

# Neighborhood Comparison: Manhattan Vs Toronto



Diversity in venues.... Places to visit....

**Let's get some data and play around!!**

Good bars? Good coffeeshops within walking distance?

Where do I go for a relaxing weekend?



## Study Objective :

- ❑ New York City in USA and Toronto in Canada are one of the two very vibrant and happening cities in the world
- ❑ People tend to visit these two cities most of the time (especially people staying within few hours drive would tend to visit these places over the weekend)
- ❑ Both the cities have various similarities and dissimilarities in terms of venues that people often visit for relaxing
- ❑ In order to facilitate the decision of choosing one over the other for a weekend visit, a data based comparison of different venues/places in the respective neighborhoods is quite beneficial
- ❑ Many articles have been published on the same topic in a qualitative manner, however, since the data are available for both the cities, it would be of great importance and value if we can provide a simple comparison of the two cities and make some simple recommendations

# Data Details:

## NY (Manhattan Data)

We will use the data from our previous modules that we completed. The NY data can be accessed from the below link:

[https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)

The above link has data set which contains data from 5 boroughs and 306 neighborhoods of the NY city. In addition, it also contains the longitude and latitude of each neighborhood.

## Toronto Data

**Step1:** We scrape the data from Wikipedia as we have done this in one of the assignments in week 3. The following link contains the borough and neighborhood information for Toronto but not the longitude and latitude:

[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

**Step2:** We use the geospatial data from the below link to synchronize the co-ordinates of the neighborhoods (longitude and latitude) which we already have from Step 1. This is what we have done in the same previous assignment in week3: [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)



# Data Analysis (Manhattan)

- Data from IBM repository
- Data is already pre-processed with required information
- Foursquare API to explore neighborhoods
- Data is selected only for the Manhattan neighborhoods
- Dig at the various unique venues
- Use K-clustering to cluster the data into 5 clusters
- Use folium to visualize the Manhattan clusters

```
In [88]: manhattan_venues_total=manhattan_venues.shape[0]
print("Size of the dataframe is ",manhattan_venues.shape)
manhattan_venues.head(50)
```

Size of the dataframe is (3324, 7)

```
Out[88]:
```

|   | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue                     | Venue Latitude | Venue Longitude | Venue Category |
|---|--------------|-----------------------|------------------------|---------------------------|----------------|-----------------|----------------|
| 0 | Marble Hill  | 40.876551             | -73.910660             | Arturo's                  | 40.874412      | -73.910271      | Pizza Place    |
| 1 | Marble Hill  | 40.876551             | -73.910660             | Bikram Yoga               | 40.876844      | -73.906204      | Yoga Studio    |
| 2 | Marble Hill  | 40.876551             | -73.910660             | Tibbett Diner             | 40.880404      | -73.908937      | Diner          |
| 3 | Marble Hill  | 40.876551             | -73.910660             | Starbucks                 | 40.877531      | -73.905582      | Coffee Shop    |
| 4 | Marble Hill  | 40.876551             | -73.910660             | Dunkin'                   | 40.877136      | -73.906666      | Donut Shop     |
| 5 | Marble Hill  | 40.876551             | -73.910660             | Blink Fitness Riverdale   | 40.877147      | -73.905837      | Gym            |
| 6 | Marble Hill  | 40.876551             | -73.910660             | TCR The Club of Riverdale | 40.878628      | -73.914568      | Tennis Stadium |

## Clustering of Manhattan Neighborhood

Step 1: Run k-means to cluster the neighborhood into 5 clusters.

