

# Introduction

## Problem: Knowing Where to Live in Toronto

This project is intended to help anyone with plans to move to the Toronto area, but does not know anything about Toronto, and would like to live in a vibrant, convenient neighborhood with lots of amenities.

The goal is to source, all in one place, a specific, intuitive number that can quickly help someone looking to move to the Toronto area, with no previous knowledge, whether a potential zip code is right for them or not.

## Plan: Create a Neighborhood Convenience Score for each Toronto area Postal Code

In order to determine the most convenient zip codes in Toronto, we are going to source information on the number of local amenities available within a short distance, such as restaurants, theaters, grocery stores and subway stations.

We will then normalize the data, effectively ranking each neighborhood based on the total number of each amenity, add those normalized numbers, and create an overall rating, called the Neighborhood Convenience Score or "NCS".

## Intended Audience: People moving to Toronto and Real Estate Brokerages

This project is intended to help anyone who is looking to move to or within the greater Toronto area. Conversely, this data could also be of use to real estate developers who want to expand into new territories, or build an action plan to increase the value of the real estate they already own, by understanding the convenience of their own neighborhoods, and how they can go about improving their score.

# Datasets

For this project I will be utilizing the Foursquare API to source information about various neighborhoods, towns and boroughs in and around Toronto.

I will also use [this dataset on Toronto postal codes](#) found in the following Wikipedia page in order to source a list of various neighborhoods in Toronto.

## Methodology

To determine the NCS for each district, I will source information on local amenities, specifically restaurants, gyms, banks, pharmacies and parks. I will source the total number of each type of amenity, then normalize the data (the neighborhood with the most restaurants will receive a '1' and a zip code with 2 restaurants receives a score of  $2/\text{Highest number of restaurants in a zip code}$ ).

This will give me normalized values for each type of amenity, which I will add together to create the Neighborhood Convenience Score for each zip code. Some amenities, specifically grocery stores, will be given a higher value, because they are more important to have in your neighborhood, and sell the greatest number of common items. They are simply more convenient to live nearby than most other venues.

Banks and pharmacies will also be given a slight lesser weighted importance, because they are more rare than things like restaurants, and more important than parks.

In addition to these methods, I have also considered things like the total number of unique venue types each neighborhood has. A neighborhood with 80 total venues, but 70 unique venue types, may be more desirable than a neighborhood with 95 venues, if it only has 40 unique venue types, and therefore has less variety.

For example, A neighborhood with 5 distinct types of restaurants (Japanese, Indian, pizza place, gastropub, Mexican) would be more convenient and nice to live in than a neighborhood that had 10 restaurants, but they were all pizza places.

I have added these methods in order to make sure that I am not simply creating the NCS to reflect which neighborhoods are the most densely packed. A small neighborhood with a nice variety of stores, a bank, a pharmacy and 2+ grocery stores would have a great NCS, while a densely packed neighborhood without some of these essential services, but packed with bars and restaurants, may not score as high.

## Results

As a result of this statistical analysis, I have listed 96 unique neighborhoods/areas in and around Toronto in order of convenience. Because postal codes do not align exactly with neighborhoods, many neighborhoods are combined into single postal codes, or in one case, in lieu of a name, PO Boxes are given.

Here is a list of the 10 postal codes which received the highest Neighborhood Convenience Score:

1. Kensington Market, Chinatown, Grange Park
2. Stn A PO Boxes
3. Thorncliffe Park
4. Dufferin/Dovercourt Village
5. Harbourfront East, Union Station, Toronto Islands
6. Leaside

7. Christie
8. St. James Town
9. Downsview
10. St. James Town, Cabbagetown

And here are the 10 postal codes which received the lowest NCS:

1. Roselawn
2. West Deane Park, Princess Gardens, Martin Grove
3. Malvern, Rouge
4. Humberlea, Emery
5. Scarborough Village
6. Old Mill South, King's Mill Park, Sunnylea, Humber Bay, Mimico NE, The Queensway East, Royal York South East, Kingsway Park South East
7. Birch Cliff, Cliffside West
8. Agincourt
9. The Beaches
10. Hillcrest Village

Apart from the lowest 4 postal codes (1-4 of the preceding list), every single postal code received some score, meaning they each returned venues when prompted by the Foursquare API.

## Discussion

It should be noted that the Neighborhood Convenience Score is not the end all be all of deciding where you want to live. Just because a neighborhood is particularly convenient, does not mean it is necessarily the best place to live.

Other factors such as safety are extremely important, and the NCS does not take crime statistics into account at all.

I also relied on my own valuation of how important it is to have a grocery store in your neighborhood versus a wide variety of restaurants, as well as pharmacies and banks. Others may view these things differently. Someone who doesn't know how to cook and orders out for almost all their meals, may very well want to live in a neighborhood with more restaurants and fewer or no grocery stores.

Upon completing this course and mentioning it to a friend, I learned that there is a website called [Walkscore](#) that has a similar function, and after running my results through their site, it seems that many areas have similar ratings in both places, which gives me reason to think that the NCS is relatively accurate.

Still, in order for this to be a truly valuable tool, I think I would need to factor in crime ratings in each neighborhood. I was interested in this project because I do not live in Toronto and don't know anything about it, and I was eager to learn how well I could create a formula that tells you where the best places to live in the city are. From speaking to someone who lives in

the city of Toronto, it seems that this formula does have some success, but there are also several neighborhoods in the top 10 which are very undesirable to live in, because they have a high density of drug use, crime, or homeless populations. The NCS has not taken these things into account, and in order to do so would need access to a greater array of datasets.

## **Conclusion**

This has been a fun project to undertake. I wanted to approach this project by doing something related, but also completely different than what we have learned so far. I didn't want to just mimic the same projects we have already done, because I didn't find them to be extremely interesting, and I wanted to venture out on my own and see how I fared.

I think, because of that, this project is probably less structured than it could have been if I had chosen a more standard style in order to complete it.

I did not rely on any machine learning or inferential statistics. The data that I used was very basic, but it did achieve what I aimed for it to achieve. This is generally how I like to operate, and I think I was able to accomplish something that could be taken, improved upon, and deployed in the real world.