Apartment Prices Index & Venues Data Analysis of Toronto

Jun Xiong

2021/05/03

Introduction

Toronto is the largest city in Canada which has 2.93 million population and covers an area of 630 square kilometers. As one of the most multicultural urban areas in the world. Each year tens of thousands of newcomers from around the globe move to Toronto and settle down. The city is divided into 35 districts by the Toronto Real Estate Board (TREB).

As a newcomer, we want to find a convenient area to live with relatively lower cost. Therefore, we can create a map and information chart where the real estate index is placed on Toronto and each district is clustered according to the venue which represent the convenient level.

Data Description

To consider the problem we can list the data as below:

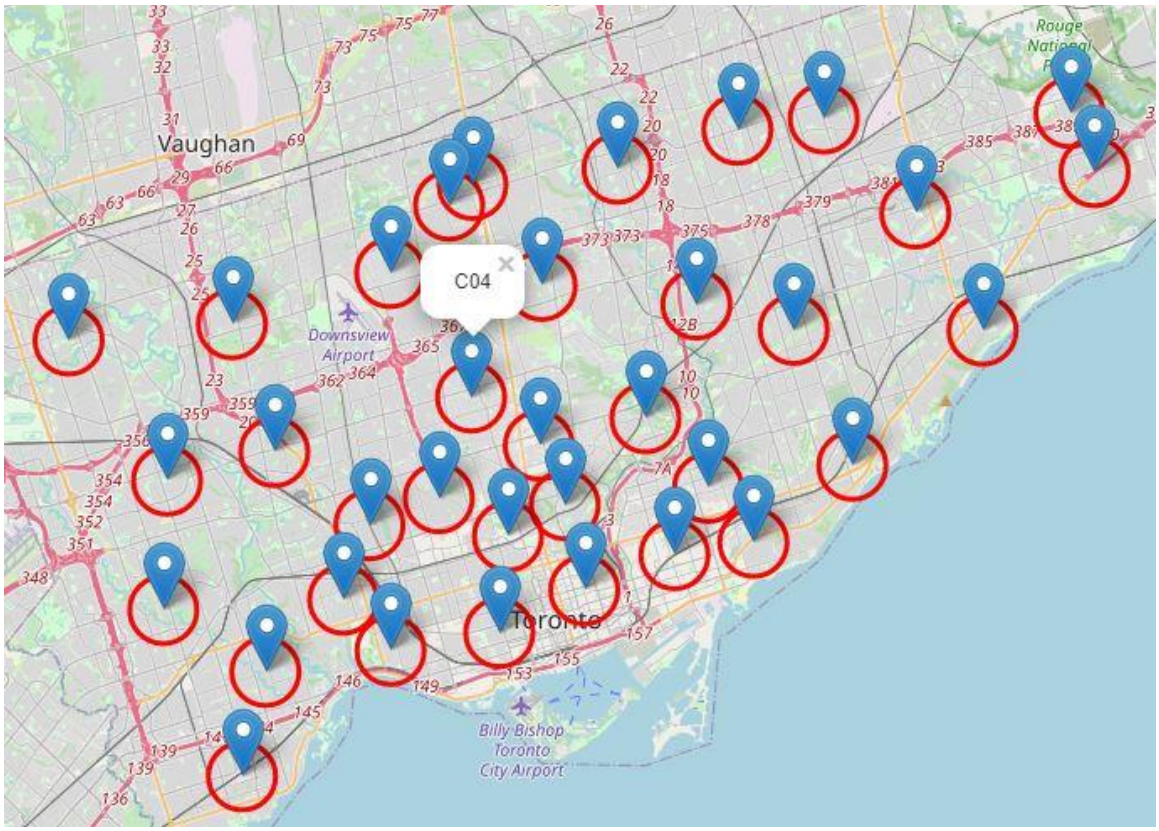
1. I found the list of Toronto Multiple listing service (MLS) districts and neighborhoods within each district on Wikipedia.
2. I used Google Maps API geocoding to find the coordinate of each neighborhood and calculated the coordinates of districts by calculating the central point of all the neighborhoods within the district.
3. I used Foursquare API to get the venues of districts.
4. I found the price index of districts on TREB.

