Coursera Capstone Project

IBM Applied Data Science Capstone

Determining the Best Neighborhoods to Open a Restaurant in Toronto and New York

By: Jeffrey Lu
Date: August 2019

Introduction

This project aims to identify the neighborhoods in Toronto and New York where it makes the most sense to open a restaurant. More specifically, the restaurants owners that would be most interested in this project would be those whose businesses are affected by the weather. Imagine an ice cream shop whose revenues are affected by how hot or humid it feels outside or a pizzeria that recently realized that its gradual decline in in-house traffic has been due to people's recent drive to grill outside due to the summer weather. How about a Hotpot restaurant that observes revenue increases when the temperature outside experiences a slight drop? These are just a few examples of weather-sensitive restaurants that would be able to benefit from this project.

To provide more insight on how the weather affects the restaurant industry, a study published by Milos Buijsic, assistant professor at the Department of Human Sciences at Ohio State University, suggests that the weather not only affect's a restaurant's volume of customers, but also how much they enjoy their experience at a restaurant. According to the National Restaurant Association, "More than 90 percent of restaurant operators indicate that changes in local weather conditions affect their sales and customer counts."

In addition to weather data, we will also be using location data from Foursquare in order to determine which location is the most popular for restaurants. Industrial organization, a field of economics that studies the theory of the firm, suggests that in a given geographical plane, similar firms will eventually start to clump together in at a pareto-optimal focal point in order to optimize revenue. This is why you will see stores that sell similar products located in very close proximity to each other. With both weather, location, and venue data, we will help restaurant owners decide whether or not they should open their store in New York or Toronto.

Data and Methodology

We will be using a 5-day forecast of the temperature, humidity, and weather description (cloudy, clear sky, light rain etc.) for our project. The data will come from OpenWeatherMap's API. There will be forecast observations recorded every 3 hours so in order to obtain the total average 5-day temperature/humidity, we will need to average 40 data points for each category (5 days consists of 120 hours/ 3 observations per hour = 40 data points per location). We will be appending the average 5-day temperature and humidity as well as the most frequent weather condition and the frequency of the "clear sky" weather condition to our Neighborhoods dataset. We have decided that the "clear sky" weather condition would be the most favorable condition for restaurants so it would be important to know the frequency of this weather condition.

The Neighborhoods dataset was created by first obtaining a list of postal codes that we web scraped from the following Wikipedia page: "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M". We then used geocoder to obtain the Latitude and Longitude coordinates of these postal codes. The New York location dataset was obtained by downloading NYU's New York dataset through the following link:" https://cocl.us/new_york_dataset".

After the weather data obtained and appended to the neighborhood's dataset, we will start to incorporate Foursquare's location dataset. We will use Foursquare to obtain a list of all venues in various neighborhoods to determine which neighborhood has the highest restaurant frequency as well as total restaurants. We will assume that locations with the highest restaurant frequency as well as total restaurants will be the locations that are the most favorable places to open up the restaurant.

For the sake of the restaurant owners that is leveraging this project, we have included a clustering analysis on the neighborhoods of Toronto and New York so that these owners can gain more insight into these neighborhoods and decide what area they should position their restaurant.

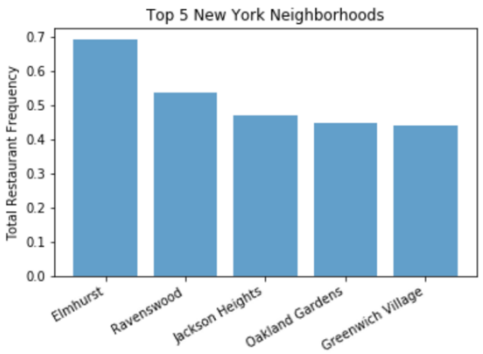
Results and Discussion

We first start by analyzing the neighborhoods in Toronto. We sorted the Toronto dataframe by the highest total restaurant frequency and with the number of total restaurants being greater than or equal to 10. This was done to ensure that the high restaurant frequency was not due to a lack of other venues. Taking a look at the weather data, it seems like the most frequent weather condition is "clear sky", which is a good sign. The clear sky condition frequency, average 5-day temperature and humidity were relatively the same across the entire dataset, which is not surprising since these neighborhoods are geographically very close to each other and thus the weather is more or less likely to be very similar.

