

# Toronto House Hunting

## Introduction

A house, or simply a home is a place everyone longs for at the end of hectic working day. Today with the ravages of the COVID-19 still fresh in our minds, the home has been turned into the new office of the twenty first century. Today we work and stay at home. It therefore becomes very important to carefully choose the best place that might work as both as a home and a work place, and for those working outside of their homes, it's also paramount to choose a safe neighborhood with good commuting distance.

Toronto like many developed cities in the world, house hunting is a full time job. Even with agents, at least you will have to go in for viewing. Also many listing websites do not provide enough details for one to make an informed decision. For example which neighborhood is safe. Though provide a map view of the properties, these lack comparisons amongst neighborhoods. As a new young professional

As a young professional with a limited budget, it becomes hard to find a good house that make a good home in a convenient neighborhood without having to arrange hundreds of viewings. This assignment, I will cluster house/apartment give a commute distance, crime rate of the neighborhoods, and the amenities around neighborhoods to filter house and reduce of the number of viewing.

## Dataset

The first step will be to write a scraper to collect some basic information:

1. Property address
2. Property price
3. Number of rooms
4. Number of bathrooms
5. Property type

To obtain this data, I will scrap [LINK](#) that list hundreds of properties across Canada.

## Enhancing the Data

1. Use geopy library, I will retrieve the latitude and longitude coordinates using the property address.
2. Using the foursquare API, I will retrieve the amenities around each property
3. I will also use the crimes data to see which neighborhood has the highest crime rates.

## Methodology

### Data cleaning

The properties dataset obtained contained upto 346 rows and five columns see figure below;

	Address	number_of_bed	number_of_bath	property_kind	price
0	186 Gooch Ave, Toronto, Ontario	3 BedsBds	2 BathsBa	Residential	0
1	7 Grenville St, Toronto, Ontario	2 BedsBds	2 BathsBa	Residential	838000
2	51 East Liberty St E, Toronto, Ontario	1 BedBd	2 BathsBa	Residential	659000
3	35 Mariner Terr, Toronto, Ontario	1 BedBd	1 BathBa	Residential	749900
4	4968 Yonge St, Toronto, Ontario	2 BedsBds	1 BathBa	Residential	598000

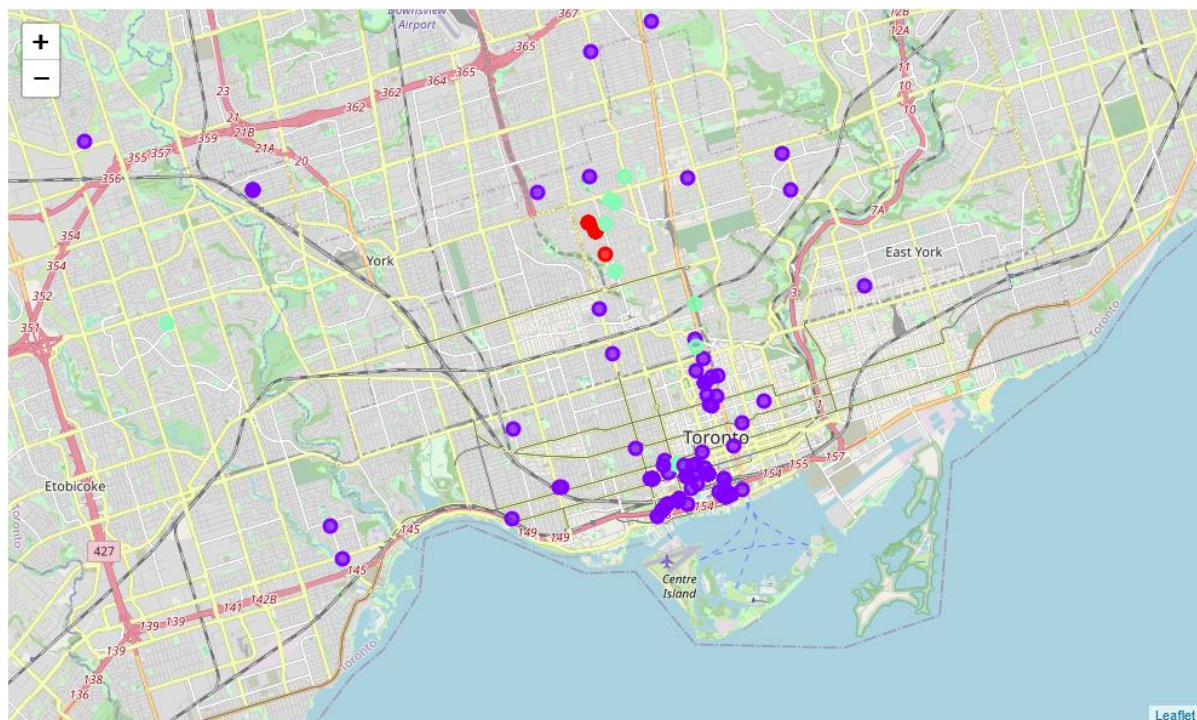
Let's clean up our data abit, we will perform the following operations

- Remove the text in the **number\_of\_bed** and **number\_of\_bath** columns to make sure they numeric
- Convert the **property\_kind** column to numeric categories
- Drop rows where the price is not listed or is zero

	Address	number_of_bed	number_of_bath	property_kind	price
1	7 Grenville St, Toronto, Ontario	2	2	1.0	838000
4	4968 Yonge St, Toronto, Ontario	2	1	1.0	598000
5	131 Markham St, Toronto, Ontario	2	3	1.0	1089000
6	115 Long Branch Ave, Toronto, Ontario M8W0A9	2	3	2.0	799000
7	523B ROYAL YORK RD, Toronto, Ontario M8Y2S5	3	4	3.0	1269900

## Discussion

In the map, the purple points represent the low cost houses, the green points represent the average cost houses and the red points represent the expensive houses.



## Conclusion

This way, we are able to narrow down the house from 364 house to few average house that we can start viewing.