

Where to invest in Toronto

Alejandro Sánchez Meroño

December 13, 2020

1.- Introduction/Business Problem

A real estate investment company wants to start rehabilitation and rental projects of offices, commercial premises, apartments, and studios in different cities. The pilot project will be in Toronto. This company wants to target a young and professional public and is looking for locations where a good quality of life is combined with a large number of venues and services.

As a first step in the search for properties by this company, we have been asked to carry out a preliminary study that analyzes the number of venues and the ease of transport and communications in three aspects: ease of walking around the neighborhood, ease of moving around by bicycle, and the quality of public transport.

With the four factors combined, we will obtain a single score that will serve us to classify the different neighborhoods and that this company can use to prioritize different areas where to acquire properties and offices.

We will use Foursquare's API to obtain the values of each neighborhood, and WalkScore's API, which is a real estate-oriented service whose free version, for specific coordinates (in the U.S.A., Canada, or Australia) provides different scores for ease of movement on foot, by bicycle or public transport.

Using a commercial license of WalkScore would allow us to consider more aspects, like population density, parks, businesses... But this approach will be enough for the preliminary study.

2.- Data

Source: Wikipedia

We start from the list of Toronto's neighborhoods and postal codes that appear in Wikipedia:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- Toronto neighborhoods: Strings
- Postal codes: Strings

Source: ARCGIS

We will use ARCGIS API, through the Geocoder library, to get the values of longitude and latitude for every postal code:

- Longitude: Float
- Latitude: Float

Source: Foursquare

We will use the latitude and longitude values together with the Foursquare API to get a list of venues for every neighborhood:

- Venues: Strings

Source: WalkScore

And finally, with the objective of getting a more complete valuation of every neighborhood using the indexes of Walking, Public Transport and Bike riding easiness from the WalkScore API.

- WalkScore index: Integer
- Transit Score index: Integer
- Bike Score index: Integer

Example of an answer from WalkScore for the neighborhood of Regent Park:

```
{ "status": 1 , "walkscore": 90 , "description": "Very Walkable"
, "updated": "2020-10-04 02:40:47.082396" , "logo_url":
"https://cdn.walk.sc/images/api-logo.png" , "more_info_icon":
"https://cdn.walk.sc/images/api-more-info.gif" ,
"more_info_link": "https://www.redfin.com/how-walk-score-works"
, "ws_link": "https://www.walkscore.com/score/Regent-
Park/lat=43.65512/lng=-
79.36264/?utm_source=codigonooverbal.com&utm_medium=ws_api&utm_ca
mpaign=ws_api" , "help_link": "https://www.redfin.com/how-walk-
score-works" , "snapped_lat": 43.6545 , "snapped_lon": -79.3620
, "transit" : {"score": 94, "description": "Rider's Paradise",
"summary": "17 nearby routes: 10 bus, 7 rail, 0 other"} , "bike"
: {"score": 99, "description": "Biker's Paradise"} }
```

It has a Walkscore of 90 (Very Walkable), a Transit Score of 94 (Rider's Paradise) and a Bike Score of 99 (Biker's Paradise).

In contrast, this is the result for The Beaches:

```
{ "status": 1 , "walkscore": 74 , "description": "Very Walkable"
, "updated": "2020-11-18 22:41:38.250355" , "logo_url":
"https://cdn.walk.sc/images/api-logo.png" , "more_info_icon":
"https://cdn.walk.sc/images/api-more-info.gif" ,
"more_info_link": "https://www.redfin.com/how-walk-score-works"
, "ws_link":
"https://www.walkscore.com/score/loc/lat=43.67709/lng=-
79.29547/?utm_source=codigonooverbal.com&utm_medium=ws_api&utm_ca
mpaign=ws_api" , "help_link": "https://www.redfin.com/how-walk-
score-works" , "snapped_lat": 43.6770 , "snapped_lon": -79.2960
, "transit" : {"score": 80, "description": "Excellent Transit",
"summary": "23 nearby routes: 18 bus, 5 rail, 0 other"} , "bike"
: {"score": 53, "description": "Bikeable"} }
```

Where we get: WalkScore: 74 (Very Walkable), Transit Score: 80 (Excellent Transit) and Bike Score: 53 ("Bikeable").

