Find a community to live around University of Toronto

Sheng Gu

January 1, 2021

1. Introduction

1) Background

In the real world, people need to move from a city to another, even from one country to another country. Teenagers might have to go to another city/country for study, even for short term such as 3-5 years. Adults might have to move to another city/country for job opportunities. Moving to another city/country is always a big decision, everybody wants to live in a safe and convenient community.

2) Objective

To find a safe and convenient community to live in a new city/country, a lot of people consider following things:

A) The driving/commuting distance to the study/work place will affect the consideration.
B) Necessities: The community should be convenient for living, such as it should be close to grocery stores, gas stations, restaurants, shopping centers, …
C) Safety: Safety is above everything else. Nobody wants to live in a community with a lot of crime, find communities with lower crime rate is always a big consideration.

Of course, people might have more considerations such as close to their relatives or friends. In this report, the above 3 consideration are listed as the common considerations. Data analysis will be based on those.

3) Interest group

Different group of people might give different weights for those common considerations.

A) Teenagers for study in University of Toronto: they would think the driving/commuting distance, safety would be the most important considerations. Since the fast food restaurants are always located near the colleges or university.
B) Adults with job opportunity in University of Toronto might think all those 4 considerations are important equally.

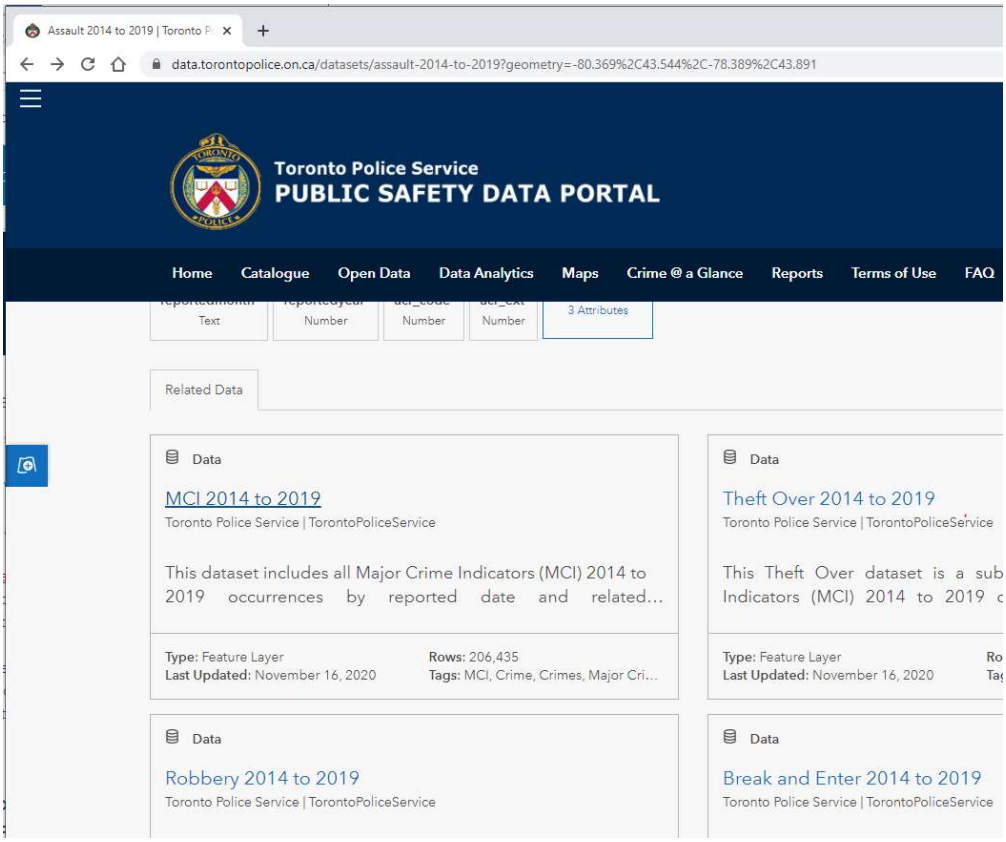
2. Data Sources and explanation

1) Safety data

Toronto Police Service has public safety data portal (https://data.torontopolice.on.ca/). The data portal provides following data:

A) MCI (Major Crime Indicators) 2014 to 2019: include Assault, Break and Enter, Robbery data.
B) Theft Over 2014 to 2019: a subset of the MCI.
C) Robbery 2014 to 2019: a subset of the MCI.
D) Break and Enter 2014 and 2019: a subset of the MCI.

