Introduction

It is said that moving is one of the most stressful events in life. It can be especially daunting when relocating to a different country, or even continent. Each person has a different level of tolerance to things like crime rate, local rent prices, noise pollution etc. The problem becomes even harder to solve when any dietary preferences and work commute come into the mix.

Especially for immigrants, where to live in the city can be a huge choice and it can change at least their first months in a new place. A typical immigrant may not know all the ins and outs of the city they are moving to and the process of learning them can take vast amounts of time.

The goal of the model is to help choose a good neighbourhood or area to relocate to in the city of Toronto. It does so by recommending an area given a set of preferences and their importance to the individual. The preferences include lifestyle/dietary choices, crime rate, rent prices and transport in the neighbourhood. The model calculates a score for each area and compares it to each neighbourhood’s score based on publicly available data.

Data

The project uses several datasets in conjunction with maps to better visualize the answers.

First, the city map is split into neighbourhoods based on Wikipedia list of postal codes in Canada[[1]](#footnote-1). The list is limited to the city of Toronto only. This data is later connected to a static list of coordinates of postal codes originally taken from Geocoder API[[2]](#footnote-2).

The neighbourhood data is then joined with Foursquare API information about most common establishments in each neighbourhood, which will determine the nature of each neighbourhood. There are X (300+, fill in later) categories of establishments in the Toronto area, based on Foursquare data.

The crime data is sourced from Toronto Police Service public safety data portal through the site’s API. There are 6 categories of crimes represented in the set and the slice that will be used represents the 2018 rate (Occurrences per 100,000 population) of each crime in every neighbourhood.

The transport data is collected from Walk Score API, and will be calculated from the center of the neighbourhood, making it less than ideal, but still a good indicator of the area’s transport and commute options. The results will include public transport, walking and biking scores to fit the user’s needs. The website itself offers more comprehensive transit checks, including calculating the commute time from and to a specific place.

The price data will be averaged from rental listings in the Multiple Listing Service. It is the system used by realtors in Toronto. Using only MLS data misses out on a large number of (often cheaper) rentals, but guarantees safety from scams often found on sites like Kijiji and Craigslist often used to find rentals or room sharing.

All the previously listed data sources will be given weight based on the perceived importance of each characteristic and then listed as a ranking of top neighbourhoods based on selected characteristics.

(TBC)

1. <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> [↑](#footnote-ref-1)
2. http://cocl.us/Geospatial\_data [↑](#footnote-ref-2)