

Where is my new store in York Region

Powered by Data Science

I have a problem

- ▶ I want to open a new grocery store in York Region. But what neighborhood I should choose?
- ▶ If I open it at a rural area where I don't have enough potential customers, I can't make the business to run properly.
- ▶ If I open it at a great area but there already have enough similar stores, I may not be able to get enough revenue.
- ▶ I want to find a neighborhood where -
 - (1) have a fair amount of households
 - (2) doesn't have enough competitors

I believe I can have the best business at there!

My Plan

What I am going to do are -

- ▶ 1. Get York Region Data
- ▶ 2. Explore and Segment York Region
- ▶ 3. Cluster Neighborhoods York Region
- ▶ 4. Make the decision

I will utilize my data science skills!

Data Acquisition

Here -

- ▶ York Region neighborhood data can be obtained from Wikipedia.
- ▶ The latitudes and longitudes data can be acquired from the Geocoder Python data or Google.
- ▶ The location & venues data can be retrieved from Foursquare Places API.

All free!

Data Quality

Pros & cons always!

For example, Wiki pages may update often without telling you 😊

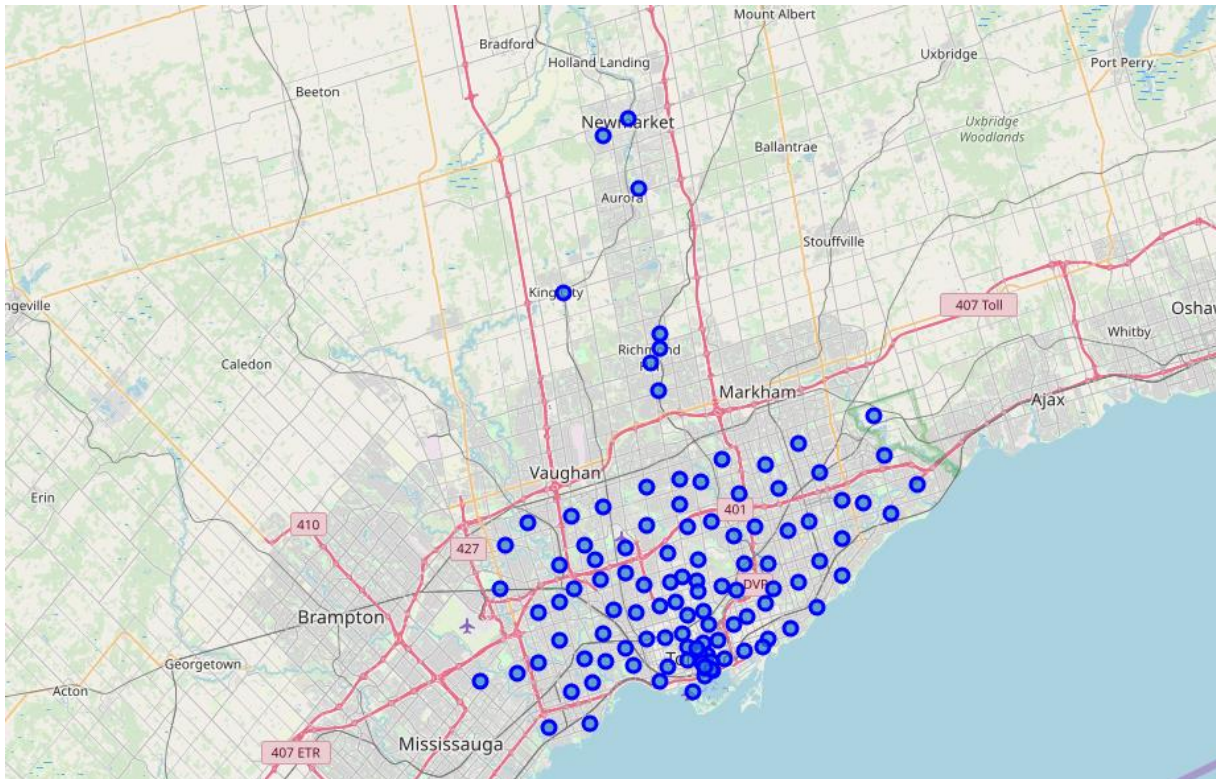
We need to make sure the data to be analyzed is clean with high quality.

