Segmenting and Clustering Neighborhoods in NYC, USA and Toronto, Canada to effectively predict the best possible location for the new store of Melaleuca Inc.

Arsalan Aziz

April 19, 2019

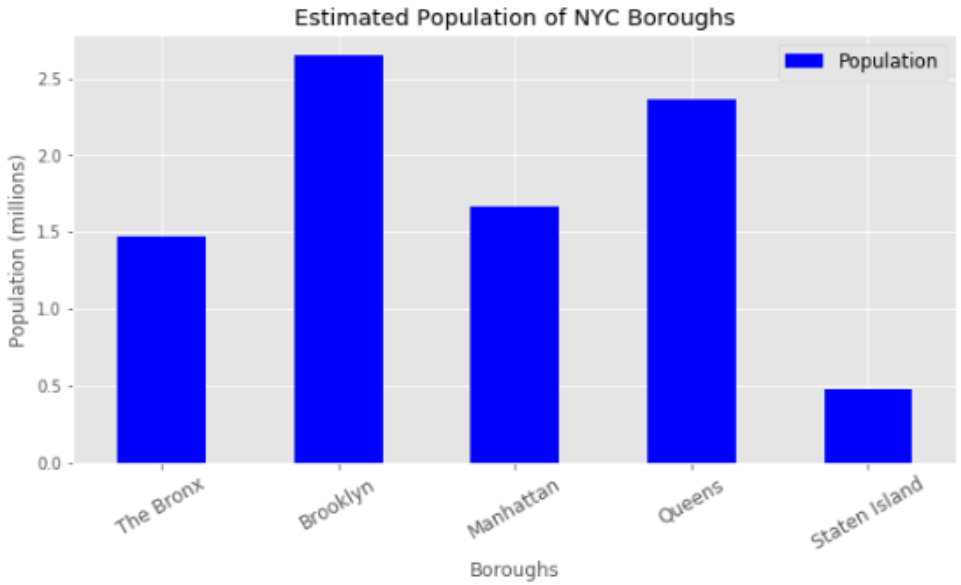
Predicting the most suitable location between NYC and Toronto to open a new retail store is the most valuable for executives of melaleuca Inc

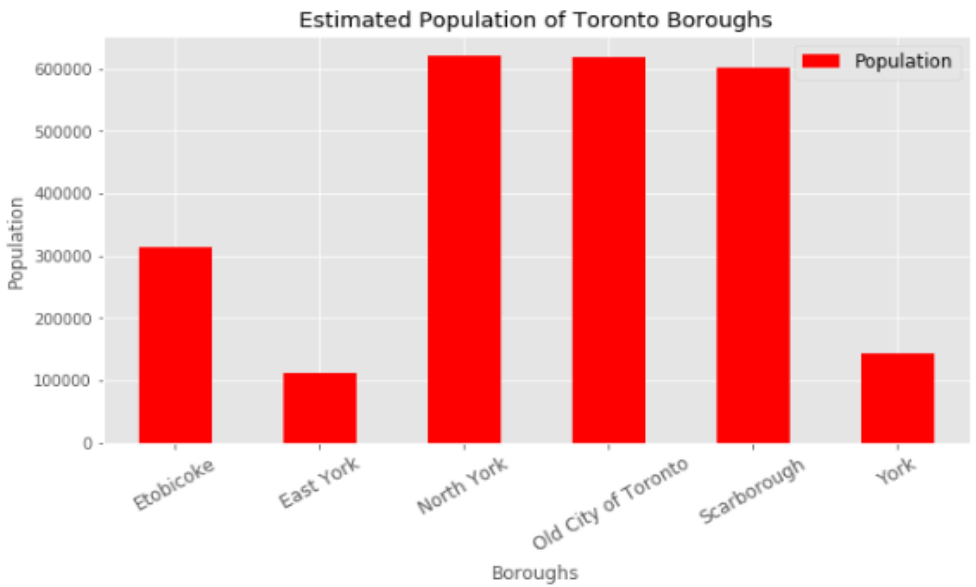
- Melaleuca wants to expand its business and they wanted to open a retail store in one of the major metropolitan cities in North America either NYC or Toronto
- Melaleuca Executives would be very interested in accurate prediction of best possible location and neighborhood in one of those cities for the competitive advantage and business value.
- This will be a huge breakthrough for them because the population density is high and there is great amount of awareness about health and Wellness in those areas.
- Others who are interested in melaleuca such as investors and loyal customers to the brand will be very interested to know the expansion of the business and the planning and preparation behind it.
- This project aims to Analyze and predict which particular city and neighborhood inside the city will be the best location for the new store of melaleuca that will help the retailer grow the business in that better location and city.

