**Finding the Most Affordable Homing in Toronto**

Dhruv Jain

April 11, 2020

# **Introduction**

## **Background**

Buying a housing is one of the toughest decisions in a human’s life. Houses don’t come at a cheap price, and it is usually a long-term investment. A person should carefully choose where to buy a house, because if someone doesn’t like their house, it isn’t easy nor cheap to move to a new one. Buying the right house can be a more difficult choice than choosing which university to attend.

The cost of houses in Toronto in 1985 was $109,094. In 2015, that number rose to $566,696. In February 2020, the cost of the same house is $852,900, growing a whopping 60% in the past 5 years. The change in house prices between February 2019 and February 2020, itself, was 10.15% [1][2]. If the same trend continues, the housing market will break the 1 million mark within the next 2 years (5 years max, if the market falls due to the COVID-19 pandemic).

Therefore, selecting the right house is a very important decision and it would be better if the person lives in the neighborhood where whatever they want is at their disposal and the price of his house is affordable.

## **Problem**

This project aims to categorize the neighborhoods based on the venues and services around it and choose the best neighborhood according to the affordability of the houses and the venues available around it.

## **Interest**

People who want to buy a house in Toronto and want to compare the neighborhoods on the services they have would be very interested. Researchers who are finding reasons for the growth in market may also be interested.

# **Data Acquisition and Cleaning**

## **Data Sources**

The housing price of the neighborhoods (HOUSES) in Toronto were obtained from [here](https://toronto.listing.ca/real-estate-prices-by-community.htm), and the coordinates of the neighborhoods (LOCATION) in Toronto were obtained from [here](https://open.toronto.ca/dataset/neighbourhoods/). Foursquare was used to get the data of the venues in the neighborhoods.

## **Data Selection and Cleaning**

The data was download and scraped from these websites and combined into a table. It was found that one neighborhood has various property types and the properties has various house types. The data scraped from the Real Estate website had a lot of missing data, mainly because some neighborhoods do not have a certain type of property or a certain type of house. For example:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Neighborhood | House Type | 1 Bedroom | 2 Bedroom | 3 Bedroom |
| A | Condos | 10 | 20 | 30 |
| A | Detached Homes |  |  | 100 |
| B | Condos | 30 | 60 |  |
| B | TownHouses |  | 50 | 75 |

*Fig 1: Example of the Data structure*

To reduce the loss of data when rows with No available values would be drop, only houses with 2 and 3 bedrooms were chosen. This was mainly because people who are looking to buy houses would mostly comprise of individuals and families and they would prefer to live in a 2/3 bedroom house rather than a smaller or larger house. And also, it is the most common type of housing layouts.

Secondly, not all neighborhood names in the HOUSES dataset matched with the ones in LOCATION. A script was written to extract the neighborhood names that were not an exact match and it was examined. The names which were similar in both datasets (only differed in symbols and spacing) were changed in the HOUSES (to match the ones in LOCATION). Dissimilar names were researched manually, and it was found that the neighborhoods were same, only the names differed, so that was changed according.

It was also found that some neighborhoods in the HOUSES dataset were further divided into 2 parts (for example: Neighborhood Rouge was divided to RougeE10 and RoughE11). The average of the house prices in these 2 neighborhoods were taken, assigned to the first neighborhood, and the other one was dropped from the dataset. Then the first neighborhood’s name in the HOUSES dataset was changed corresponding to the same one in the LOCATION dataset.

After the data was processed and cleaned, it was found that the dataset had 140 neighborhoods and the total entries of houses were 454, implying that every 1 out of 4 neighborhoods have 4 types of properties, (being Condos, Condos Townhouses, Townhouses and Detached Homes) while the rest only have 3. Foursquare data will be used to collect the venues around the particular neighborhoods.

References:

[1] - <https://business.financialpost.com/personal-finance/mortgages-real-estate/now-and-then-do-canadian-homes-really-cost-that-much-more-than-30-years-ago>

[2] - <https://creastats.crea.ca/en-CA/>