As a restaurant owner, looking at the below list of neighborhoods, it would seem that the North York borough with the neighborhoods of Bedford Park, Lawrence and Manor East would be the best place to open up a restaurant. However, the neighborhoods of Church and Wellesley, Chinatown, Grange Park, and Kensington Market of Downtown Toronto also seem to be fantastic. Even with a slightly lower restaurant frequency, there is a greater number of total restaurants, which should inspire confidence that restaurants do thrive in these neighborhoods.

Out[148]:

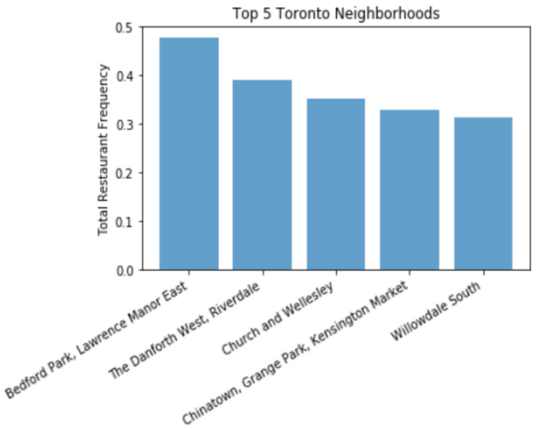
	Borough	Neighborhood	PostalCode	Latitude	Longitude	Total_Restaurant	Total_Restaurant_Frequency	Average 5-Day Temperature	Average 5-Day Humidity	Most Frequent Weather Condition	Clear Sky condition frequency
62	North York	Bedford Park, Lawrence Manor East	M5M	43.733283	-79.419750	11.0	0.478261	64.266912	77.925	clear sky	0.375
41	East Toronto	The Danforth West, Riverdale	M4K	43.679557	-79.352188	16.0	0.390244	64.268458	77.925	clear sky	0.375
52	Downtown Toronto	Church and Wellesley	M4Y	43.665860	-79.383160	29.0	0.353659	64.262875	77.925	clear sky	0.375
67	Downtown Toronto	Chinatown, Grange Park, Kensington Market	M5T	43.653206	-79.400049	33.0	0.330000	64.264000	77.925	clear sky	0.375
22	North York	Willowdale South	M2N	43.770120	-79.408493	11.0	0.314286	64.267450	77.925	clear sky	0.375



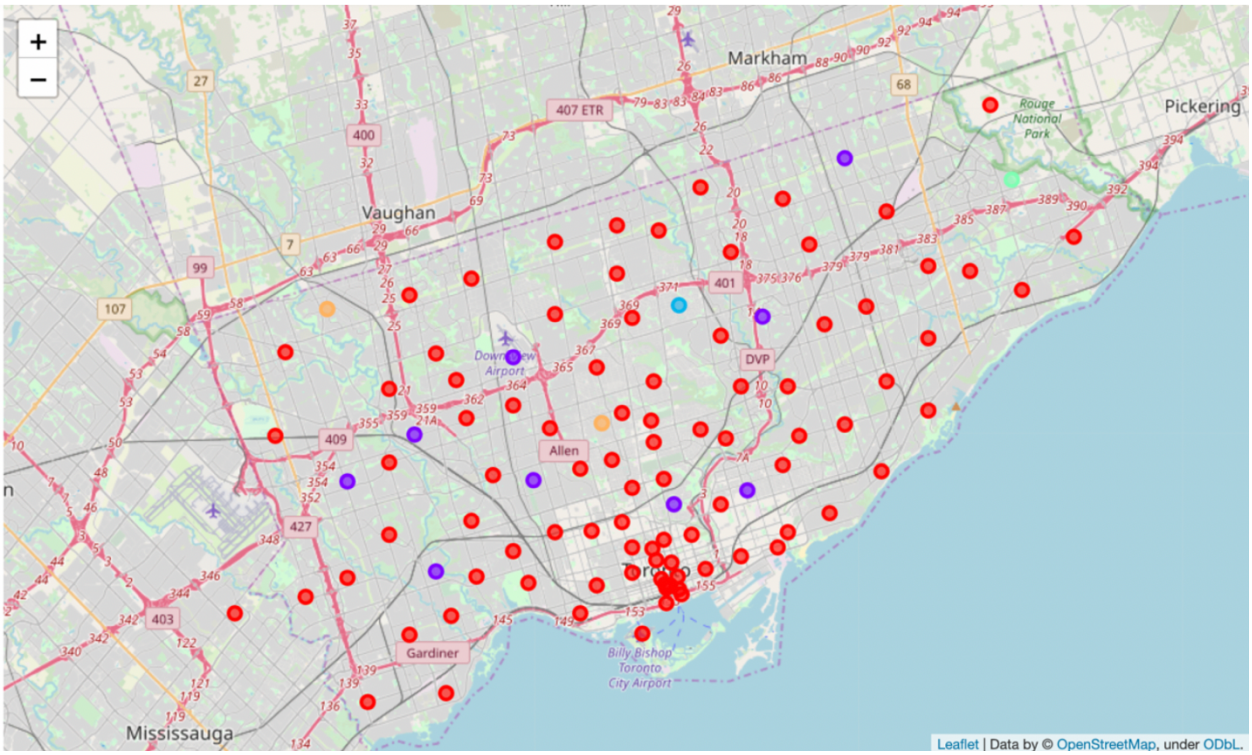
Looking at the New York dataset, it seems like Elmhurst, Ravenswood, Jackson Heights, Oakland Gardens, and Greenwich Village all seem to be good spots when considering opening up a restaurant in New York. It seems that the Queens borough is the best place to open a restaurant, with Manhattan coming in at 2nd place. Temperature, humidity, and weather condition forecasts seemed to be very close across the dataset. Greenwich Village did seem to experience higher rates of the clear sky weather condition (0.525), which could make it a slightly more appealing neighborhood for restaurants to thrive.

