The Battle of Neighborhoods

—New York and Toronto

Introduction

Problem Background:

Both New York City and the city of Toronto are the financial capitals of their respective countries. They provide a lot of business opportunities and business friendly environment. Also, they have attracted many different players into the market. As global hubs of business and commerce, these two market are highly competitive. Thus, any new business venture or expansion needs to be analyzed carefully. The insights derived from comparison analysis will give good understanding of business environment which help in strategically targeting the market. This will help in reduction of risk and the return on investment will be reasonable.

Problem Description:

The company need to choose the correct location to start its first venture. If this is successful, they can replicate the same in other locations. First move is very important, thereby choice of location is very important. It's important to understand how New York City similar and dissimilar with Toronto. The average income and population density of boroughs of each city are needed to be explored. In order to learn about the competitors in New York City and Toronto, we need to explore the most common 20 venue categories of every neighborhood in each city.

Introduction

Target Audience:

To recommend the correct location, the company has appointed me to lead of the Data Science team. The objective is to compare New York city and Toronto and recommend to the management which city and its neighborhood will be the best choice to start a restaurant. The management also expects to understand the rationale of the recommendations made. This would interest anyone who wants to start a new restaurant in New York city or Toronto.

Success Criteria:

The success criteria of the project will be a good recommendation of borough/Neigborhood choice to the company based on large population, average income, lack of such restaurants in that location and nearest suppliers of ingredients.

Data

Data 1:

Borough, Neighborhood, Latitude and Longitude of New York city data from https://cocl.us/new_york_dataset;

Borough, Neighborhood, Latitude and Longitude of Toronto data: merge data from Wikipedia https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and geospatial coordinates of each neighborhood from http://cocl.us/Geospatial_data

Data 2:

Demographics of Toronto neighbourhoods:

https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods

Boroughs of New York City: https://en.wikipedia.org/wiki/Boroughs_of_New_York_City

Data 3:

Venues data from the Foursquare API were imported into two data frame for Toronto and New York city.

Methodology and Analysis

The overall strategy is based on mapping the relevant Toronto and New York data in order to facilitate the choice of places for new restaurant. The information will be consolidated into one map where one can see the details of each cluster and each point.

Web-scraping of boroughs and neighborhoods list of two cities from the Wikipedia is used to consolidate data-frame information which was saved as csv files for convenience and to simply the report. Geodata was obtained by coding a program to use Nomination to get latitude and longitude of the city. Maps with popups labels allow quick identification of location and making the selection very easy.

For every neighborhood, the most common 20 venue categories are taken into account and rest are eliminated. The KMeans clustering algorithm is used to cluster all the neighborhoods, once for Toronto and once for New York. K-means clustering aims to partition all observations into k clusters in which each observation belongs to the cluster with the nearest mean. It uses iterative refinement approach. The clusters are determined as 5 for both cities and the cluster labels are assigned to each neighborhoods.

The average income and population density of Toronto and New York are imported from Wikipedia and converted to data frame. All imported web data are preprocessed including cleaning and modification.

