**Real Estate Neighborhood Recommendations:**

**City of Toronto**

Sergio Conde

07/28/2019

**Introduction:**

A common marketing technique is customer segmentation. Identifying the type of customers you have and their needs. However, once this is done, companies must identify the best product for customers. For real estate companies, this means identifying homes that tailor to customer needs. But how can companies narrow down such homes?

Data is the answer! By collecting data from cities and their neighborhoods, real estate companies can create better informed product recommendations for their clients. In this report, I will focus on the city of Toronto and attempt to identify neighborhoods that tailor to specific real estate customers.

**Methodology:**

**I) Approach:**

I will be using neighborhood data of Toronto found on Wikipedia, and venue data extracted from the Foursquare API. Since Wikipedia only has zip codes and I need coordinates for the Foursquare API, I will need to use a geo location python library to obtain coordinates of each neighborhood. Once coordinates are obtained, I can use Foursquare to get venues in each neighborhood. With venue data, I can then use a clustering machine learning algorithm to separate neighborhoods into 3 distinct segments. Don't want too few or too many segments, so I found 3 to be an appropriate amount for the possible real estate customers to look at.

**II) Exploratory Analysis:**

With clean and correctly formatted data, I went ahead and plotted each neighborhood in Toronto. I observed that there is a stronger neighborhood concentration towards the downtown Toronto area. As you get farther from downtown neighborhoods become more spread out. Thus, one can expect higher venue concentration around downtown than from the outskirts. As a matter of fact, 61% of neighborhoods have less than 50 venues and thus less busier neighborhoods seem to be slightly more common.

In total there are 4361 venues collected from FourSquare, with 338 unique venue categories in all of Toronto. It is evident that park venues are very common in Toronto, which is great to see! It always nice to live in a green city. Other than parks, however, other top venues consist of coffee shops, cafes, and pizza places. In fact, coffee shops are in almost every neighborhood’s top five venues! The people of Toronto must be big foodies.

To further explore the venue distribution throughout the city, I made a choropleth map using venue counts per neighborhood. The map showed, as expected, that there is more overall venue activity around the downtown area. Although, there are randomly distributed busy neighborhoods throughout Toronto. Another thing that stands out is that busy venue neighborhoods seem to have calmer neighborhoods around them. Super busy neighborhoods (99+ venues) are usually not found next to each other.

**Results:**

The results of the clustering algorithm gave me tree possible neighborhood recommendations for real estate customers:

* The first cluster (0) seems to contain tourist friendly neighborhoods. This cluster contains frequent hotels, bars, coffee shops to keep tourists going, greater food venue diversity, and entertainment such parks, zoos, and theaters. It is perfect for customers looking to invest in airbnb, hotel, and other tourist related properties. It could also be perfect for people looking for a fast-paced lifestyle, like young adult singles.
* Cluster 1 in the other hand represents small town atmosphere type of neighborhoods. They have the life essentials such as supermarkets, pharmacies, and food places, nothing crazy out here. Thus, it is perfect for customers looking for homes with a small-town feel.
* Cluster 2 has an interesting mix. It contains neighborhoods with lots of food places. They have entertainment here and there such as parks, bars and shopping malls, but not as common as cluster 0. They have some frequent grocery stores, and pharmacies. These neighborhoods are a mix of small town and tourist friendly neighborhoods! If you live in the outskirts of a major city, like me, these neighborhoods would definitely seem familiar. These mix-use residential areas could be perfect for customers looking for a calm, but with entertainment nearby, type of neighborhood. Perfect for families. This cluster is also relevant to Airbnb and real estate investors looking to provide a local/residential type of vacation to its customers due to semi-tourist neighborhood status!

**Discussion:**

With this data, real estate companies in Toronto can rapidly pinpoint best fit neighborhoods based on their client needs! It gives them an idea on where to first look for their customers dream place. They can also use the data to help potential business owners locate the best neighborhoods to set up shop based on competition (types of venues nearby) or foot traffic (amount of venues).

Although this is a good place to start, I can see this project being perfected in the future. It can be improved by narrowing down to smaller neighborhood blocks for recommendation. Currently neighborhood recommendations are based on postcode values, which can be fairly large areas. This improvement could be possible if the venue data could be enhanced with better venue location services, as the FourSquare API free version is very limiting to the amount of venues that can be obtained. I believe that the lack of enough venue data per neighborhood was the biggest source of bias in my data.

Things I learned:

* Need better data cleaning techniques as wrongly formatted data would appear out of nowhere and forced me to correct it on the spot. Should have done better analysis of the data as soon as I got it from geolocation server, or foursquare to spot error sooner. But as my first data science project, I guess this is all part of the learning process.

**Conclusion:**

Data can help businesses find insight into different type of markets, products, and customers, as shown in my neighborhood data analysis of the city of Toronto. Using FourSquare, geolocation libraries, and publicly available data of the city of Toronto, I was able to find insights on the types of venues available per neighborhood in the city. I was able to cluster these neighborhoods based on their available venues by using a K-Means Clustering machine learning algorithm. These 3 clusters are of the following designation:

* Cluster 0: Tourist Neighborhoods
* Cluster 1: Small Town Neighborhoods
* Cluster 2: Semi-Tourist Residential Neighborhood

Using these groups of neighborhoods, real estate companies in Toronto can make better neighborhood recommendations to their clients, be it for personal home usage or for real estate business investment.