Extensive Study on Venues Around Neighborhoods of Toronto and New York City

Vismay Shah

3rd June 2020

Objective

- To identify the similarities and dissimilarities between the neighborhoods of Toronto and New York City in terms of venues exist around them using K Means Clustering algorithm

- To identify the business opportunities related to food industry in each neighborhood based on the venues exist around it, most common venues of that cluster it fall under, and most common venues of the borough it belongs to

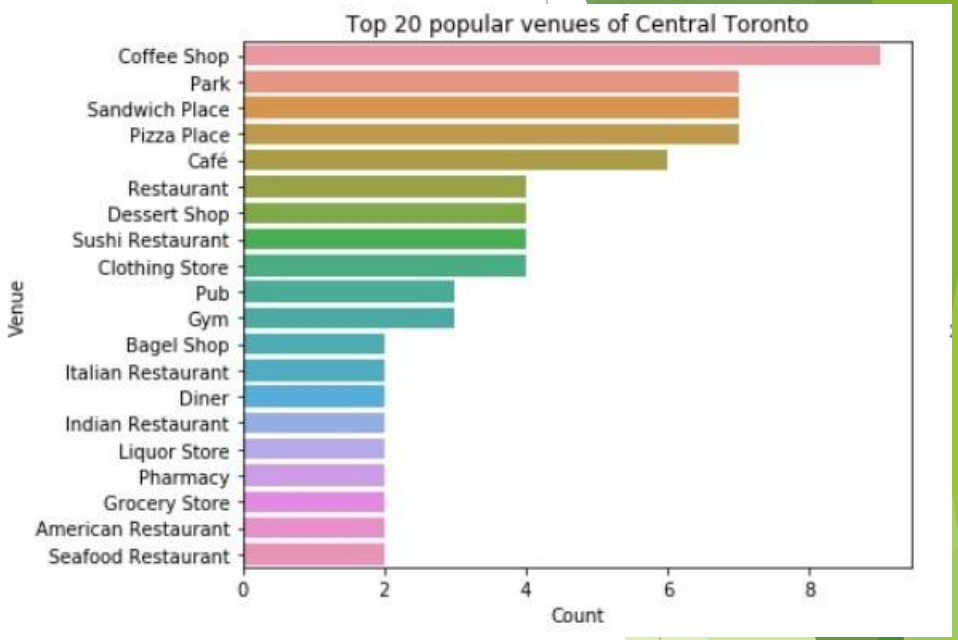
Data

- The list of Toronto's boroughs and neighborhoods was extracted from Wikipedia page which has it in Tabular format. The link of the page is https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. Latitude and longitude of all those neighborhoods were provided by IBM in ".csv" format
- The dataset of NYC's boroughs and neighborhoods was imported in JSON format using this link https://cocl.us/new_york_dataset. From the dataset imported in the JSON format, useful features like name of the borough, name of the neighborhood, and their latitude and longitude were obtained and put them into pandas dataframe
- Foursquare location data was used to explore the venues exist around each neighborhood of both the cities

Methodology

- With the information of latitude and longitude of each neighborhood and help of Foursquare location data, venues that exist within 500 meters (0.3 miles) of each neighborhood were explored and extracted into the dataframe
- The limit of number of venues to be explored was set to be 100 so no more than 100 venues will show up in the dataframe for any neighborhood
- After getting the required data, all the venue categories were converted into dummy variables (0's and 1's)
- groupby() with sum() was implemented on the dataframe and most popular venues of each borough were analyzed to get the big picture of how venues are located in particular region
- Then K means clustering was applied on the dataframe to see the difference and similarities of neighborhoods of Toronto and New York City
- After that detailed analysis was done on each neighborhood falling under a particular cluster and suggestion were provided to start a new business or expanding the business mostly in the food industry