- ▶ Data has to be clean enough - i.e. no duplicates, be consistent etc.
- ▶ Data has to be accurate - i.e. York Region only.
- ▶ Data may have to be refined before the analysis - i.e. I need to use coarse-level venue categories from the original Foursquare data.

Data Visualization

It is so cool to have a good view!

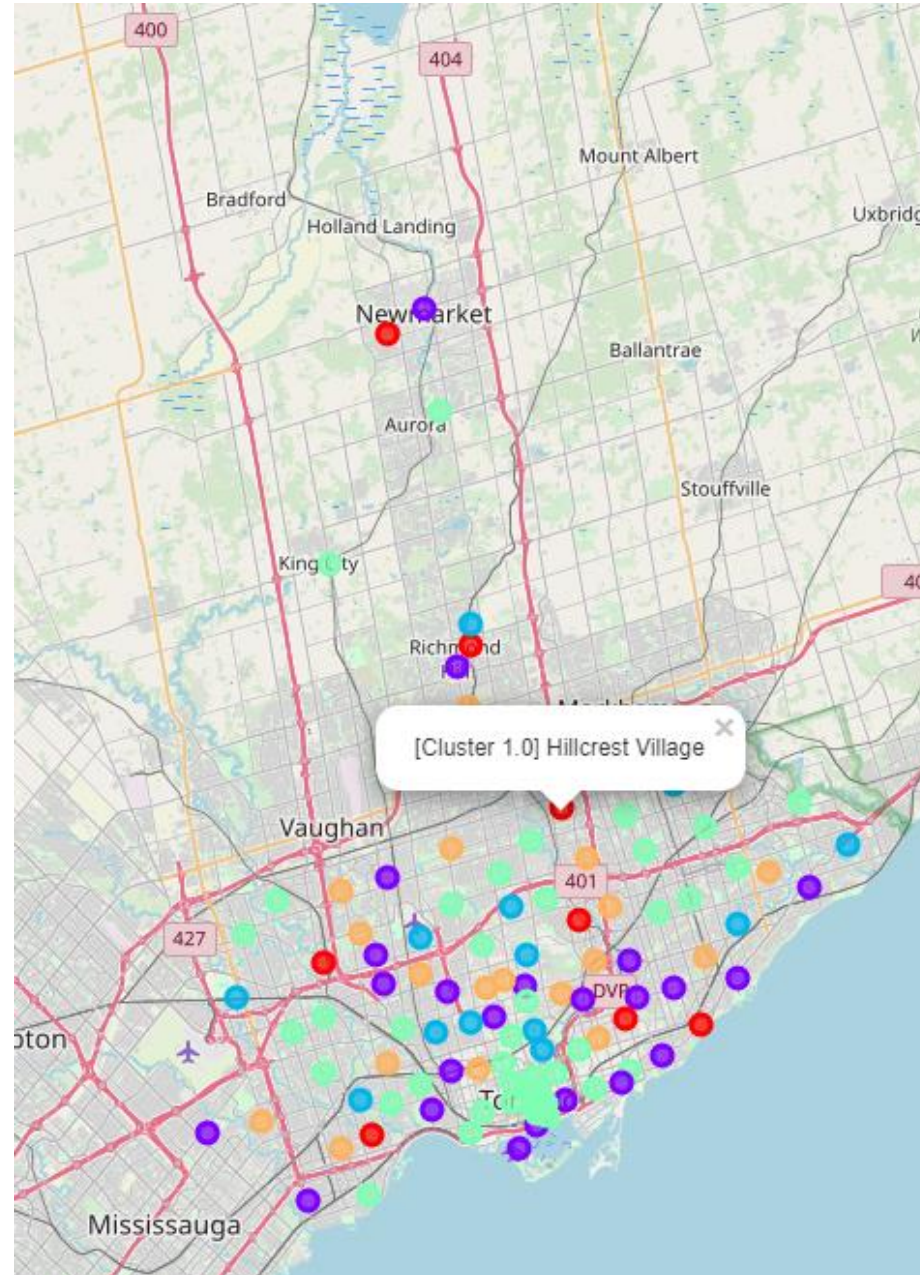
Check this out -



Data Mining

I only have interests on the hot spots.

So let me divide them with k-means - total 5 clusters.



Data Analysis

I only want to choose from Cluster 1 boroughs.

Any Cluster 1 borough has 'Shop' as the '1st Most Common Venue'?

```
In [236]: toronto_merged[(toronto_merged['Cluster Labels'] == 0) & (toronto_merged['1st Most Common Venue'] == 'Shop')]
```

Out[236]:

Postcode	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	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 M
----------	---------	---------------	----------	-----------	----------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-------

Any Cluster 1 borough has 'Shop' as the '2nd Most Common Venue'?

```
In [237]: toronto_merged[(toronto_merged['Cluster Labels'] == 0) & (toronto_merged['2nd Most Common Venue'] == 'Shop')]
```

Out[237]:

Postcode	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	
36	M4C	East York	Woodbine Heights	43.695344	-79.318389	0.0	Sport	Shop	Others	Transportation	Food	Entertainment

Any Cluster 1 borough has 'Shop' as the '3rd Most Common Venue'?

```
In [238]: toronto_merged[(toronto_merged['Cluster Labels'] == 0) & (toronto_merged['3rd Most Common Venue'] == 'Shop')]
```

Out[238]:

Postcode	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
91	M8Y	Etobicoke	Humber Bay,King's Mill Park,Kingsway Park Sout...	43.636258	-79.498509	0.0	Sport	Transportation	Shop	Others	Food
97	M9M	North York	Emery,Humberlea	43.724766	-79.532242	0.0	Sport	Transportation	Shop	Others	Food
105	L3X	Newmarket	Newmarket Southwest	44.046400	-79.487400	0.0	Sport	Food	Shop	Transportation	Others

Any Cluster 1 borough has 'Shop' as the '4th Most Common Venue'?

```
In [239]: toronto_merged[(toronto_merged['Cluster Labels'] == 0) & (toronto_merged['4th Most Common Venue'] == 'Shop')]
```

Out[239]:

Postcode	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Comr	
26	M3B	North York	Don Mills North	43.745906	-79.352188	0.0	Sport	Food	Transportation	Shop	Others	En
107	L4B	Richmond Hill	Richmond Hill Southeast	43.887501	-79.428406	0.0	Sport	Transportation	Others	Shop	Food	En

