Battle of the Neighbourhoods

Open new restaurant in Toronto

April 11, 2019

1. Introduction

The entrepreneur is planning to open a new restaurant in Toronto, but he is not sure which location would be most appropriate for his new venue. We noticed that the Toronto already has a lot of restaurants in town, but we need to help this entrepreneur to find this location.

2. Data

We had to discover the most important factors that contribute to the restaurant's success. We can expect these factors to be among the following list: neighbourhood wealth, accessibility, crime rates, visibility, competition, etc. We should use the datasets from Toronto Opendata website to address some of these considerations.

From there, we were able to get the city's average housing prices list. We will be working with Get Wellbeing Toronto - Economics data set that includes average house price by neighbourhood. Also, we were using the "Foursquare" location data to retrieve the food venues. We would use Foursquare location data in conjunction with the average house price by neighbourhood to determine the best possible location for a new restaurant.

We started with the dataset called "Get Wellbeing Toronto – Economics" which included the average house price by neighbourhood This dataset contains the child case spaces, debt risk score, home prices, local employment and social assistance recipients for 138 neighbourhoods. Due to the scope of this project we were only interested in home prices data.

Next, we've got the Postal code data for the Toronto. To create this dataframe we had to parse https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M html page using Beautiful Soup. After removing all rows where we had not assigned values in borough column and replacing the not assigned values in neighbourhood with corresponding values of borough we end up with the dataframe for 211 postal codes.

3. Methodology

To analyse the acquired data, we would combine the average house price from "Get Wellbeing Toronto – Economics" dataframe with the Neighbourhood postal code dataset to get house prices per postal codes. Then we would get the venues from food category using the "Foursquare" location data. We would cluster the combined data and would try to determine the best possible location for the new restaurant.

Starting from "Get Wellbeing Toronto – Economics" dataframe we added the average house prices column to postal code dataframe, matching the values to the respective neighbourhoods. We converted the average house prices to the units of millions. And removed the postal code rows that had no value for the average house price. This gave us the complete data for 54 postal codes.

[8]:

	PostalCode	Borough	AvHomePrice	Neighborhood
0	M1B	Scarborough	0.360725	Rouge, Malvern
1	M1C	Scarborough	0.529278	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	0.347395	Guildwood, Morningside, West Hill
3	M1G	Scarborough	0.316584	Woburn
4	M1J	Scarborough	0.356096	Scarborough Village

```
[9]: df_loc.shape
[9]: (56, 4)
```

Figure 1. Postal code dataframe with average home prices

We imported Geospatial_data file containing the coordinates by postal code from http://cocl.us/Geospatial_data We found the respective coordinates by the Postal code in Geospatial_data dataframe and add it to Neighbourhood dataframe. To limit the dataframe to Toronto only postal codes we filter out the borough's that contain "Toronto". This gave as a 14 row dataframe.

	PostalCode	Borough	AvHomePrice	Neighborhood	Latitude	Longitude
0	M4E	East Toronto	0.751945	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	0.677840	The Danforth West, Riverdale	43.679557	-79.352188
2	M4N	Central Toronto	1.098110	Lawrence Park	43.728020	-79.388790
3	M4T	Central Toronto	1.265389	Moore Park, Summerhill East	43.689574	-79.383160
4	M4W	Downtown Toronto	1.265389	Rosedale	43.679563	-79.377529
5	M4X	Downtown Toronto	0.537025	Cabbagetown, St. James Town	43.667967	-79.367675
6	M5A	Downtown Toronto	0.484444	Harbourfront, Regent Park	43.654260	-79.360636
7	M5H	Downtown Toronto	0.617042	Adelaide, King, Richmond	43.650571	-79.384568
8	M5P	Central Toronto	0.957688	Forest Hill North, Forest Hill West	43.696948	-79.411307
9	M5T	Downtown Toronto	0.477989	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049
10	M6H	West Toronto	0.502736	Dovercourt Village, Dufferin	43.669005	-79.442259
11	M6J	West Toronto	0.619434	Little Portugal, Trinity	43.647927	-79.419750
12	M6P	West Toronto	0.615948	High Park, The Junction South	43.661608	-79.464763
13	M6R	West Toronto	0.540739	Parkdale, Roncesvalles	43.648960	-79.456325
14	M6S	West Toronto	0.644205	Runnymede, Swansea	43.651571	-79.484450

Figure 2. Postal code dataframe with average home prices and geographical coordinates

