IBM Data Science

Capstone Project – Toronto Supermarket

Mustapha Ziade

I. Introduction

Peter just graduated from university and wants to start his own business. In fact, Peter wants to expand his family business of Supermarket. While his family business is operating in his hometown, Peter decides to expand the business in Toronto, where he studied. With his expertise in this industry, in addition to his knowledge of the market, and his relationship with suppliers, Peter is expecting the new business to be very successful. However, he is not sure in which area to open the supermarket. This project is intended to help Peter identify the best location for his business. Each neighbourhood on Toronto will be studied separately. There will be two main criteria to determine the score of neighbourhoods; competition which will be the number the supermarket in the area, and the demand which will be the population of the area. In the end, the neighbourhood with the best score will be the one to be advised for opening the business.

II. Data

As there are two criteria to determine the score (competition and demand), two data sources will be needed:

First, the competition is determined by the number of supermarkets in a neighbourhood. This information will be collected from Foursquare location data. Venues will be searched for the term “Supermarket” while having the latitude & the longitude fixed to the middle of each neighbourhood and having a radius of 1000. This data will tell how many nearby supermarkets each area has.

Second, the demand is determined by the population of the neighbourhoods. This data will be retrieved from Wikipedia. The following link is the page that will be used to get the name of neighbourhoods and their population.
https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods
The below table is part of the main table on the Wikipedia page.

Name	FM	Census Tracts	Population	Land area (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit Commuting %	% Renters	Second most common language (after English) by name	Second most common language (after English) by percentage	Map
Toronto CMA Average		All	**5,113,149**	**5903.63**	**866**	**9.0**	**40,704**	**10.6**	**11.4**			
Agincourt	S	0377.01, 0377.02, 0377.03, 0377.04, 0378.02, 0378.08, 0378.14, 0378.23, 0378.24	44,577	12.45	3580	4.6	25,750	11.1	5.9	Cantonese (19.3%)	19.3% Cantonese	
Alderwood	E	0211.00, 0212.00	11,656	4.94	2360	-4.0	35,239	8.8	8.5	Polish (6.2%)	06.2% Polish	
Alexandra Park	OCoT	0039.00	4,355	0.32	13,609	0.0	19,687	13.8	28.0	Cantonese (17.9%)	17.9% Cantonese	

From the table, the name of neighbourhood and their population is the only data needed. The rest will be dropped.

III. Methodology

IV. Results

V. Discussion

VI. Conclusion