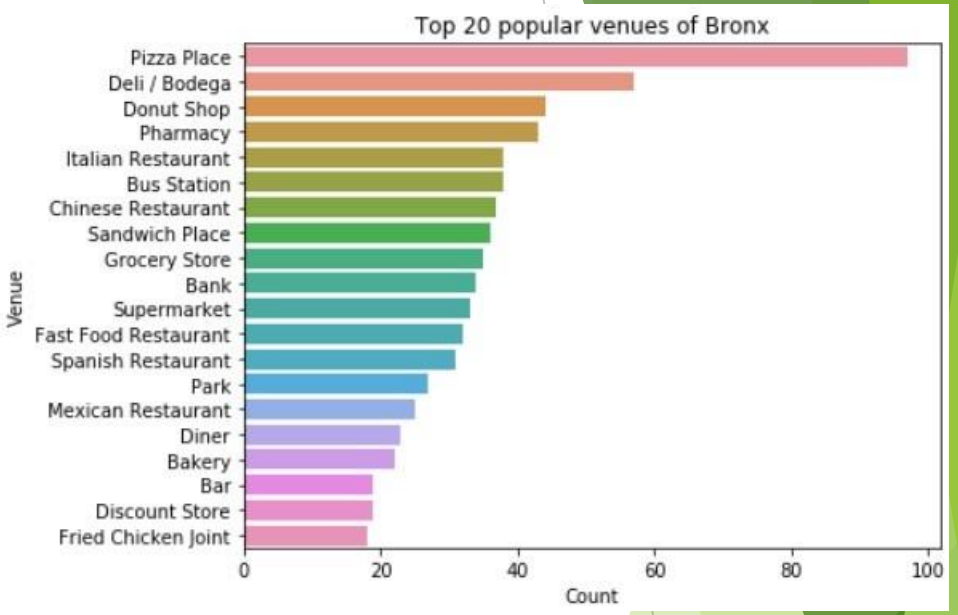
Exploratory Data Analysis (Toronto)



- There are 10 boroughs exist in the dataset
- Central Toronto, East Toronto, West Toronto, Downtown Toronto, Mississauga, North York, East York, York, Scarborough, Etibicoke
- Venues such as coffee shops, cafe, pizza places, parks, sandwich places, pharmacy, grocery store, bakery, and gym exist almost in all the boroughs
- Pubs and bars do exist in few of the boroughs
- There are variety of restaurants as well in which Italian restaurants dominate
- The less common venues were liquor stores, diner, baseball field, juice bar, pet store, spa, yoga studio, beer store,
- Downtown Toronto was tightly packed with coffee shops, cafe, pubs, bars, bakery, and variety of restaurants, and hotels. Pizza places were found to be less popular than the above venues

Exploratory Data Analysis (NYC)

- There are 5 boroughs exist in the dataset
- Bronx, Brooklyn, Manhattan, Queens, Staten Island
- Plenty of pizza place, deli/bodega, pharmacy, park, sandwich place, park, bar, and bakery were observed
- Variety of restaurants were also found in which Italian, Chinese, and Mexican being most popular ones
- Donut shop, grocery store, bank, and bagel shop found to be popular in all the boroughs except Manhattan



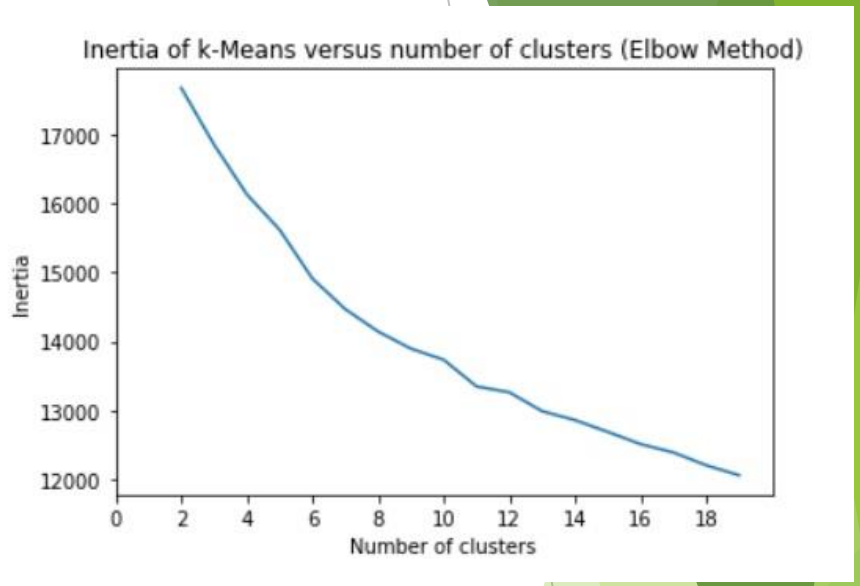
- Queens found to have highest variety of cuisines available including Chinese, Korean, Mexican, Italian, Latin American, and Thai
- Overall, most of the boroughs found to have similar venues except Manhattan
- Pharmacy, bank, donut shop, bagel shop or such venues do not fall in top 20 venues of Manhattan