Data acquisition and cleaning

- The demographics related data like population and average income in different boroughs for New York City were picked up from Wikipedia source https://en.wikipedia.org/wiki/Demographics_of_New_York_City.
- In order to understand the overall economy of the city we analyzed the worth of some of the big companies in the NYC using the source https://en.wikipedia.org/wiki/Economy_of_New_York_City.
- The weather and the number of crimes recorded in NYC, were basically obtained from Kaggle.
- The demographic data like population and average income for Toronto were picked up from Wikipedia https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods and
- The weather related data of Toronto was picked up from https://www.wunderground.com/history/monthly/ca/toronto/CYTZ/date/2019-1.
- Similarly, the number of crimes committed in Toronto was picked up from http://data.torontopolice.on.ca/datasets/neighbourhood-crime-rates-boundary-file- and to analyze the economy, the data related to businesses were picked up from https://en.wikipedia.org/wiki/List_of_largest_public_companies_in_Canada_by_profit.
- FourSquare data was used to get the venues and their categories around the neighborhood.
- There was some data cleaning and data preprocessing performed before we analyzed and build the clustering and segmentation model.

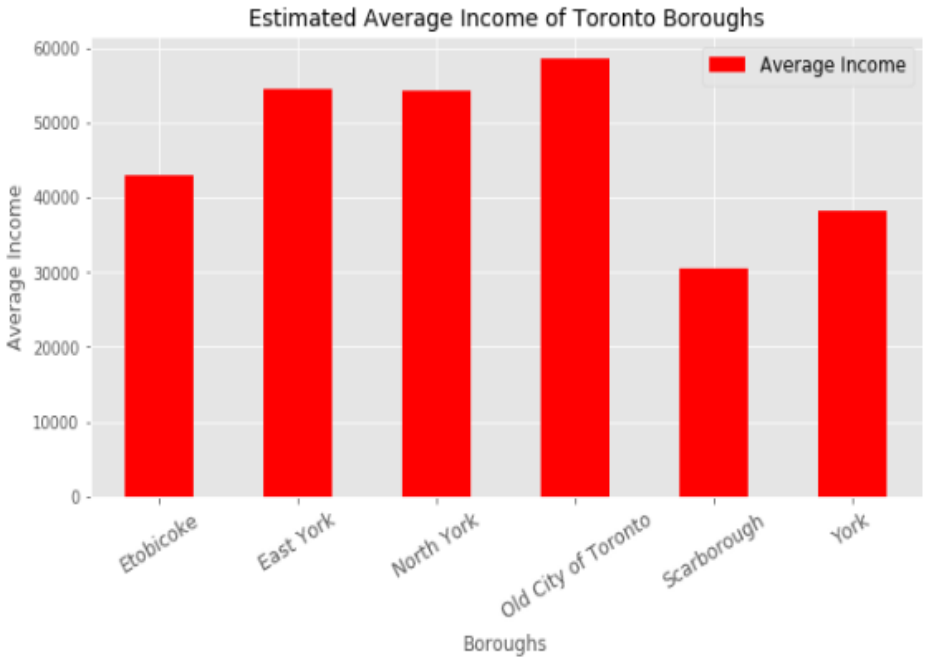
Relationship between NYC and Toronto Populations