Decision Making

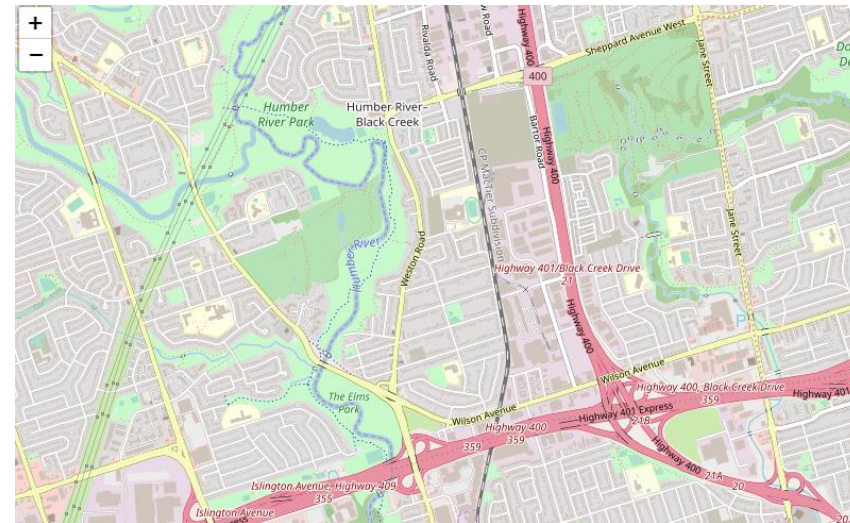
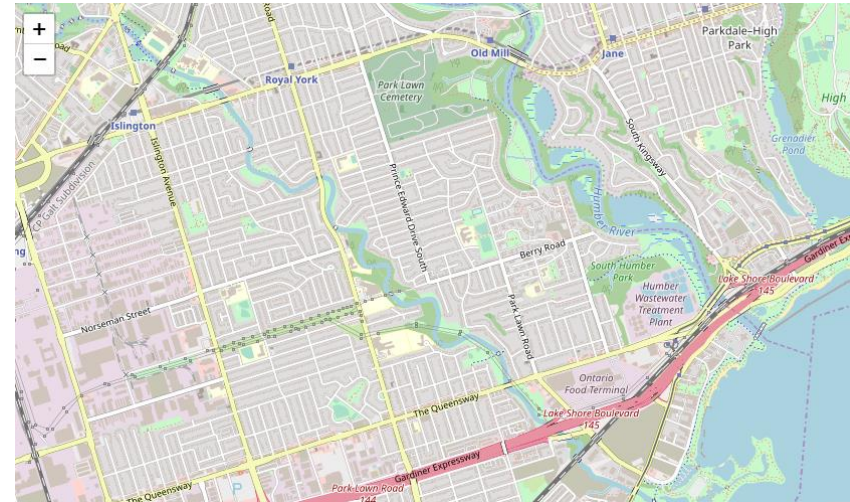
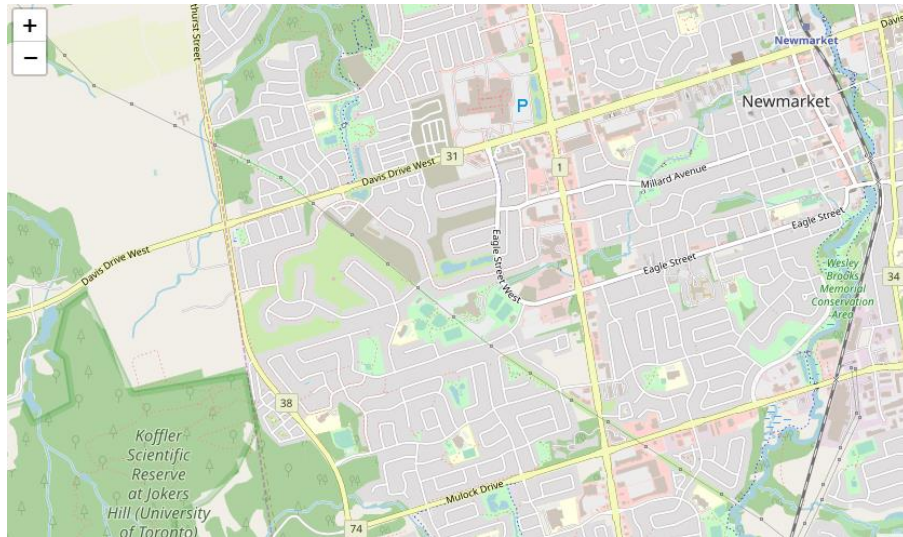
Based on above analysis, I want to choose a borough in Cluster 1 who has 'Shop' as the '3rd Most Common Venue'.

I found there are 3 boroughs meet these 2 requirements -

- ▶ M8Y
- ▶ M9M
- ▶ L3X

Candidates Comparison

Let's check the maps.



I've made my decision!

I finally chose "L3X - Newmarket Southwest" as the location of my new store!

Thank you, Data Science!