Application of K Means Clustering

- Elbow method was used to find the best value of k (number of clusters) which did not prove helpful

- Silhouette Score was used as well to get better idea about choosing best value ok k but that did not help either

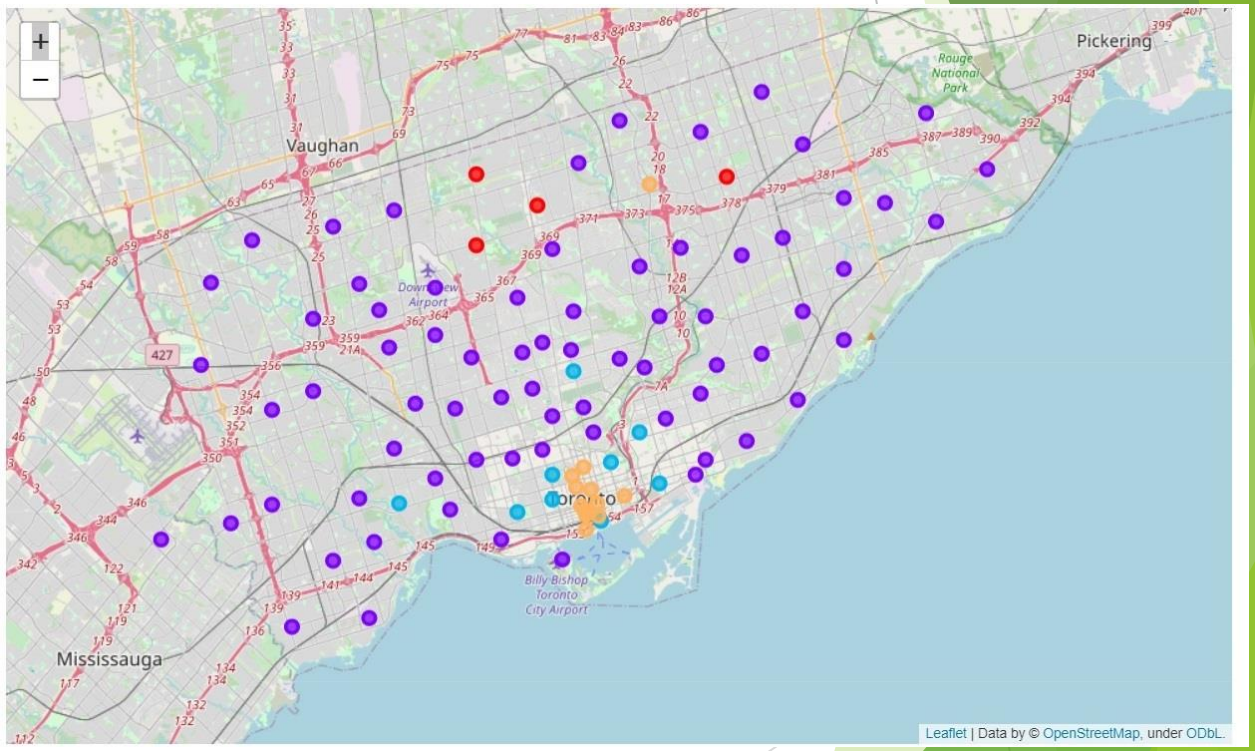
- k = 9 was selected after observing the distribution of the cluster labels for various k values



Clustered Neighborhoods of Toronto

The distribution of the neighborhood in each cluster

- Cluster 0 (red dots)- 4 neighborhoods
- Cluster 1 (violet dots)- 72 neighborhoods
- Cluster 3 (blue dots)- 9 neighborhoods
- Cluster 7 (orange dots) – 13 neighborhoods