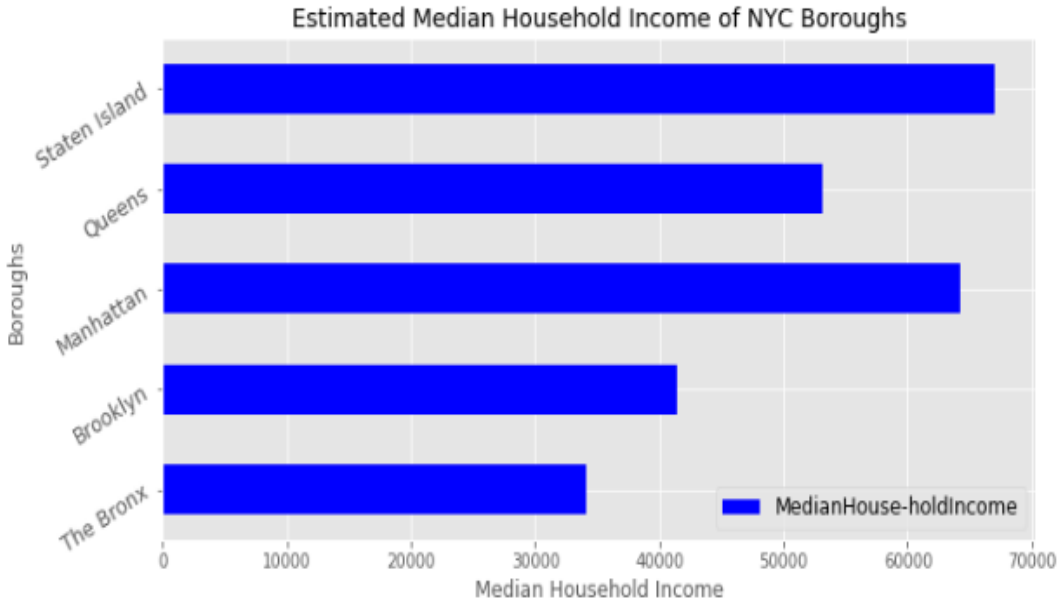




- In general the population of NYC is higher as compare to Toronto. The population of NYC is around 8.623 million and the population of Toronto is around 2.732 million,
- This is also evident from our analysis as well but we wanted to do the population comparison down to the boroughs level to get the better understanding of picking the appropriate place for attracting more traffic to the new store location of Melaleuca.

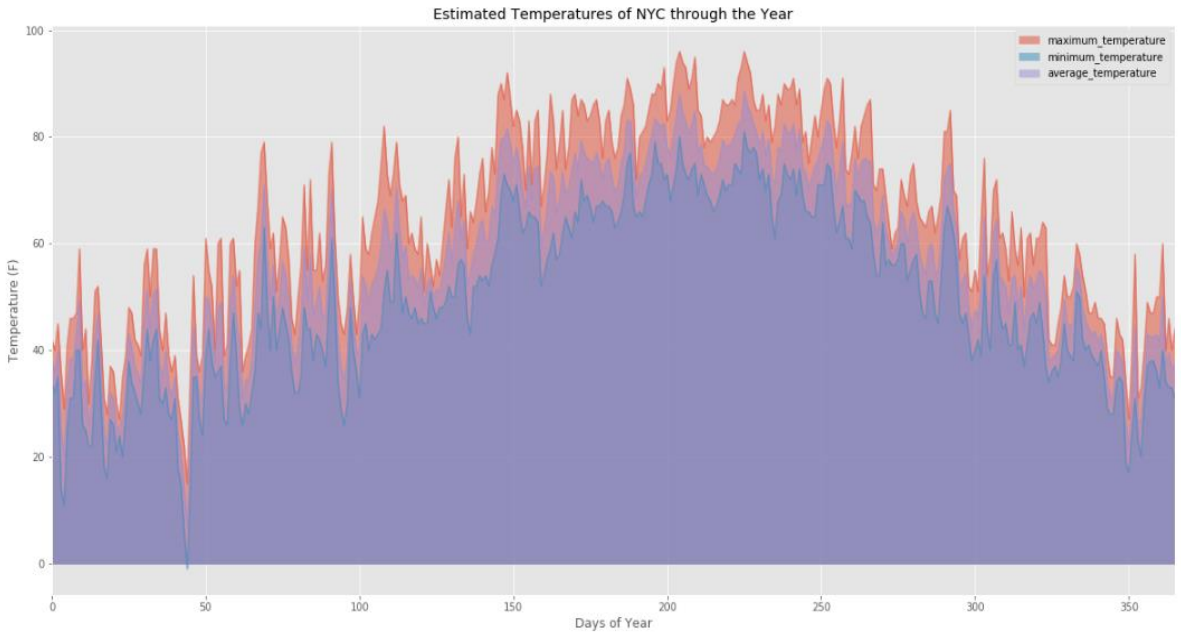
Relationship between NYC and Toronto Average Income

