Find the suitable locations for opening new restaurant in the Toronto

Le Thanh Lam

February 6, 2020

1. Introduction

1.1 Background

Toronto, city, capital of the province of Ontario, southeastern Canada. It is the most populous city in Canada, a multicultural city, and the country's financial and commercial centre. Its location on the northwestern shore of Lake Ontario, which forms part of the border between Canada and the United States, and its access to Atlantic shipping via the St. Lawrence Seaway and to major U.S. industrial centres via the Great Lakes have enabled Toronto to become an important international trading centre.

1.2 Problem

Our company is XYZ JSC and I am one of data scientists in this company. Our company is planning to open new food restaurant Toronto and ask me to answer question for them: which location is the best place to open new restaurant in Toronto?

1.3 Interest

Perhaps where to open restaurant in the Toronto depends multiple factors to have successful result of business. And most important factor is location to open restaurant. It depends on some follow factor: density of population, status distribution of restaurants based on locations... In this project we are going to discuss and solve how to recommend a the best stuitable location for opening new restaurant in Toronto for the company XYZ JSC.

2. Data acquisition and cleaning

2.1 Data sources

- Data set of borough in Toronto at url https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M wiki page to get all the information about the neighborhoods present in Toronto. This page has the postal code, borough & the name of all the neighborhoods present in Toronto.
  We will scraping data from this url and use following data of districts of Torronto: postcode, boroughm, neigbourhood

- Dataset from url “https://cocl.us/Geospatial_data” csv file to get all the geographical coordinates of the neighborhoods in Toronto We will scraping data from this url and use following data properties for each location of Torronto: latitude, latitude,postcode. Combine this step with step b we could determine location lat,lng for a district based on postcode
- To get information about the distribution of population at url: https://en.m.wikipedia.org/wiki/Demographics_of_Toronto#Ethnic_diversity) to evaluate density of population each district in Toronto We will use this information like useful to evaluate some idea usefully for demanding people for a retaurant.
- Data location from API Foursquare.com to get information about current restaurant in values of Toronto.
- Data API is regarding to Location Data on Internet

I use data via api to get essetial information base on location pupolcation of a given location

2.2 Data cleaning

Data downloaded or scraped from multiple sources were combined into one table: via scraping from wikipedia and via online api.

There are several problems with the datasets. There were a lot of missing values or incorrect values from data. First, some location is named by short name whileas other were identified by full name . So it leads difficultly when using name to combine between data population and.

Second, some name of location is empty in popuplcation dataaset. Therefore, these item with empty names were removed.

Thirdly, the data are heterogeneous and complete in terms of location data, there are some places with some data while other is no, so I have to use some free api on internet to fill the missing values.

Finally, restaurant data from API FourSquare has too many categories and including non-food restaurant (such as car repair, golf…) so I have to define some keywords to remove non-food restaurant types in order to be fit requirements of Company XYZ JSC

2.3 Feature selection

Before cleaning data there are 1217 samples and 10 features in the data. And analizing above I cleaned data and get final datasets consit 157 restaurants at 12 locations with 5 features

3. Methodology

We will download the data source in Toronto from Wikimedia, standardize the features, then based on FourSquare's API to perform calculations based on the number of restaurants and the number of people in each location to find out. number of residents and restaurants at each location.

To be sure of the potential location, I use statistical methods : visuallization and query data based on pupolcation and restaurants at each location. On by graph and correlation graph to see the current correlation between quantity and population at each location. And finally I visuallize the number of current restaurants on the map to see how they are distributed at each location (to avoid the case of price for a number of restaurants in some cases: many locations are few populated but large in size or densely populated but narrow in area)

Finally, combined with data query visuallizations: select the 3 most potential locations for opening new restaurant based on a reasonable percentage of population and restaurants.

4. Result

4.1 Number of restaurants and pupolation each location:

After downloading the data, standardizing the data in Toronto city locations, we have a dataset: including number of the restaurants in each location as well as the corresponding total of population in each location.

