

IBM Coursera Data Science Capstone:

Opening A Coffee Shop in Manhattan

August 5, 2020

A. Rains

Introduction

Often called, the city “that never sleeps” New York City is a city known around the world for food, fashion, and culture. Manhattan, a borough, is known as one of the most expensive real estate areas in the world based and an epicenter for tourists and residents alike. For a city that never sleeps, there must be major opportunities to start a 24/7 coffee shop within Manhattan.

The goal of this project will be to help identify where to open our new business venture: '24/7 Caffeine Queen.' Our target audience, or the interested party, is the group of business investors who have grouped together to open this new venture.

Data

Ideally, we would like to open our shop up where there are not too many coffee shops nearby to limit competition, but also where there may be several bookstores based on previous reports that foot traffic to bookstores drives customers to purchase beverages nearby. ([Article 1](#), [Article 2](#))

To make a recommendation to our business investors, we will use Foursquare API data on venues within Manhattan neighborhoods to determine which have bookstores and/or coffee shops. This information comes from Foursquare's [categories of venues](#).

From their Developer resources, there are three categories of bookstores listed:

1. Bookstores
2. College Bookstores
3. Used Bookstores

There are also four categories including coffee:

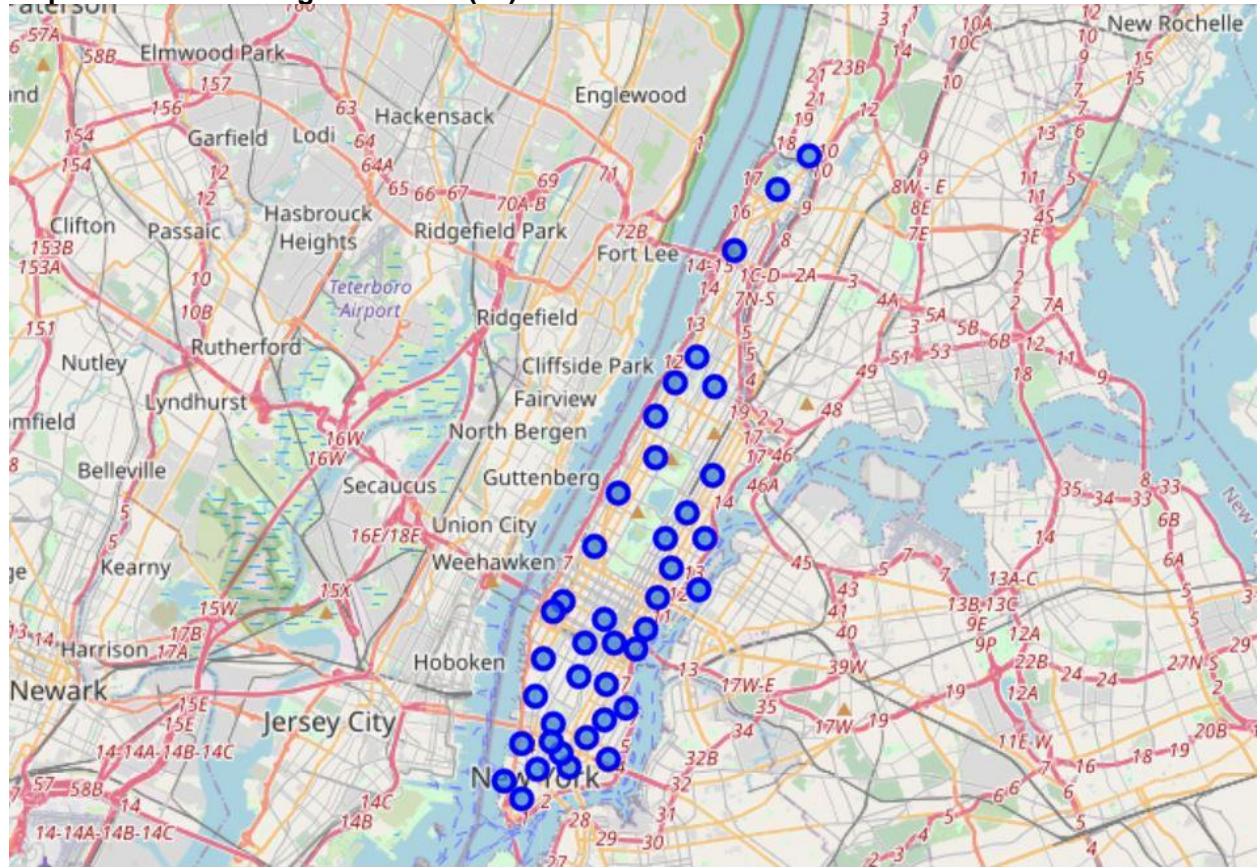
1. Coffee Shops
2. Coffee Roasters
3. Corporate Coffee Shops
4. Turkish Coffeehouse

