

NEW BUSINESS LOCATION RECOMMENDATION SYSTEM

The location used for this project is Toronto city, and the business is a fast food chain

BY MJIDIBA

INTRODUCTION

Hi everyone. In this project we will tackle one of the problems mentioned in the instructions for the assignment, which is a recommendation system of where to open a new business.

If someone is looking to open a business in the city of Toronto, how can we recommend the best neighborhood to open his business in? The idea is that we want to open a business in a neighborhood where that venue type is least common to maximize profits and avoid competition.

For this problem, we will take as an example an investor who wants to open a fast food chain in the city of Toronto. We will give at the end our recommendation of where to open the restaurant, and where not to open it.

DATA DESCRIPTION

The data used for this project will be obtained from Foursquare (<https://foursquare.com/>) and from a Wikipedia page containing the list of postal codes for Canada (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). The Information about the neighborhoods as well as popular venues within them will be obtained using Foursquare API.

METHODOLOGY

First of all, we will extract the information from the Wikipedia page, which is a table containing 3 columns: Postcode, Borough and Neighborhood. We will be using for this the 'BeautifulSoup' library. Then we will perform some modifications so as to have one line for each postal code, meaning that we will group neighborhoods of the same postal code, and fill the empty ones with their corresponding Borough

Next, we will search for coordinates for each postal code in Toronto and add them as two separate columns to the table. We will be using for this an already prepared csv file that contains each postal code with the corresponding coordinates.

This what looks our data like when plotting it in the notebook using the folium library

First of all, using the Foursquare API, we were able to extract the venues within 1km of each set of neighborhoods with a certain postal code. Then, we only left the venues with 'Fast Food Restaurant' as category. The resulting dataframe is the following table:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	Parkwoods	43.753259	-79.329656	KFC	43.754387	-79.333021	Fast Food Restaurant
102	Queen's Park	43.662301	-79.389494	Wendy's	43.662820	-79.383379	Fast Food Restaurant
116	Rouge, Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
130	Woodbine Gardens, Parkview Hill	43.706397	-79.309937	Harvey's	43.710730	-79.308838	Fast Food Restaurant
133	Woodbine Gardens, Parkview Hill	43.706397	-79.309937	Harvey's	43.708136	-79.314105	Fast Food Restaurant
191	Ryerson, Garden District	43.657162	-79.378937	Bourbon St. Grill	43.655135	-79.380823	Fast Food Restaurant
258	Flemingdon Park, Don Mills South	43.725900	-79.340923	Harvey's	43.726618	-79.340997	Fast Food Restaurant
446	Caledonia-Fairbanks	43.689026	-79.453512	KFC	43.690647	-79.456326	Fast Food Restaurant
601	Bathurst Manor, Downsview North, Wilson Heights	43.754328	-79.442259	Dairy Queen	43.755680	-79.440166	Fast Food Restaurant
747	Dovercourt Village, Dufferin	43.669005	-79.442259	McDonald's	43.667942	-79.439046	Fast Food Restaurant

There is a little problem that we are facing. When we just created the dataframe, we had 2240 venues, but when we only left fast food restaurants, we only had 38 venues. This means that we have now at most 38 neighborhoods in our dataset.

So, instead of looking for the neighborhoods for the least competition, we will be looking for the neighborhoods with the most intense competition to avoid. We will be looking for the neighborhoods with the highest number of fast food chains present in it. The result is:

Neighborhood	
Fairview, Henry Farm, Oriole	5
L'Amoreaux West, Steeles West	3
Woodbine Gardens, Parkview Hill	2
Bedford Park, Lawrence Manor East	2
Church and Wellesley	2
Stn A PO Boxes 25 The Esplanade	2
Humber Bay Shores, Mimico South, New Toronto	1
Willowdale South	1
The Beaches West, India Bazaar	1
Ryerson, Garden District	1
Name: Venue, dtype: int64	

So, I would recommend to this investor to avoid the neighborhoods Fairview, Henry Farm, Oriole, L'Amoreaux West and Steeles West, because they contain a total of 8 fast food chains.

CONCLUSION

I think that the methodology that we used in this project is rather simple and still lacks a lot of things. First of all, we should not only examine the fast food chains, but also the restaurants that are present in the area. Another thing that we could have done is see which type of fast food our investor will be selling, and give more value to the restaurants that have the same business. For example, if the investor is selling burgers, we would put more emphasis on fast food restaurants like 'Burger king' and 'McDonald's'. One last thing we could have is to examine the population within each neighborhood, higher population meaning higher profits.

This methodology can also be applied to other businesses like starting up a hotel or a supermarket, we just need to change the filtered category we used in the notebook. This methodology however is not applied to smaller cities, where there are less venues in each

neighborhood. If we have an average less than one similar venue in each neighborhood like in our case, it will not be reflective of the true reality.