	Neighborhood	Restaurant	Population	Lat	Lng
0	Beaches-East York	12	911315.0	43.676357	-79.293031
1	Davenport	4	100420.0	41.523644	-90.577637
2	Eglinton-Lawrence	14	114530.0	43.719265	-79.429765
3	Etobicoke Centre	10	254835.0	43.679833	-79.538993
4	Etobicoke North	32	202675.0	43.741093	-79.589225

4.2 Visualization to see relationship between population and number of restaurants at each location by bar graph

To see in visualization, we will model the bar graph to see the correlation and show the most potential locations

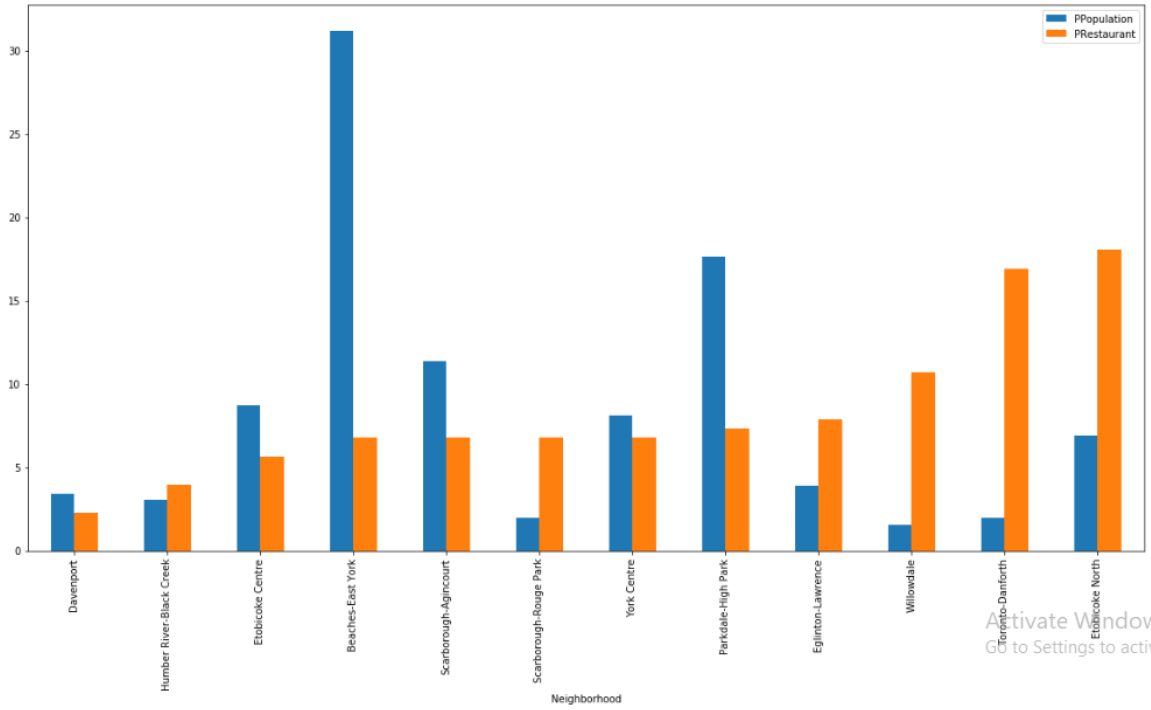

Figure 1. Graph for percentage of number of restaurants and pupolations each location

4.3 See correlation in visuallization to see list of potential locations in Toronto

We use the graph to see the correlation between the population and the number of restaurants in each location, we can see the current distribution of restaurants by locations is not same with the number of people. The number of locations is high, some places are highly populated but the number of restaurants is very low, these are the places XYZ company needs to consider the potential for restaurant investment.

