1. Introduction and Business Problem

A costumer is considering opening a Pizza Restaurant in New York City ,he also have a pizza restaurant in Toronto and look for alike neighborhoods in NY to create it . I am gathering information regarding the feasibility of such business, and the areas where it would have the biggest chance of success. First, I need to know general information about the city:

“Because of its favorable location, excellent port facilities, and large population, New York City is the leading wholesale and retail trade center in the United States. New York is also a leader in communications, the hotel and restaurant business, building construction, and manufacturing.” (Source: https://www.theusaonline.com/cities/newyork.htm)

After some search online, it looks like opening a restaurant in NY is a great idea. I will proceed to analyze the area or areas where there are similar neighborhoods

2. Data

To achieve this, we'll use an aggregation of the two datasets of both cities and perform the clustering on all the neighborhoods in Toronto an NY

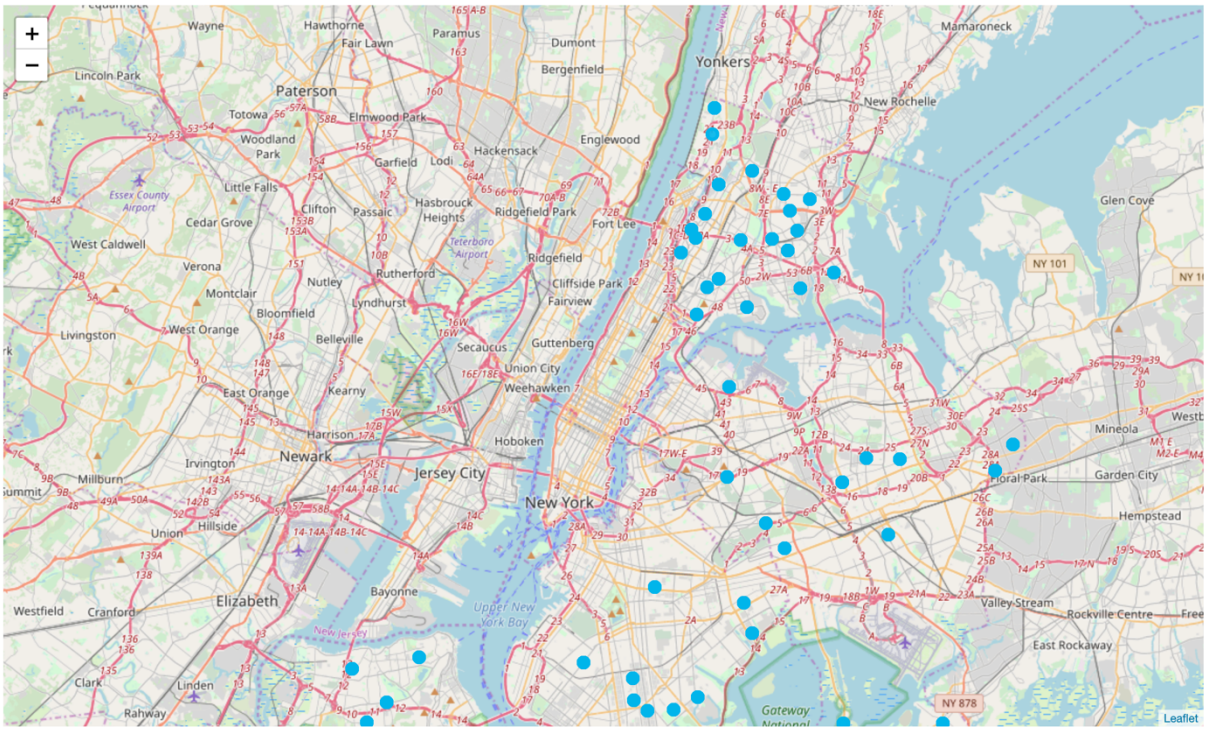
We're likely to find many neighborhoods from the same city in one cluster but any other neighborhood from the other city brings a great information.

3. Methodology

In this project I use the methodology described in Week 3 of the Applied Data Science Capstone, I use Foursquare API to obtain location data, in order to analyze a specific business in NY. Next, this data is clustered using K-means algorithm to analyze the most similar and / or dissimilar neighborhoods and find where would it be the most suitable area for establishing a new Pizza restaurant. After processing the neighborhood data, converting addresses into their respective latitude and longitude values, I create a map with Foursquare API to visualize with Folium library the current Pizza restaurant business, as shown in Figure

4. Result

From this data it looks like the best location to establish a Pizza restaurant would be Cluster 2, as shown in Figure 3.