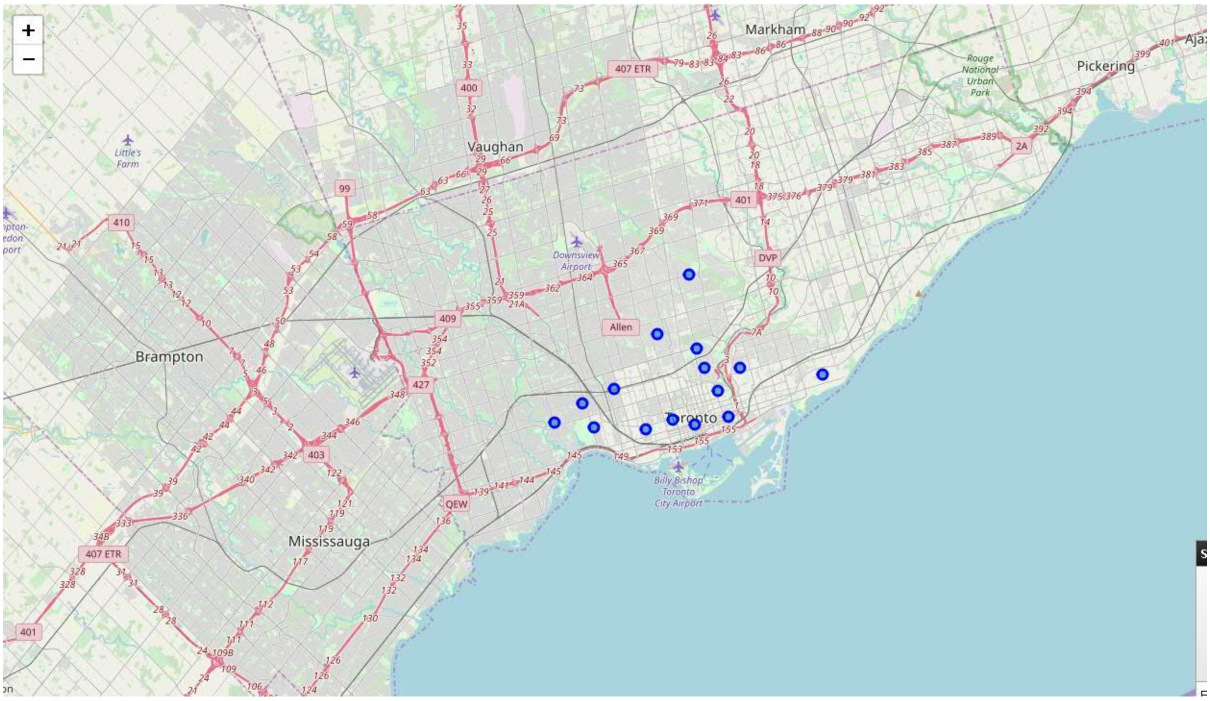

Figure 3. Toronto map with Neighbourhood markers

On Figure 4c is the choropleth map showing the Average house prices for Toronto. Unfortunately, we were unable to get the average price data for all Toronto neighbourhoods, but we even though we can see three areas where the average house price is above 1 million - around Rosedale, Summerhill East and Lawrence Park.

