[{'id': '4bf58dd8d48988d199941735', 'name': 'C...	v-1590240510	False	1B Lengkong Lima	1.326653	103.913678
4	4d2e77f54377224bc10f0f38	Rosemount International School/ Rosemount	[{'id': '4f4533814b9074f6e4fb0107', 'name': 'N...	v-1590240510	False	25 Ettrick Terrace	1.312413	103.922502

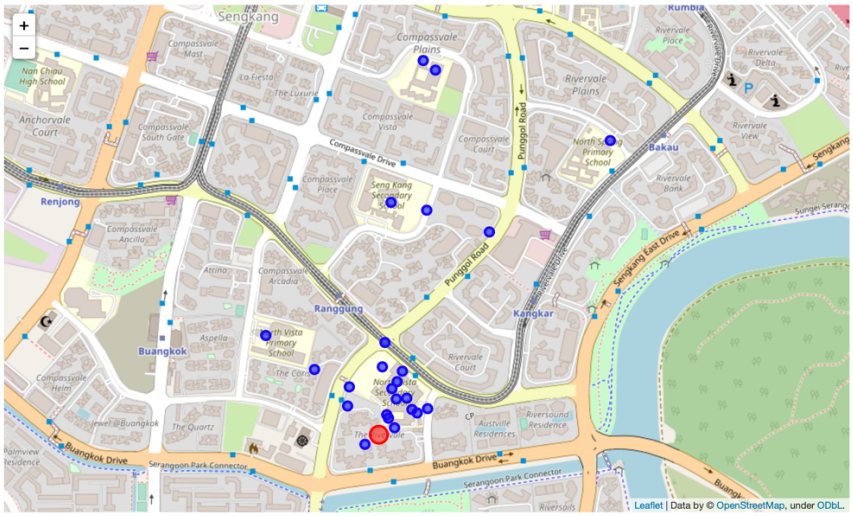
4.2 FourSquare API and Clustering

After using FourSquare API to make an Explore call to gather the nearest schools for each of our homes.

A preliminary Folium plot was then created for both locations.

“My place”

“Her place”



6. CONCLUSION

Based on the above clusters above, I am able to confidently share this information with my fiancée:

1. Nearest schools within 1km of both places
2. Nearest popular places within the vicinity of our places

7. Info Gap and Further Analysis

In order to take this analysis a step further, additional information could be gathered in these other areas:

- Proximity to each location in terms of distance (km)
- How congested are these areas. (E.g. No. of residents per sq meter)
- Availability of late night amenities when working late