Methodology

To start with the data, I used python **pandas** library to collect the data from **Wikipedia**, and use **Google Maps API** to find the coordinates:

	District Number	Neighbourhoods Included	lat	lng
0	C01	Downtown, Harbourfront, Little Italy, Little P...	43.649651	-79.407145
1	C02	The Annex, Yorkville, South Hill, Summerhill, ...	43.678110	-79.403615
2	C03	Forest Hill South, Oakwood–Vaughan, Humewood–C...	43.689419	-79.431893
3	C04	Bedford Park, Lawrence Manor, North Toronto, F...	43.719138	-79.418671
4	C06	North York, Clanton Park, Bathurst Manor	43.755783	-79.451779

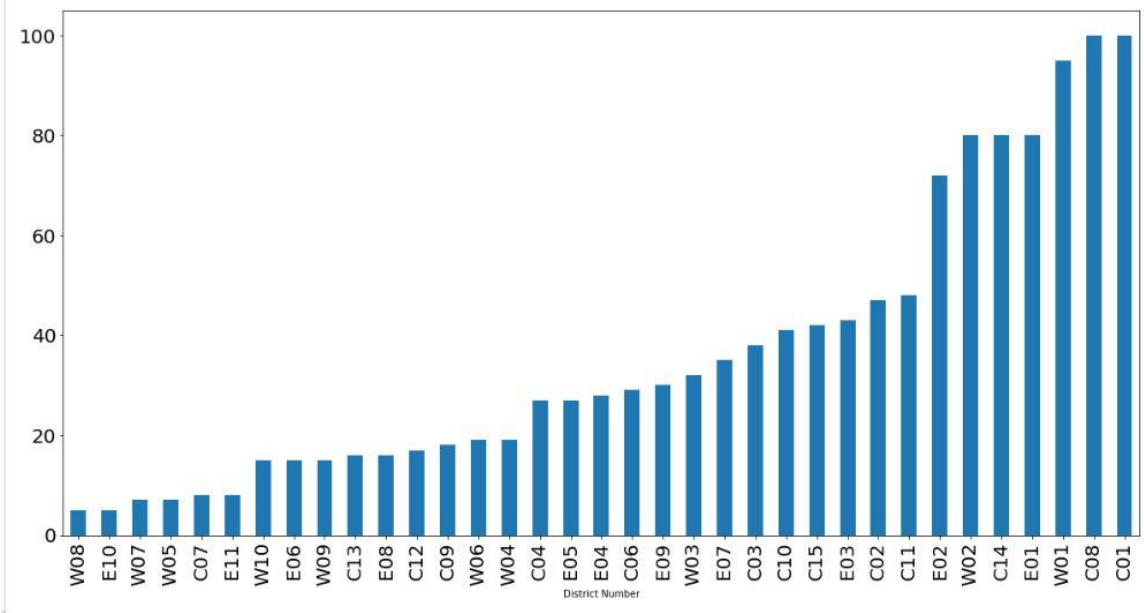
I used python **folium** library to visualize these districts on the map. The district number pops up, and the red circle represent the 1000-meter radius from the coordinates which we will input later to find the venues:



I utilized the **Foursquare API** to explore the districts and segment them. I input the limit as **100 venue** and the radius **1000 meter** for each district:

	District Number	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	C01	43.649651	-79.407145	#Hashtag Gallery	43.651830	-79.408103	Art Gallery
1	C01	43.649651	-79.407145	North of Brooklyn Pizzeria	43.646945	-79.406561	Pizza Place
2	C01	43.649651	-79.407145	Cumbrae's	43.646248	-79.408922	Butcher
3	C01	43.649651	-79.407145	Sud Forno	43.646208	-79.408986	Bakery
4	C01	43.649651	-79.407145	Dufflet Pastries	43.646306	-79.408456	Dessert Shop

We noticed that the venues in Toronto districts are highly schismatic, 2 districts have 100 venues while 6 districts have less than 10 venues:



This is since some districts are in urban area and close with each other, but this result is unbiased since the radius is the same which indicates that some districts are more convenient than others.