The data has reported date, latitude and longitude, neighborhood information. When comparing the crime data, we could set up trending for a same community from month by month. Also, we could display the crime place (latitude/longitude) on a map and find the area/zone with lower crime rate even inside a same neighborhood.



2) **Living necessity data:**

Once we find the latitude and longitude of each community, it will be easier to find the close grocery stores, restaurants using Foursquare Interface.

3. Methodology and Exploratory Data Analysis

3.1 Latitude and Longitude of University of Toronto

When you google University of Toronto, the St. George Campus will jump on the top of searching result page. Take St. George Campus and its latitude and longitude as the latitude and longitude of University of Toronto.

3.2 Latitude and Longitude of each community

In the Toronto Police Safety Data, every crime was recorded with latitude and longitude. To make things simple, I take the average of all latitudes and longitudes in the same community as the latitude and longitude of the community.

3.3 Distance between community and University of Toronto

It is hard to find the driving/commuting distance from a community to University of Toronto. Even inside the same community, a house that is close to the driving road might have a big difference from anther house which is located the center of the community.

To keep things simple, we assume that the average of driving/commuting distance is the spherical distance from one point (latitude, longitude) to another.

In this case, we will find the latitude and longitude for each community and University of Toronto, and then calculate the distance.

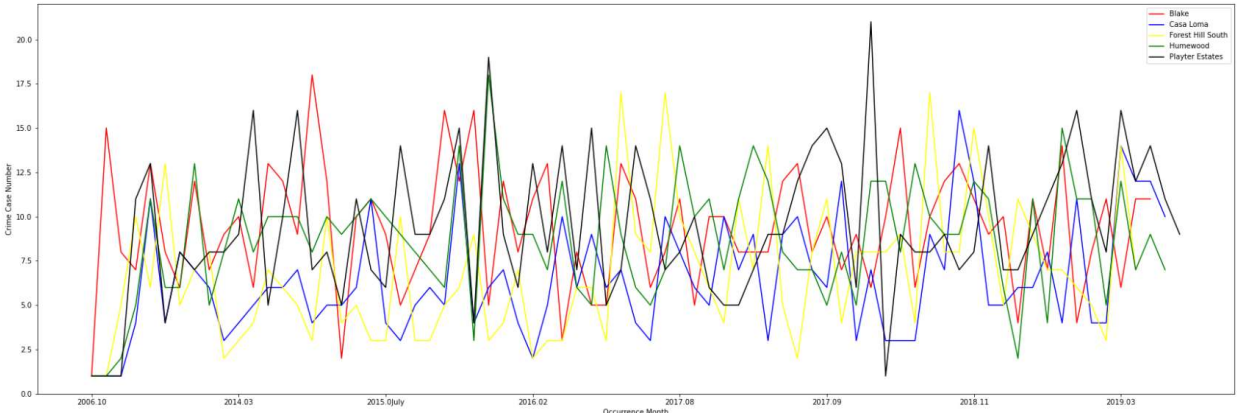
A couple of tools can calculate the distance between two points based on latitude/longitude. One tool can be seen at here. In Python, mpu (Martins Python Utilities) is used the calculate the Haversine distance.

3.4 Grocery stores and Restaurants

Foursquare API is used to get the number of grocery stores and restaurants. We suppose a community will be more convenient, if it is close to more grocery stores and restaurants.

3.5 Crime Data trending

Toronto Police public safety data has occurrence date, and crime type (Assault, Break and Enter, Robbery). Once the shortlisted communities are chosen, to check the trends, I aggregated the crime data on a monthly base and plotted them on the chart for easy visual check.



