# BATTLE OF THE CITIES

**J.R. Gutierrez**

## Intro / Business Problem

John and his wife Libby want to take a vacation to either New York and Toronto. However they aren't sure what would offer them the right experience for what they are seeking. New York and Toronto are two very powerful, diverse cities in their respective countries. While they are very similar, they also differ greatly in their food, hotels and tourists attractions. John and Libby don't find enough information online to satisfy these concerns.¶

## Background

John happens to be a data scientist, and will explore New York and Toronto by segmenting and clustering the most exciting neighborhood for each city using the Foursquare API and K-Means algorithm. These clusters of data will include the restaurants, hotels, parks, galleries and other tourist attractions. The neighborhoods selected include Manhattan from New York and Downtown Toronto from Toronto.

## Data

John will use the ML algorithm K-means to cluster the neighborhoods with similar objects, with a focus on locating the hottest tourist spots and venues of each. This way the cities can be analyzed and compared adequately.

For downtown Toronto, John will extract data from a wikipedia page to breakdown the neighborhoods within. Then he we will clean and wrangle data as needed, like removing duplicates and "Not assigned" values. John will use Foursquare API to retrieve the coordinates of Downtown Toronto and explore its neighborhoods. The neighborhoods are then further characterized and clustered.

Manhattan will go through much of the same process. John will retrieve a saved data file which is already explored through foursquare API in which he extracted and sorted all the boroughs of New York. Afterwards, Manhattan's neighborhoods and its venues will be characterized and clustered as well.

# Methodology

John visualized the data in various stages through a machine learning algorithm called K-means that clusters data. In the first stage, John clusters neighborhoods in downtown Toronto and Manhattan so that he can better understand their general whereabouts and nearby venues. He then clusters downtown Toronto and Manhattan again so that he can more specifically group the clusters by which venues are nearest them. He then analyzes the proximity of these various tourist attractions, and assigns them names accordingly. This way John and Libby can better the clusters they'd like to visit during their vacation.¶

## One Hot Encoding

Before clustering, John analyze cities through one hot encoding (giving '1' if a venue category is there, and '0' in case of venue category is not there). On the basis of one hot encoding, he calculated the mean of the frequency of occurrence for each category and picked top ten venues on that basis for each neighborhood. It means the top venues are showing the foot traffic or the more visited places.¶

## Exploring Neighborhoods in Downtown Toronto

John clustered the neighborhoods of Downtown Toronto into 5 clusters using K-Means modeling. These were the clusters that were created and the venues that were assigned to each:

Cluster 1 (Coffee Shops, Cafes, Restaurants & Grocery Stores)

Cluster 2 (Gastropubs)

Cluster 3 (Cafes)

Cluster 4 (Coffee Shop, Cafe, Park & Japanese Restaurant)

Cluster 5 (Seafood, steakhouse, Hotel & Cafe)

See

### Exploring Neighborhoods in Manhattan

John clustered the neighborhoods of Manhattan into 5 clusters using K-Means modeling. These were the clusters that were created and the venues that were assigned to each:

Cluster 1 (Residential)

Cluster 2 (Commercial Places)

Cluster 3 (Tourist Areas & Hubs)

Cluster 4 (Center Activity)

Cluster 5 (Cultural & Going Out Places)

## RESULTS

After clustering the data of the respective neighborhoods of both Downtown Toronto and Manhattan, John and Libby can more clearly see which of the neighborhoods they would like to see based on the tourist attractions nearby. The neighborhoods are very similar in features like restaurants, movie theaters, opera shows, diners, clubs, museums, parks and more. The biggest differences reside in their historical places and monuments, due to their diverse histories.

## Discussion

When John and Libby compared the cities and their varied neighborhoods, they observed that while downtown Toronto had exciting venues like a nearby airport, a harbor with ferry services and gardens nearby, Manhattan had a more exciting amount of museums, restaurants and cafes of different cultures to choose from-- all of which were very close by.

The tipping point for John and Libby was the Rock Club and the French Restaurant in Noho. Both John and Libby are avid rock climbers. In addition, they are lovers of French food. So by clustering this data, they decided they'd have a day where they go rock climbing and then have dinner after at a French Restaurant.

# Conclusion

Clustering the data for both Downtown Toronto and Manhattan proved to be a great success. In successfully clustering their respective neighborhoods, this couple managed to uncover the amazing tourist attractions that exists in both places. However, Manhattan's Noho especially caught their attention since they are avid fans of French bistro and rock climbing. They have already bought their tickets;)

(Due to a bad Gateway on https://labs.cognitiveclass.ai/ I wasn't able to upload images and tables from the notebook jupyter). To see these tables and the coding that produced them, please visit this link:
https://github.com/jrgutz13/Coursera_Capstone/blob/master/Final_Battle_of_the_Cities.ipynb