Out[149]:

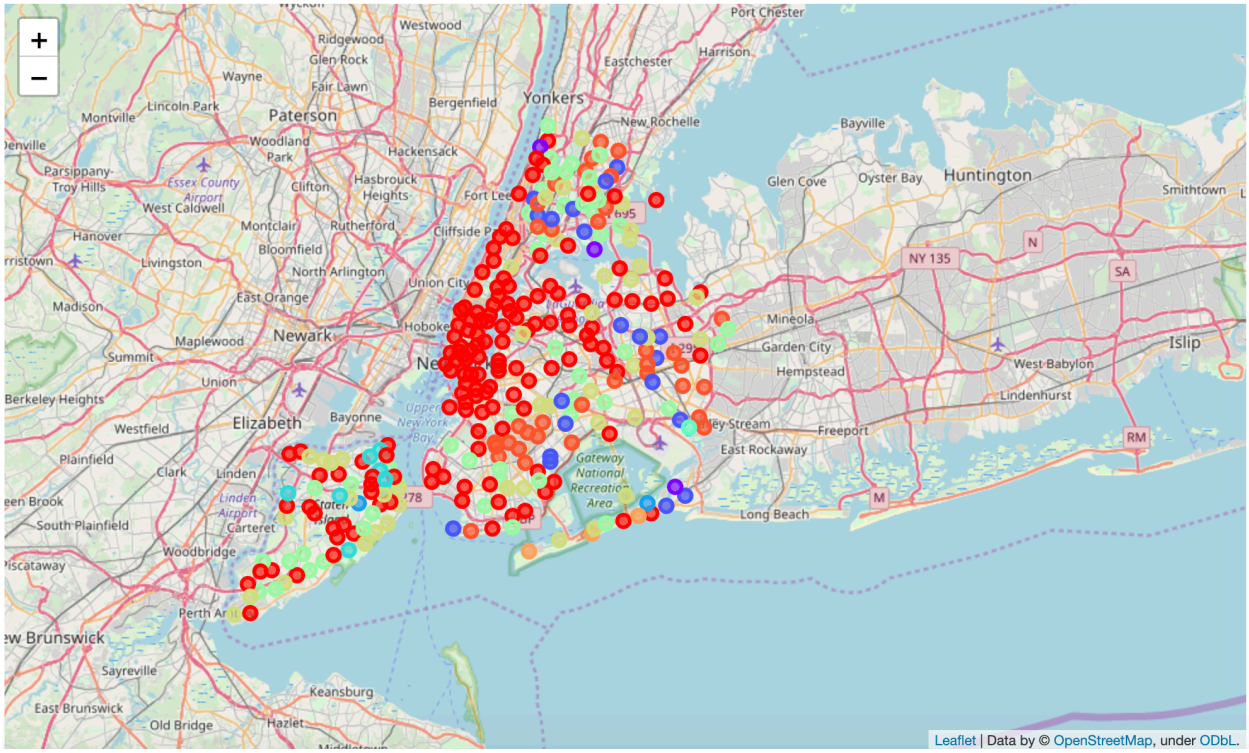
	Borough	Neighborhood	Latitude	Longitude	Total_Restaurant_Frequency	Total_Restaurant	Average 5-Day Temperature	Average 5-Day Humidity	Most Frequent Weather Condition	Clear Sky condition frequency
132	Queens	Elmhurst	40.744049	-73.881656	0.690476	29.0	75.596711	51.555263	clear sky	0.500
186	Queens	Ravenswood	40.761705	-73.931575	0.551724	16.0	75.056460	53.320635	clear sky	0.500
131	Queens	Jackson Heights	40.751981	-73.882821	0.469136	38.0	75.110167	52.870833	clear sky	0.500
161	Queens	Oakland Gardens	40.745619	-73.754950	0.448276	13.0	75.226285	53.005556	clear sky	0.500
117	Manhattan	Greenwich Village	40.726933	-73.999914	0.440000	44.0	73.786000	56.441667	clear sky	0.525
153	Queens	Little Neck	40.770826	-73.738898	0.428571	21.0	75.041922	53.134804	clear sky	0.500
100	Manhattan	Chinatown	40.715618	-73.994279	0.420000	42.0	74.448562	54.656250	clear sky	0.500
273	Manhattan	Turtle Bay	40.752042	-73.967708	0.400000	40.0	74.845900	53.585000	clear sky	0.500
84	Brooklyn	Clinton Hill	40.693229	-73.967843	0.400000	36.0	75.441312	51.978125	clear sky	0.500



