Predicting the best location for a second office

By: Kennedy Chinyamutangira

December 27, 2020

1. Introduction

1.1 Background

The company I work for is currently located in Times Square, New York. The firm is growing at a good rate. Given the expansion of our services and client base it may soon be time to seek additional office space. Our company employs professionals from the Tristate area (New York, New Jersey and Connecticut) and our clients are generally spread out across Manhattan and across the other four boroughs in New York.

1.2 Problem Statement

Determining the optimal location for expanding the company's operation in New York City requires careful consideration of a number of options. The company may choose to increase the space it leases in the current building, if this is possible, or may consider opening up a second office location in New York City. Opening up a second office and ensuring that this office provides the same amenities and facilities that the current office provides is important in order to ensure that staff in different locations are getting similar conditions and surrounding environments to achieve the same level of productivity. This requires considering the accessibility of the location to public transit, proximity to other venues that the staff frequent outside the office such as gyms, restaurants for breakfast and lunch meals or evening entertainment venues following a hard day's work in the office. In the search for a possible second location it would be ideal if the company could identify other locations that have a similar profile to the company's existing location.

Target Audience

This report is prepared for the use of the company's management to assist them in planning for future expansion. The report seeks to aid the decision making process of management in choosing a future office location as part of the expansion plan.

2. Data

2.1 New York Neighborhoods Data

For this dataset we will use the New York neighborhood data file used in the IBM Data Science course labs that contains the 5 boroughs in New York City and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. This available at the URL: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json and will be read as a JSON file into a pandas dataframe.

2.2 New York Public Transit Data

Most of the company's employees use the New York City subway for transportation to the office. The proximity of the office location to a number of subway lines will be ideal. Data on the location of subway station will be sourced from the New York Metropolitan Transportation Authority (MTA). This data is available on the MTA's website at http://web.mta.info/developers/developer-data-terms.html#data. A URL to a file containing subway station data is available at this location and this file is available in .csv format and will be read directly into a pandas dataframe using the **read_csv** function. The URL is as follows: https://atisdata.s3.amazonaws.com/Station/Stations.csv.

2.3 Foursquare data to explore venues around New York neighborhoods

The Foursquare API will be used to explore venues surrounding the New York neighborhoods in the Manhattan neighborhoods from the neighborhoods from the above dataset. The current office is located in Manhattan and a second location in the Manhattan area is considered appropriate as this borough is central enough and has more options for commercial real estate that meets the profile of the current office location and is also suitable office for a professional services firm.

3. Methodology

We will use the Manhattan neighborhood data and the Foursquare location data to explore the most common venues in each neighborhood. The data will be summarized by taking the mean of the frequency of occurrence of each category. To assess the most common categories this will be summarized to show the top ten venues for each location. An extract of this table is shown as Figure 1 below:

Figure 1

	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	Nearby Stations	Nearby Lines	Coffee Shop	Clothing Store	Park	Hotel	Memorial Site	Gym	Plaza	Playground
1	Carnegie Hill	Nearby Lines	Nearby Stations	Coffee Shop	Café	Pizza Place	Wine Shop	Bar	Bookstore	Gym	Gym / Fitness Center
2	Central Harlem	Nearby Lines	Nearby Stations	Art Gallery	Bar	Seafood Restaurant	French Restaurant	Fried Chicken Joint	Chinese Restaurant	Gym / Fitness Center	American Restaurant
3	Chelsea	Nearby Lines	Nearby Stations	Coffee Shop	Bakery	Art Gallery	Wine Shop	American Restaurant	Italian Restaurant	Ice Cream Shop	French Restaurant
4	Chinatown	Nearby Stations	Nearby Lines	Chinese Restaurant	Bakery	Cocktail Bar	American Restaurant	Vietnamese Restaurant	Optical Shop	Spa	Bubble Tea Shop
5	Civic Center	Nearby Stations	Nearby Lines	Coffee Shop	Cocktail Bar	Spa	Hotel	Gym / Fitness	French Restaurant	Park	American Restaurant

