The Choice | NYU v/s UofT

CAPSTONE PROJECT

Coursera IBM Data Science Professional Certificate

By Emma K

May 2019

BUSINESS PROBLEM

- Make a decision whether New York City or Downtown Toronto have more art related venues near the New York University (NYU) and University of Toronto (UofT) campus sites.

- The analysis should also provide a recommendation on neighbourhoods based on the same criteria to be considered by the subject for residence.

Image source: Wikipedia

DATA DESCRIPTION

- FourSquare explorer API to search for Art related venues around NYU and UofT in their respective cities of New York City and Toronto.
- Publicly available dataset that contains neighbourhoods data of Manhattan (and New York City) at https://cocl.us/new_york_dataset

APPROACH

- Convert addresses into their equivalent latitude and longitude values.
- Use the Foursquare API to explore neighbourhoods in New York City and Toronto
- Use the Foursquare API explore function to get the most common venue categories in each city
- Group the neighbourhoods into clusters using k-means clustering algorithm

DATA PREPARATION

- Nominatim module to convert an address into latitude and longitude values
- Foursquare API requests with Art related venues in the vicinity of each university (within a radius of 1000m).
- The resulting json response was analyzed based on the ‘venues’ component, then normalized into a dataframe.
- Venue categories selected:

  – Performing Art Venues

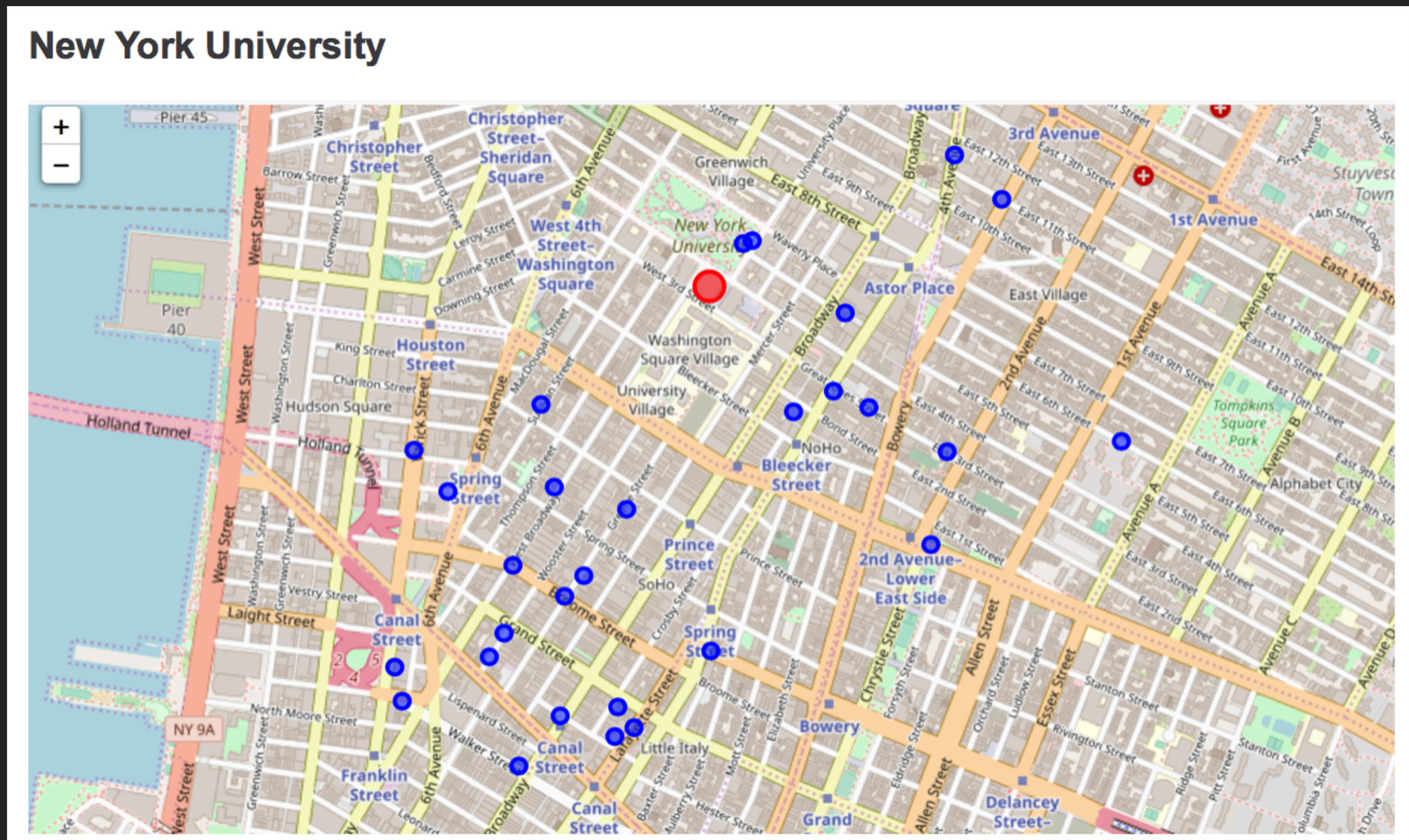
  – Art Galler

  – Arts & Crafts Store

Image source: Wikipedia

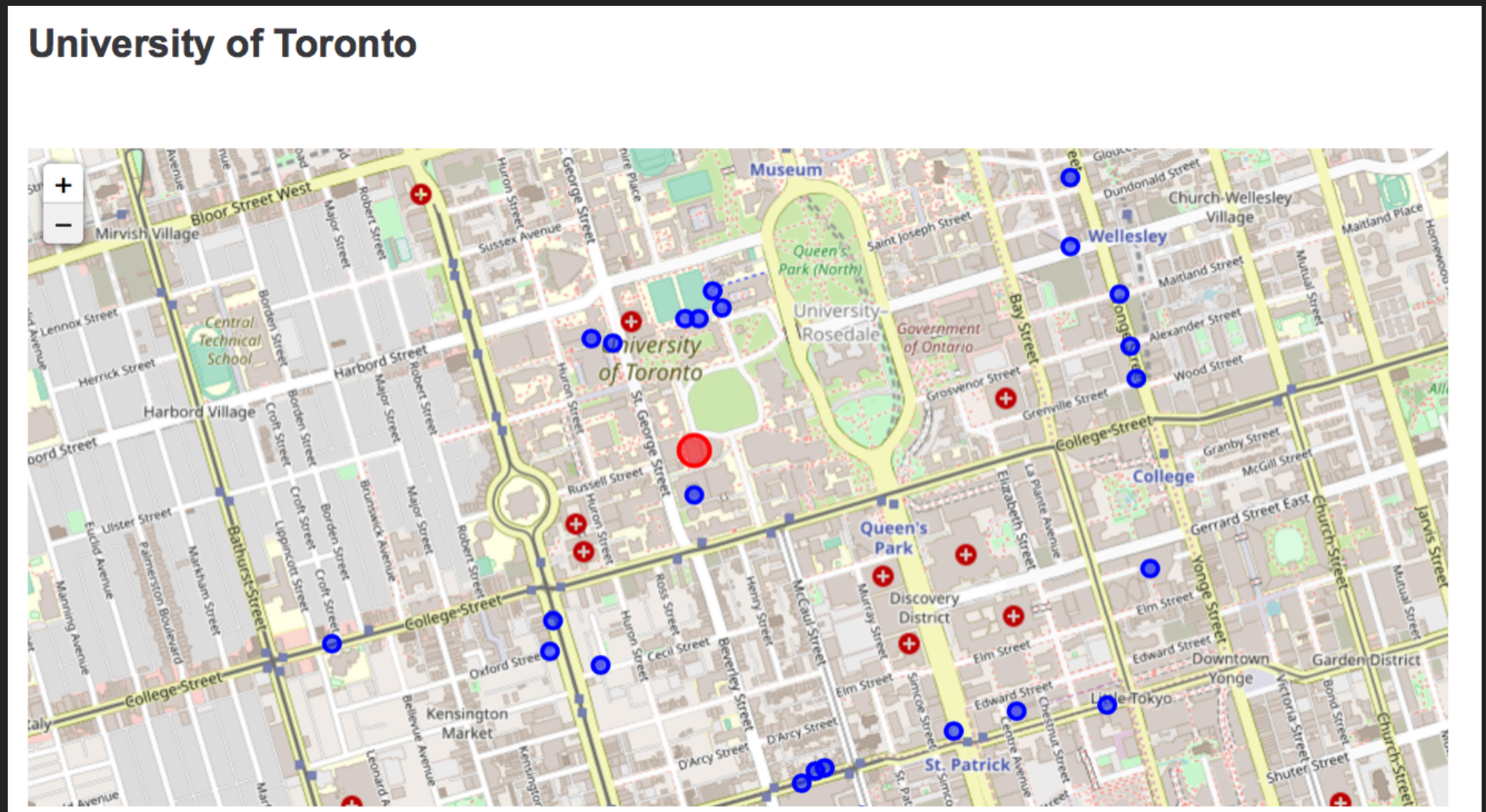
EXPLORATORY ANALYSIS

- Red marker indicates the campus, and the blue markers are the art venues.