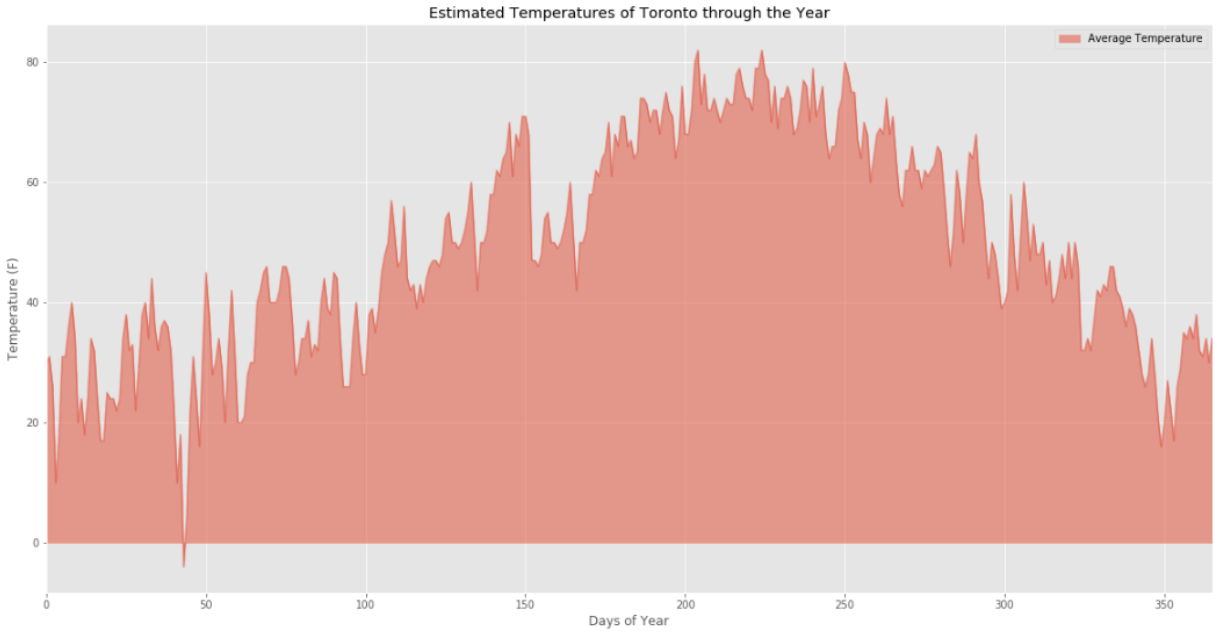




- The average house hold income for the top NYC boroughs has the income of more than 60 thousand dollars as compared to the top boroughs of Toronto where the average household income is around 50 thousand dollars.
- This shows there is more opportunity for businesses around NYC area to acquire customers because the population over there is wealthier and they are more open to try new brands and products thus creating good environment for businesses like Melaleuca to invest and grabs customer attention.

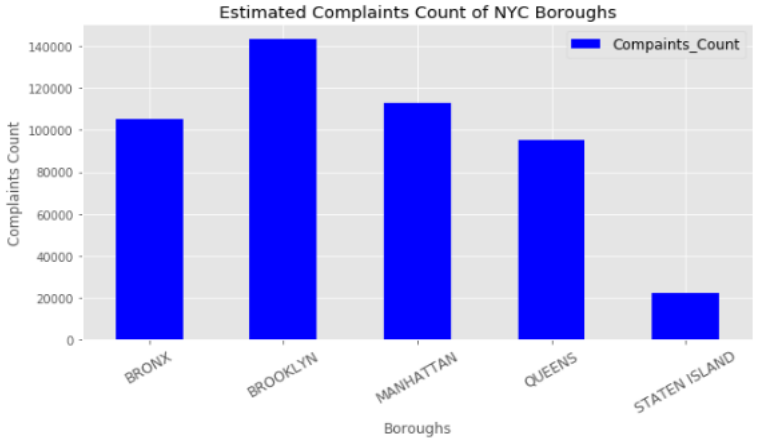
Relationship between NYC and Toronto Weather conditions