This analysis will only use the *Coffee Shop* category for our segmenting analysis combined with *Bookstores* and *Used Bookstores* to explore each neighborhood. This information will also be populated on a visual map with the clustering technique *k-means*.

Below is an example map created depicting the neighborhoods within Manhattan. We will use similar clustering technique on this visual map with Folium to depict the highest concentration of bookstores and coffee shops to aid in our decision. With this information, our business investors, our target

audience, can prioritize their real estate searches and determine the top two neighborhoods they should consider for opening the coffee shop venture.

Map 1: Manhattan Neighborhoods (39)



Methodology

Exploratory Data

First, we imported geodata to create a visual map of Manhattan using Folium within Jupyter notebook utilizing Python language to depict the neighborhoods (visual shown above). Three python dataframes were created with these inputs: Manhattan bookstores, Manhattan coffee shops, and Manhattan neighborhoods. These dataframes were then analyzed to make our recommendation through performing machine learning analysis.

Inferential Statistics

Then, geodata was pulled for the top 10 venues in each neighborhood as well as total number of coffee shops and bookstores within each neighborhood for exploration purposes. By compiling this subset of information, it allows for neighborhoods to be grouped together based on common venue types, which can then be fed into a machine learning model for further examination.

Machine Learning

For the analysis, k-Means modeling was utilized to segment the neighborhoods based on similarity. This approach of machine learning was used to cluster similar neighborhoods together to help refine the overall search for location for the coffee shop venture. Additionally, data normalization was used on the coffee shops & bookstores dataframes to calculate a rating per each neighborhood.

Results

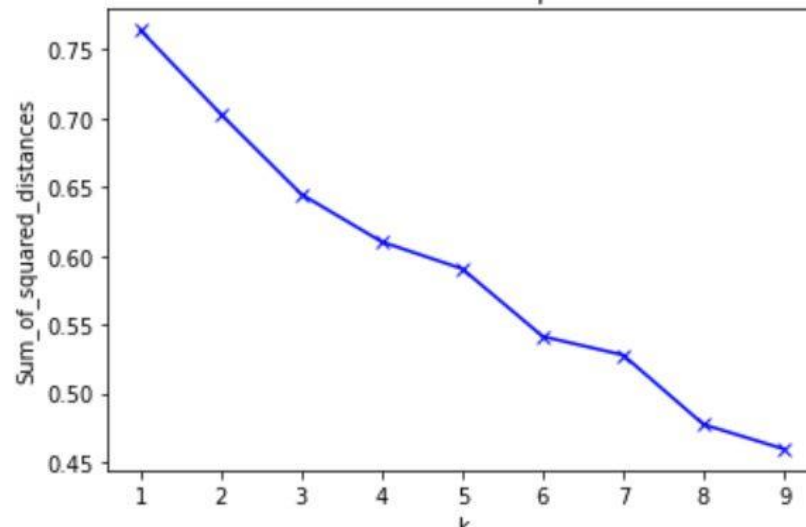
Geodata from Foursquare API yielded the following information about the number of coffee shops and bookstores within all 39 Manhattan neighborhoods:

Table 1: Total Number of Coffee Shops & Bookstores with Neighborhood Cluster

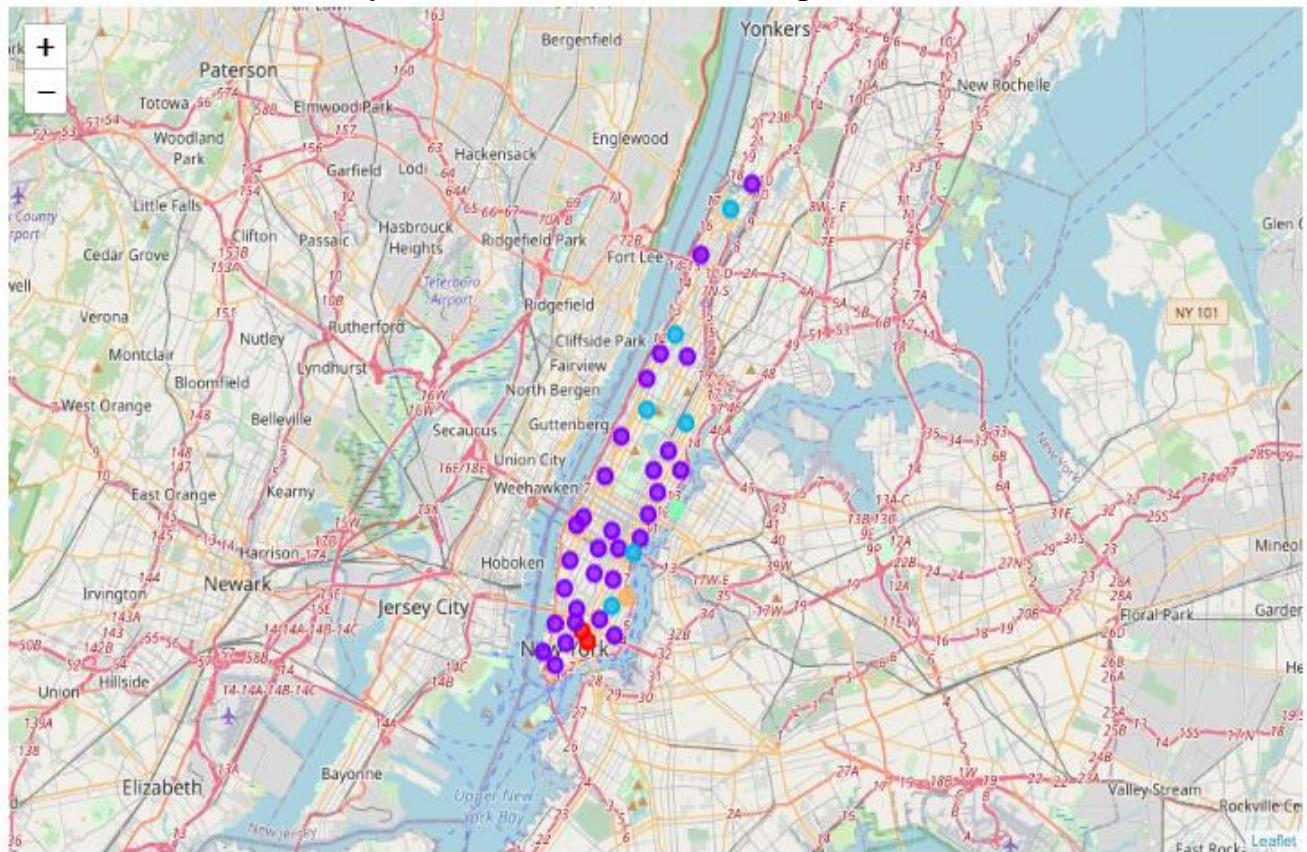
	Cluster Labels	Count of Bookstores	Count of Coffee Shops
Marble Hill	1	0.0	0.0
Chinatown	0	11.0	48.0
Washington Heights	1	0.0	0.0
Inwood	2	2.0	7.0
Hamilton Heights	2	0.0	0.0
Manhattanville	1	0.0	0.0
Central Harlem	1	2.0	6.0
East Harlem	2	0.0	0.0
Upper East Side	1	10.0	34.0
Yorkville	1	2.0	19.0
Lenox Hill	1	5.0	40.0
Roosevelt Island	3	1.0	3.0
Upper West Side	1	6.0	27.0
Lincoln Square	1	3.0	26.0
Clinton	1	4.0	44.0
Midtown	1	18.0	50.0
Murray Hill	1	18.0	50.0
Chelsea	1	6.0	48.0
Greenwich Village	1	21.0	48.0
East Village	2	13.0	48.0
Lower East Side	1	0.0	0.0
Tribeca	1	2.0	25.0
Little Italy	0	19.0	48.0
Soho	1	16.0	49.0
West Village	1	15.0	43.0
Manhattan Valley	2	1.0	15.0
Morningside Heights	1	4.0	27.0
Gramercy	1	2.0	31.0
Battery Park City	1	2.0	28.0
Financial District	1	3.0	50.0
Carnegie Hill	1	6.0	34.0
Noho	1	23.0	50.0
Civic Center	1	13.0	49.0
Midtown South	1	21.0	50.0
Sutton Place	1	3.0	36.0
Turtle Bay	1	9.0	48.0
Tudor City	2	9.0	43.0
Stuyvesant Town	4	0.0	0.0
Flatiron	1	17.0	50.0
Hudson Yards	1	0.0	0.0

