**Analysis on Venue Categories in Toronto Downtown and Manhattan¶**

**Background**

New York, like Toronto,With every new wave of immigration, a new food scene inevitably follows. Now neighborhoods such as Coney Island, Corona, Flushing, Jackson Heights and Sunset Park are being revitalized the same way with new foods,New hangout spots from Ethiopian to Malaysian style.The New wave of immigration has led to explosion of population. There is a need to have new city planning.

**Problem**

Here in this study we are analyzing about different venues in Manhattan ,NY(lively neighborhoods with huge population) and Toronto Downtown (CA) neighborhoods.The ever growing population in Toronto has led to the expansion of Toronto Downtown area by 7000 sq km.Therefore a Canadian based real estate company ABC wants to invest on commercial properties to make Toronto more livelier and enjoyable experience.

**Data Acquisition and Cleaning:**

To solve this problem we need to information on all the venues categories that are in Manhattan and Toronto For eg,
1.Coffee Shop
2.Restaurants(Italian,Caribbean, etc)
3.Theatre
4.Parks
5.Fitness Centers,etc Above mentioned venue datasets are found using the foursquare api link. Now,to find all the neighborhoods,latitude,longitude in Manhattan ,New York: https://cocl.us/dataset is used.Borough,Neighborhood in Toronto are found in :https://en.wikipedia.org/wiki/.List_of_postal_codes_of_Canada:_M .Latitude ,Longitude of Toronto is found in "http://cocl.us/Geospatial_data" The above data can show us how many venues in Toronto downtown is a subset of venues in Manhattan.The venues that are not in Toronto and if its a frequented category ,can form a good recommendation.

Now,we need to create a data-frame for Newyork that contains Neighborhood,Latitude,Longitude and Boroughs values.Filter this data-frame by Manhattan.Like mentioned above create a data-frame for Toronto and then filter by Downtown Toronto borough containing the same columns as Manhattan data-frame.
**Merge** the **Manhattan data-frame**, **Downtown Toronto data-frame** as shown below

| | Borough | Neighborhood | Latitude | Longitude | City |
|---|---|---|---|---|---|
| 0 | Manhattan | Marble Hill | 40.876551 | -73.910660 | NY |
| 1 | Manhattan | Chinatown | 40.715618 | -73.994279 | NY |
| 2 | Manhattan | Washington Heights | 40.851903 | -73.936900 | NY |
| 3 | Manhattan | Inwood | 40.867684 | -73.921210 | NY |
| 4 | Manhattan | Hamilton Heights | 40.823604 | -73.949688 | NY |

## Data Analysis

Find the nearby venue categories of each neighborhoods in downtown Toronto and Manhattan .Now find mean  of each venue by neighborhood .This gives the most frequented category in each neighborhood.Below  table gives the top 5 frequented venue category.

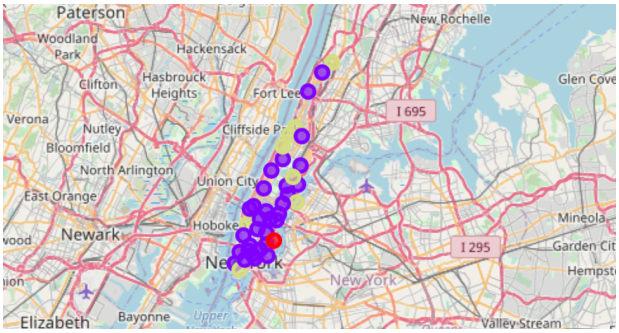
| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 0 | Battery Park City | Park | Hotel | Memorial Site | Gym | Coffee Shop |
| 1 | Berczy Park | Coffee Shop | Cocktail Bar | Seafood Restaurant | Bakery | Café |
| 2 | CN Tower / King and Spadina / Railway Lands / ... | Airport Service | Airport Lounge | Airport Terminal | Bar | Plane |
| 3 | Carnegie Hill | Coffee Shop | Yoga Studio | Bakery | Japanese Restaurant | Bookstore |
| 4 | Central Bay Street | Coffee Shop | Café | Italian Restaurant | Sandwich Place | Bar |

## Predictive Model

KMeans  algorithm is used for the below reasons:
1) Fast, robust and easier to understand.
2) Gives best result when data set are distinct or well separated from each other.
3)It is an unsupervised algorithm.
In our analysis , k-means algorithm is used to cluster the neighborhood into 4 clusters.The result should give us either neighborhoods of downtown Toronto and Manhattan  in one single cluster which means they have similarity between each other in terms of venue category or  they don't come together in any cluster ,then its  not similar.The kmeans algorithm is run on top5 frequented venues data-frame.It divides the neighborhoods of Manhattan and Downtown Toronto into 4 cluster by frequented venue categories , which  is indicated by cluster labels column.

