

Capstone Project - The Battle of the Neighborhood

Applied Data Science Capstone by IBM/Coursera

Xiaoling Zhuang

Apr 21, 2021

1. Introduction

With the development of the global economy, more and more people change their places of residence due to the needs of study and work. In this project, we will take one of our clients as an example to discuss how to apply data analysis to find a satisfactory residence.

In this case, suppose our client currently lives in the downtown area of Toronto, Ontario, Canada. He recently accepted a job offer in Manhattan. When he is satisfied with his current neighborhood, he is looking for similar neighborhoods in Manhattan, where the venues and living environment are similar. We will use data science techniques to help clients identify the most identical neighborhoods in Manhattan to his current residential area.

This project is not only suitable for our customers, but also for customers/readers who are interested in finding similar neighborhoods around the world.

2. Data Description

2.1 Data Sources

The data will be collected from two areas. First, the Venues data of the targeted neighborhoods are from Foursquare. We will collect the data using Foursquare API. Second, the geographic location data of the Manhattan dataset can be downloaded from https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json.

2.2 Data Pre-processing

The data preparation will be conducted into 3 parts:

2.2.1: Collecting the Venues data of the Client 's Current Neighborhood in Toronto

We will first utilize the Foursquare API to collect the venues data of downtown Toronto. Then, the data will be pre-processing and convert into a clean dataframe that contains frequency of venues in the Downtown area of Toronto.

2.2.2: Explore the Neighborhoods in Manhattan, New York City

We will explore Manhattan Neighborhoods by: Collecting Manhattan venues information using Foursquare API; Applying Folium to map out the neighborhoods in Manhattan; Cleanup the data frame which matches the downtown Toronto data frame.

2.2.3: Combine the data frames of Toronto and NYC

At the end, the data frames of Downtown Toronto and neighborhoods of Manhattan will be combined into one data frames. Then the data set is ready for analysis.

3. Methodology

3.1 Explore the Client 's Current Neighborhood in Toronto

There are 67 different venue categories in the downtown Toronto neighborhood. The data visualization tool is used here to demo the downtown Toronto venues frequency. The histogram bellowed highlights the 6 most popular venues in this neighborhood. The most popular venue category is Coffee Shop (n=10). Combining with the Café venue number (n=4), the total venues number in downtown Toronto regarding the coffee type of venue is 14 out of 100. The second and fourth popular venue categories are Gastropub (n=4) and Pizza Place (n=3), respectively.

