

Capstone Project
IBM Data Science Professional Certificate

Project Presentation
The Battle of Manhattan and Downtown Toronto Neighborhoods
by Jelena Rota

Basel, August 2020

Table of Content

1. Introduction
2. Data
3. Methodology
4. Results
5. Discussion
6. Conclusion

1. Introduction

- New York City and Toronto are financial capitals of their respective countries and remarkably diverse north American cities.
- They obviously have their similarities, but do they have similar neighborhoods as well?
- This is what I would like to answer by comparing and grouping into clusters the neighborhoods of Manhattan and Downtown Toronto, central boroughs of New York City and Toronto.

2. Data

- Data set that needs to be prepared has 59 rows and 4 columns (borough name, neighborhood name, latitude, and longitude), and represents 40 Manhattan and 19 Downtown Toronto neighborhoods.
- Sources that were used for data set preparation are New York University, Wikipedia, and Cognitive Class websites.

3. Methodology

Methodology section is divided into following paragraphs presenting the project steps:

- A) Importing libraries
- B) Downloading and transforming New York City data
- C) Preparing Manhattan data
- D) Downloading and transforming Toronto data
- E) Preparing Downtown Toronto data
- F) Merging Manhattan and Downtown Toronto data
- G) Exploring neighborhoods in Manhattan and Downtown Toronto with Foursquare
- H) Clustering Manhattan and Downtown Toronto neighborhoods with k-means
- I) Presenting clusters

A) Importing libraries

In the first step of the project necessary libraries, packages and services were imported.

B) Downloading and transforming New York City data

New York City data were downloaded from Cognitive Class website (JSON file) and then transformed by looping through and filling in the data frame one row at a time with information on borough, neighborhood, latitude, and longitude.

C) Preparing Manhattan data

Manhattan data set was prepared by selecting Manhattan as “Borough”. It has 40 different neighborhoods. Here are the top 5 rows:

	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

D) Downloading and transforming Toronto data

Toronto data were taken from Wikipedia page (List of postal codes of Canada) and Cognitive Class website (Geospatial coordinates for Toronto), and present combination of both.

E) Preparing Downtown Toronto data

Downtown Toronto data set was prepared by selecting Downtown Toronto as "Borough". It has 19 different neighborhoods. Here are the top 5 rows:

	Borough	Neighborhood	Latitude	Longitude
0	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
2	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	Downtown Toronto	St. James Town	43.651494	-79.375418
4	Downtown Toronto	Berczy Park	43.644771	-79.373306

F) Merging Manhattan and Downtown Toronto data

Manhattan and Downtown Toronto data sets were merged with the append function. Total number of neighborhoods is 59 (40 Manhattan and 19 Downtown Toronto neighborhoods).

G) Exploring neighborhoods in Manhattan and Downtown Toronto with Foursquare

- Neighborhoods in Manhattan and Downtown Toronto were explored with a function that gets nearby venues using the Foursquare API.
- Each neighborhood was then analyzed using one hot encoding, rows were then grouped by neighborhood and by taking the mean of the frequency of occurrence of each venue category. This data set was used for k-means clustering. Here are the top 5 rows:

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquariu
0	Battery Park City	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.013699	0.0	0.0
1	Berczy Park	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.0
2	CN Tower, King and Spadina, Railway Lands, Har...	0.0	0.0	0.0	0.0	0.058824	0.058824	0.058824	0.117647	0.176471	0.117647	0.000000	0.0	0.0
3	Carnegie Hill	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.011628	0.0	0.0
4	Central Bay Street	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.0

