Applied Data Science Capstone:

A Comparison between Amenities surrounding the Top 5 Most Expensive and Top 5 Least Expensive Areas in Greater Toronto Area

By

Yun En Tan

2019-07-21

Table of Contents

1.0 Introduction

The housing cost in Greater Toronto Area (GTA) has had a significant increase in 2017. Although in the past two years, the federal and provincial government have come up with some policies to prevent further increase in housing cost, the current price of housing in GTA is still significantly higher than pre-2017 era. This has caused young millennials to be priced out of the housing market. In this report, we will be looking into the surrounding areas of GTA's neighbourhood which are among the top 5 most expensive and top 5 least expensive based on Toronto Real Estate Board (TREB)'s Q1 2019 data. With this comparison, we hope that young millennials will be able to find a place in GTA which are relatively affordable but having the same neighbourhood's amenities comparing to the expensive neighbourhood in GTA.

2.0 Data

First and foremost, we used the dataset from Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) which we saw in Week 3. This dataset contains the postal codes in GTA area.

Next, we also used the geospatial coordinates (http://cocl.us/Geospatial_data) which we saw in Week 3, which contains the latitudes and longitudes based on postal codes of GTA area. These latitudes and longitudes will then be used to explore venues using Foursquare Place API.

Next, we used the Toronto geojson file from https://github.com/ag2816/Visualizations/blob/master/data/Toronto2.geojson. This geojson file contains the boundaries of latitudes and longitudes based on postal codes in GTA area.

The next dataset used will be a mapping between Toronto Real Estate Board (TREB)'s district to postal code. This mapping is performed manually and as best as possible since TREB district isn't exactly based on postal codes. This dataset is needed to map the housing prices to postal code.

Lastly, the Q1 2019's quarterly average housing price per TREB district is retrieved from http://www.trebhome.com/index.php/market-news/community-reports.

3.0 Methodology

We first begin by plotting a heat map using python folium library to show the housing prices in GTA.

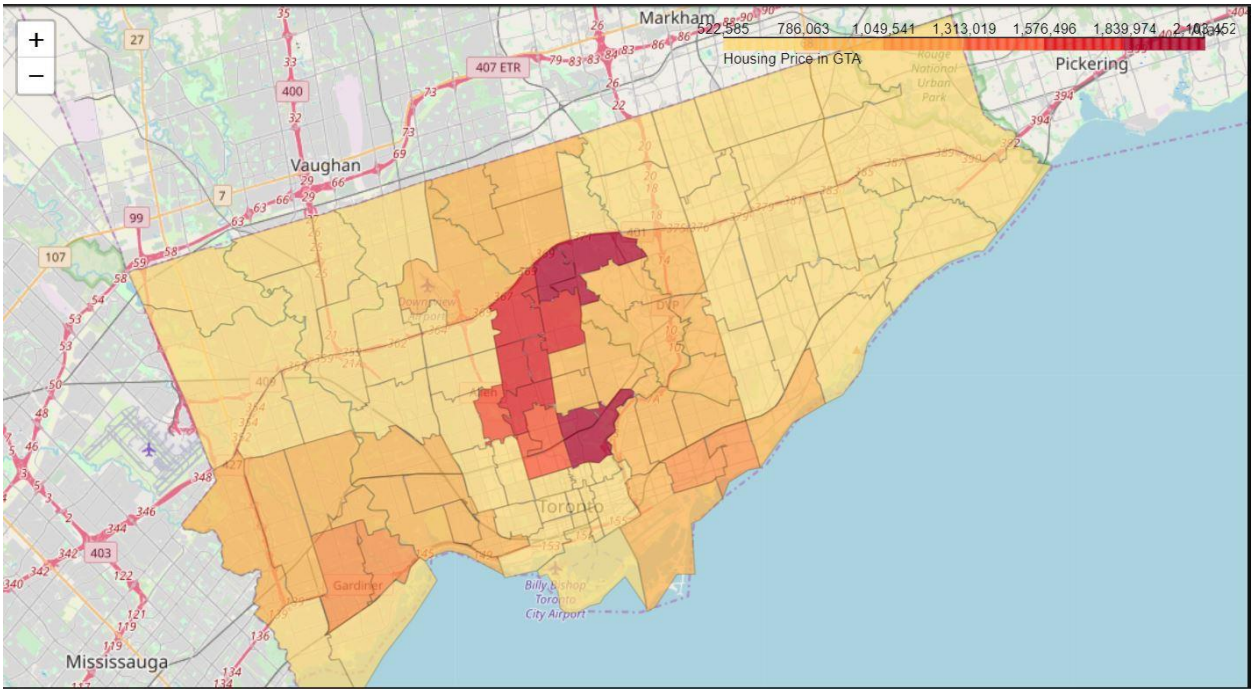

Figure 1: Heat Map of Housing Price in GTA

We observed that the center regions of GTA is the most expensive region in GTA, whereas the northwest and east of GTA are generally cheaper. This plot is generated using the mapping between TREB district and postal codes in GTA, with the geojson file.