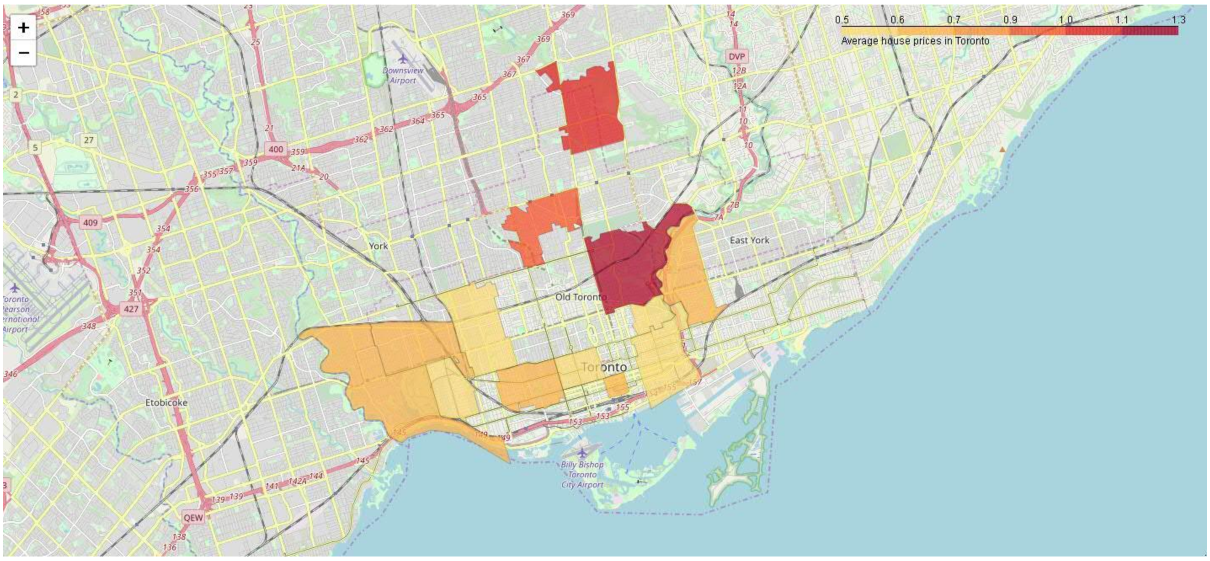

Figure 4. Average house price in Toronto

Using “Foursquare” API we got the get nearby venues from Food category for all Toronto neighbourhoods. This gave us all restaurants in the selected area. We created a new dataframe that had all the venues for each of the neighbourhoods. The resulting dataframe, had 73 unique categories.

Neighborhood	AvHomePrice	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Adelaide, King, Richmond	95	95	95	95	95	95	95
Cabbagetown, St. James Town	28	28	28	28	28	28	28
Chinatown, Grange Park, Kensington Market	60	60	60	60	60	60	60
Dovercourt Village, Dufferin	9	9	9	9	9	9	9
Forest Hill North, Forest Hill West	5	5	5	5	5	5	5
Harbourfront, Regent Park	24	24	24	24	24	24	24
High Park, The Junction South	15	15	15	15	15	15	15
Lawrence Park	1	1	1	1	1	1	1
Little Portugal, Trinity	43	43	43	43	43	43	43
Moore Park, Summerhill East	1	1	1	1	1	1	1
Parkdale, Roncesvalles	8	8	8	8	8	8	8
Rosedale	1	1	1	1	1	1	1
Runnymede, Swansea	27	27	27	27	27	27	27
The Beaches	4	4	4	4	4	4	4
The Danforth West, Riverdale	35	35	35	35	35	35	35

Figure 5. Number of venues in Food category were returned for each neighbourhood

We grouped the rows by neighbourhood and by taking the mean of the frequency of occurrence of each category. Then we sorted the venues in descending order and create the new dataframe that contain the top 10 venues for each neighbourhood.

	Neighborhood	AvHomePrice	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide, King, Richmond	0.617042	Restaurant	Sandwich Place	Café	American Restaurant	Asian Restaurant	Salad Place	Deli / Bodega	Steakhouse	Thai Restaurant	Burger Joint
1	Cabbagetown, St. James Town	0.537025	Pizza Place	Restaurant	Café	Chinese Restaurant	Italian Restaurant	Sandwich Place	Gastropub	Indian Restaurant	Japanese Restaurant	Diner
2	Chinatown, Grange Park, Kensington Market	0.477989	Café	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Mexican Restaurant	Dim Sum Restaurant	Comfort Food Restaurant	Burger Joint	Caribbean Restaurant
3	Dovercourt Village, Dufferin	0.502736	Bakery	Pizza Place	Café	Portuguese Restaurant	Middle Eastern Restaurant	Fast Food Restaurant	Brazilian Restaurant	Vietnamese Restaurant	Dim Sum Restaurant	Diner
4	Forest Hill North, Forest Hill West	0.957688	French Restaurant	Sushi Restaurant	Restaurant	Mexican Restaurant	Sandwich Place	Vietnamese Restaurant	Donut Shop	Deli / Bodega	Dim Sum Restaurant	Diner
5	Harbourfront, Regent Park	0.484444	Café	Bakery	Restaurant	Mexican Restaurant	Breakfast Spot	Seafood Restaurant	Greek Restaurant	Italian Restaurant	Sandwich Place	Japanese Restaurant
6	High Park, The Junction South	0.615948	Mexican Restaurant	Café	Thai Restaurant	Fried Chicken Joint	Gastropub	Diner	Steakhouse	Bakery	Fast Food Restaurant	Italian Restaurant
7	Lawrence Park	1.098110	Dim Sum Restaurant	Vietnamese Restaurant	Gastropub	Deli / Bodega	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Falafel Restaurant
8	Little Portugal, Trinity	0.619435	Asian Restaurant	Vietnamese Restaurant	New American Restaurant	Bakery	Pizza Place	Café	Cuban Restaurant	Vegetarian / Vegan Restaurant	French Restaurant	Restaurant
9	Moore Park, Summerhill East	1.265389	Restaurant	Vietnamese Restaurant	Eastern European Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Falafel Restaurant
10	Parkdale, Roncesvalles	0.540739	Breakfast Spot	Cuban Restaurant	Restaurant	Eastern European Restaurant	Burger Joint	Deli / Bodega	Italian Restaurant	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant
11	Rosedale	1.265389	Japanese Restaurant	Vietnamese Restaurant	Falafel Restaurant	Dim Sum Restaurant	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Fast Food Restaurant
12	Runnymede, Swansea	0.644205	Pizza Place	Café	Sushi Restaurant	Italian Restaurant	Diner	Restaurant	Fish & Chips Shop	Burrito Place	Food	Sandwich Place
13	The Beaches	0.751945	Pizza Place	Burger Joint	Asian Restaurant	BBQ Joint	Dim Sum Restaurant	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant
14	The Danforth West, Riverdale	0.677840	Greek Restaurant	Sushi Restaurant	Italian Restaurant	Pizza Place	Café	Restaurant	American Restaurant	Breakfast Spot	Caribbean Restaurant	Japanese Restaurant