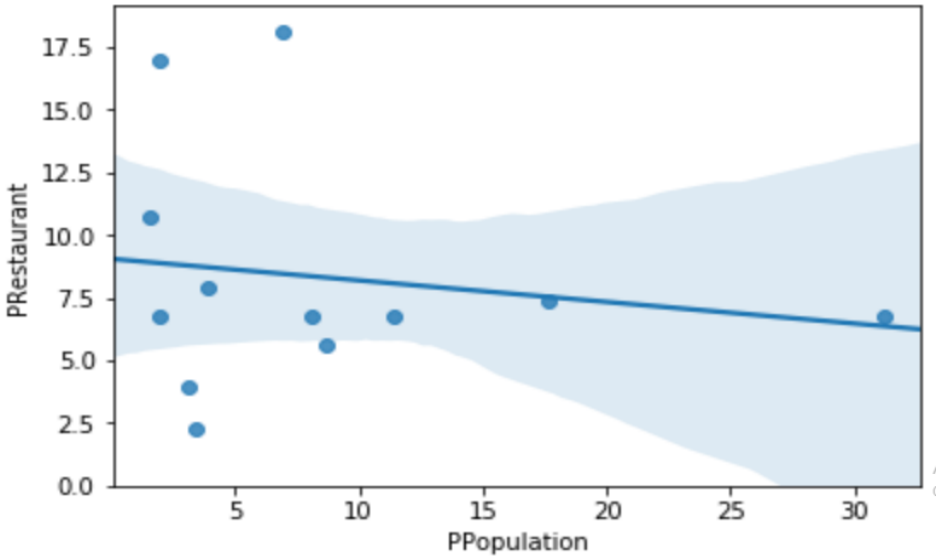

Figure 2. Correlations between Percentage of Restaurant and Populcations

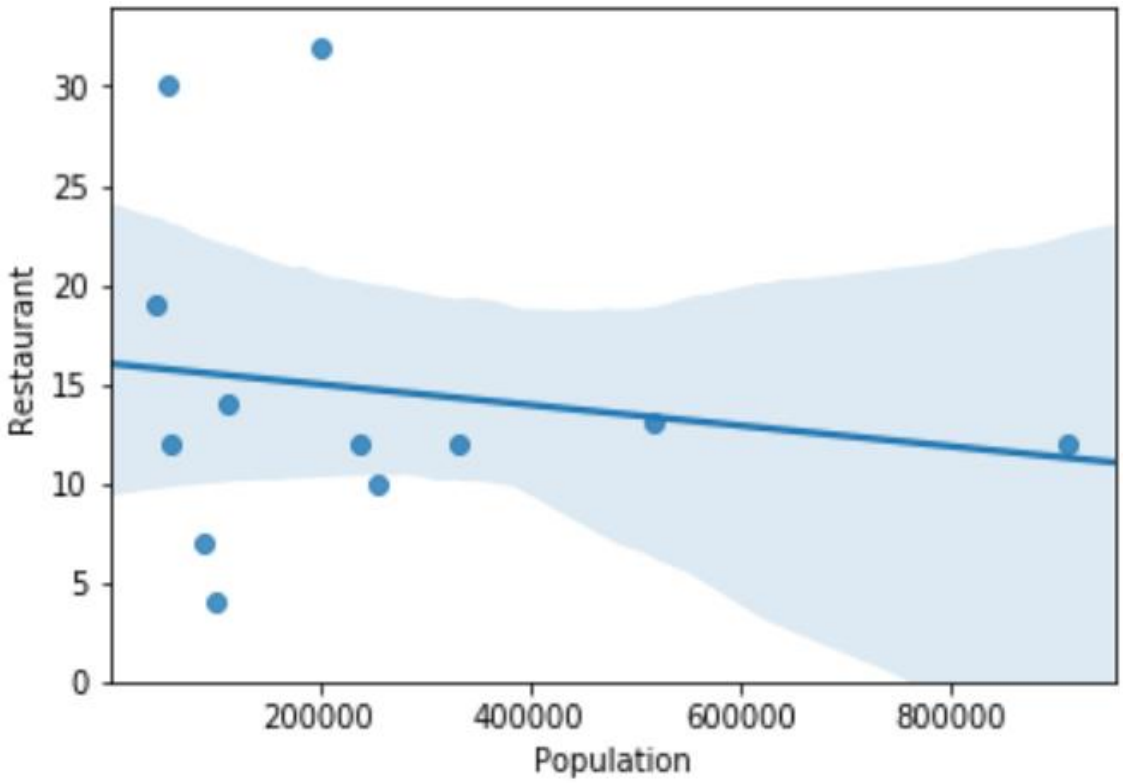

Figure 2. Correlations between Number of Restaurants and total of Populcations each locations

4.4 See location restaurans in visuallization on the map

Because the population can be not ratio with the area of each location, so I use the mapping method to combine the results of my evaluations. We can clearly see that the number of restaurants is not evenly distributed on the map, in some locations completely empty with the number of restaurants.

