

# **Recommendation System for Apartment Seekers in the GTA**

Troy Gloyn

July 27, 2020

## **Introduction**

Finding an ideal location for a house or an apartment in the general Toronto area can be a real time-consuming and tedious task that often results in numerous listings falling short of customer preferences. The purpose of this project was to speed up the process of seeking that ideal living location by using Foursquare API venue location data and a recommendation system that inputs user preferences to generate more customized location suggestions for potential buyers/renters.

The advantages of having such a recommendation system can dramatically reduce potential options that a customer must look through, saving time for both the seller/renter of the property and the customer. Furthermore, it gives customers more personalized options by having their desired venues in close proximity to the property.

## **Data**

To generate this recommendation system, a list of neighborhood's in Toronto was first web scraped from a Wikipedia website ([https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)). This information was necessary as it provides all possible options for someone looking for a house/apartment in Toronto. It was then required to gather geospatial data for each neighborhood, which was obtained from a CSV file ([http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)). The geospatial data of each neighborhood was then used in API calls to Foursquare, obtaining a list of all possible venues within those neighborhood locations (limit=100, radius=500m).

## **Methodology**

After importing necessary packages and retrieving both the neighborhood and geospatial data, the information was merged into a single data frame for easier processing. The neighborhoods were clustered into 10 respective Boroughs using the geospatial data, and mapped, providing more organization for subsequent recommendation system analyses. As mentioned above, location data for each neighborhood was used to make Foursquare API calls, retrieving venue information such as name, latitude, longitude, and category. Since category type was the information most pertinent to prospective buyers/renters, it was used to characterize what each neighborhood offered. The categories were manipulated into binary code information, containing a '1' if such as category was present in a particular neighborhood or a '0' if it was absent. The information was summed across each neighborhood to get a score for each category present.

Since there was no specific user being referred to, a random dataset containing the weight from 0-1 (0 being not important, and 1 being necessary) of each category was generated as a reference for the recommendation system. The user weighted dataset was multiplied by each neighborhood row category, summed, and then divided by the user sum to achieve an average similarity score of each neighborhood to the user's preferences. The similarity score data was sorted into descending order to view the top 5 most similar neighborhoods to the user data. This information is then used to guide the hypothetical buyer/renter to look for properties within these locations.

## **Results**

The results indicate for this particular randomly generated user that the most ideal location for a house/apartment would be in the Harbourfront East/Union Station/Toronto Islands neighborhoods (similarity score=0.426), followed by Commerce Court/Victoria Hotel (similarity score=0.425), and Stn A PO Boxes (similarity score=0.418).

## **Discussion**

The results provided illustrate the potential use of this recommendation system. The generated user would likely benefit from seeking places to live in the top 5 most recommended neighborhoods for him/her. The recommendation system can be further extended to include other venue information such as ratings, which would allow for better quality venues to be filtered for the user.

## **Conclusion**

The recommendation system was able to provide a hypothetical user with likely neighborhoods they would enjoy living in by utilizing that user's weighted preferences score for each category of venue, and comparing it to the score for the venues from each neighborhood. This process would have the potential to limit the amount of time required by prospective property buyers/renters to find a place that fits within their desired preferences, in addition to giving them a more customized location, thus improving the efficiency of the process for both the customer and seller of the property.