	Neighborhood Latitude	Venue Longitude	Venue Category	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	
3	43.725882	-79.315635	Hockey Arena	2	Hockey Arena	French Restaurant	Coffee Shop	Portuguese Restaurant	Pizza Place	Egyptian Restaurant	Eastern European Restaurant	Electronics Store	Factory	
4	43.725882	-79.313103	Coffee Shop	2	Hockey Arena	French Restaurant	Coffee Shop	Portuguese Restaurant	Pizza Place	Egyptian Restaurant	Eastern European Restaurant	Electronics Store	Factory	
5	43.725882	-79.312785	Portuguese Restaurant	2	Hockey Arena	French Restaurant	Coffee Shop	Portuguese Restaurant	Pizza Place	Egyptian Restaurant	Eastern European Restaurant	Electronics Store	Factory	
6	43.725882	-79.317418	French Restaurant	2	Hockey Arena	French Restaurant	Coffee Shop	Portuguese Restaurant	Pizza Place	Egyptian Restaurant	Eastern European Restaurant	Electronics Store	Factory	
7	43.725882	-79.312860	Pizza Place	2	Hockey Arena	French Restaurant	Coffee Shop	Portuguese Restaurant	Pizza Place	Egyptian Restaurant	Eastern European Restaurant	Electronics Store	Factory	
115	43.706397	-79.312913	Gastropub	2	Pizza Place	Fast Food Restaurant	Bank	Pet Store	Gastropub	Gym / Fitness Center	Athletics & Sports	Pharmacy	Bus Line	Ir
116	43.706397	-79.309279	Gym / Fitness Center	2	Pizza Place	Fast Food Restaurant	Bank	Pet Store	Gastropub	Gym / Fitness Center	Athletics & Sports	Pharmacy	Bus Line	Ir
117	43.706397	-79.312825	Pharmacy	2	Pizza Place	Fast Food Restaurant	Bank	Pet Store	Gastropub	Gym / Fitness Center	Athletics & Sports	Pharmacy	Bus Line	Ir
118	43.706397	-79.312270	Bank	2	Pizza Place	Fast Food Restaurant	Bank	Pet Store	Gastropub	Gym / Fitness Center	Athletics & Sports	Pharmacy	Bus Line	Ir
119	43.706397	-79.313130	Pizza Place	2	Pizza Place	Fast Food Restaurant	Bank	Pet Store	Gastropub	Gym / Fitness Center	Athletics & Sports	Pharmacy	Bus Line	Ir
120	43.706397	-79.308838	Fast Food Restaurant	2	Pizza Place	Fast Food Restaurant	Bank	Pet Store	Gastropub	Gym / Fitness Center	Athletics & Sports	Pharmacy	Bus Line	Ir
121	43.706397	-79.312196	Pet Store	2	Pizza Place	Fast Food Restaurant	Bank	Pet Store	Gastropub	Gym / Fitness Center	Athletics & Sports	Pharmacy	Bus Line	Ir
122	43.706397	-79.313274	Intersection	2	Pizza Place	Fast Food Restaurant	Bank	Pet Store	Gastropub	Gym / Fitness Center	Athletics & Sports	Pharmacy	Bus Line	Ir
123	43.706397	-79.313957	Pizza Place	2	Pizza Place	Fast Food Restaurant	Bank	Pet Store	Gastropub	Gym / Fitness Center	Athletics & Sports	Pharmacy	Bus Line	Ir
124	43.706397	-79.313808	Bus Line	2	Pizza Place	Fast Food Restaurant	Bank	Pet Store	Gastropub	Gym / Fitness Center	Athletics & Sports	Pharmacy	Bus Line	Ir
125	43.706397	-79.314105	Fast Food Restaurant	2	Pizza Place	Fast Food Restaurant	Bank	Pet Store	Gastropub	Gym / Fitness Center	Athletics & Sports	Pharmacy	Bus Line	Ir

5. Discussion

There are some areas with more competition for the same kind of business, and through clustering we have analyzed the areas that could have a potential advantage, based on an inferior level of competition. The selected area to establish the business is Inwood

neighborhood, and as a second option I would choose the areas of Upper East Side, Yorkville, Lenox Hill, Roosevelt Island, and Sutton Place neighborhoods. However, this analysis could be improved if more data were collected and taken into account, such as level of tourism in the area, information about residential and business areas, and income level.

6. Conclusion

Clustering with K-means algorithm is a great way to analyze unlabeled data, with low computational cost and fast insights, which can be the beginning of further exploration. One hot encoding is a very helpful method to convert categorical data into numerical data that can be easily used, in this case with K-means. And Foursquare API has proved to be very efficient for many tasks such as conversion of addresses into their respective latitude and longitude.

This project can serve to obtain a fast general idea of the potential success of a business, in a certain city, in a certain area, by analyzing data and explore the number of similar business in this area. However, probably with more data it would be possible to obtain more accurate results.