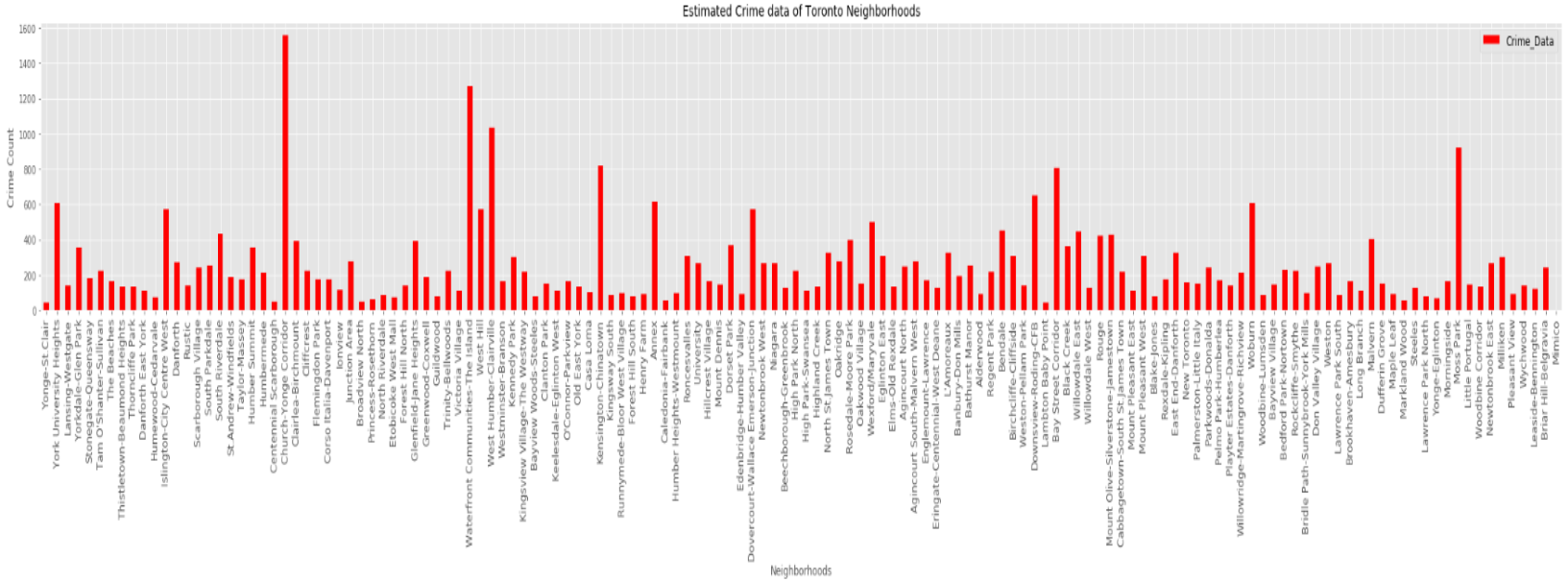


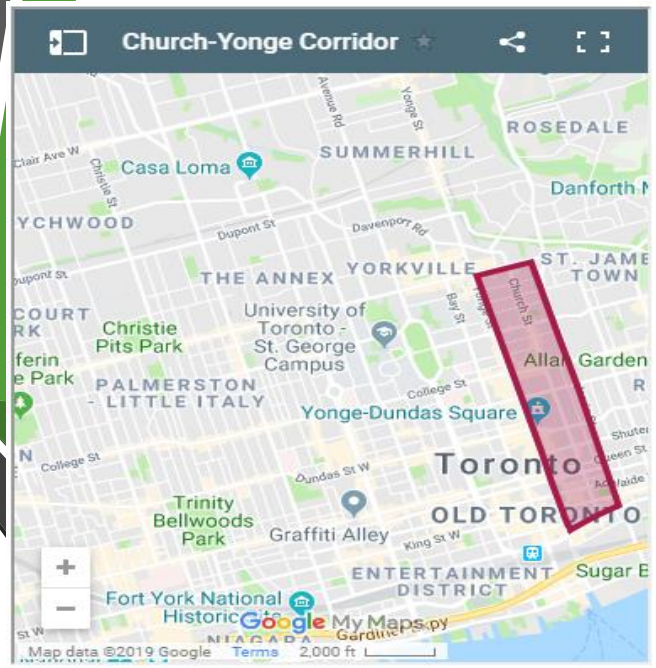


- When we compare NYC with Toronto we can see that Toronto has extreme weather conditions as compare to NYC. The NYC temperature remains in high 80(F)'s in summer and around 30(F)'s in winter where as for Toronto the temperature drops below 20(F)'s in winter sometimes goes below freezing.
- There is a very high chance that the weather condition will have impact on traffic to stores and possibly will affect the business of Melaleuca. NYC on other hand has high probability that more traffic will be attracted to stores with good favorable weather conditions.