Figure 6. Top 10 venues for each neighbourhood.

We cluster the Neighborhoods into 5 clusters by runing *k*-means to cluster the neighborhood into 10 clusters. We created a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

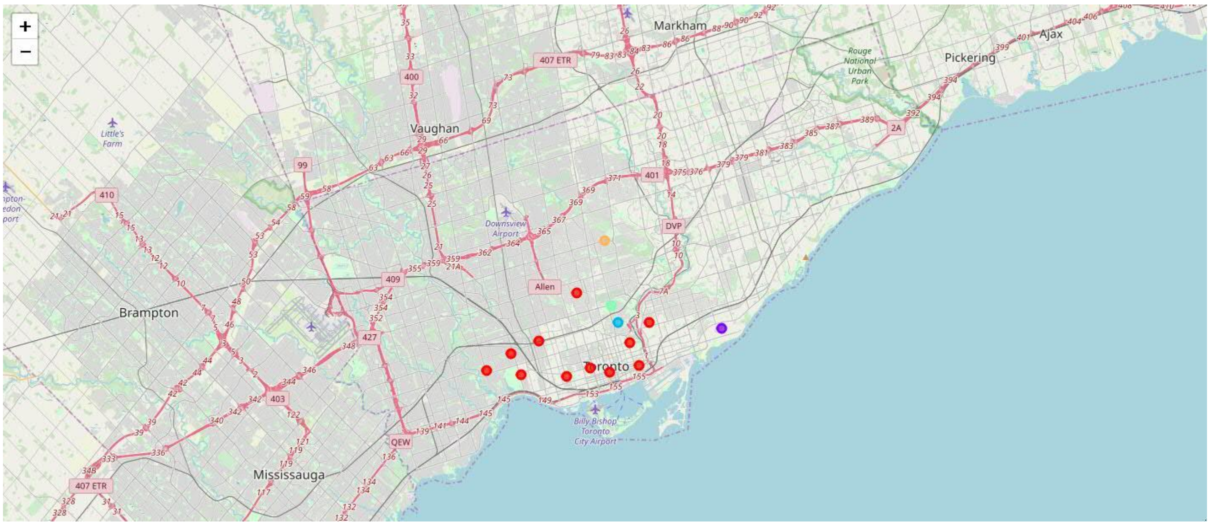

Figure 7. Map of resulting clusters.

4. Results

It appears that most of the neighbourhoods are located in the 1st cluster. In the 1st cluster, the most common venues in the neighbourhoods are cafes and pizza places. Also, it appears that three clusters are around the neighbourhoods with the highest average house prices.

