Predicting Best Neighborhoods for Opening a New Restaurant

Author: Betul Dincer
Date: May 23, 2020

Introduction

This project is targeted for those who are interested in opening a restaurant in Toronto and our goal is to detect the best neighborhood for a client to open a restaurant in Toronto city. After learning the major cuisine of the restaurant from our client, we want to analyze all the neighborhoods in the city to understand the distribution of different types of restaurants. This way, we can detect the neighborhoods where the similar restaurants are located and offer these neighborhoods as the best possible locations to open a new restaurant. Our goal is to maximize the number of potential customers for the new restaurant and we assume that if we select neighborhoods where similar restaurants are common, we can reach to maximum number of customers. For instance, if we want to open an Italian restaurant, we want to detect the neighborhoods where Italian restaurants are very common which would indicate that people in the neighborhoods like Italian food and they would potentially be interested in a new Italian restaurant.

Data

In order to solve our problem, we need two different data sources: the neighborhood information of Toronto and the count and types of restaurants in each neighborhood. The neighborhood information is crucial for us to determine the optimal neighborhoods to open a restaurant. Getting the location and type of each restaurant will enable us to detect the most popular restaurant types in each neighborhood. Using this information along with the appropriate machine learning technique, we will be able to detect the best neighborhood for a client to open a restaurant at. We show the Toronto neighborhoods on map in Figure 1.
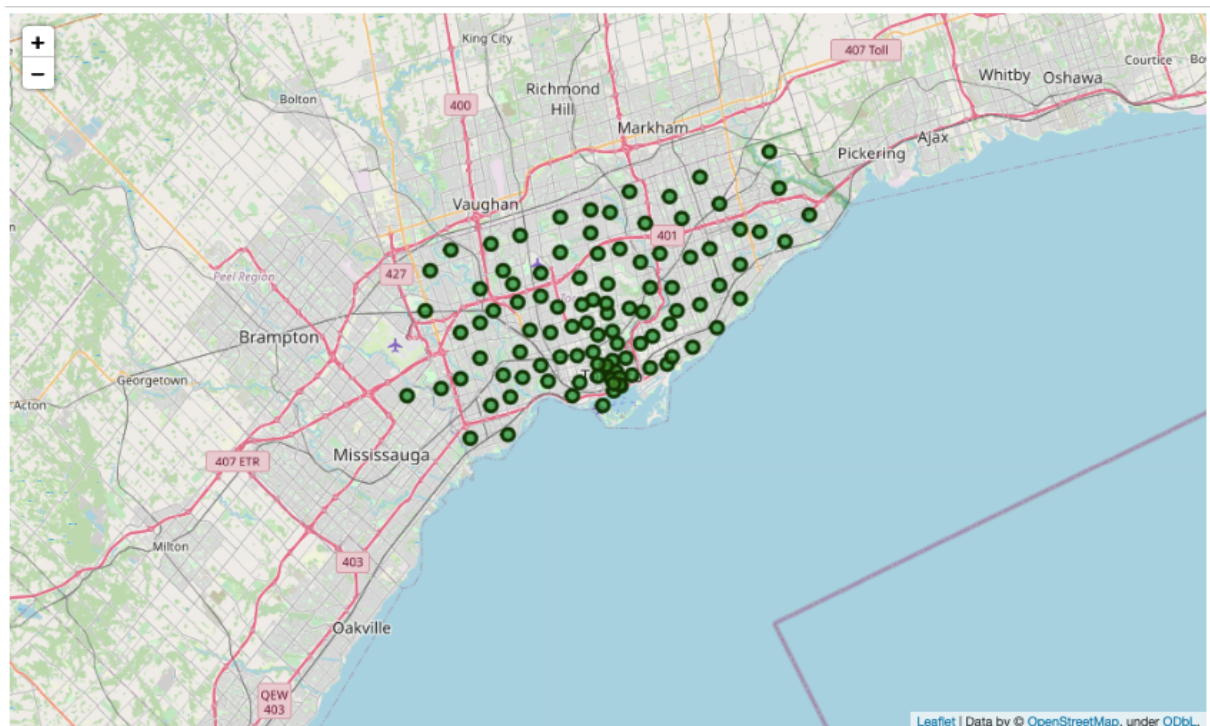

Figure 1 Map of Toronto neighborhoods.

Methodology

First the data needed is downloaded from Wikipedia which was in a table format and transformed to a data frame, and then using the geospatial, the coordinate of Toronto is extracted, and a neighborhood map is drawn. And then using Foursquare API all the information about all the venues in all neighborhoods was gathered. Especially information about the restaurants and their categories were our key points. Creating a restaurant-based feature matrix, for each neighborhood, the percentage of restaurants of each category is calculated.

Then the neighborhoods are grouped in terms of restaurant categories by applying k-means clustering. We experimented with different k values and generated different maps with group labels to see the results.