EXPLORATORY ANALYSIS

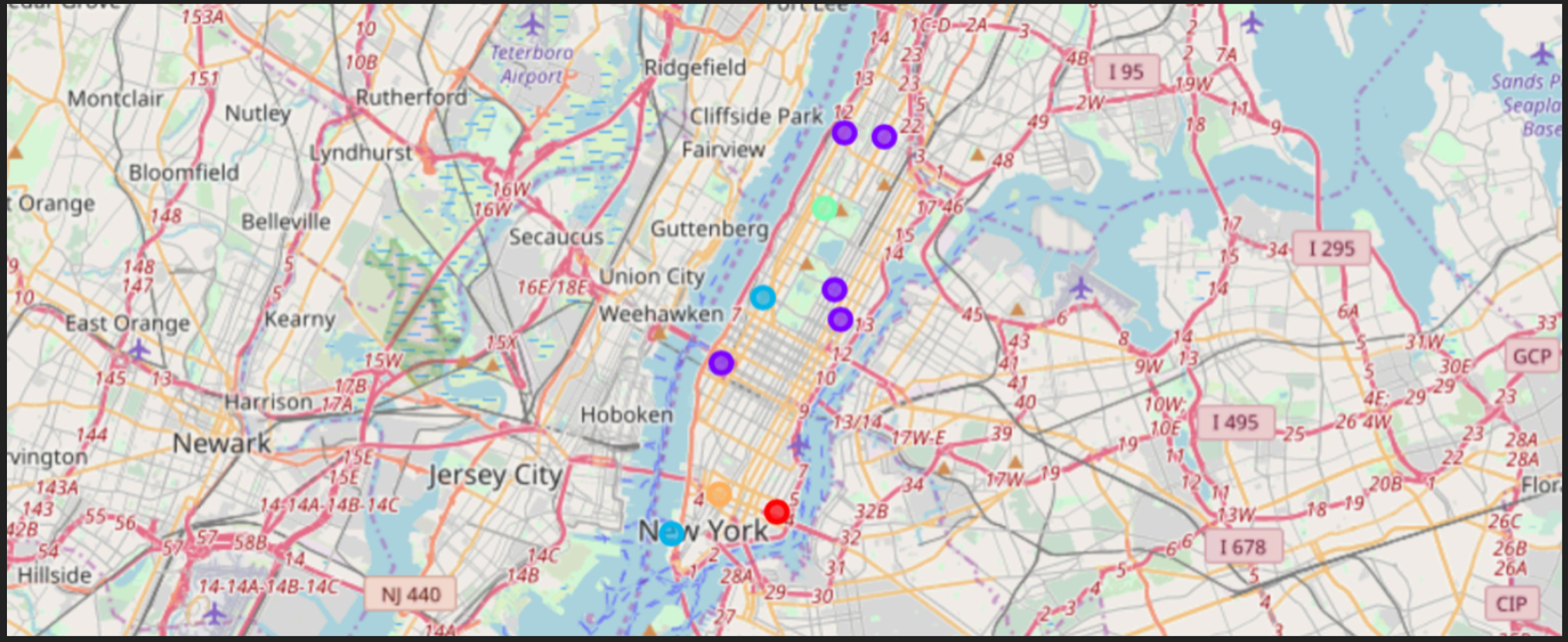
- Red marker indicates the campus, and the blue markers are the art venues.



MACHINE LEARNING

- Clustering is a technique for finding similar grouping in a data, also referred to as clusters. It attempts to group by similarity, but not driven by a known goal/purpose. Clustering is an unsupervised learning, as you don't have prescribed labels in the data.

- The algorithm was run for 5 clusters, here's a visualization of all on the map where each colour represents a cluster.



RESULTS

- There is no shortage of art related venues, galleries and art supply stores around both campus sites.
- NYU has a total of 18 and UofT has 16, very comparable.
- UofT campus has 6 art venues within 500 m radius compared to NYU with only 4.
- 4 out of 6 venues are within the UofT campus and only 1 out of 5 was in NYU campus.
- This insight indicates that our choice towards NYU is the better one for the client.
- The only cluster that had all three types of art venues: Performing Art Venue, Art Gallery and Arts supply store was Cluster #2 with 5 different neighbourhood choices for our client to choose from.

CONCLUSION

- This analysis depended on the accuracy of the data returned by Foursquare API.

- The client's request for having art related venues nearby was assumed to be venues of categories belonging to either a Performing Arts, Art Gallery and Arts & Crafts stores. The results could have been different if other venue categories were considered.

- A similar analysis on the Toronto neighbourhoods could also be done to improve the level of confidence in the recommendation.

- The analysis could be improved by adding data on demographics and rent prices for the resulting neighbourhoods.