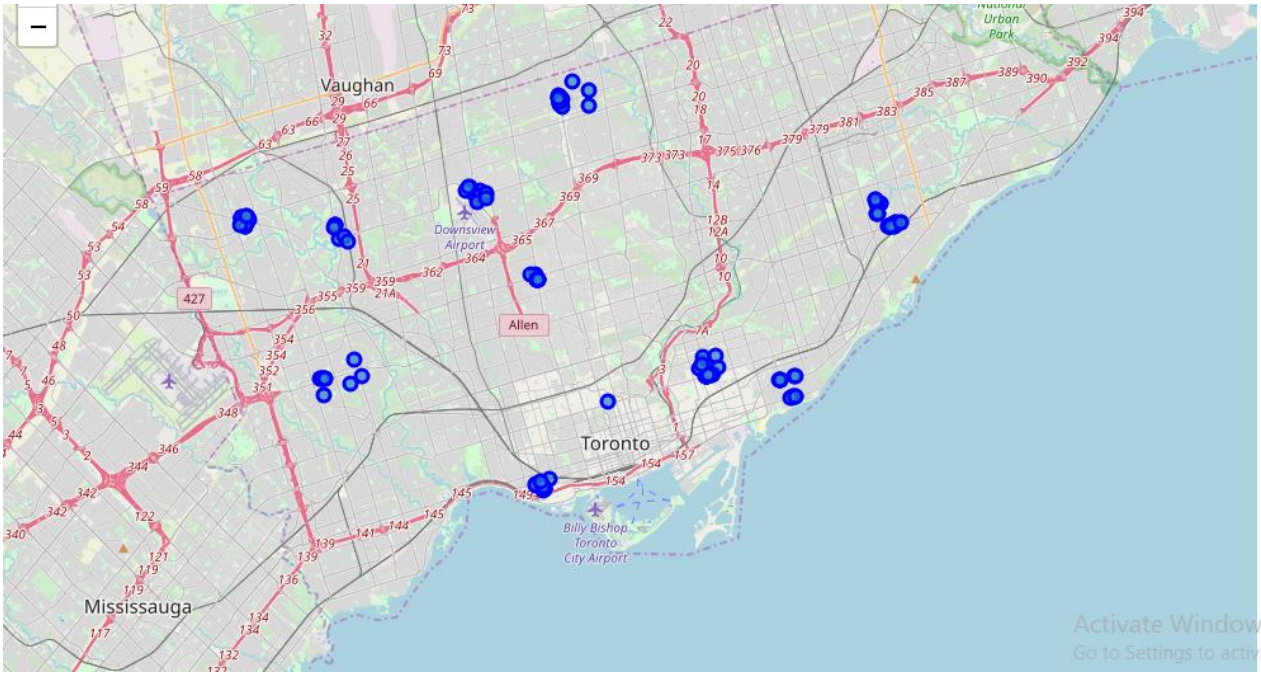

Figure 2. Distribution of restaurants in Toronto on the maps.

4.5 Use query data to see list of potenital locations in Toronto

The final result is a direct data query from the analyzed data, as can be seen in graphs and locations, we have the 5 most potential places sorted in the following order, with the top one being Beaches.
-East York with the highest population ratio and relatively small restaurant rate

	PPopulation	PRestaurant
Neighborhood		
Beaches-East York	31.194835	6.779661
Parkdale-High Park	17.679407	7.344633
Scarborough-Agincourt	11.392962	6.779661
Etobicoke Centre	8.723148	5.649718
York Centre	8.115042	6.779661

5. Result

Through the above analysis, we have some following result:

- Restaurants is distributed not same in locations in Toronto
- Beaches-East York takes most population but there is few restaurants
- Parkdale-High Park takes second population but numbe of restaurants is greater than Beaches-East York but it still takes a few of restaurants in Toronto

Company XYZ JSC should consider 5 locations to open new restaurants:

	PPopulation	PRestaurant
Neighborhood		
Beaches-East York	31.194835	6.779661
Parkdale-High Park	17.679407	7.344633
Scarborough-Agincourt	11.392962	6.779661
Etobicoke Centre	8.723148	5.649718
York Centre	8.115042	6.779661

6. Discussion

Through the above analysis, we have analyzed the best location to invest in XYZ JSC based on the graph of two characteristics: population and number of restaurants in each location. Thereby finding the best 3 places by both forms of visual analysis and data analysis

If we can gather more specific features (income, area, population age, etc.) in detail for each location, by this method we can find even more accurate and detail result , the restaurant type is more suitable for each ethnic if it is possible

7. Conclusions

In this project, I analyzed the evaluation sites that are the most potential for XYZ JSC based on population and number of restaurants using statistical methods: query data and data visualization.