Cluster 0

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 0, toronto_merged.columns[[2] + list(range(5, toronto_merged.shape[1]))]]
```

	Neighborhood	Cluster Labels	AvHomePrice	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	The Danforth West, Riverdale	0	0.677840	Greek Restaurant	Sushi Restaurant	Italian Restaurant	Pizza Place	Café	Restaurant	American Restaurant	Breakfast Spot	Caribbean Restaurant	Japanese Restaurant
5	Cabbagetown, St. James Town	0	0.537025	Pizza Place	Restaurant	Café	Chinese Restaurant	Italian Restaurant	Sandwich Place	Gastropub	Indian Restaurant	Japanese Restaurant	Diner
6	Harbourfront, Regent Park	0	0.484444	Café	Bakery	Restaurant	Mexican Restaurant	Breakfast Spot	Seafood Restaurant	Greek Restaurant	Italian Restaurant	Sandwich Place	Japanese Restaurant
7	Adelaide, King, Richmond	0	0.617042	Restaurant	Sandwich Place	Café	American Restaurant	Asian Restaurant	Salad Place	Deli / Bodega	Steakhouse	Thai Restaurant	Burger Joint
8	Forest Hill North, Forest Hill West	0	0.957688	French Restaurant	Sushi Restaurant	Restaurant	Mexican Restaurant	Sandwich Place	Vietnamese Restaurant	Donut Shop	Deli / Bodega	Dim Sum Restaurant	Diner
9	Chinatown, Grange Park, Kensington Market	0	0.477989	Café	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Mexican Restaurant	Dim Sum Restaurant	Comfort Food Restaurant	Burger Joint	Caribbean Restaurant
10	Dovercourt Village, Dufferin	0	0.502736	Bakery	Pizza Place	Café	Portuguese Restaurant	Middle Eastern Restaurant	Fast Food Restaurant	Brazilian Restaurant	Vietnamese Restaurant	Dim Sum Restaurant	Diner
11	Little Portugal, Trinity	0	0.619435	Asian Restaurant	Vietnamese Restaurant	New American Restaurant	Bakery	Pizza Place	Café	Cuban Restaurant	Vegetarian / Vegan Restaurant	French Restaurant	Restaurant
12	High Park, The Junction South	0	0.615948	Mexican Restaurant	Café	Thai Restaurant	Fried Chicken Joint	Gastropub	Diner	Steakhouse	Bakery	Fast Food Restaurant	Italian Restaurant
13	Parkdale, Roncesvalles	0	0.540739	Breakfast Spot	Cuban Restaurant	Restaurant	Eastern European Restaurant	Burger Joint	Deli / Bodega	Italian Restaurant	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant
14	Runnymede, Swansea	0	0.644205	Pizza Place	Café	Sushi Restaurant	Italian Restaurant	Diner	Restaurant	Fish & Chips Shop	Burrito Place	Food	Sandwich Place

Cluster 1

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[2] + list(range(5, toronto_merged.shape[1]))]]
```

	Neighborhood	Cluster Labels	AvHomePrice	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	The Beaches	1	0.751945	Pizza Place	Burger Joint	Asian Restaurant	BBQ Joint	Dim Sum Restaurant	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant

Cluster 2

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[2] + list(range(5, toronto_merged.shape[1]))]]
```

	Neighborhood	Cluster Labels	AvHomePrice	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	Rosedale	2	1.265389	Japanese Restaurant	Vietnamese Restaurant	Falafel Restaurant	Dim Sum Restaurant	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Fast Food Restaurant

Cluster 3

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 3, toronto_merged.columns[[2] + list(range(5, toronto_merged.shape[1]))]]
```

	Neighborhood	Cluster Labels	AvHomePrice	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Moore Park, Summerhill East	3	1.265389	Restaurant	Vietnamese Restaurant	Eastern European Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Falafel Restaurant

Cluster 4

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 4, toronto_merged.columns[[2] + list(range(5, toronto_merged.shape[1]))]]
```

	Neighborhood	Cluster Labels	AvHomePrice	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Lawrence Park	4	1.09811	Dim Sum Restaurant	Vietnamese Restaurant	Gastropub	Deli / Bodega	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Falafel Restaurant

Figure 8. Clusters.

5. Discussion

Where should we open a new restaurant? Housing price maps show that the Lawrence Park cluster (4) neighbourhood might be a good candidate. This area looks like a quite densely populated area, so we expect the region to have a lot of foot and car traffic, so good visibility. This neighbourhood has also reasonable average house prices.

6. Conclusion

This is only a first-order solution to the question 'Where to open a new restaurant in Toronto?' Using public datasets, we were able to partially address one of the factors that we have mentioned at the beginning - average house prices. There certainly is lot of room for improvement. For example, we have to factor in crime rates, competition etc. Toronto Opendata website should have other datasets that we might use to further improve the results.