Source: GitHub

We will use the Toronto GeoJSON file created by GitHub user BlizzWiz. By the way, his blog post (<https://blizzwiz.github.io/2020/06/25/Draw-folium-map-of-Toronto/>) shows very clearly the methodology to create a GeoJSON file in case we need it for another location.

This GeoJSON file will allow us to draw the boundaries of every neighborhood in a folium map.

The original BlizzWiz file only included the Postal Code, and the State ("Ontario"), besides the geographical coordinates, so I modified it to include the name of the neighborhoods as well, for ease of visualization of the map. I enclosed the final GeoJSON file (toronto_m2.geojson) on my GitHub repository (https://github.com/alsanme/Coursera_Capstone/raw/main/toronto_m2.geojson).

3.- Methodology

The target variable of our analysis is a value or score that tells us how appropriate each neighborhood is for our client to develop the planned investments in.

As the potential real estate developments are both homes and offices and commercial premises, we will not pre-screen the neighborhoods, we will consider them all as potential candidates, since, a priori, the value of some of the variables we will consider could compensate (in positive or negative) for the others. Also, this makes it possible for us to discover "surprising" neighborhoods outside the downtown area (and possibly cheaper ones).

We are talking about "neighborhoods" throughout this project, when in fact we should be talking about postal codes, many of which include several neighborhoods. However, using postal codes as a geographic unit makes things much easier for us: It allows us to easily web-scrape the Wikipedia page, reduces the number of rows in our main working dataframe, and allows us to better visualize the maps, adapting an existing GeoJSON file, and making sure that there are no areas that are too small.

Regarding the results of this study, using every neighborhood would allow us to atomize the results more, achieving greater accuracy. But this would only be useful if we could also have more data, such as population, social services, or traffic. We are confident that our client will want to invest more (pay me more!) in a second phase of the study to address all these details.

As this is a preliminary study, aimed at establishing a planning methodology that can be developed in other locations, we will address only four significant variables, which indicate us:

- Number of venues that exist in the neighborhood.
- Comfort factor for pedestrians, which measures the ease of moving on foot from one point to another in the neighborhood ("Walk Score").
- Comfort factor for bicyclists, which measures the ease of getting around on a bicycle ("Bike Score").
- Neighborhood public transportation quality factor ("Transit Score").

Once compiled the geographical information of the neighborhoods, we will obtain the information from Foursquare and WalkScore APIs, will analyze and visualize the results individually for each of the 4 aspects to consider, and finally will combine them into one single Global Score which will be an indicator of the goodness of every neighborhood regarding potential investments.

4.- Results and Discussion

After compiling and managing all the data, we have found a list of neighborhoods that seem appropriate for developing the business intended, which are the following:

Neighborhood	Score
Garden District, Ryerson	98
Richmond, Adelaide, King	97
Commerce Court, Victoria Hotel	97
Toronto Dominion Centre, Design Exchange	96
First Canadian Place, Underground City	96
St. James Town	93
Church and Wellesley	91
Kensington Market, Chinatown, Grange Park	89
Central Bay Street	88
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	88

Among these, there is a privileged location, which is Garden District - Ryerson, that scores in the Top-10 of every category (Venues, Walk, Bike and Public Transport).

If we need more detailed and accurate work, we should look for an alternative to Foursquare's API, as the information we get is very outdated. A visual inspection (via Google Maps and Google Street View) of some of the neighborhoods with very few establishments shows that there are restaurants and other establishments that are not on Foursquare.

A future continuation of this project would require specific information by neighborhood (not by zip code), considering the population of each area, and obtaining more variables to take into account to build the Global Score, such as the presence of parks, hospitals and health centers, schools, etc.

If we had economic information on real estate, rentals and yields of stores and offices, as well as satisfaction indexes of customers, professionals and residents, we could apply Machine Learning techniques to determine which of the factors considered are more determining in the selection of an appropriate location to invest.

5.- Conclusion

Data Science techniques related to geographic and social information available in repositories (public or paid) can provide a great help when making investment and entrepreneurial decisions. With the information currently available, it is possible to perform a much more in-depth analysis for any major city in the world.

To be more precise, if we did not do this type of analysis before deciding on an investment, we would have a high probability of failure. For example, in the case of a city with very high quality of life standards such as Toronto, choosing a neighborhood at random would lead us with a very high probability to choose areas with scores below 50.