| | Borough | Neighborhood | Latitude | Longitude | City | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Manhattan | Marble Hill | 40.876551 | -73.910660 | NY | 3 | Sandwich Place | Coffee Shop | Gym | American Restaurant | Yoga Studio |
| 1 | Manhattan | Chinatown | 40.715618 | -73.994279 | NY | 1 | Chinese Restaurant | Bakery | Cocktail Bar | Coffee Shop | Optical Shop |
| 2 | Manhattan | Washington Heights | 40.851903 | -73.936900 | NY | 1 | Café | Pizza Place | Bakery | Chinese Restaurant | Grocery Store |
| 3 | Manhattan | Inwood | 40.867684 | -73.921210 | NY | 1 | Mexican Restaurant | Pizza Place | Café | Bakery | Restaurant |
| 4 | Manhattan | Hamilton Heights | 40.823604 | -73.949688 | NY | 3 | Pizza Place | Coffee Shop | Café | Deli / Bodega | Mexican Restaurant |

The Clusters are visualized in the map as seen below shows distribution of clusters in manhattan to the left and Downtown Toronto to the right

Below is a sample cluster output got for cluster 4:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **29** | Financial District | NY | 3 | Coffee Shop | Hotel | Pizza Place | Café | |
| **30** | Carnegie Hill | NY | 3 | Coffee Shop | Yoga Studio | Bakery | Japanese Restaurant | |
| **40** | Marble Hill | NY | 3 | Sandwich Place | Coffee Shop | Gym | American Restaurant | |
| **44** | Hamilton Heights | NY | 3 | Pizza Place | Coffee Shop | Café | Deli / Bodega | Me |
| **45** | Manhattanville | NY | 3 | Coffee Shop | Seafood Restaurant | Italian Restaurant | Park | Me |
| **51** | Roosevelt Island | NY | 3 | Park | Playground | Bubble Tea Shop | Farmers Market | |
| **57** | Chelsea | NY | 3 | Art Gallery | Coffee Shop | Ice Cream Shop | Italian Restaurant | |
| **66** | Morningside Heights | NY | 3 | Park | Bookstore | Coffee Shop | American Restaurant | |
| **69** | Financial District | NY | 3 | Coffee Shop | Hotel | Pizza Place | Café | |
| **70** | Carnegie Hill | NY | 3 | Coffee Shop | Yoga Studio | Bakery | Japanese Restaurant | |
| **80** | Regent Park / Harbourfront | Tor | 3 | Coffee Shop | Park | Pub | Bakery | |
| **81** | Queen's Park / Ontario Provincial Government | Tor | 3 | Coffee Shop | Sushi Restaurant | Yoga Studio | Park | |
| **82** | Garden District / Ryerson | Tor | 3 | Clothing Store | Coffee Shop | Café | Middle Eastern Restaurant | I |
| **83** | St. James Town | Tor | 3 | Café | Coffee Shop | American Restaurant | Gastropub | |
| **84** | Berczy Park | Tor | 3 | Coffee Shop | Cocktail Bar | Seafood Restaurant | Bakery | |
| **85** | Central Bay Street | Tor | 3 | Coffee Shop | Café | Italian Restaurant | Sandwich Place | |
| **86** | Christie | Tor | 3 | Grocery Store | Café | Park | Coffee Shop | |
| **87** | Richmond / Adelaide / King | Tor | 3 | Coffee Shop | Café | Restaurant | Gym | |

## Observation

**Cluster 1** : Contains Venues such as boating/ferry ,baseball court etc from Manhattan
**Cluster 2**:.The most frequently occurred categories are restaurants(Italian,Mexican,Chinese,African),theaters,Gyms,Yoga studios seems to be the frequented places. The venues are mainly from neighborhoods of Manhattan
**Cluster 3**: Contains Venues such as Park,Playground,etc from Downtown Toronto.
**Cluster 4**: Contains venues in  neighborhoods from Manhattan and Toronto.Coffee shop and cafe seems to be extremely frequently occurred category for the respective neighborhoods.Adding to the above art gallery,park,sandwich store seems to be a frequented category for Manhattan neighborhood.

## Discussions

There are lot more varieties of restaurants ranging from Italian,middle eastern to African in Manhattan whereas in Toronto Downtown it seems one dimensional with too many coffee shops and cafes,though there are many restaurants, it lacks variety. But to me a safe bet would be a art gallery or Sandwich shop as all the manhattan neighborhood that had cafe and coffee shop as a frequented venue ,the art gallery and  Sandwich shop was also a frequented Venue.These venues on the other hand complements well with coffee shops and cafes

## Conclusion

I did not have any datas of number of people who visited the venues each day(popular??).Though both the places have diverse population it is also better to know the

ethnic demographics of each neighborhood to come to a better conclusion.The above venue is the best possible recommendation with the data available.