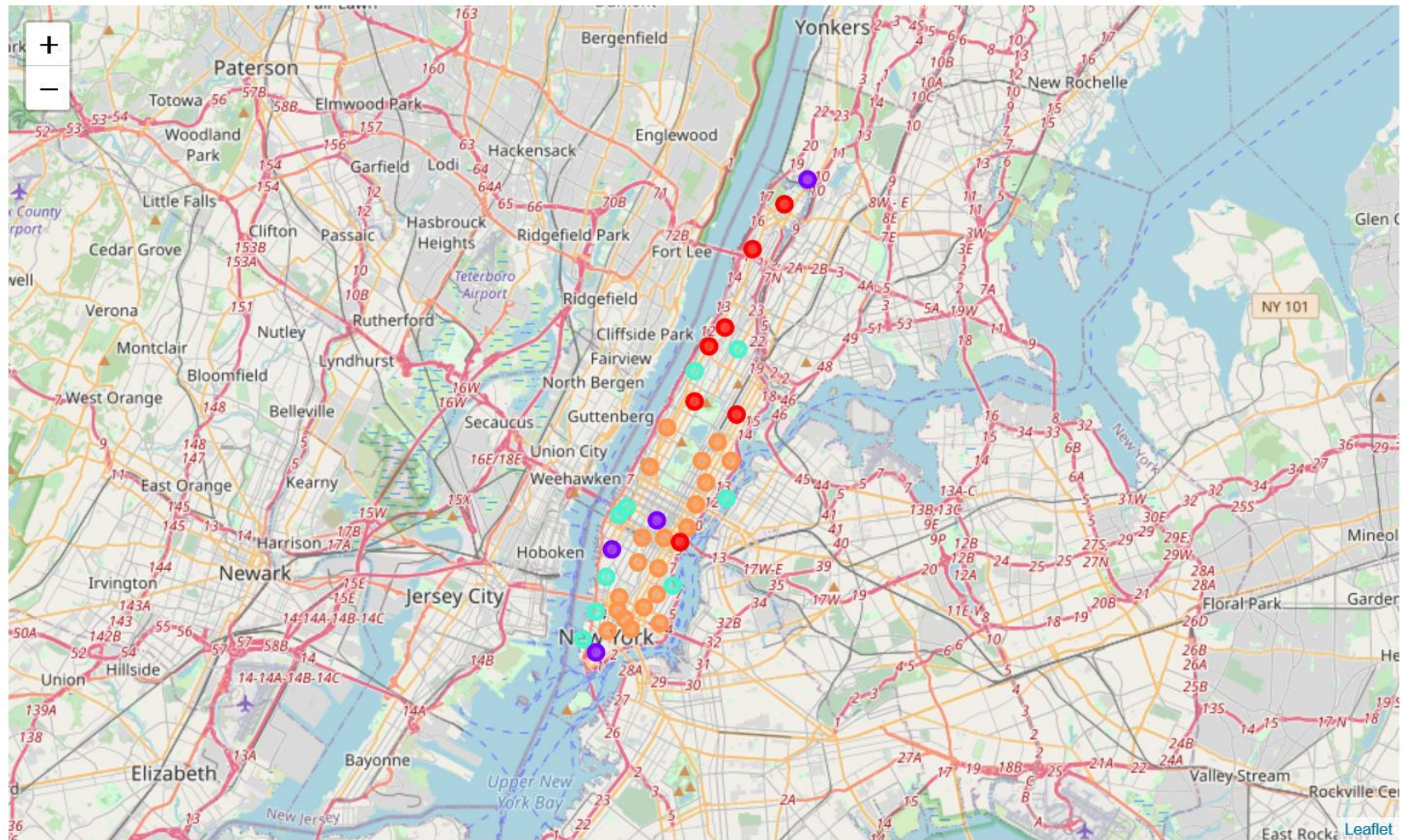
H) Clustering Manhattan and Downtown Toronto Neighborhoods with k-means

- Data set with all 59 neighborhoods and the frequency of occurrence of each venue category was analyzed using k-means clustering algorithm.
- k-means clustering is a method of vector quantization, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster.
- The number of clusters (k) was set to 6. The resulting clusters were presented in maps and tables (see chapter 4).

