Data Description

1. Data Acquisition and Pre-Processing

1.1 Data Sources

Data 1 : Web Scrapping

To get our first data sources which we can get data from Wikipedia using web scrapping, Xpath Python Library, then wrangling it into pandas Data Frame. And also for coordinate data, which we get from foursquare Application Programming interface (API) in JSON format and the combine it with csv data from Toronto open data. Once the data are already tidied, we can see the pattern that we want to research on.

In [8]: df.head(12)

Out[8]:

	Postal code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.752935	-79.335641
1	M4A	North York	Victoria Village	43.728102	-79.311890
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.650964	-79.353041
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.723265	-79.451211
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.661790	-79.389390
5	M9A	Etobicoke	Islington Avenue	43.667481	-79.528953
6	M1B	Scarborough	Malvern, Rouge	43.808626	-79.189913
7	M3B	North York	Don Mills	43.748900	-79.357220
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.707193	-79.311529
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657491	-79.377529
10	M6B	North York	Glencairn	43.707279	-79.447500
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650023	-79.554089

Data 2 : Foursquare API :

For location and geo tagging we collect data from foursquare and geopy and combine into map below. The data contain latitude longitudes and venues type.

	Postal code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	M3A	North York	Parkwoods	43.752935	-79.335641	3	Construction & Landscaping	Food & Drink Shop	Park	Eastern European Restaurant	Food
1	M4A	North York	Victoria Village	43.728102	-79.311890	0	Park	Pizza Place	Portuguese Restaurant	French Restaurant	Coffee Shop
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.650964	-79.353041	0	Pub	Mediterranean Restaurant	Café	Athletics & Sports	Performing Arts Venue
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.723265	-79.451211	0	Clothing Store	Furniture / Home Store	Men's Store	Bookstore	Food Court
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.661790	-79.389390	0	Coffee Shop	Sushi Restaurant	Café	Juice Bar	Bookstore