From our k-means clustering, we determined that 5 clusters was the ideal number of clusters using the elbow method. The visual graph of that analysis as well as the 5 neighborhoods is shown below.

Plot of Elbow Method for Optimal k
Elbow Method For Optimal k



Map 2: 5 Clusters of Manhattan Neighborhoods



Here are the results of the clustering:

Table 2: Clustered Neighborhoods

Cluster Group Number	Neighborhoods
0	Little Italy, Chinatown
1	Marble Hill, Washington Heights, Central Harlem, Manhattanville, Upper East Side, Yorkville, Lenox Hill, Upper West Side, Lincoln Square, Clinton, Midtown, Murray Hill, Chelsea, Greenwich Village, Tribeca, Lower East Side, Soho, West Village, Gramercy, Morningside Heights, Battery Park City, Financial District, Noho, Carnegie Hill, Civic Center, Midtown South, Sutton Place, Turtle Bay, Flatiron, Hudson Yards
2	Inwood, Hamilton Heights, East Harlem, East Village, Manhattan Valley, Tudor City
3	Roosevelt Island
4	Stuyvesant Town

Once the neighborhoods were clustered from our k-means exercise, the sum data was normalized to calculate a rating for each neighborhood. This then allowed us to narrow down the neighborhood recommendations.

Table 3: Normalized Data, Neighborhoods Rated

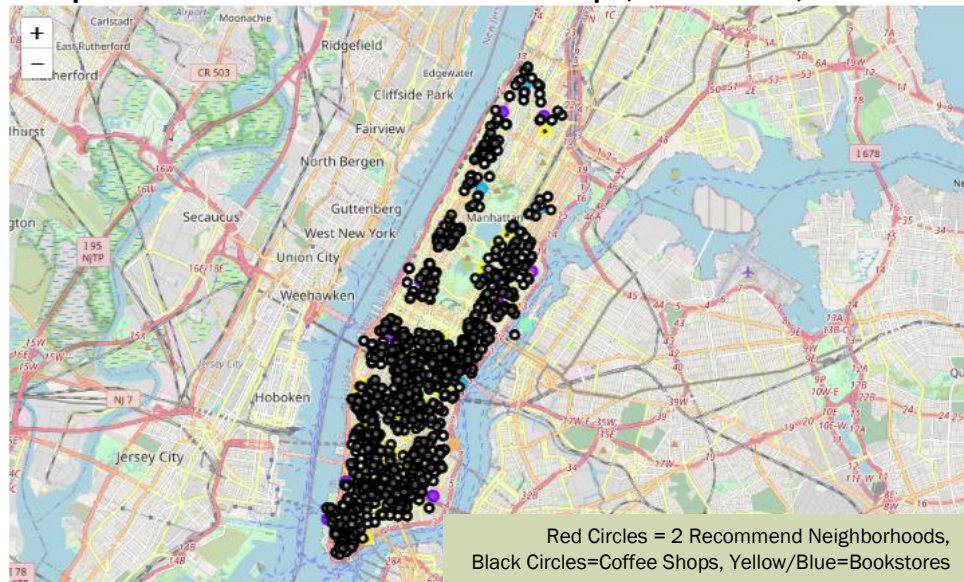
Cluster Labels	Count of Bookstores	Count of Coffee Shops	Count of Neighborhoods	Normalized Bookstores	Normalized Coffee Shops	Rating	Percentage B to C
0	30.0	96.0	2	0.129870	0.099792	0.012960	31.250000
1	231.0	962.0	30	1.000000	1.000000	1.000000	24.012474
2	25.0	113.0	6	0.108225	0.117464	0.012713	22.123894
3	1.0	3.0	1	0.004329	0.003119	0.000014	33.333333
4	0.0	0.0	1	0.000000	0.000000	0.000000	NaN