Results

By observing the maps and analyzing each cluster we can see the neighborhoods in each cluster which are categorized according to the restaurant types, so for anyone who is interested in opening a new restaurant, it is easy to decide where to open one. We show the summary of our results in Figure 2. We also investigated our results by examining the clusters we have learned. We provide multiple examples from the clusters we learned.
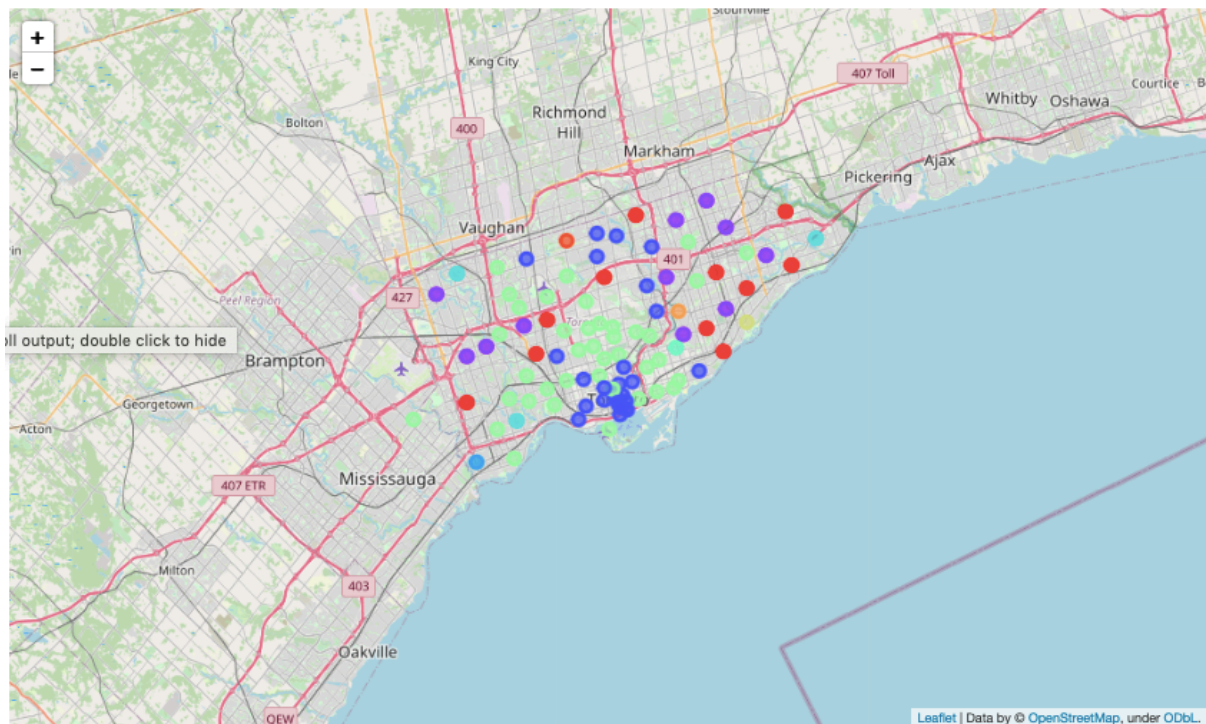

Figure 2 Map of Toronto neighborhoods colored by the clusters we learned.

Cluster 4 contains neighborhoods where Italian restaurants are very common. The listed neighborhoods can be good options for opening a new Italian restaurant.

Neighborhood	Cluster Label	Postal Code	Borough	Most Common Restaurant	Second Most Common Restaurant	Third Most Common Restaurant
Humber Summit	4	M9L	North York	Italian Restaurant	Jewish Restaurant	Korean Restaurant
Old Mill South, King's Mill Park, Sunnylea, Hu...	4	M8Y	Etobicoke	Italian Restaurant	Jewish Restaurant	Korean Restaurant
Rouge Hill, Port Union, Highland Creek	4	M1C	Scarborough	Italian Restaurant	Indian Restaurant	Indonesian Restaurant

Cluster 1 contains neighborhoods where Chinese restaurants are very common. The listed neighborhoods can be good options for opening a new Chinese restaurant.

Neighborhood	Cluster Label	Postal Code	Borough	Most Common Restaurant	Second Most Common Restaurant	Third Most Common Restaurant
Agincourt	1	M1S	Scarborough	Chinese Restaurant	Jewish Restaurant	Korean Restaurant
Kennedy Park, Ionview, East Birchmount Park	1	M1K	Scarborough	Chinese Restaurant	Latin American Restaurant	Mediterranean Restaurant
Kingsview Village, St. Phillips, Martin Grove ...	1	M9R	Etobicoke	American Restaurant	Hotpot Restaurant	Indian Chinese Restaurant
Milliken, Agincourt North, Steeles East, L'Amo...	1	M1V	Scarborough	Chinese Restaurant	Korean Restaurant	Latin American Restaurant
North Park, Maple Leaf Park, Upwood Park	1	M6L	North York	Chinese Restaurant	Jewish Restaurant	Latin American Restaurant
Parkview Hill, Woodbine Gardens	1	M4B	East York	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Jewish Restaurant
Parkwoods	1	M3A	North York	Caribbean Restaurant	Indian Chinese Restaurant	Indian Restaurant
South Steeles, Silverstone, Humbergate, Jamest...	1	M9V	Etobicoke	Caribbean Restaurant	Indian Chinese Restaurant	Indian Restaurant
Steeles West, L'Amoreaux West	1	M1W	Scarborough	Chinese Restaurant	Korean Restaurant	Latin American Restaurant
Westmount	1	M9P	Etobicoke	Chinese Restaurant	Jewish Restaurant	Korean Restaurant
Woburn	1	M1G	Scarborough	Chinese Restaurant	-	

Discussion

According to the results we get, there are categories of restaurants like Asian or fast food which are so common, it can be difficult to output a true optimal location. Also, some categories are very rare like Turkish so it is again difficult to decide on the optimum neighborhood.

Conclusion

By using the present data on the internet, and using the data science tools so many analysis and decisions can be made. Here in this project we used the data about Toronto neighborhoods from Wikipedia, using Geospatial we got the coordinates and were able to draw the maps and then using Foursquare API we got information about restaurants in all neighborhoods of Toronto and then using k-means clustering we were able to cluster the neighborhoods according to the restaurant categories. It is now easy to analyze all the maps and tables and decide on the place to open the restaurant.