

Analyzing Toronto Neighbourhoods

by Sushant Deshpande

In this project, we'll analyze Toronto's most multicultural neighbourhoods, East Toronto and West Toronto. By doing so, we'll see what types of venues are popular in each neighbourhood. This will help a group of investors understand what type of restaurant they would like to invest in. In order to achieve this, We'll perform the following:



Get the postal codes from Wikipedia using web scraping method in python.



Clean up the data and merge it with the geopositional data provided to us by Coursera.



Use FourSquare to get the venues in our desired neighbourhood.



Use python to find to 10 most popular venues and display it graphically.



STEP 1 - Importing the data

We begin with grabbing the postal codes from Wikipedia's webpage using python's web scraping method.

Downloaded data looks like this:

```
Out[11]:
```

	0
0	[]
1	[M1A, Not assigned, Not assigned\n]
2	[M2A, Not assigned, Not assigned\n]
3	[M3A, North York, Parkwoods\n]
4	[M4A, North York, Victoria Village\n]
5	[M5A, Downtown Toronto, Harbourfront\n]
6	[M5A, Downtown Toronto, Regent Park\n]
7	[M6A, North York, Lawrence Heights\n]
8	[M6A, North York, Lawrence Manor\n]
9	[M7A, Queen's Park, Not assigned\n]

This data does not look very clear, all the postal codes, borough and neighbourhoods are together. There is no head to our data frame, hence we need to clean it. We'll do this in step 2.



STEP 2 - Cleaning the data

In this step, we clean the data so that it more easy readable and understandable by us as well as python.

We split the data, drop unnecessary rows and columns and rename the the column index. We then delete empty cells

```
Out[12]:
```

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M5A	Downtown Toronto	Regent Park
4	M6A	North York	Lawrence Heights

More than one neighbourhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that M5A is listed twice and has two neighbourhoods: Harbourfront and Regent Park. These two rows will be combined into one row with the neighbourhoods separated with a comma.

```
Out[13]:
```

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Now we need to combine our data with Geospatial_Coordinates.csv file provided to us. This'll get us coordinates for each of the postal codes. These coordinates will be used in the next step with FourSquare

Out[14]:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476



STEP 3 - FourSquare

First, we split the data in to 2 categories, **East Toronto** and **West Toronto**. With FourSquare, we get the list of venues that are close to a selected postal code.

EAST TORONTO

Out[27]:

	name	categories	lat	lng
0	Tori's Bakeshop	Vegetarian / Vegan Restaurant	43.672114	-79.290331
1	Beaches Bake Shop	Bakery	43.680363	-79.289692
2	The Fox Theatre	Indie Movie Theater	43.672801	-79.287272
3	Ed's Real Scoop	Ice Cream Shop	43.672630	-79.287993
4	The Beech Tree	Gastropub	43.680493	-79.288846

WEST TORONTO

Out[28]:

	name	categories	lat	lng
0	The Greater Good Bar	Bar	43.669409	-79.439267
1	Parallel	Middle Eastern Restaurant	43.669516	-79.438728
2	Blood Brothers Brewing	Brewery	43.669944	-79.436533
3	Planet Fitness Toronto Galleria	Gym / Fitness Center	43.667588	-79.442574
4	Happy Bakery & Pastries	Bakery	43.667050	-79.441791

When we run the analysis for the whole of East and West Toronto, we can see that Four Square tells us that East Toronto has **133 unique categories** and West Toronto has **162 unique categories**.

We analyze each neighbourhood using one hot encoding. Next, we group rows by neighbourhood and by taking the mean of the frequency of occurrence of each category and print each neighbourhood along with the top 5 most common venues.

Now we cluster our neighbourhoods using k-means cluster algorithm.



STEP 4 - Python

Now we can see **10 venues** for each neighbourhood.

EAST TORONTO

Out[45]:

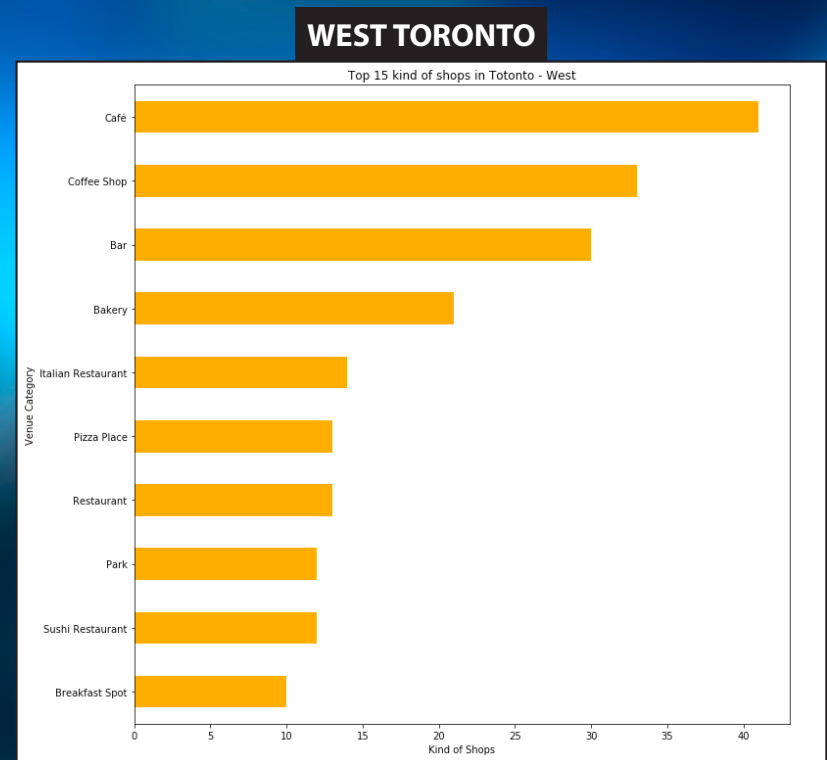
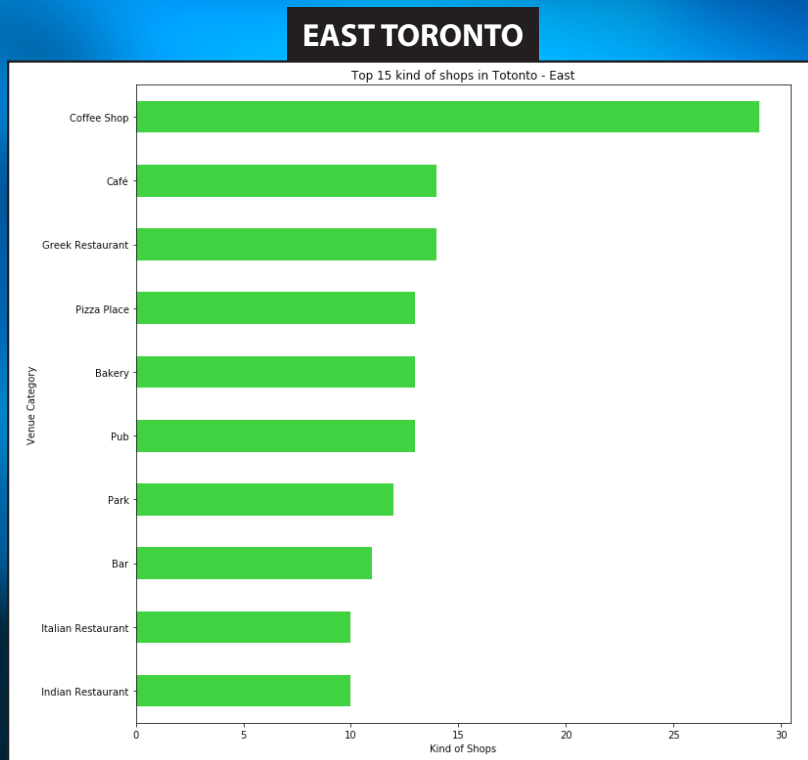
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Business reply mail Processing Centre969 Eastern	Park	Coffee Shop	Pizza Place	Sushi Restaurant	Italian Restaurant	Burrito Place	Brewery	Thai Restaurant	Breakfast Spot	Comic Shop
1	Studio District	Coffee Shop	Bar	Café	Bakery	American Restaurant	Italian Restaurant	Vietnamese Restaurant	Diner	Brewery	Sandwich Place
2	The Beaches	Coffee Shop	Pub	Pizza Place	Breakfast Spot	Beach	Japanese Restaurant	Bar	Bakery	Burger Joint	Indian Restaurant
3	The Beaches West, India Bazaar	Indian Restaurant	Coffee Shop	Sandwich Place	Beach	Café	Bakery	Italian Restaurant	Burrito Place	Spa	Restaurant
4	The Danforth West, Riverdale	Greek Restaurant	Coffee Shop	Pub	Café	Fast Food Restaurant	Pizza Place	Ice Cream Shop	Spa	Sandwich Place	Diner

WEST TORONTO

Out[46]:

	Neighborhood	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	Brockton, Exhibition Place, Parkdale Village	Café	Coffee Shop	Restaurant	Bar	Furniture / Home Store	Vegetarian / Vegan Restaurant	Bakery	Soccer Stadium	Sandwich Place	Gift Shop
1	Dovercourt Village, Dufferin	Café	Park	Coffee Shop	Supermarket	Sushi Restaurant	Bar	Bakery	Breakfast Spot	Convenience Store	Pharmacy
2	High Park, The Junction South	Bar	Café	Coffee Shop	Italian Restaurant	Convenience Store	Thai Restaurant	Sandwich Place	Sushi Restaurant	Antique Shop	Bakery
3	Little Portugal, Trinity	Bar	Café	Restaurant	Bakery	Coffee Shop	Cocktail Bar	Men's Store	Italian Restaurant	Pizza Place	Vegetarian / Vegan Restaurant
4	Parkdale, Roncesvalles	Bar	Coffee Shop	Sushi Restaurant	Breakfast Spot	Pizza Place	Café	Thai Restaurant	Gym	Eastern European Restaurant	Bakery

Now, we can display this result using a bar graph for ease of understanding.



As is clear from the data analysis, Cafes are the the most popular venues in West Toronto with **41 locations** and in East Toronto, there second most popular with **14 locations**. It is clear that they are the new emerging trend. As an investor, it would be advisable to invest in a Cafe Shop in East Toronto. Since there are on 14 there compared to 41 in West Toronto, opening one up in East Toronto would be a wise move.