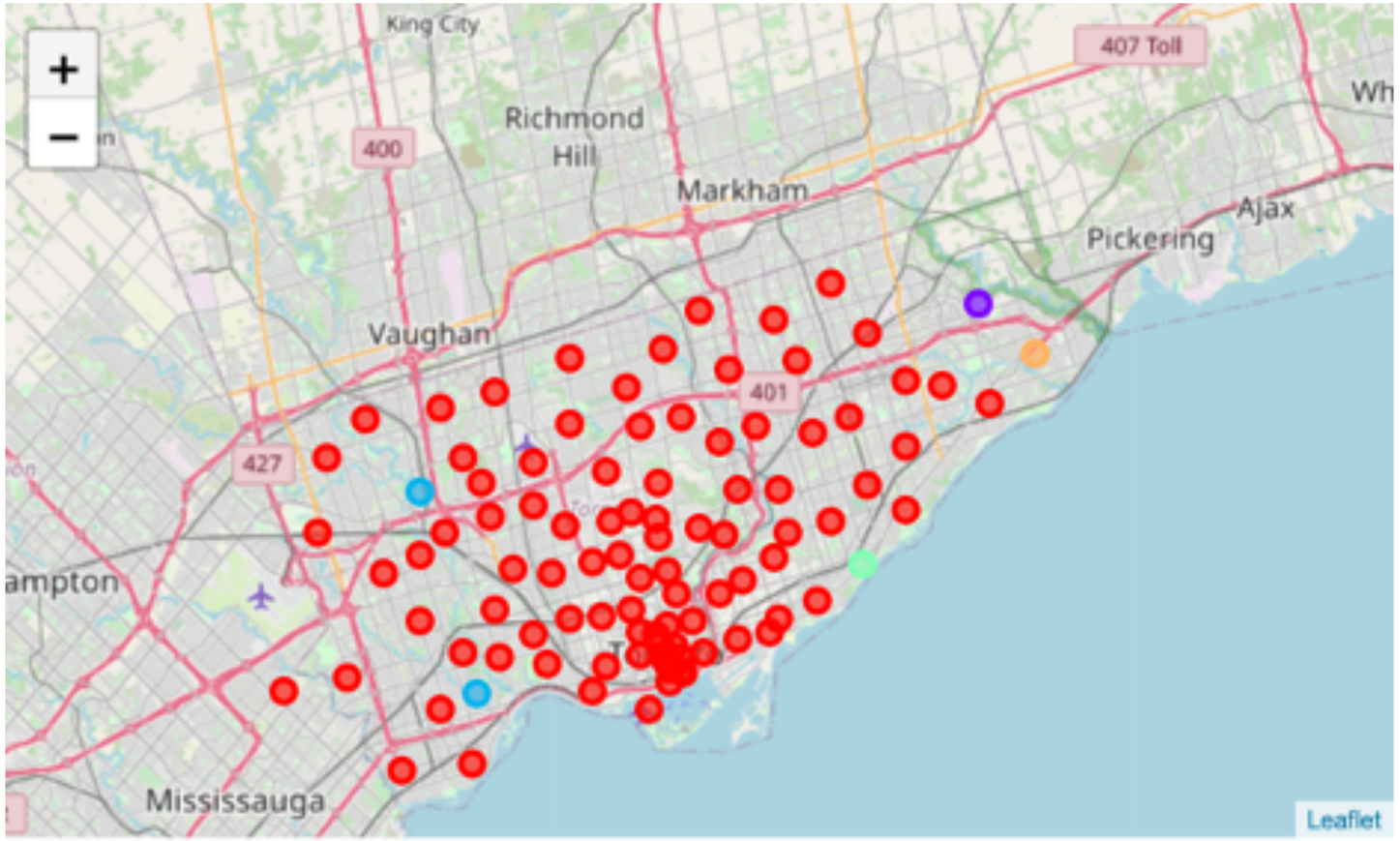
Results and Discussion

10 most popular venues in each neighborhoods of Toronto

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	10th Most Common Venue
Bus Stop	Food & Drink Shop	Music Venue	Church	College Rec Center	College Gym	College Auditorium	College Arts Building	Coffee Shop	Cocktail Bar
Hockey Arena	Coffee Shop	Drugstore	College Rec Center	College Gym	College Auditorium	College Arts Building	Cocktail Bar	Clothing Store	Climbing Gym
Coffee Shop	Bakery	Café	Breakfast Spot	French Restaurant	Cosmetics Shop	Health Food Store	Historic Site	Distribution Center	Chocolate Shop
Clothing Store	Furniture / Home Store	Miscellaneous Shop	Boutique	Coffee Shop	Accessories Store	Event Space	Diner	Chinese Restaurant	College Auditorium
Coffee Shop	Hobby Shop	Gym	Japanese Restaurant	Italian Restaurant	Beer Bar	Burger Joint	Burrito Place	Café	General Entertainment

Results and Discussion

The clusters of Toronto is shown in the following map, it can been seen that Toronto has one big cluster and other smaller one. We can say that the other four clusters are insignificant compared to the main big one. So the Toronto city seems have a very uniform neighborhood type.