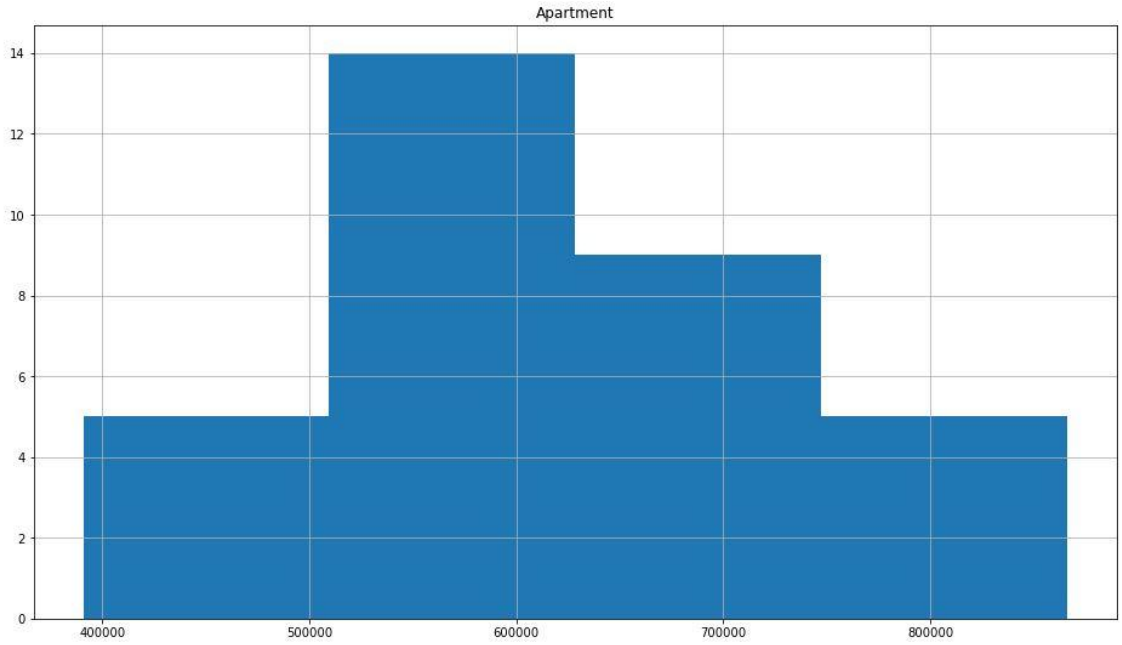
Eventually we found 1264 venues in 236 unique categories in our result, and I summarized the top 5 common venues for each district:

	District Number	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	C01	Café	Bakery	Bar	Cocktail Bar	Pizza Place
1	C02	Café	Italian Restaurant	Sandwich Place	Coffee Shop	History Museum
2	C03	Italian Restaurant	Restaurant	Indian Restaurant	Convenience Store	Grocery Store
3	C04	Bank	Pharmacy	Skating Rink	Coffee Shop	Department Store
4	C06	Coffee Shop	Park	Pizza Place	Grocery Store	Mobile Phone Shop