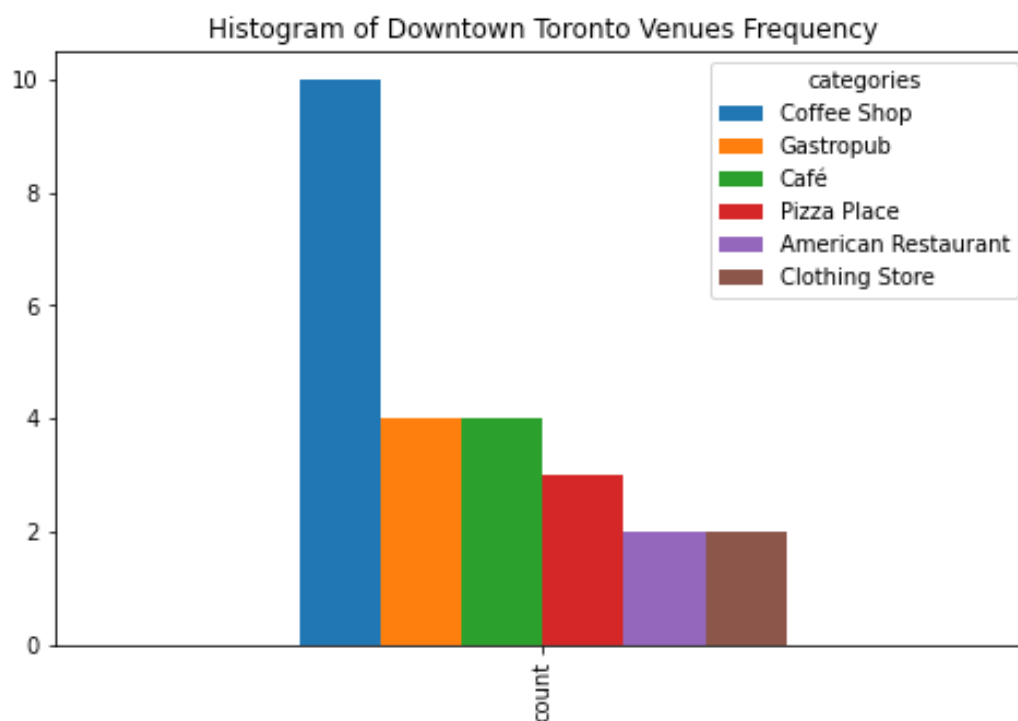


Table 1

3.2 Explore the Manhattan Neighborhood in NYC

In the Manhattan dataset, there are 40 neighborhoods. I created a map of Manhattan with neighborhoods superimposed on top using folium.

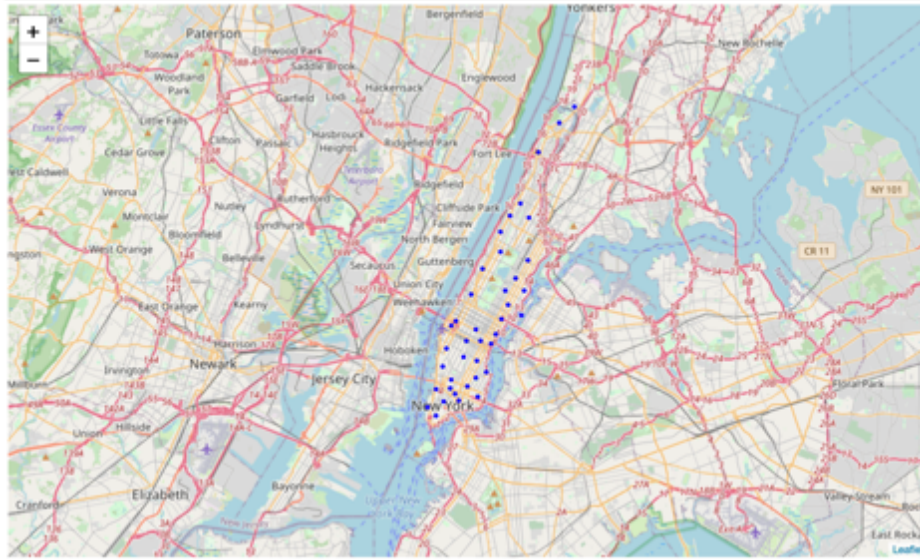


Table 2. Manhattan Neighborhood Distribution

Through Foursquare API, we collected the venues data of each Manhattan neighborhood and then convert the frequency to the orders of top venues. As shown in Table 3, each neighborhood has a different level of venue frequency. This would become the best data support for our cluster analysis.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Park	Coffee Shop	Hotel	Gym	Gourmet Shop	Shopping Mall	Sandwich Place	Plaza	Pizza Place	Cosmetics Shop
1	Carnegie Hill	Coffee Shop	Café	French Restaurant	Gym	Bookstore	Pizza Place	Grocery Store	Bakery	Cosmetics Shop	Italian Restaurant
2	Central Harlem	American Restaurant	Art Gallery	French Restaurant	Restaurant	Dessert Shop	Park	Pizza Place	Gym	Cosmetics Shop	Grocery Store
3	Chelsea	Coffee Shop	Bakery	American Restaurant	Art Gallery	French Restaurant	Italian Restaurant	Park	Thai Restaurant	Bookstore	Café
4	Chinatown	Bakery	American Restaurant	Dessert Shop	Sandwich Place	Bubble Tea Shop	Coffee Shop	Vietnamese Restaurant	Spanish Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant

Table 3. Top Venues in Manhattan Neighborhood

3.3 Cluster Analysis

Finally, we combined the data frames of Downtown Toronto and Neighborhoods of Manhattan and then apply cluster analysis. In this approach, we are able to find out the cluster where the

Downtown Toronto is and find out the most similar neighborhoods in Manhattan. We imported the KMeans functions from Sklearn and then added the cluster labels to the combined data frame. As shown in table 4, there are totally 6 clusters which are represented in different colors.

