

# MANHATTAN VS TORONTO: WHERE SHOULD I TRAVEL TO?

---

Coursera Capstone Project

By Christopher Luu

# THE PROBLEM

- ❖ Ever wondered whether to travel to America or Canada?
- ❖ How about exploring their two main cities of New York or Toronto?
- ❖ This analysis presents an analysis on Manhattan (Borough of New York) and a subset of Toronto providing interested travellers a suggestion for their next destination!

# BACKGROUND

- ❖ The two major cities of New York and Toronto are both tempting and exciting places to be
- ❖ Deciding on only one of the two provides an ultimatum and that is an opportunity foregone until you save enough money to travel again
- ❖ By understanding the most common venues at these cities, an meaningful decision can be concluded ensuring you made the right decision for travel

# DATA USED

- ❖ Manhattan data available from: [https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork\\_data.json](https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json)
- ❖ Toronto data available from: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- ❖ Geospatial data available from: [https://cocl.us/Geospatial\\_data](https://cocl.us/Geospatial_data)

# DATA USED

- ❖ The data sourced on previous slide was wrangled, cleaned and formatted to the tables on the right
- ❖ Both present the following features:
  - ❖ Borough
  - ❖ Neighbourhood
  - ❖ Latitude
  - ❖ Longitude
- ❖ Both wrangled data sets are suitable for clustering in conjunction with Foursquare Location Data

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

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

# METHODOLOGY

1. Source Data
2. Data Wrangling
3. Mapping Initial Data with Folium
4. Foursquare API
5. Clustering with K-means
6. Mapping Clusters
7. Analysis
8. Conclusion

# METHODOLOGY

- ❖ Key data wrangling was on Toronto data
- ❖ Analysing data of similar shapes more standardised for comparison
- ❖ Tables on right show process of combining East, West, Central and Downtown Toronto Borough into 1 main borough – Toronto Main
- ❖ Toronto Main: 1 Borough, 39 Neighbourhoods
- ❖ Manhattan: 1 Borough, 40 Neighbourhoods
- ❖ Data wrangling and processing successful

```
In [138]: test = merged_table['Borough'].unique()
test
```

```
Out[138]: array(['North York', 'Downtown Toronto', 'Etobicoke', 'Scarborough',
                'East York', 'York', 'East Toronto', 'West Toronto',
                'Central Toronto', 'Mississauga'], dtype=object)
```

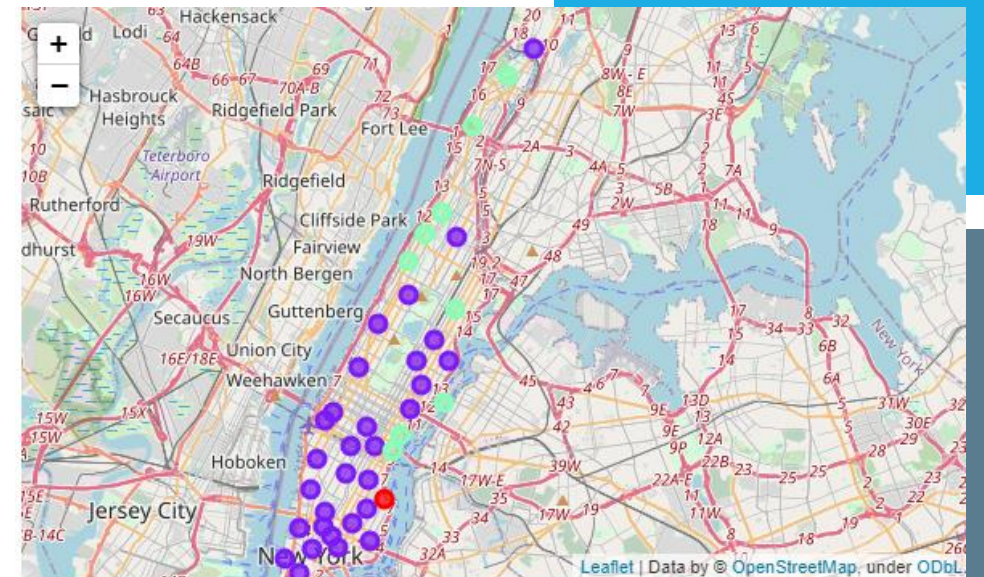
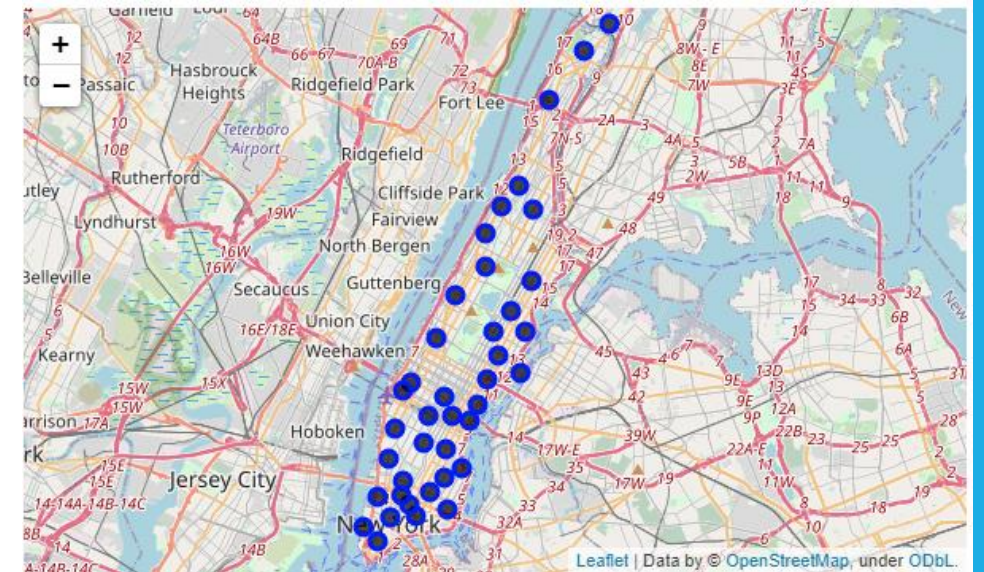
```
In [139]: #identifying best borough to use
for i in test:
    a = merged_table[merged_table['Borough']==i]
    print('The dataframe {} has {} boroughs and {} neighborhoods.'.format(i,1
```

```
The dataframe North York has 1 boroughs and 24 neighborhoods.
The dataframe Downtown Toronto has 1 boroughs and 19 neighborhoods.
The dataframe Etobicoke has 1 boroughs and 12 neighborhoods.
The dataframe Scarborough has 1 boroughs and 17 neighborhoods.
The dataframe East York has 1 boroughs and 5 neighborhoods.
The dataframe York has 1 boroughs and 5 neighborhoods.
The dataframe East Toronto has 1 boroughs and 5 neighborhoods.
The dataframe West Toronto has 1 boroughs and 6 neighborhoods.
The dataframe Central Toronto has 1 boroughs and 9 neighborhoods.
The dataframe Mississauga has 1 boroughs and 1 neighborhoods.
```