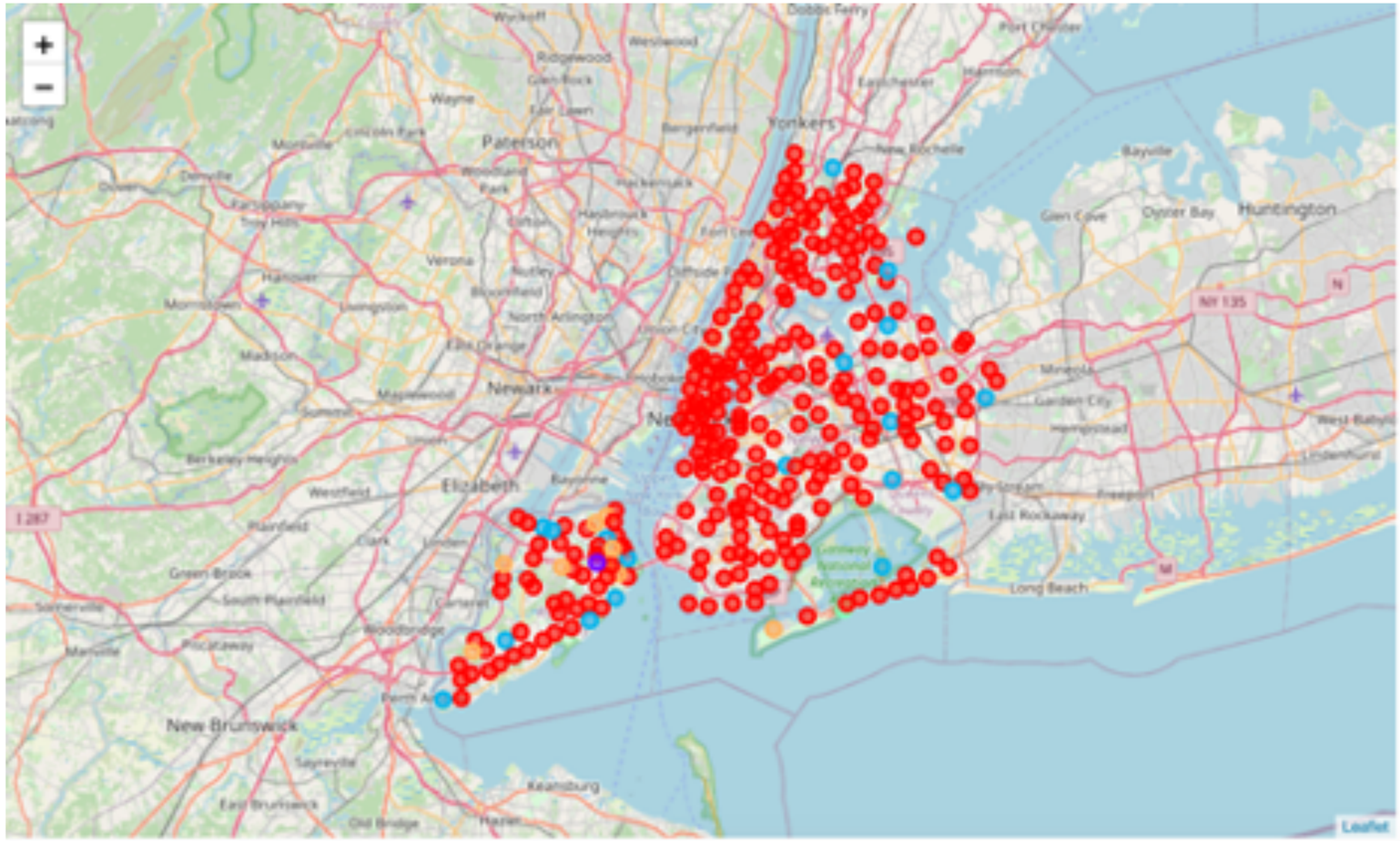
Results and Discussion

10 most popular venues in each neighborhoods of New York City

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	10th Most Common Venue
Gas Station	Laundromat	Deli / Bodega	Ice Cream Shop	Donut Shop	Dessert Shop	Cocktail Bar	Circus	Climbing Gym	Clothing Store
Bus Station	Fast Food Restaurant	Accessories Store	Discount Store	Ice Cream Shop	Chinese Restaurant	Grocery Store	Baseball Field	Liquor Store	Bagel Shop
Caribbean Restaurant	Deli / Bodega	Diner	Bus Station	Convenience Store	Juice Bar	Fast Food Restaurant	Food & Drink Shop	Bowling Alley	Donut Shop
Bus Station	Nail Salon	College Arts Building	Climbing Gym	Clothing Store	Club House	Cocktail Bar	Coffee Shop	College Academic Building	College Basketball Court
Bank	Gym	Home Service	Bus Station	Baseball Field	Food Truck	College Arts Building	Club House	Cocktail Bar	Coffee Shop