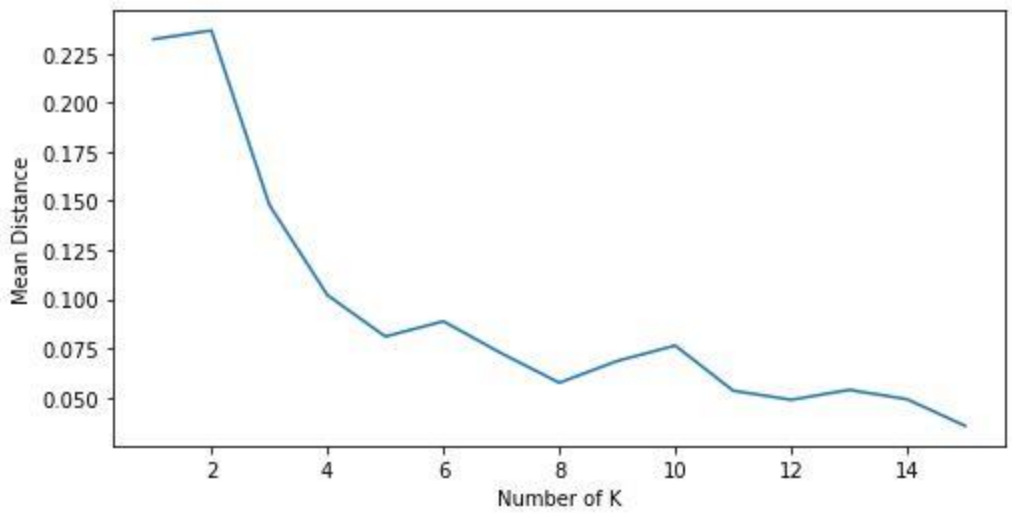
Toronto Real Estate Board has the price index for each districts and different type of properties:

District Number	Composite	Single Family Detached	Single Family Attached	Apartment
W01	1207400	1603000	1257100	659500
W02	1269700	1469500	1187200	717800
W03	956000	1039600	955400	562000
W04	856400	1033900	914100	521900
W05	740200	1057300	855800	440700

In this case, we will use the price index of apartments for our analysis, and the distribution looks normal:

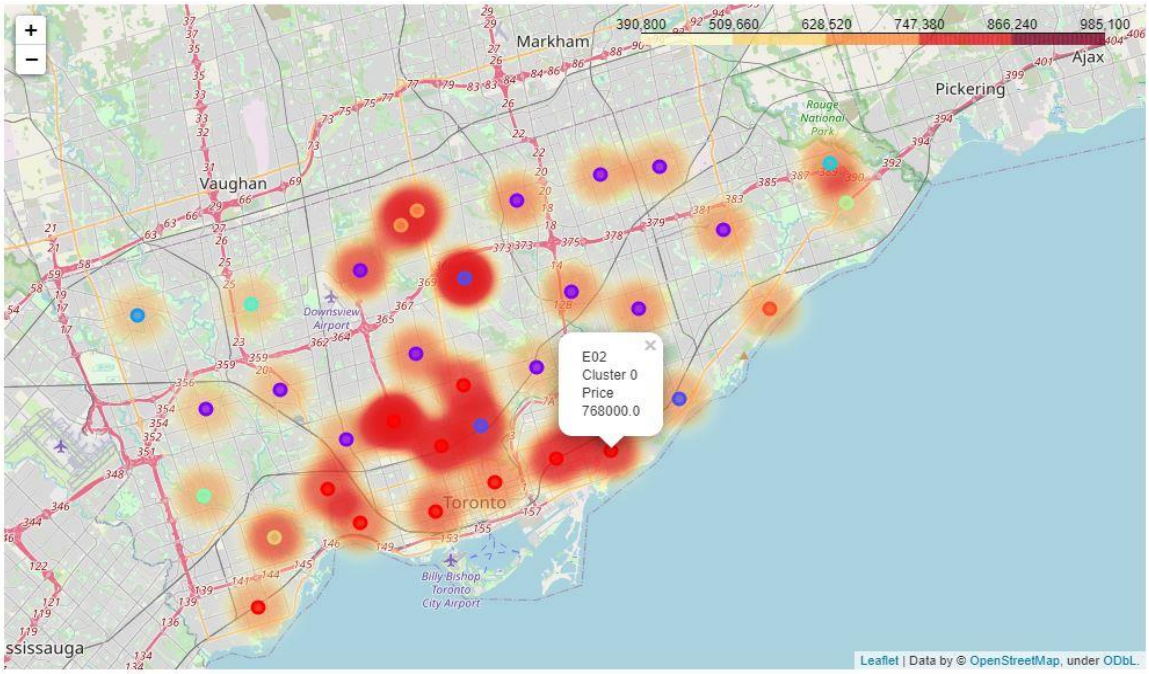


Then I implemented K-Means algorithm to cluster the districts by the venues. To find the optimal K Value, I firstly attempted to use elbow method. This is the chart shows the mean distance with different K Value:

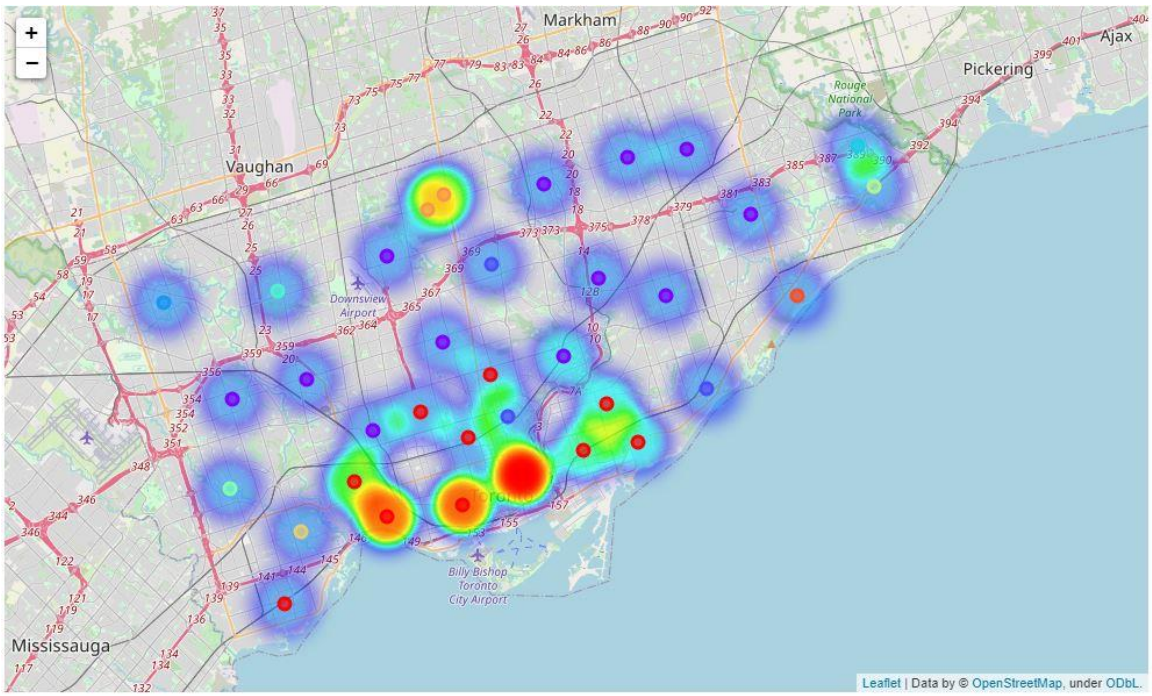


The elbow method suggested to choose 5 as K value. However, since 6 districts have less than 10 venues which makes them outliers, the model will be influenced. I eventually selected 11 as my K value after a few attempts.