From the normalization and review, it was determined that cluster group 0 is the most ideal location to consider opening up a coffee shop venture business. This is based on the rating of 0.012 (a multiple of bookstores and coffee shops). We also see that there is a higher percentage of bookstores to coffeeshops, 31.5%. The two neighborhoods within that cluster are: Little Italy and Chinatown.

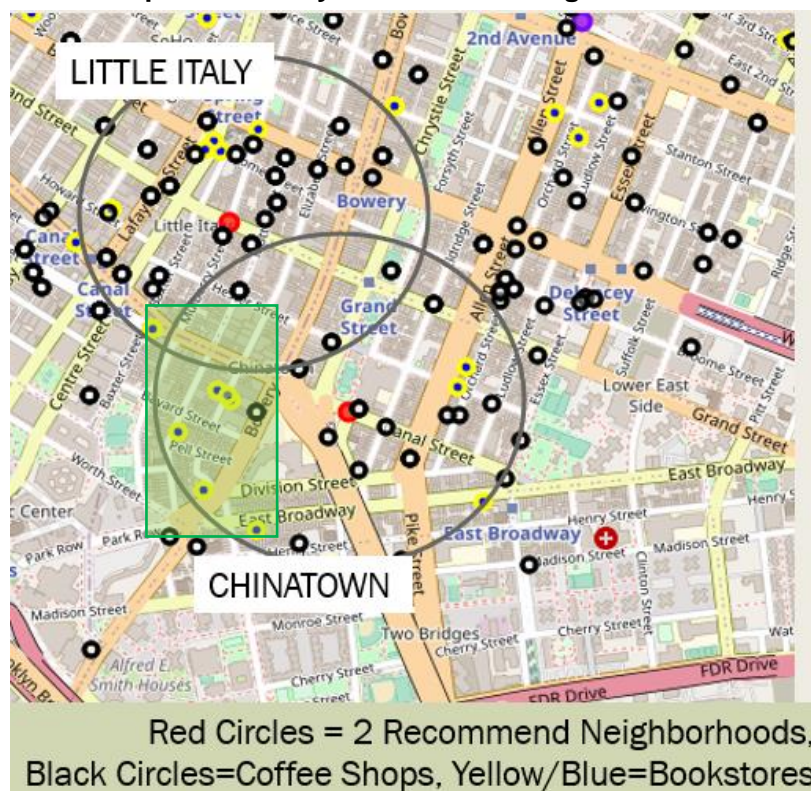
Discussion

From the analysis, it is quite clear that Manhattan alone has quite a few coffee shops! Upon further review, the best location and the final recommendation would be to open a coffee shop in either Little Italy or Chinatown neighborhoods. Map 3 below show Manhattan with all neighborhoods, coffee shops and bookstores. Map 4 shows the area of recommendation for opening the coffee shop. Future recommendations would be to open a Premium Foursquare API account to also view venue check-ins in these two specific neighborhoods to determine typical volumes of customers and the size of the competition. Additionally, investigating average commercial real estate rental prices would be ideal to determine which street as well as building out the business case.

Map 3: Manhattan Folium With Coffee Shops, Bookstores, & Clusters



Map 4: Little Italy & Chinatown Neighborhoods



Conclusion

In conclusion, it is recommended that a coffee shop should be opened in Little Italy or Chinatown. This is based off the number of bookstores to number of coffee shops and using k-means for clustering. The green shaded area in Map 4 shows quite a few bookstores with very little coffee shops and represents a large area of opportunity.