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M5A	Toronto Main	Regent Park, Harbourfront	43.654260	-79.360636
1	M7A	Toronto Main	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
2	M5B	Toronto Main	Garden District, Ryerson	43.657162	-79.378937
3	M5C	Toronto Main	St. James Town	43.651494	-79.375418
4	M5E	Toronto Main	Berczy Park	43.644771	-79.373306
5	M5G	Toronto Main	Central Bay Street	43.657952	-79.387383

# RESULTS

- ❖ Top image shows pre-clustered Manhattan data
- ❖ Bottom image shows post-clustered Manhattan data
- ❖ 3 Clusters
  1. Red (Cluster 0)
  2. Purple (Cluster 1)
  3. Green (Cluster 2)
- ❖ Number of clusters  $K = 3$ , 3 Clusters generated





# RESULTS

- ❖ Top image shows pre-clustered Toronto data
- ❖ Bottom image shows post-clustered Toronto data
- ❖ 1 Clusters
  1. Red (Cluster 0)
- ❖ Number of clusters  $K = 3$ , 1 Clusters generated
- ❖ Not enough diversity in venue data



# RESULTS

## Manhattan Cluster 0

Park 1  
Name: 1th Most Common Venue,

## Manhattan Cluster 1

Italian Restaurant	8
Coffee Shop	6
Bar	3
Café	2
Gym / Fitness Center	1
Hotel	1
Theater	1
Gym	1
American Restaurant	1
Park	1
Clothing Store	1
Art Gallery	1
Korean Restaurant	1
Chinese Restaurant	1
Plaza	1

Name: 1th Most Common Venue,

❖ 1<sup>st</sup> Most Common  
Venue per  
neighbourhood counted  
by frequency of  
apperance

❖ Food shops high  
frequency in  
Manhattan

❖ Spa only in Toronto

## Manhattan Cluster 2

Park	3
Café	2
Coffee Shop	2
Pizza Place	1
Mexican Restaurant	1

Name: 1th Most Common Venue,

## Toronto Cluster 0

Venue CategorySpa 19  
Name: 1st Most Common Venue,



# DISCUSSION

- ❖ Not enough diversity in Toronto data to warrant more clusters
- ❖ Larger data set would give more accurate comparison between the two cities
- ❖ Current analysis suggests:
  1. Food loving travellers go to Manhattan.
    - ❖ More diversity in food venues per 500m radius from each neighbourhood
  2. Relaxation travellers go to Toronto
    - ❖ Spa venues most common amongst the neighbourhoods.
    - ❖ High saturation may imply high demand

# CONCLUSION

- ❖ K-Means clustering performed on Manhattan and Toronto in conjunction with Foursquare location data
- ❖ Food travellers recommended to Manhattan
- ❖ Relaxation travellers recommended to Toronto
- ❖ Dataset used not extensive enough to recommend anything further
- ❖ For improved suggestions, more in-depth analysis required for future work including more venues and locations.