

# Data Science Capstone Final Project

Tarini Ghosh

# Using Four-Square API to get venue type information for each neighborhood

- Using Four-Square API, I downloaded the venuetypes for each neighborhood belonging to the two cities
- The types of venues were then categorized as 'found in both cities', 'found only in Toronto' and 'Found only in New York'
- Let's first focus on those 'found in both cities'.
- Even within them there were differences.

# Top 20 venue-types having higher prevalence in either cities

## New York

	Toronto	New York	LogFoldChange
['Bus Station']	0.001947	0.027586	-2.650811
['Mobile Phone Shop']	0.000974	0.010345	-2.363129
['Martial Arts Dojo']	0.000974	0.006897	-1.957663
['Shoe Store']	0.002921	0.013793	-1.552198
['Supermarket']	0.005842	0.025862	-1.487660
['Deli / Bodega']	0.009737	0.036207	-1.313306
['Supplement Shop']	0.001947	0.006897	-1.264516
['Baseball Field']	0.001947	0.006897	-1.264516
['Moving Target']	0.000974	0.003448	-1.264516
['Metro Station']	0.003895	0.012069	-1.130985
['Playground']	0.002921	0.008621	-1.082195
['Market']	0.001947	0.005172	-0.976834
['Latin American Restaurant']	0.005842	0.015517	-0.976834
['Discount Store']	0.008763	0.022414	-0.939094
['Bank']	0.014606	0.032759	-0.807758
['Chinese Restaurant']	0.014606	0.032759	-0.807758
['Food Truck']	0.003895	0.008621	-0.794513
['Mexican Restaurant']	0.009737	0.018966	-0.666679
['Pharmacy']	0.018500	0.034483	-0.622662

## Toronto

	Toronto	New York	LogFoldChange
['Cheese Shop']	0.004869	0.001724	1.038069
['Art Gallery']	0.009737	0.003448	1.038069
['Café']	0.029211	0.010345	1.038069
['Indian Restaurant']	0.010711	0.003448	1.133379
['Bagel Shop']	0.005842	0.001724	1.220390
['Breakfast Spot']	0.017527	0.005172	1.220390
['Clothing Store']	0.012658	0.003448	1.300433
['Thai Restaurant']	0.012658	0.003448	1.300433
['Farmers Market']	0.006816	0.001724	1.374541
['Sporting Goods Shop']	0.007790	0.001724	1.508072
['Yoga Studio']	0.008763	0.001724	1.625855
['Greek Restaurant']	0.008763	0.001724	1.625855
['Sushi Restaurant']	0.017527	0.003448	1.625855
['Lounge']	0.008763	0.001724	1.625855
['French Restaurant']	0.009737	0.001724	1.731216
['Beer Bar']	0.009737	0.001724	1.731216
['Steakhouse']	0.011685	0.001724	1.913538
['Dessert Shop']	0.011685	0.001724	1.913538
['Bookstore']	0.014606	0.001724	2.136681
['Japanese Restaurant']	0.021422	0.001724	2.519673

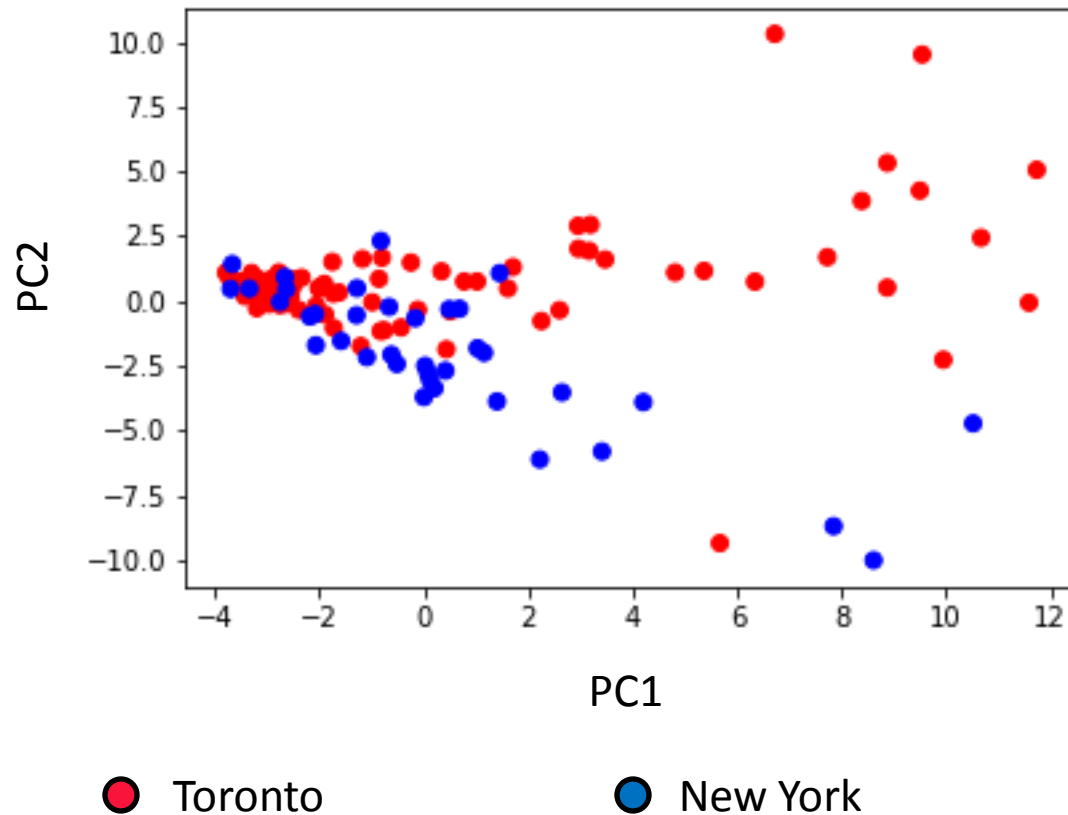
# Kinds of venues specifically high in either cities displayed specific trends