Results and Discussion

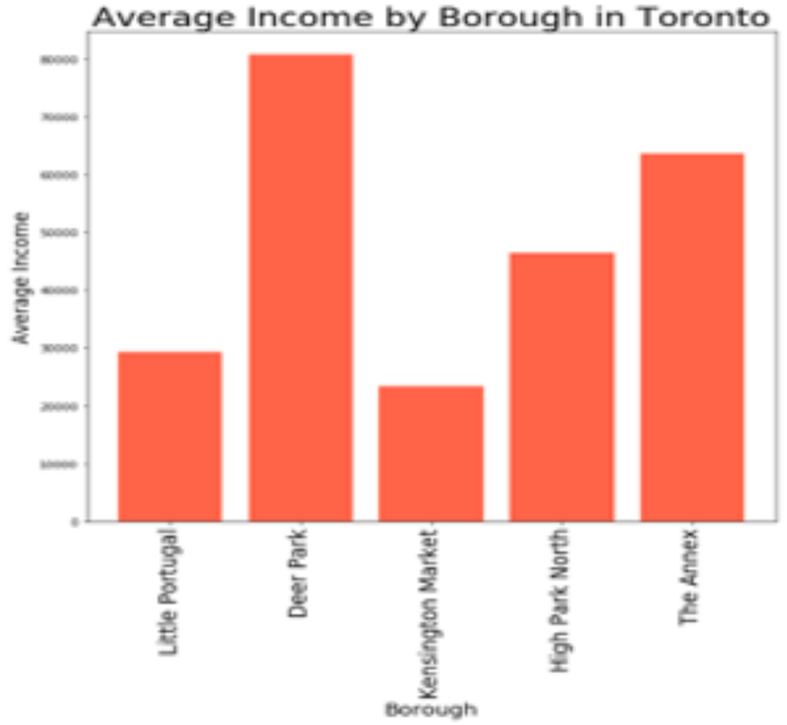
To compared with neighborhoods of Toronto, we mapped the clusters of New York. There are two big clusters and some mid size clusters. So we can see New York has much more varieties. The segmentations between New York and Toronto different.

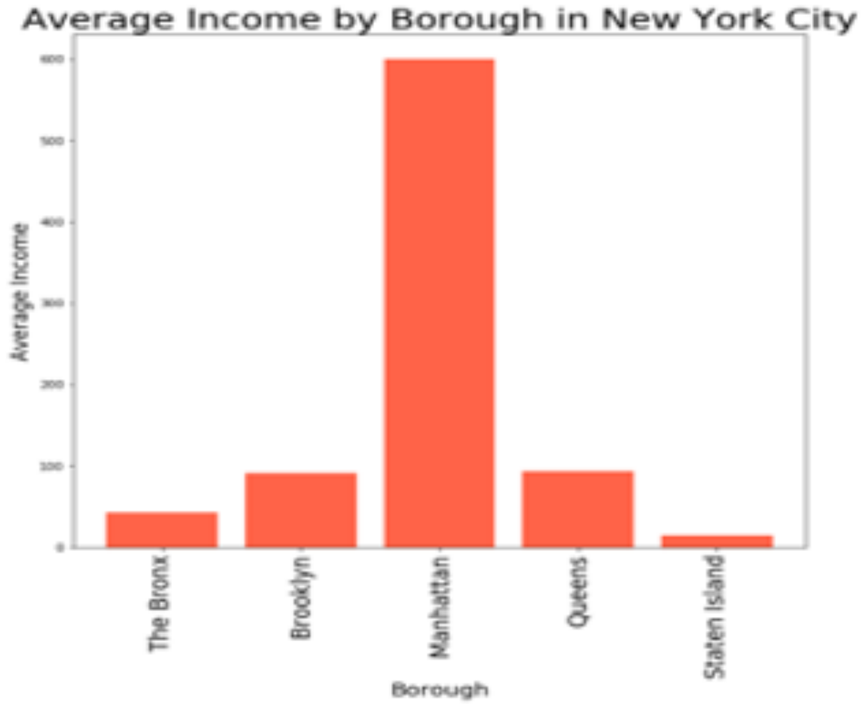


Results and Discussion

From the average income chart of Toronto and New York, it is clear that Deer Park has highest average income in Toronto than other boroughs. The Annex and High Park North have middle average income, others are poorer but not very far from them.