Relationship between NYC and Toronto Crime

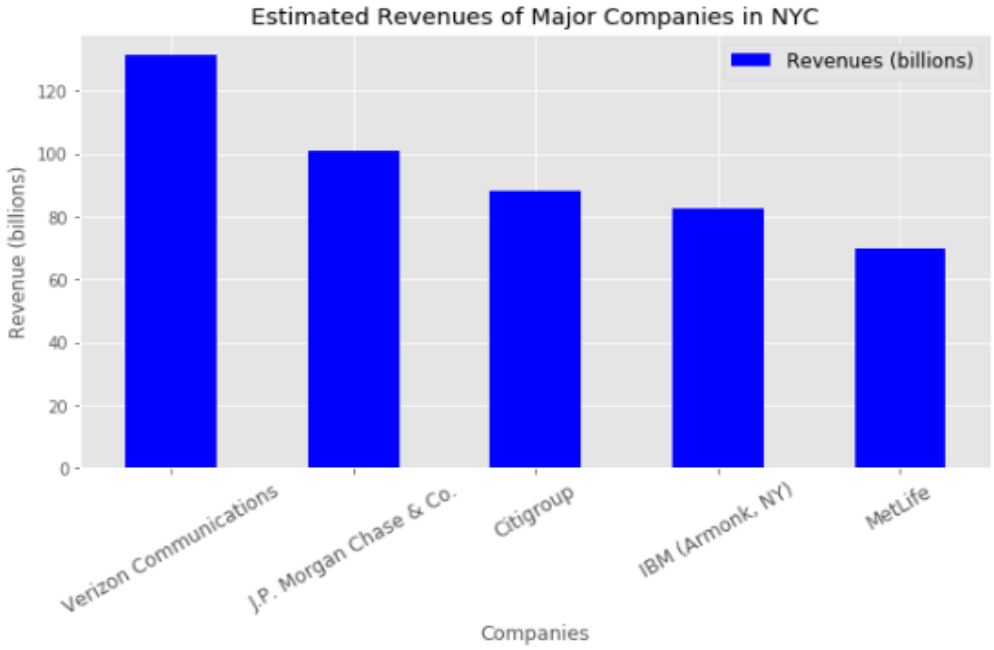


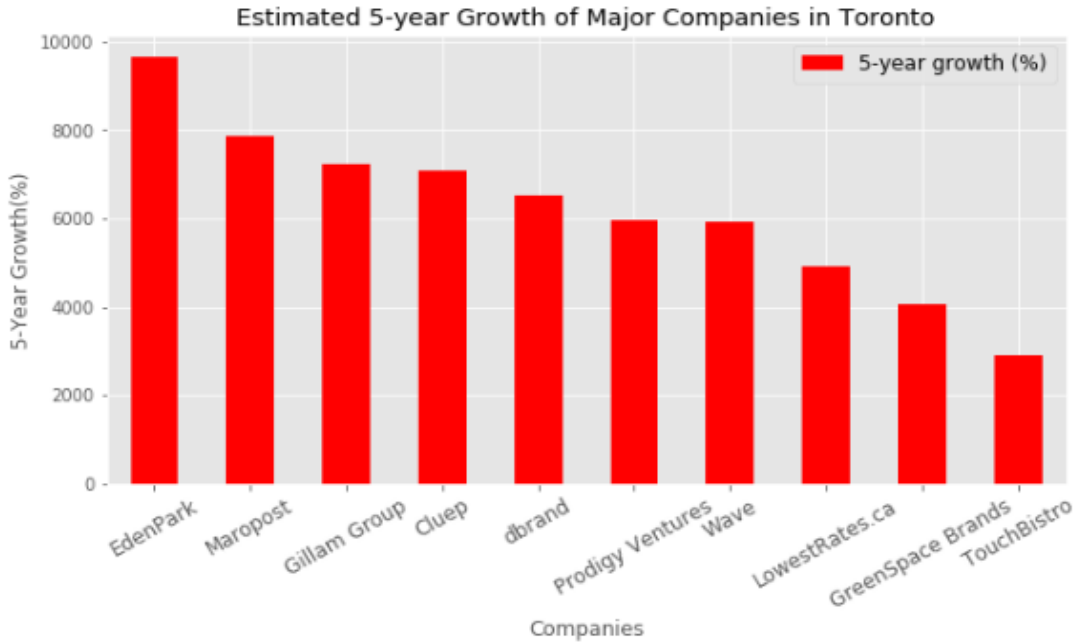




- Overall we observed that the number of crimes recorded in NYC is much higher as compare to Toronto in NYC for one full year the number of crimes recorded were 478,579 where in Toronto the number of crimes recorded were 34,930.