Step 2: Create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

Step 3: Visualize the new cluster

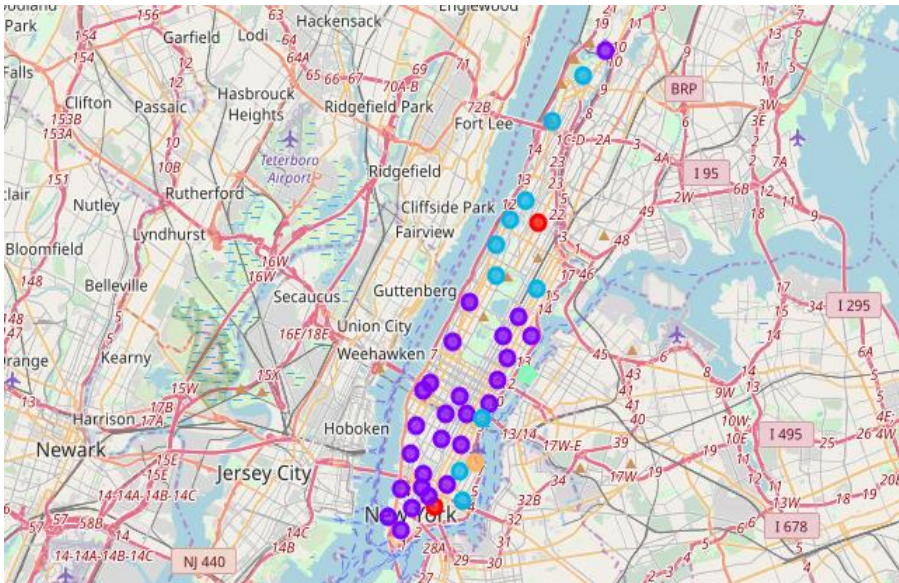
```
In [21]: # set number of clusters
kclusters = 5

manhattan_grouped_clustering = manhattan_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(manhattan_grouped_clustering)

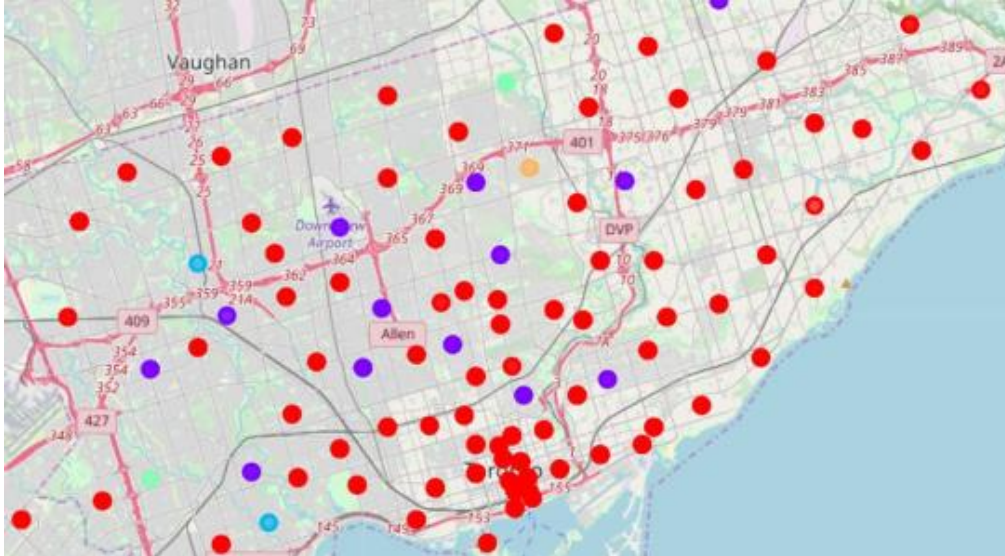
# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
Out[21]: array([2, 1, 1, 2, 1, 2, 2, 0, 1, 1], dtype=int32)
```



# Data Analysis (Toronto)

- Data from Wikipedia
- Data is pre-processed with to avail required information
- Foursquare API to explore neighborhoods
- Dig at the various unique venues
- Use K-clustering to cluster the data into 5 cluster
- Use folium to visualize the Toronto clusters



```
In [44]: print(toronto_venues.shape)
toronto_venues_total=toronto_venues.shape[0]
toronto_venues.head()
```