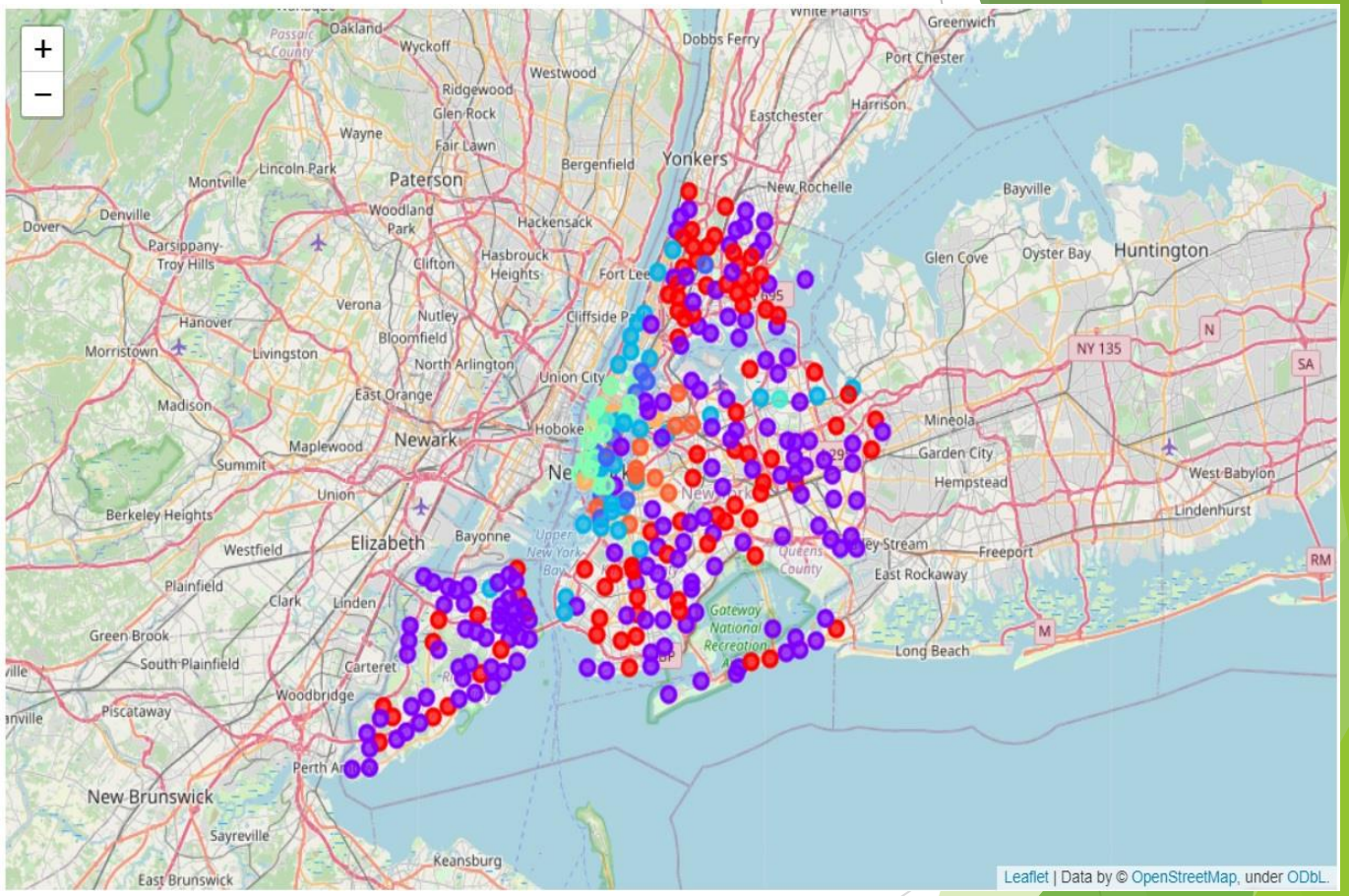


Clustered Neighborhoods of NYC

The distribution of the neighborhoods

In each cluster

- Cluster 0 (red dots)- 82
- Cluster 1 (violet dots)- 152
- Cluster 2 (light blue dots)- 10
- Cluster 3 (dark blue dots)- 30
- Cluster 4 (bluish dots)- 2
- Cluster 5 (light green dots)- 14
- Cluster 6 (lemon yellow dots)- 1
- Cluster 7 (light orange dots)- 3
- Cluster 8 (orange dots) - 11



