Property Partner Final Report

Eshan Ratnayake

Cousera Capstone Project

December 2, 2018

Introduction

Canada is a country well known for its multiculturalism and diversity. Toronto is one city in particular where one can find communities, food, and venues of various ethnic groups. Immigrating to a country like Canada however can be a very stressful and intimidating task, specifically choosing the best location to purchase a residence. There are many factors that one must look into when buying a house, and it can be tough to make a decision while analyzing all these factors. To make things easier for new immigrants to Canada, I've proposed a solution that can help make their lives easier. What I will be doing is taking 3 factors (price, distance to work, and nearby venues) into account to help people figure out the best location to purchase a house.

Data

The data that I will be using will be from a dataset I've obtained from Kaggle (https://www.kaggle.com/mnabaee/ontarioproperties).

	Address	AreaName	Price ($)	lat	lng
0	86 Waterford Dr Toronto, ON	Richview	999888	43.679882	-79.544266
1	#80 - 100 BEDDOE DR Hamilton, (	Chedoke Park B	399900	43.25	-79.904396
2	213 Bowman Street Hamilton, ON	Ainslie Wood East	479000	43.25169	-79.919357
3	102 NEIL Avenue Hamilton, ON	Greenford	285900	43.227161	-79.767403
6	#1409 - 230 King St Toronto, ON	Downtown	362000	43.651478	-79.368118
7	254A Monarch Park Ave Toronto,	Old East York	1488000	43.686375	-79.328918

This data set shows the address, area name, prices and latitude/longitude coordinates of different properties that were sold in Ontario. The housing prices in different areas is a key factor to new immigrants to familiarize themselves with so they can determine which ones best suit their budget. Distance from work is another factor that needs to be considered to avoid long commutes. I will also be leveraging the Foursquare API to analyze all the nearby venues for different neighbourhoods. Using this API, I will filter out venues with keywords that match the ethnic background of the immigrant. I believe if a neighbourhood has many venues that are similar to an immigrant's country of origin, and can help them feel more comfortable, and also help find individuals from their ethnic community to associate with.

Methodology

In the scenario I chose to use in my solution, there is a gentleman from China named Ping who is immigrating to Canada. He got a job at the TD Tower in Downtown Toronto and is looking for a good location to buy a house. He wants to explore the housing prices in different areas within Toronto, and see which areas are close by to his work. In addition, he would like to find a place that is close to a lot of Chinese venues, as he will feel more comfortable, and will find it easier to find people from China.

The first thing I did was clean up the csv file that I downloaded from Kaggle. Next, filtered out the data to only show entries of properties that were sold in Toronto. I did that by filtering out the address and area name for the word 'Toronto'. Then I calculated the average housing price for each neighbourhood and averaged out the latitude and longitude. Then, I used the Foursquare API to find all the venues that were close to each of those neighbourhoods. After, I parsed the 'Venue Category' and 'Venue Name' attributes to see if words like 'China', 'Mandarin', and 'Cantonese' were found, and added a column of the number of Chinese venues per area. Then, I computed the distance from each neighbourhood to the TD Tower. Once that was completed, I computed the Euclidean distance of the Average Area Price and the Distance to Work. I normalized the Avg Price by dividing it by 1000. Finally, I used the number of Chinese venues as a scale factor, and divided the Euclidean Distances by their corresponding number of Chinese venues.

Results

According to the final dataset that was generated, the best neighbourhood for the Chinese gentleman to live in is Amesbury, as it has the lowest score.

Out[28]:

	AreaName	Area Latitude	Area Longitude	Avg Price	# of Chinese Venues	Distance to Work (km)	Euclidean Distance	Final Score
0	Amesbury	43.70	-79.48	79450.00	1.00	10.90	80.19	80.19
1	Agincourt	43.79	-79.28	425287.29	4.00	18.68	425.70	106.42
2	Tam O'Shanter	43.78	-79.30	453432.50	4.00	17.61	453.77	113.44
3	Bridlewood	43.80	-79.32	344406.89	3.00	18.37	344.90	114.97
4	Steeles	43.81	-79.32	481273.50	3.00	20.19	481.70	160.57

Discussion

According to the data, I would recommend choosing Amesbury at it appears to have the best balance between price, distance from work, and accessibility to Chinese venues. However, despite Agincourt having a high Euclidean distance, it is still worth considering it as an option because there are 4 Chinese venues in that neighbourhood while Amesbury only has 1. Amesbury, however, is a lot cheaper and closer to work.

Conclusion

Overall, I believe Amesbury and Agincourt are the best options for Ping to choose. Despite having the best score, one thing the Amesbury severely lacks is Chinese venues. Agincourt is being more expensive and further away from work, there are a lot more Chinese venues in this area.