I) Presenting clusters

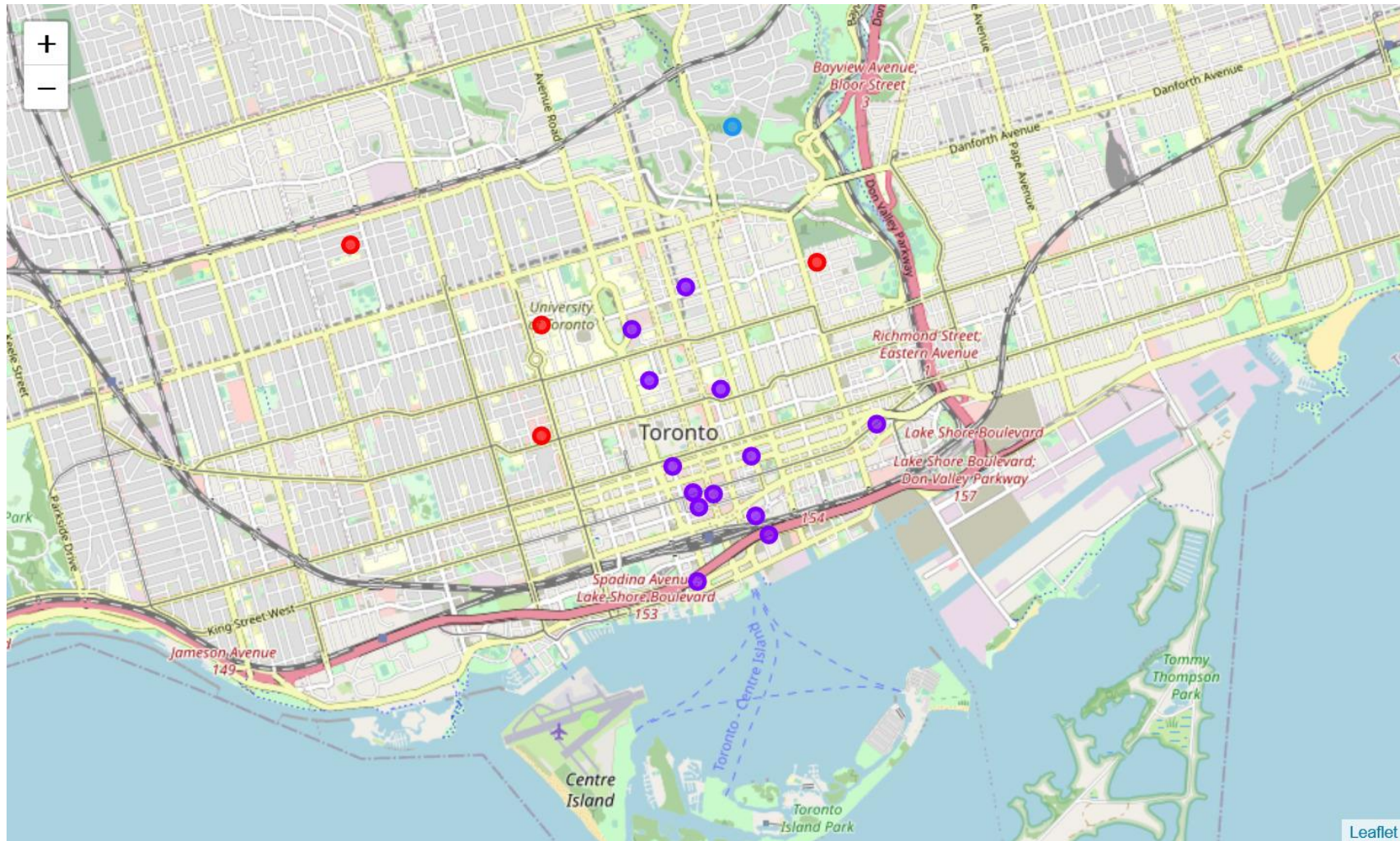
- As mentioned earlier, the clusters were presented in tables and maps in Manhattan and Downtown Toronto area using Folium (see chapter 4).
- Folium is a powerful library which visualizes data that has been manipulated in Python on an interactive leaflet map. It enables both the binding of data to a map for choropleth visualizations as well as passing rich vector/raster/HTML visualizations as markers on the map.

4. Results



Map of clusters in Manhattan

4. Results



Map of clusters in Downtown Toronto

5. Discussion

- **Cluster 1** (red color) presents a mix of Manhattan and Downtown Toronto neighborhoods, which means that Manhattan and Downtown Toronto have similar neighborhoods in respect to their top venues.
- These are seven Manhattan neighborhoods (Washington Heights, Inwood, Hamilton Heights, Manhattanville, East Harlem, Manhattan Valley and Tudor City) and four Downtown Toronto neighborhoods (Christie, University of Toronto, Kensington Market and St. James Town/Cabbage Town).
- **Cluster 2** (violet color) also presents a mix of Manhattan and Downtown Toronto neighborhoods.
- These are seventeen similar neighborhoods in total, four Manhattan neighborhoods (Marble Hill, Midtown, Chelsea, and Financial District) and thirteen Downtown Toronto neighborhoods (Reagent Park, Queen's Park, Garden District, St. James Town, Berczy Park, Central Bay Street, Richmond, Harbourfront East, Toronto Dominion Centre, Commerce Court, Stn A PO Boxes, First Canadian Place, Church and Wellesley).

6. Conclusion

Clusters 1 and 2, that are result of k-means clustering algorithm, show that Manhattan and Downtown Toronto have similar neighborhoods.

Therefore, the answer to the initial question whether Manhattan and Downtown Toronto have similar neighborhoods is positive, when one considers top venue categories, extracted by Foursquare API in each of 59 neighborhoods.