- For Toronto we used the Toronto's neighborhood data and observed that the number of crimes committed around church-Yonge corridor neighborhood was much higher as compare to other neighborhood.

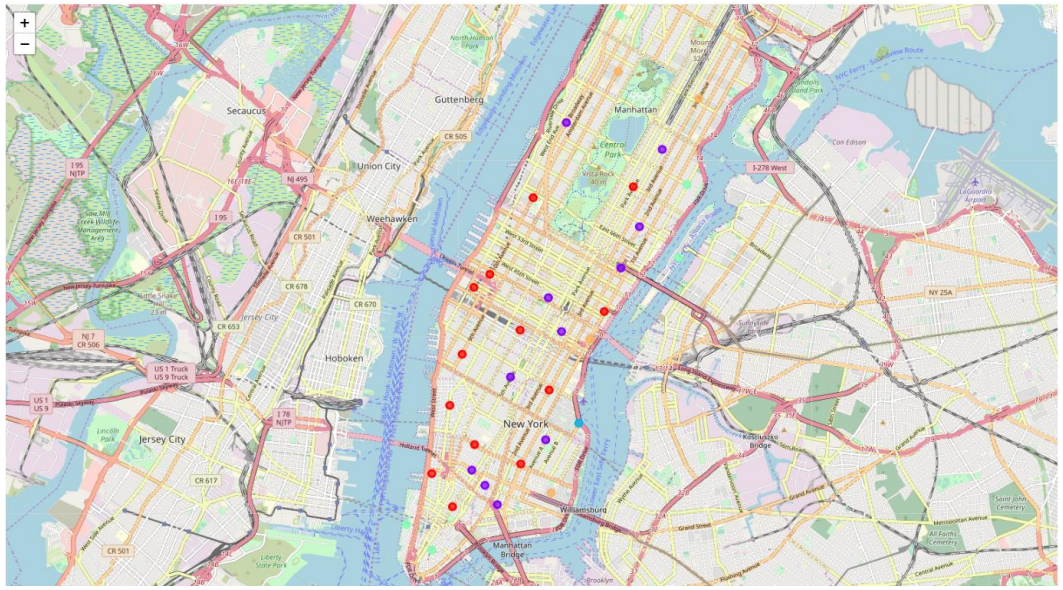
Relationship between NYC and Toronto Economy