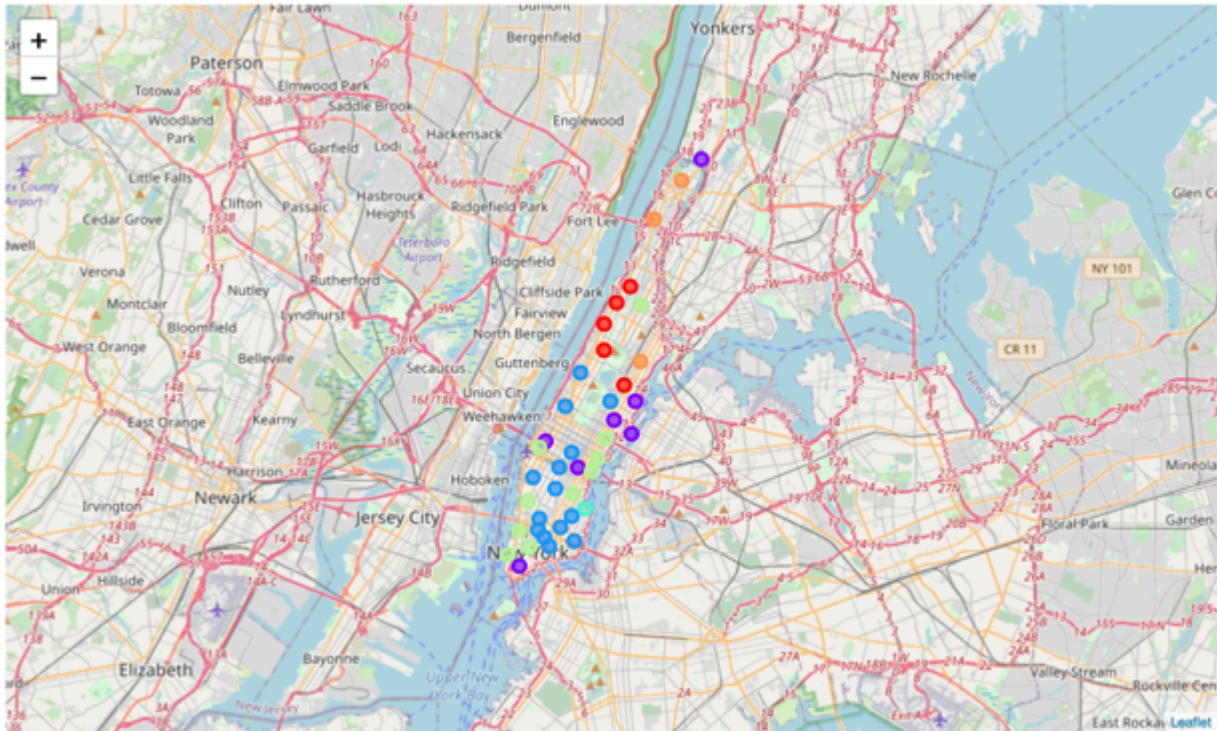


Table 4. Manhattan Neighborhoods Clusters

4. Results

The results show that Downtown Toronto is classified as cluster 0. There are five neighborhoods in Manhattan that are similar to the Downtown Toronto: Carnegie Hill, Hamilton Heights, Manhattan Valley, Manhattanville, and Morningside Heights. Similar to downtown Toronto, these neighborhoods have a higher frequency of Café/coffee venues (see table 5). There are also have park and Farmers Market nearby. Our client can focus on these three neighborhoods and pick out the best fit one based on his own preferences.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Cluster
1	Carnegie Hill	Coffee Shop	Café	French Restaurant	Bookstore	Pizza Place	Gym	Grocery Store	Bakery	Cosmetics Shop	Spa	0
13	Hamilton Heights	Pizza Place	Coffee Shop	Café	Bakery	Liquor Store	Park	Sandwich Place	Historic Site	Food Truck	Hotel	0
20	Manhattan Valley	Coffee Shop	Thai Restaurant	Pizza Place	Vietnamese Restaurant	Hostel	Furniture / Home Store	French Restaurant	Farmers Market	Park	Grocery Store	0
21	Manhattanville	Coffee Shop	American Restaurant	Café	Gastropub	Liquor Store	Food Truck	Diner	Park	Ramen Restaurant	Bakery	0
25	Morningside Heights	Coffee Shop	American Restaurant	Park	Bookstore	Café	Sandwich Place	Food Truck	Farmers Market	Mediterranean Restaurant	Pizza Place	0
40	Downtown Toronto	Coffee Shop	Café	Park	Farmers Market	Thai Restaurant	Beer Bar	Sandwich Place	Gastropub	Japanese Restaurant	Plaza	0