4. Discussion and Predictive Modeling

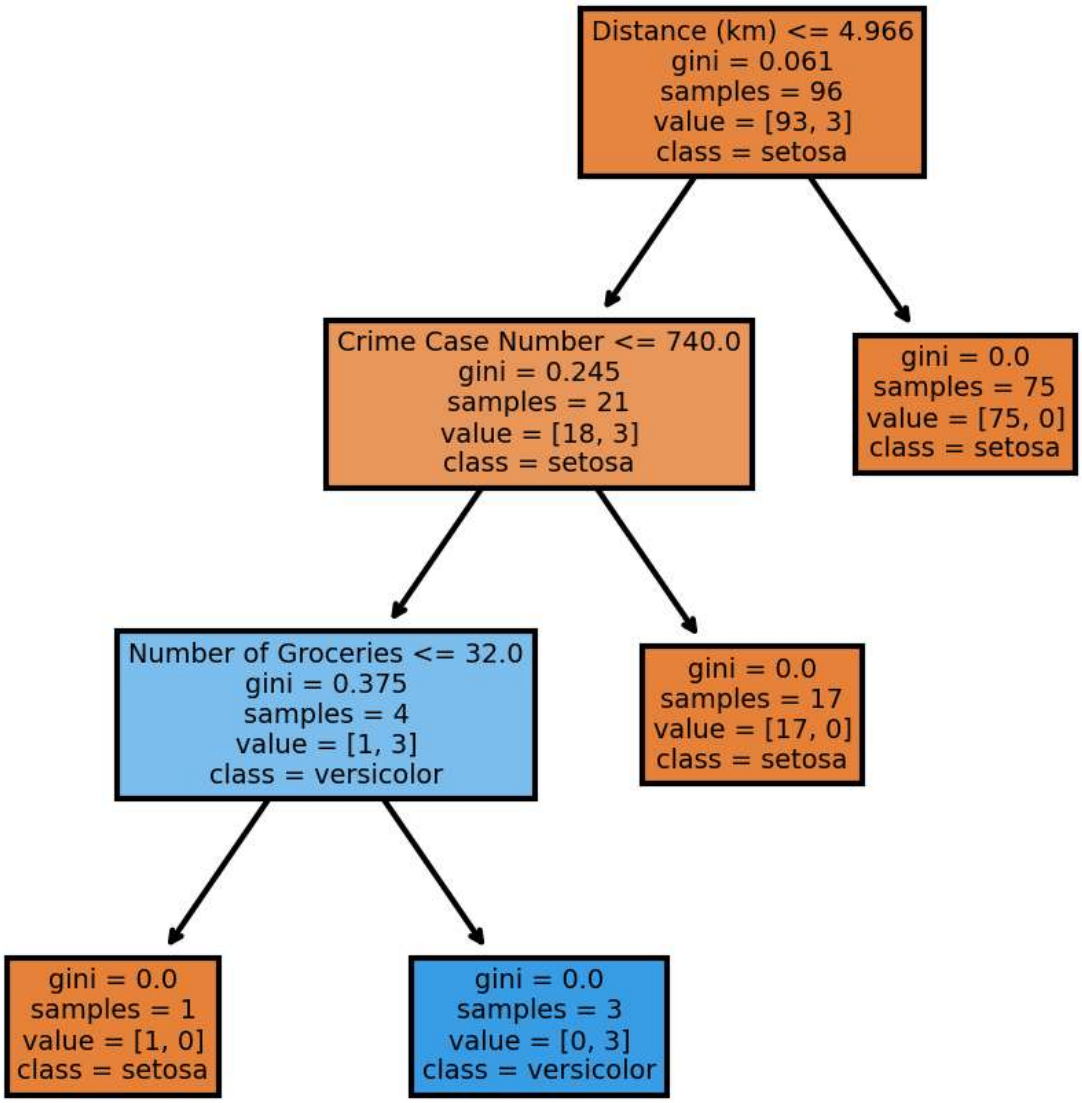
When choosing a better community to live, most people would consider following things:

1) Is a community safe to live? If a community has lower crime rate, we suppose it is safer.
2) Is a community close to the target place? Most people would like to live close to the target place, especially for those youth.
3) Is a community convenient to live? Which means it has more grocery stores to buy foods and more restaurants to dine.