(2264, 7)

```
Out[44]:
```

|   | Neighbourhood    | Neighbourhood Latitude | Neighbourhood Longitude | Venue                  | Venue Latitude | Venue Longitude | Venue Category       |
|---|------------------|------------------------|-------------------------|------------------------|----------------|-----------------|----------------------|
| 0 | Parkwoods        | 43.753259              | -79.329656              | Brookbanks Park        | 43.751976      | -79.332140      | Park                 |
| 1 | Parkwoods        | 43.753259              | -79.329656              | KFC                    | 43.754387      | -79.333021      | Fast Food Restaurant |
| 2 | Parkwoods        | 43.753259              | -79.329656              | Variety Store          | 43.751974      | -79.333114      | Food & Drink Shop    |
| 3 | Victoria Village | 43.725882              | -79.315572              | Victoria Village Arena | 43.723481      | -79.315635      | Hockey Arena         |
| 4 | Victoria Village | 43.725882              | -79.315572              | Tim Hortons            | 43.725517      | -79.313103      | Coffee Shop          |

## Clustering of Toronto Neighborhood:

Run k-means to cluster the neighborhood into 5 clusters

```
In [42]: from sklearn.cluster import KMeans
# set number of clusters
kclusters = 5

toronto_grouped_clustering = toronto_grouped.drop('Neighbourhood', 1)

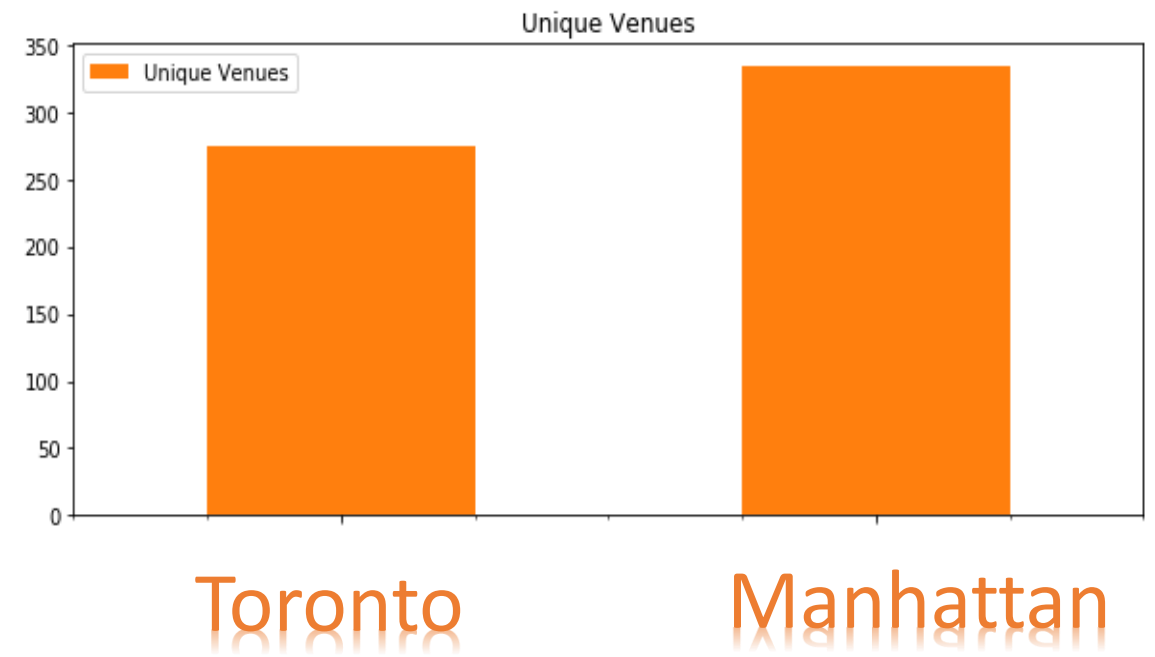
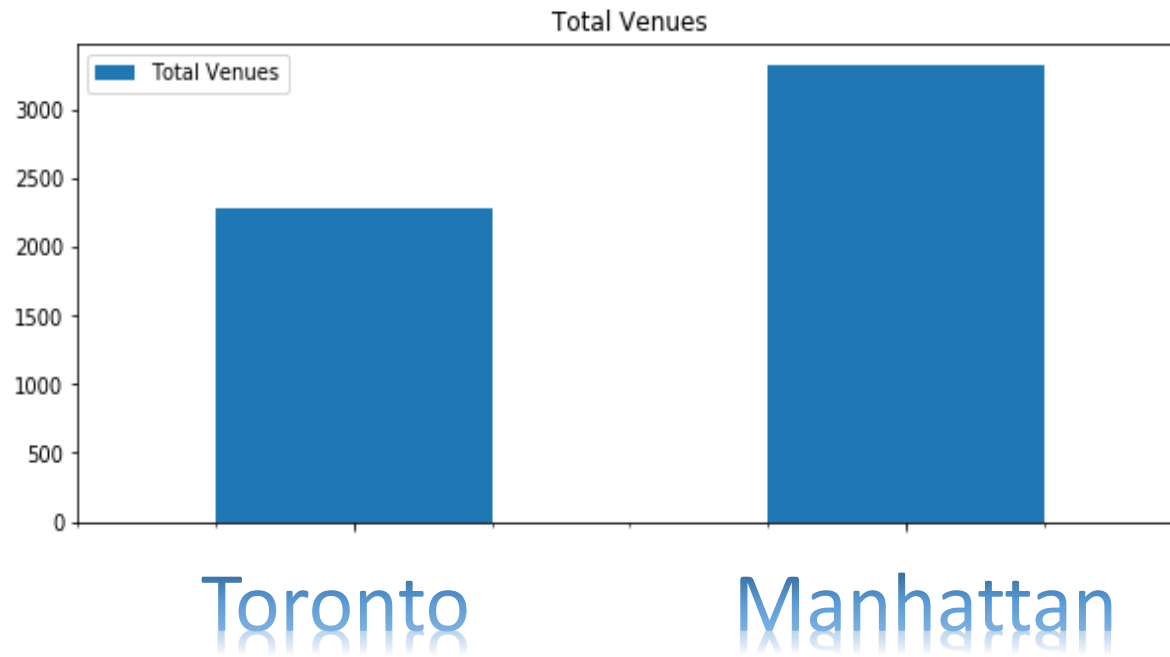
# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(toronto_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
Out[42]: array([0, 0, 4, 0, 0, 0, 0, 0, 0, 0], dtype=int32)
```