After clustering the neighborhoods of Toronto and taking a closer look, it seems that the heaviest cluster group, shown with the red color, reveals that these areas are full of restaurant's, coffee shops, and bars.



After clustering New York's neighborhoods, it seems like cluster 1 (red) consists of neighborhoods with a lot of diverse restaurants. Cluster 7 (green) features neighborhoods with an extremely high frequency of pizzerias and cluster 10 (orange) showcases neighborhoods with a high frequency for Caribbean restaurants and donut shops.



Conclusion

When selecting the best location to open a restaurant in Toronto, restaurant owners should consider Bedford Park, Lawrence Manor East, The Danforth West, Riverdale, Church and Wellesley, Chinatown, Grange Park, and Kensington Market. These locations observe the highest total restaurant venue frequency as well as maintain a solid number of total restaurants. Although the weather didn't differ that much from one neighborhood to the other, it is good to see that the most frequent weather condition was "clear sky", which should be good for any restaurant owner.

As for the best opportunity areas in New York, it seems that the neighborhoods in Queens and Manhattan are the best. Such neighborhoods such as Elmhurst, Ravenswood, Jackson Heights, Oakland Gardens, and Greenwich Village observe the highest total restaurant venue frequency and have a high number of total restaurants.

Based just off this data set, if one were to decide opening up a restaurant in either Toronto or New York, one should choose New York. The total restaurant venue frequency, total number of restaurants, the temperature, and frequency of the clear sky condition all seem to be higher and the humidity forecast seems to be lower. All of these factors seem to make New York a better place to open a restaurant.

There are definitely many other factors that we haven't considered in this dataset such as the per-capita income of the area, demographics, average cost-per-square foot in rent, local supplier costs, and the possibility of an oversaturated market. Since we are comparing different countries, we also need to consider currency fluctuations as well as where the countries are in its business cycle.

There also seems to be limitations with our weather data. The most extensive and cost-free API forecast we can obtain from OpenWeatherMap is a 5-day forecast which in theory shouldn't be sufficient enough to base any business decision off of. Although a yearly aggregate of historical weather data is preferred, we are not able to obtain it cost-free. It is also important to recognize that neighborhoods within the same city should have very similar temperatures and humidity since they are geographically very close, so it is hard to exploit weather data here. Thus, it is better to utilize weather data to analyze differences between cities and not inter-city neighborhoods. An even better use of this project would be to help project revenues with weather data.