Buying a home in Toronto

1. Introduction

1.1 Background

Since I have gotten well acquainted with Toronto in our last couple of assignments, I would like to stay there for this one too. Toronto is a large city with over 5mil of residents. In such a large city like Toronto it might be difficult to figure out what neighborhoods to concentrate on in your search to buy a new home. When searching for a house, one might want to concentrate on neighborhoods with low crime rates, average density of population, access to parks and playgrounds, as well as restaurants and coffee shops. All those criteria are usually considered with a certain budget in mind. I have visited Toronto 2 times and I have stayed with my cousins in the Forest Hill South neighborhood, so I will take that as a point of reference for a comfortable, safe, enjoyable living, and try to find other neighborhoods in Toronto that are similar to Forest Hill South, but are more affordable (have an average home price of less than $800,000).

1.2 Problem

There are about 140 officially recognized neighborhoods in Toronto. To determine neighborhoods that are like Forest Hill, these neighborhoods need to be clustered based on their safety scores, population density, access to parks, restaurants, and coffee shops. Once I determine which neighborhoods properties of interest are like those of Forest Hill South, I will be able to find neighborhoods that satisfy the budget restriction imposed in the last paragraph.

1.3 Interest

This project would be interesting to anyone who is looking to buy a house anywhere, as these methods can be transferred to finding a new house in New York, London or Tokyo. It is always good to do your own research before meeting with a real estate agent, so you can specify what exactly you are looking for. Anyone would be interested to learn more about the neighborhoods and to quickly eliminate neighborhoods that he/she is not interested in.

2. Data

2.1 Sources, cleaning and feature selection

Toronto Police has an open data access to records about various crime ratings, such as assault, robbery, homicide, auto theft and break-and-enter from 2014 to 2018. I have located the geojson file from their website (here), and extracted the crime data for the year 2018 for each neighborhood in Toronto. The file also included the information about population and size of the neighborhood as well as the coordinates of the boundaries of each neighborhood.

Features used:

1. Neighborhood
2. Assault Rate 2018
3. Auto Theft Rate 2018
4. Break and Enter Rate 2018
5. Robbery Rate 2018
6. Homicide Rate 2018
7. Population
8. Size of Hood Area
9. Population density (population/size)
10. Latitude
11. Longitude

2.2 Venues: sources, cleaning, and feature selection

I queried Foursquare API in order find venues that are most popular in each neighborhood of Toronto. The query returned 1475 venues, in 255 unique venue categories. The venue categories were converted to binary variable, using one hot encoding, and grouped by neighborhood. The resulting data set had 134 rows and 255 columns (plus the neighborhood index).

2.3 Average home price: sources, cleaning, and feature selection

The home prices data were scraped from a blog post on Toronto home prices by neighborhood for 2017, which luckily has the same format for the neighborhood names and assignment (link). There were, however, 6 neighborhoods missing from the dataset, and the missing values in those cases were replaced with average home price for all neighborhoods. In future plots, those neighborhoods can be easily spotted since their home prices have decimal points.

The average home prices were listed as strings with commas and dollar signs. Dollar signs and commas were removed, and the prices were converted to floats.

The resulting dataframe was merged with the entire dataset, to be used for plotting maps.