- Restaurants:
  - Deli, Latin American, Chinese and Mexican Restaurants were higher in New York
  - Asian (Thai, Japanese, Indian, Sushi) and European (Greek, French, Bagel shop) were higher in Toronto
- Farmers' markets, Art gallery, Yoga studio, Book store: Toronto
- Baseball field, Food trucks, Supermarkets, Discount store, Banks: New York
- Identification of such differences could help individuals interested for setting up businesses in specific cities.

# Probing which neighborhoods are similar to each other across both cities

- Using Multivariate Principal Component Analysis, we can identify similarities among neighborhoods
- The objective was to first check if we could group neighborhoods across both cities into clusters.
- This will be important for someone who wants to move from one city to another and wants to explore similar neighborhoods

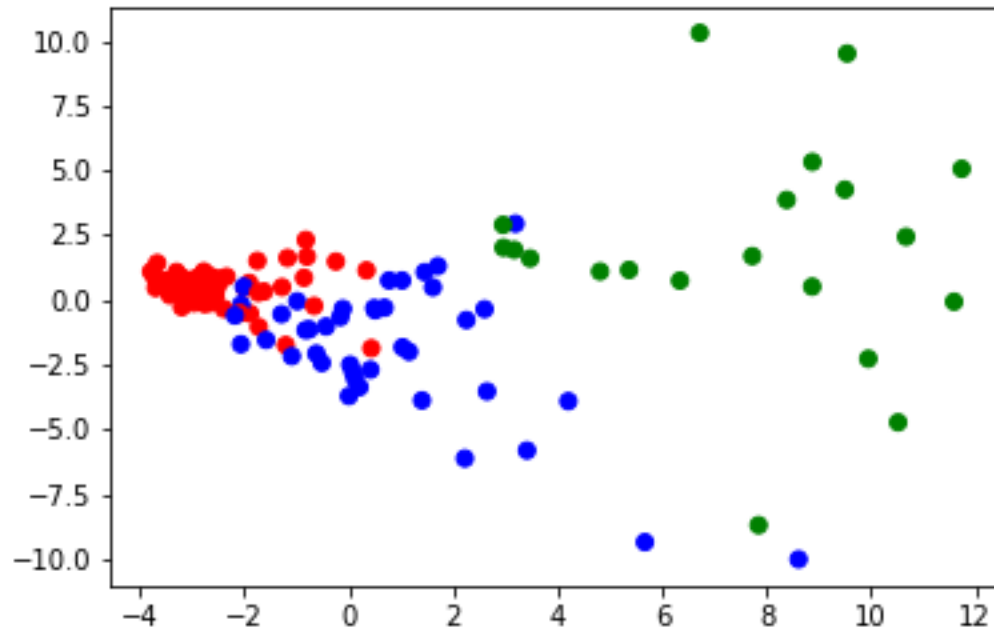
# PCA plot of neighborhoods from two cities shows three groupings



# Three groupings

- There are three distinct groups, one covering neighborhoods from both cities, the other with a larger spread and containing only neighborhoods from Toronto, the third an outlier group dominated by those from New York.
- Further validated using k-means

# K-means clustering ( $k=3$ ) also shows exact the same trend



● Cluster1 Neighborhoods  
(found in both cities)

● Cluster2 Neighborhoods  
(Toronto dominated)

● Cluster3 Neighborhoods  
(New York dominated)



# Conclusion

- The above analysis gives a rough overview of how cities can be compared among each other, to identify:
  - a. Neighborhoods that are similar (important especially for travelers and people trying to look into similar neighborhoods for moving in).
  - b. What venues are highly prevalent or less prevalent across cities and venues (important especially for individuals looking to set up business).