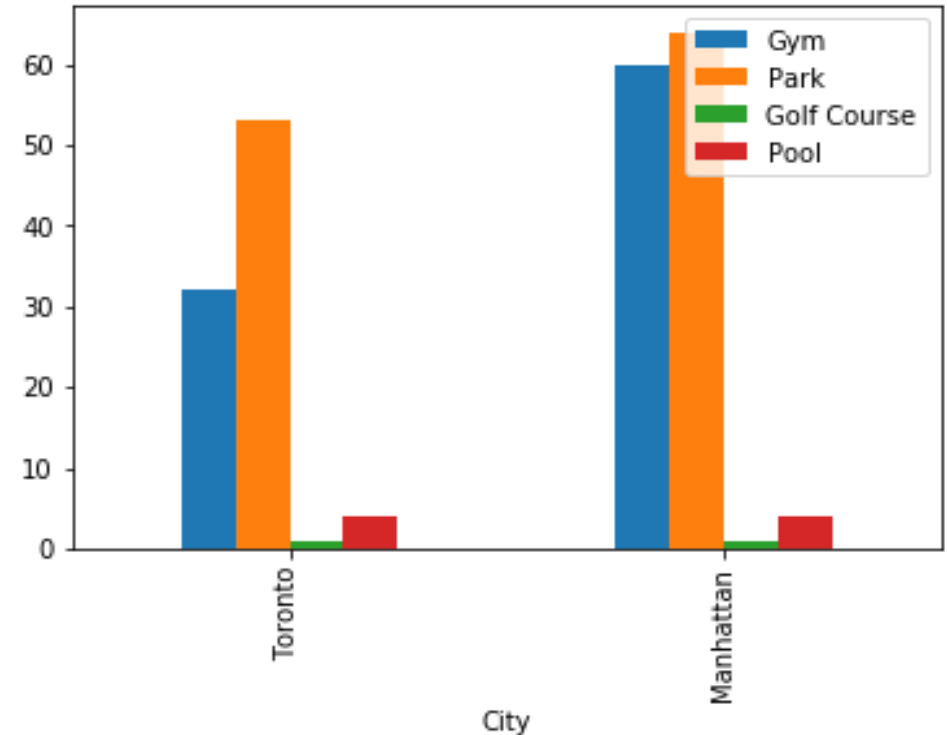
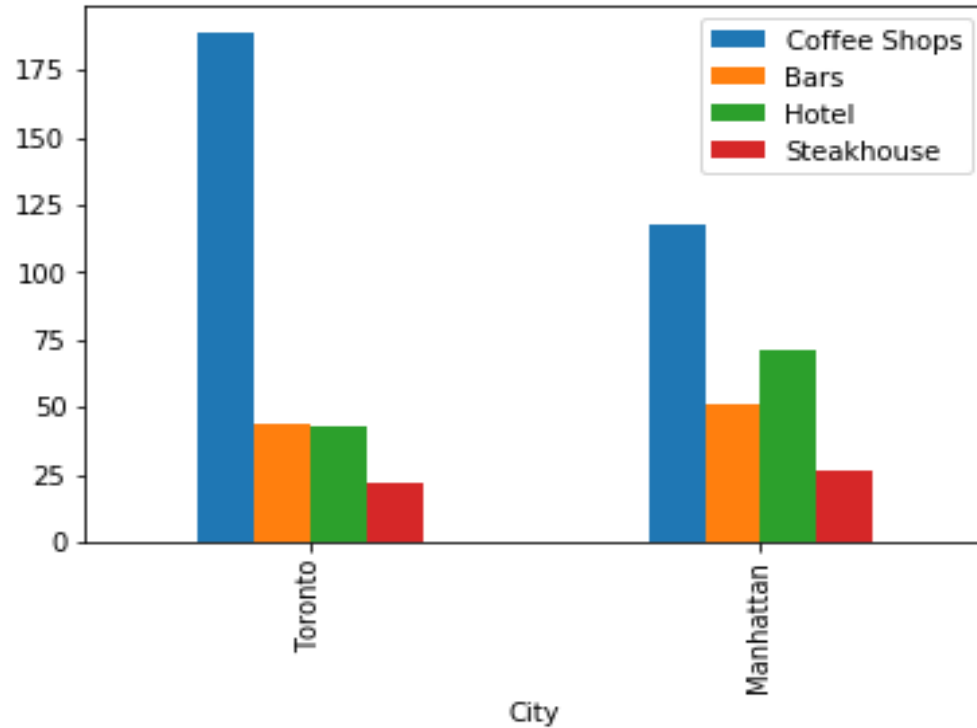
# Comparative Analysis