Second step is the visualization the price level on the map, I used a heat map to display the clusters and price level for each district, the color of the dots identified the cluster of the district:



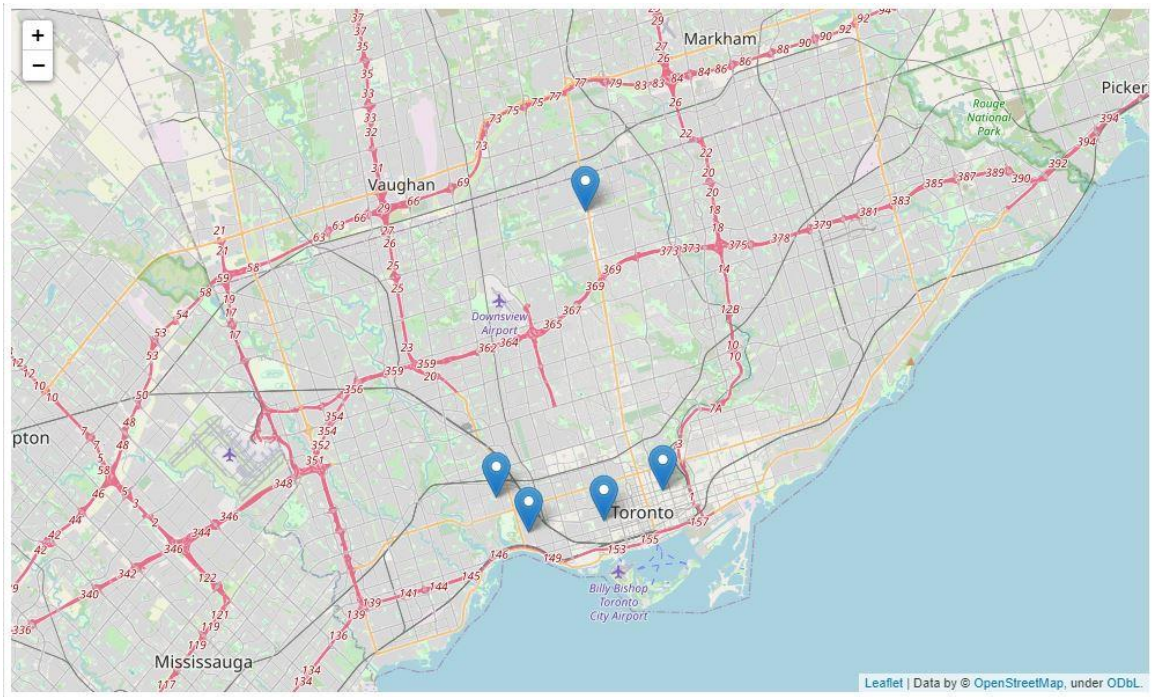
Third step is to introduce the venue-price ratio which is the number of venues divided by the price, and I used another heat map to display the ratio:



Finally, we selected the top 5 districts with the best value for money:

	District Number	Neighbourhoods Included	Venue Ratio
6	C08	Cabbagetown, St. Lawrence Market, Toronto wate...	15.787812
0	C01	Downtown, Harbourfront, Little Italy, Little P...	14.617746
25	W01	High Park, South Parkdale, Swansea, Roncesvall...	14.404852
26	W02	Bloor West Village, Baby Point, The Junction (...	11.145166
12	C14	Newtonbrook East, Willowdale East, Newtonbrook	10.920011

These are the districts on the map:



Results and Discussion

During the model selection process, we found out that the venues in Toronto districts are highly schismatic: most venues are concentrated in certain districts. This causes difficulty for K-means model since it tries to include all the outliers, so the optimal K value under the elbow is not the best fit in our data set. Therefore, I selected 11 as my K value which shows more than one cluster other than outliers (which implied about 6 outliers which is clustered with itself).

Conclusion

When we calculate the Venue Ratio, we multiplied the result by 100000 to display the difference since the number of venue and the price are not in same Magnitude. Eventually we found the top 5 districts that are most convenient with relatively cheaper price for apartments. They are C08, C01, W01. W02 and C14.

References:

- **Wikipedia**
- **TREB**
- **Foursquare API**
- **Google Map**