

2. Data Acquisition and Preparation

2.1 Data sources

Data is scraped from two web pages and also from FourSquare. From [List of postal codes of Canada](#), the postal codes of all the neighbourhoods is extracted and demographics are extracted from [Demographics of Toronto neighbourhoods](#). Location data is extracted from Geospatial_Coordinates.csv. Venues present in the neighbourhoods are extracted from FourSquare. Dataframes are:

- 1) df ([List of postal codes of Canada](#))

	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

- 2) ll_data (Geospatial_Coordinates.csv)

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

- 3) pop_den ([Demographics of Toronto neighbourhoods](#))

	Name	FM	Census Tracts	Population	Land area (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit Commuting %	% Renters	Second most common language (after English) by name	Second most common language (after English) by percentage	Map
0	Crescent Town	EY	0190.01	8157.0	0.40	20393.0	-10.0	23021.0	24.5	20.3	Bengali (18.1%)	18.1% Bengali	NaN
1	Governor's Bridge/Bennington Heights	EY	0186.00	2112.0	1.87	1129.0	4.0	129904.0	7.1	13.3	Polish (1.4%)	01.4% Polish	NaN

4) neigh_venues (FourSquare API)

	Neighbourhood_Name	Neighbourhood_Lat	Neighbourhood_Lng	Venue_Name	Venue_Category	Venue_Lat	Venue_Lng
0	Parkwoods	43.753259	-79.329656	Allwyn's Bakery	Caribbean Restaurant	43.759840	-79.324719
1	Parkwoods	43.753259	-79.329656	Brookbanks Park	Park	43.751976	-79.332140
2	Parkwoods	43.753259	-79.329656	Tim Hortons	Café	43.760668	-79.326368
3	Parkwoods	43.753259	-79.329656	Bruno's valu-mart	Grocery Store	43.746143	-79.324630
4	Parkwoods	43.753259	-79.329656	A&W	Fast Food Restaurant	43.760643	-79.326865