Similarities of Neighborhoods of Toronto and NYC

- Neighborhoods of cluster 0 (Toronto) and cluster 0 (NYC) has multiple venues exactly same such as pizza place, coffee shop, bank, pharmacy, fast food restaurant, Chinese restaurant, sandwich place, and ice cream shops
- Neighborhoods of cluster 1 (Toronto) and cluster 1 (NYC) share multiple common venues such as coffee shop, park, sandwich place, cafe, grocery store, pizza place, fast food restaurant, bakery, pharmacy, bar, and Italian restaurants
- Neighborhoods of cluster 3 (Toronto) and cluster 3 (NYC) have cafe, coffee shop, Italian restaurant, pizza place, bakery, bar, park, sandwich place in common
- Neighborhoods of cluster 7 (Toronto) and cluster 7 (NYC) share common venues such as coffee shop, cafe, hotel, Japanese restaurant, clothing store, gym, American restaurant, park, salad place, bar

Dissimilarities of Neighborhoods of Toronto and NYC

- Neighborhoods fall under cluster 2, 4, 5, 6, and 8 have different characteristics
- Neighborhood of cluster 6 is Chelsea (Manhattan) which has so many art galleries around
- Couple of neighborhoods of cluster 4 have Korean community living there so have plenty of Korean restaurants around
- Most common venues of cluster 8 are bar, coffee shops, pizza places, various different types of restaurants such as Mexican, American, Thai, Japanese, Vegan, Korean, Middle Eastern, and Italian. Cluster also has lot of wine bars, deli/bodega, and bagel shops
- Cluster 2 has Italian restaurant as the most popular venue with very huge number. The other venues seem to be more or less same as the other clusters
- Neighborhoods of Cluster 5 are all located in Lower Manhattan. Variety of restaurants, pubs, bars, wine bars, cafe, and spa fall under most common venues

Opportunities of Business

- Neighborhoods of Cluster 0
  - Toronto - coffee shops, pizza places, or sandwich places
  - NYC - pizza places, donut shops, or deli/bodega
- Neighborhoods of Cluster 1
  - Toronto - coffee shops, café, pizza places, grocery stores, or sandwich places
  - NYC – coffee shops, pizza places, donut shops, bagel shops, sandwich places, or deli/bodega
- Neighborhoods of Cluster 2
  - NYC – Not recommended
- Neighborhoods of Cluster 3
  - Toronto – pubs, bars, or sandwich places
  - NYC - pizza places, donut shops, coffee shops, bagel shops around few neighborhoods

Opportunities of Business

- Neighborhoods of Cluster 4
  - NYC – Not recommended
- Neighborhoods of Cluster 5
  - NYC – coffee shops
- Neighborhoods of Cluster 6
  - NYC – Not recommended
- Neighborhoods of Cluster 7
  - Toronto – pizza places, or sandwich places
  - NYC – Not recommended
- Neighborhoods of Cluster 8
  - NYC – pizza places, or coffee shops

Future Scope

- Increase the radius of the neighborhood and see whether the same interpretation is applicable or not
- Find the best value of k (number of clusters) in more efficient way and perform the whole analysis
- Focus on single venue or single borough of any of the cities and filter out other venues and other irrelevant data to analyze a particular borough and neighborhood in much more efficient way

Conclusion

- There are so many neighborhoods of both the cities which share the same characteristics in terms of venues surrounding the neighborhoods

- NYC has got plenty of venues surrounding almost all the neighborhoods whereas the neighborhoods of Toronto have very small number of venues surrounding the neighborhoods. So, Toronto comes out as a great prospect to start or expand any business related to food industry

References

1) https://www.toronto.ca/business-economy/invest-in-toronto/strong-economy/
2) https://www.statista.com/statistics/444906/number-of-immigrants-in-canada/
3) https://en.wikipedia.org/wiki/Toronto
4) https://www.investopedia.com/articles/investing/091114/worlds-top-financial-cities.asp
5) https://en.wikipedia.org/wiki/New_York_City