If other 2 considerations are same, which community you choose depends the current consideration's information. It is just like the decision tree. I can see for this decision-making process needs 4 features of the community:

1) Distance to University of Toronto
2) Crime Rate is high or low
3) The number of grocery stores
4) The number of restaurants

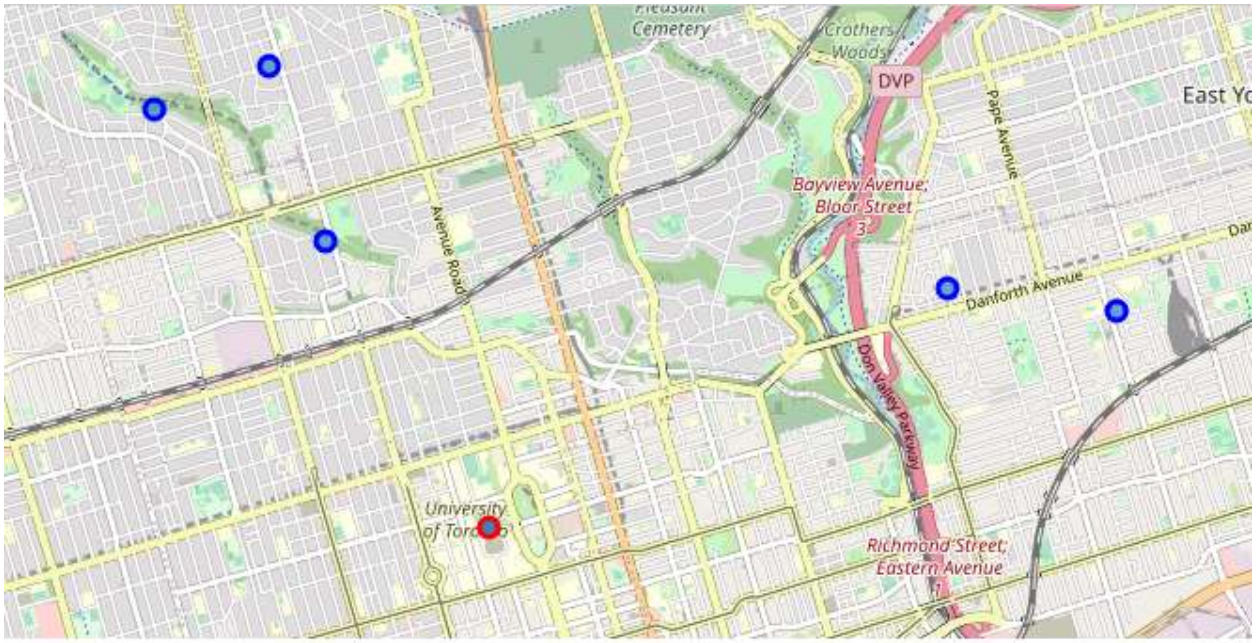
A decision tree would fit this purpose.



5. Conclusion

Basing on those 3 considerations, 5 communities are in the short-list around University of Toronto:

Community	Crime Case #	Distance to UT (km)	Grocery Stores # (within 2kM)	Restaurants # (within 2kM)
Casa Loma	480	2.418	>=50	>=50
Forest Hill South	494	3.739	40	43
Humewood	645	3.927	37	39
Blake	679	4.880	>=50	>=50
Playter Estates	707	3.807	>=50	>=50



6. Future works

House price or rental price should be considered, and the data could be found on Realtors or realtor association's web site.

The number of crime cases should not be the only factor to consider if a community is safe or not. The number of crime cases / population might make more senses.

Commuting distance is more complicated than the spherical distance from one point (Latitude/Longitude) inside the community to the target place.