- ❑ Manhattan area has 3324 venues while Toronto has 2264 venues
- ❑ Manhattan area has 339 unique venues while Toronto has 275



# Comparative Analysis

- ❑ Manhattan area has higher number of Bars, Hotels and Steakhouses in a smaller area
- ❑ Manhattan area has more fitness centers and parks in a smaller area
- ❑ Toronto area has reasonably higher number of coffee shops
- ❑ Both the areas have same number of Golf Courses and Pools



# Conclusions and Outcome

- ✓ Both Manhattan and Toronto neighborhood have very diverse and dynamic environment which make both a very nice place to spend the weekend or in general a nice tourist place.
- ✓ Manhattan has 31% more total venues compared to Toronto. A smaller area of the Manhattan neighborhoods tells us that Manhattan area has a much higher venue density, meaning, most of these venues are easily accessible if one can browse around on foot. In other words, the venues are comparatively closer in Manhattan area than the Toronto area. In addition, Manhattan has 18% more unique venues than Toronto which makes Manhattan more interesting or lucrative compared to Toronto.
- ✓ One interesting observation is Manhattan has more Bars, Steakhouses and Hotels in the smaller area compared to the bigger Toronto area. But at the same time Toronto has ~40% more coffee shops than the Manhattan area. So, this can be a very good indication of what to expect in the respective area.
- ✓ Another important outcome is Manhattan has more fitness centers and parks in a smaller area compared to Toronto. At the same time, both the neighborhoods have the same number of pools and Golf course

## **Final Remark:**

*Although both neighborhoods have their own characteristics, I would put Manhattan at a slightly higher position as a lively tourists' spot. However, as a weekend gateway or couple of days of relaxing days out, I would put Toronto at a slightly higher position because with all the necessary venues in it, it is a bit less densely populated and it seems to be more peaceful than Manhattan.*