Table 5. Similar Neighborhoods Cluster Detail

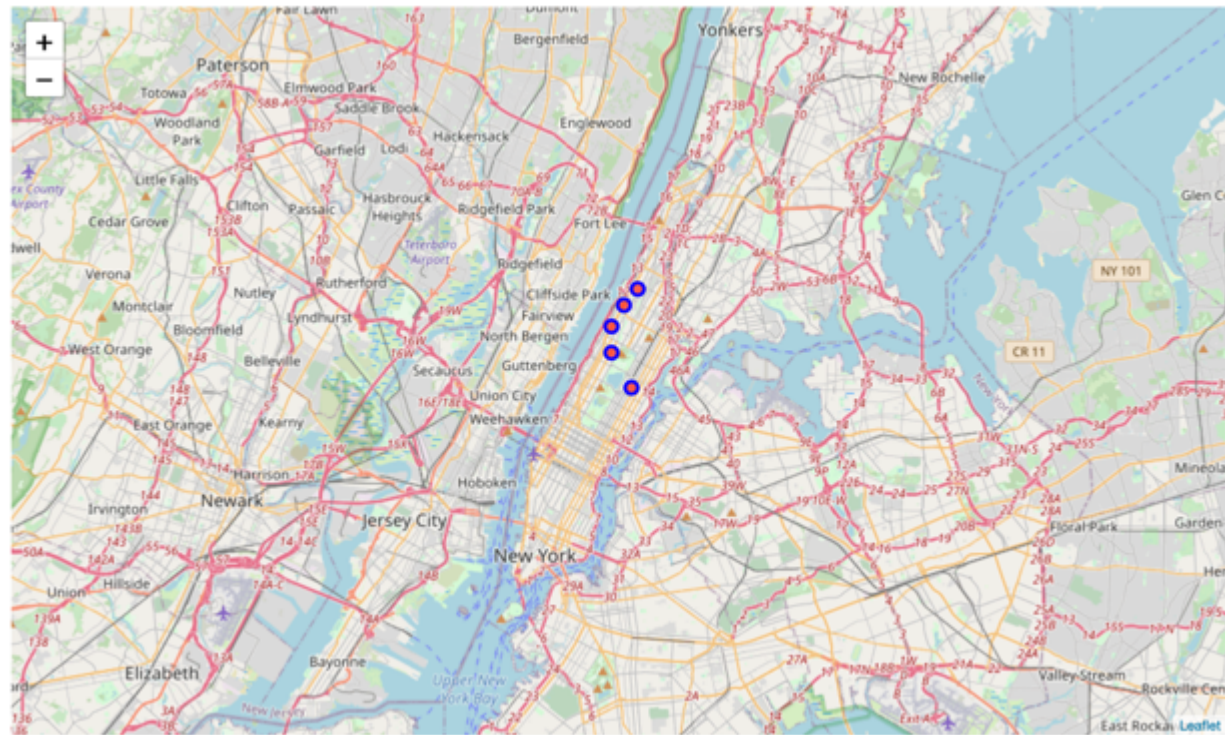


Table 5. Similar Neighborhoods Cluster Distribution

5. Discussion

In my opinion, regarding the venue's distribution, the Morningside Heights neighborhood might be the most similar one to the Downtown Toronto neighborhood. It's not only because it has the most distribution of Coffee shop and Café, but also, it's the only neighborhood that has the Farmers Market, which venues is the fourth popular venues in Downtown Toronto.

6. Conclusion

In real life, the client might have to consider other factors, such as transportation and culture, to decide his new residence. This project could narrow down his choice of neighborhood and help him make a better decision.