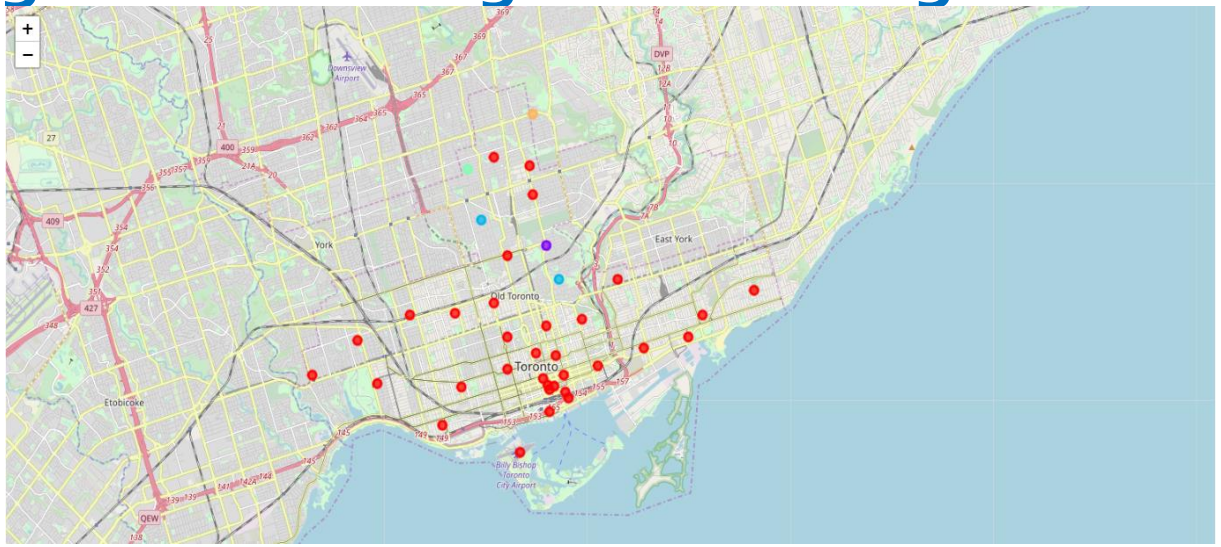
- NYC has world top companies that has the revenues in hundreds of billions. we analyzed the top 5 companies based in NYC that will give us the assessment of how the economy of the city is and it gave us a very good idea that the city is booming with a lot of opportunities for business and career for growth.
- we also did the analysis of fortune 500 business in Canada and observed that the out of fortune top 500 businesses of Canada, 112 are from Toronto and they show a very high growth where some of the businesses like Eden park and Maropost are showing almost 10,000% growth. This means that the economy is booming and in the future there will a lot more opportunities.

Segmenting and Clustering NYC Neighborhood Venues



- Cluster 1 is a Shopping Center Cluster with mix of shops like boutiques, clothing shops, Bars, restaurants and Cafe.
- Cluster 2 is more kind of recreational center where there are a lot of venues like Spa's, restaurants, hotels, nightclubs, theaters and bars etc.
- Cluster 3 is named as public visiting park because venues are more related to outdoor activities like Pak, Basketball court, baseball Field and boat or ferry etc.
- Cluster 4 will be great for tourist to stay and travel around in this area that's why we named it as tourism attraction center.
- Cluster 5 more looks like a commercial area where there is mixture of different shops, restaurants, lounges, stores and schools etc.

Segmenting and Clustering Toronto Neighborhood Venues



- Cluster 1 basically looks like a Toronto Downtown Cluster, this cluster is mainly showing all the business located in a downtown of Toronto.
- Cluster 2 basically looks like a Business Center Cluster with mix of businesses like restaurants, Bars, pet track and medical center etc. with no proper pattern.
- Cluster 3 is picnic spot cluster because there is park with a trail. There are also multiple restaurants around this cluster so it will be very convenient for the families to have an outdoor activity in this area. Pet lovers will also enjoy the spot because there is a Dog Run venue as well. In short it is very good outdoor venue for families and friends.
- Cluster 4 is given a name of festival cluster because this area is very suitable if someone who wants to arrange an event or a festival.
- Cluster 5 looks like community center cluster because there is bus line so the transport is very convenient.

Conclusion and Future Directions

- This two step approach was used by first analyzing the both cities geographic and economic situation and then building a predictive clustering model to better understand and pick the best possible neighborhood of opening a store.
- Clusters were build mainly focusing on the top 100 venues surrounding the neighborhood.
- Analysis and cluster modeling suggested that opening a new store location in NYC will be a better choice as compare to Toronto for Melaleuca Inc.
- Clusters can be further be enhanced by taking more venues under consideration.
- Increasing the number of venues also involve some challenges like the amount of data will increase by huge amount and it will take more time to process.
- With more data the venue categories will increase as well so it will be hard to extract a particular category but by increasing the number of clusters we will be able to get more granular information about the neighborhoods in each city.
- Due to high population density in NYC the moving of inventory into the city will be a challenge so special arrangements needs to be made in order to cope up with it and make sure that items doesn't get out of stock for customers are interested in
- We can further enhance the model by doing further analysis and modeling on customer behavior, shopping habits and shopping preferences in those areas. We can get the data from Kaggle or other open source data forums.