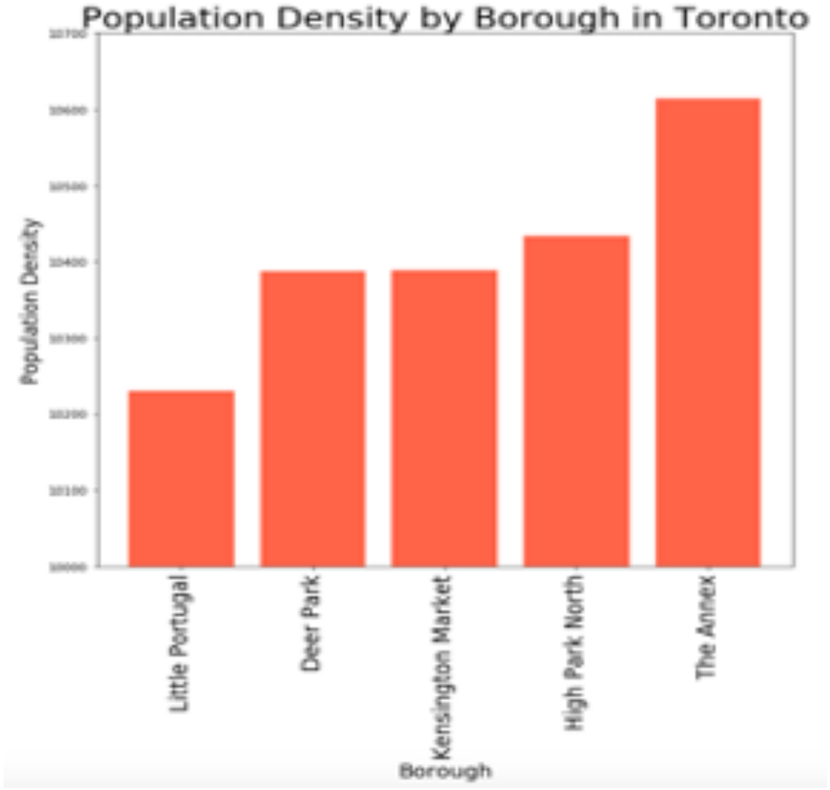
For the average income in New York, Manhattan has by far the highest average income than other boroughs. Other four are very obviously poorer than Manhattan.

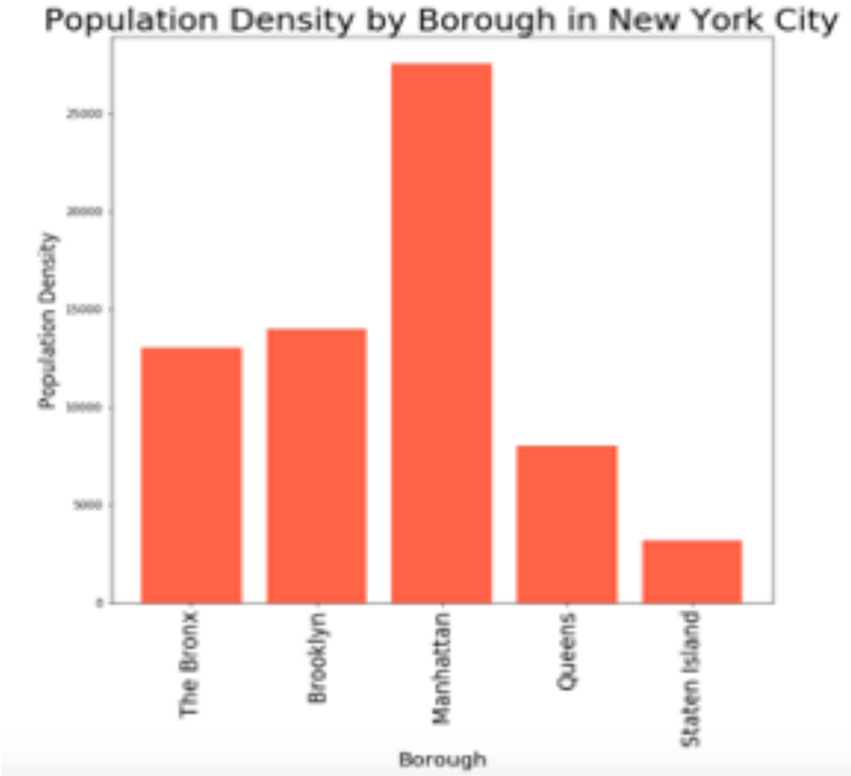




Results and Discussion

From the population density charts below, we can see the Annex has the most population density, while deep park, Kensington and High Park North have similar medium population density. In New York, Manhattan has the highest population density than other boroughs.





Conclusion

This analysis is performed on limited data, so this may be wrong or totally right. But if good amount of data is available, there is scope to come up with better results. As per the neighborhood and its most popular venues, our client can decide which type of restaurant could be built in the near future.

Overall, in terms of average income and population density, Toronto is much different from New York. New York wide gap between boroughs and everything seems centralized. On the other hand, Toronto had much more uniform distribution of population and income. I would suggest to choose Toronto as your first restaurant place for optimal profit and promising future.