Next, we determined the top 5 most expensive and top 5 least expensive TREB districts and their corresponding mapping to postal codes.

	TrebDistrict	AveragePrice
1	C02	1471235
2	C03	1495283
3	C04	1592683
7	C09	2075220
10	C12	2087953

	TrebDistrict	AveragePrice
34	W10	538084
24	E11	557377
22	E09	560882
20	E07	594407
29	W05	602487

Figure 2: Left diagram is the top 5 most expensive TREB districts, Right is the top 5 least expensive TREB districts

	TrebDistrict	PostalCode
0	C02	[M5R]
1	C03	[M6C, M4V]
2	C04	[M5M, M4N, M5N, M5P, M4R]
3	C09	[M4T, M4W]
4	C12	[M2L, M2P]
5	E07	[M1H, M1S, M1V]
6	E09	[M1G]
7	E11	[M1B, M1X]
8	W05	[M3J, M3K, M3L, M9L, M3M, M3N]
9	W10	[M9V, M9W]

Figure 3: Postal Code Mapping of the Top 5 Most Expensive and Top 5 Least Expensive TREB Districts

Now, with the postal codes from Figure 3, and using the geospatial coordinates data, we queried Foursquare API for nearby venues between 2km with a limit of 1000 venues. We then merged them back into TREB district and see the Top 5 most common venues in all 10 of the TREB districts.

	TrebDistrict	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	C02	Café	Sandwich Place	Spa	Asian Restaurant	Indian Restaurant
1	C03	Café	Park	Sandwich Place	Coffee Shop	Italian Restaurant
2	C04	Café	Park	Coffee Shop	Italian Restaurant	Indian Restaurant
3	C09	Café	Coffee Shop	Park	Farmers Market	Hotel
4	C12	Park	Café	Coffee Shop	Italian Restaurant	Bakery
5	E07	Caribbean Restaurant	Indian Restaurant	Chinese Restaurant	Coffee Shop	Park
6	E09	Park	Indian Restaurant	Caribbean Restaurant	Coffee Shop	Zoo Exhibit
7	E11	Zoo Exhibit	Park	Caribbean Restaurant	Indian Restaurant	Coffee Shop
8	W05	Café	Coffee Shop	Italian Restaurant	Clothing Store	Liquor Store
9	W10	Hotel	Steakhouse	Bakery	Indian Restaurant	Diner

Figure 4: Top 5 Most Common Venues in the Top 5 Most Expensive and Top 5 Least Expensive TREB districts

Now, we run k-means algorithm on the postal codes to cluster the postal codes into 3 clusters based on the frequency of the venue categories, and finally, we merged it back into TREB district.

4.0 Results

In this section, we will begin by showing the result of the k-means algorithm.

	Cluster Labels	TrebDistrict
0	0	[C12, W05, W10]
1	1	[C02, C03, C04, C09]
2	2	[E07, E09, E11]

Figure 5: Clustering Result

We see that the most expensive TREB districts (C02, C03, C04, C09) are all classified in Cluster 1, which is not that surprising if we refer to Figure 4 where we see that the 1st most common venues for all 4 of this TREB districts are all Café. E07, E09, E11 (i.e., the east end of GTAs) are classified together, and referring to Figure 4, we can see that these districts have high frequency of Indian restaurants and Caribbean restaurants. The leftover districts (C12, W05, W10) are classified together.

5.0 Discussion

From Figure 5, we would be prompted to say that W05 and W10 (the relatively cheaper TREB districts) has identical amenities in the surrounding with C12 (one of the most expensive TREB district), however, if we refer to Figure 4, we do see fairly high similarities in venues among W05 and C12, but W10's surroundings seem completely different than W05 and C12.

Based on the result, among the Top 5 most expensive and Top 5 least expensive TREB districts, only W05 has some similarities in surrounding area with C12.

6.0 Conclusion

In conclusion, for young millennials who are looking for surroundings with similar amenities but with a more affordable housing price, we would suggest TREB district W05.

7.0 Future directions

Recall that we used dataset containing the aggregated average housing price in each TREB district, this might not be the best idea since some districts could contain more houses and other could contain only condominiums. For a better comparison, we should further decompose and look at a specific housing type or instead of using the absolute housing price, use the relative

increment of housing price since 2016. A better result would most likely be obtained with the suggested changes in dataset.