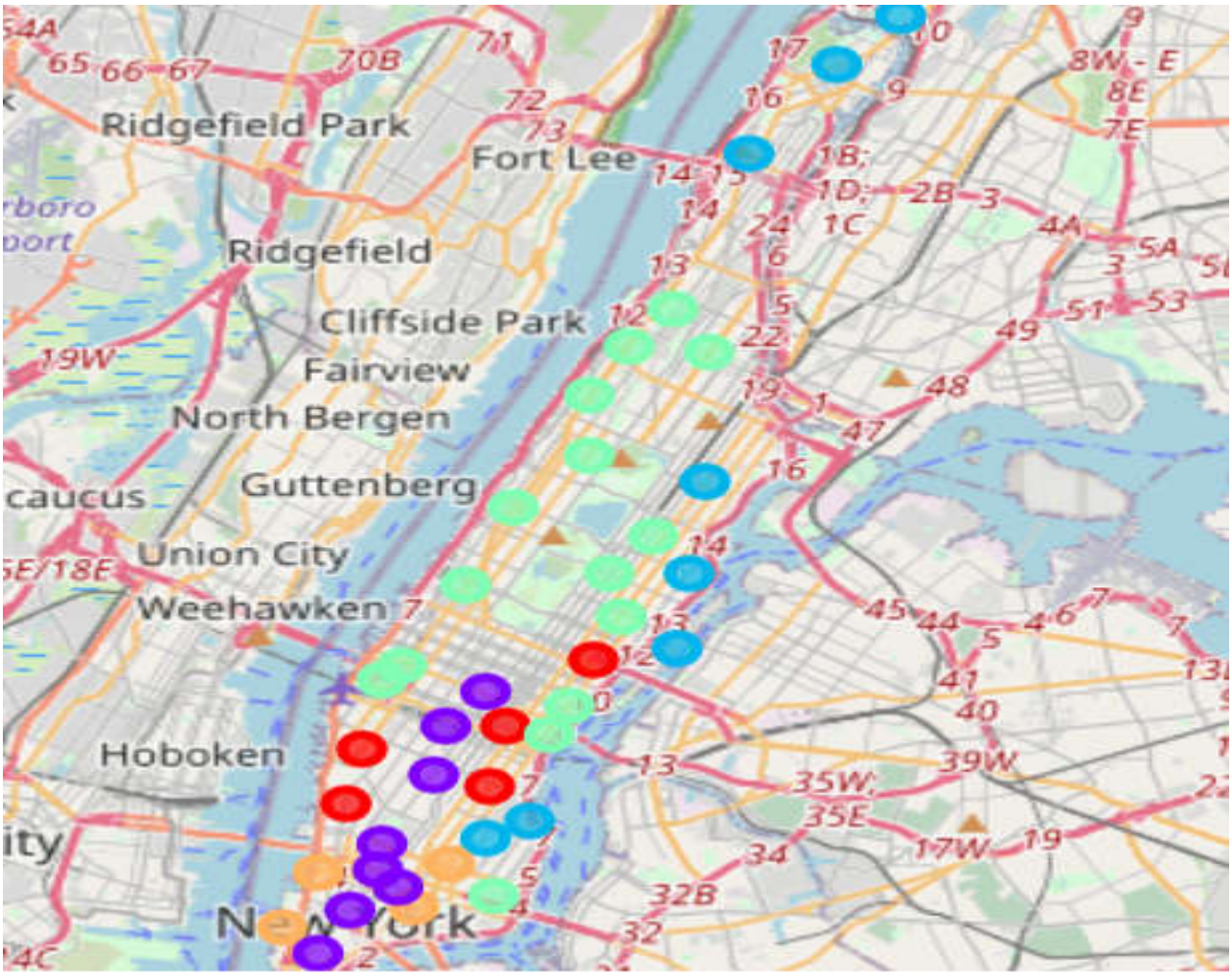
We will also use the New York City subway data to identify the nearby subway stations and subway lines that are within 1 mile of each neighborhood. 1 mile was considered a reasonable walking distance to get to the neighborhood after disembarking at a given subway station. The longitude and latitude of each station were converted to X/Y coordinates on a two-dimensional plane in order to calculate physical distance. See Figure 2 below for extract of the enriched data. The distance of the stations from each neighborhood will be used to determine the number of station and number of subway stations that are close to each neighborhood and this information will be appended to the neighborhood data and used as input into the clustering algorithm.

Figure 2

	Stop Name	Subway Lines	GTFS Latitude	GTFS Longitude	X	Y
0	Lexington Av/59 St	[N, W, R]	40.762660	-73.967258	1332.746172	-4637.833890
1	5 Av/59 St	[N, W, R]	40.764811	-73.973347	1332.210171	-4637.825398
2	57 St-7 Av	[N, Q, R, W]	40.764664	-73.980658	1331.621314	-4638.005611
3	49 St	[N, R, W]	40.759901	-73.984139	1331.434938	-4638.418884
4	Times Sq-42 St	[N, Q, R, W]	40.754672	-73.986754	1331.327953	-4638.844513

The information of each neighborhoods venues and accessibility via the subway was used to cluster the neighborhoods into 5 clusters using k-means clustering. This resulted in the following clustering map:

Figure 3



The current office is in the Midtown neighborhood of Manhattan. We will observe which cluster this neighborhood falls into and will review other neighborhoods in the same cluster and examine these for similarities with the current office location.

A recommendation will be made on the ideal office location for a possible future second office out of the neighborhoods in the same cluster as the current office based on the neighborhoods that most closely replicates the most important factors that the employees value about the existing office.

4. Results

4.1 Results Summary: Observations

Cluster 1: The following similarities were observed for neighborhoods in this cluster:

- The neighborhoods are generally residential neighborhoods in the surrounding Midtown and northern sections of Downtown Manhattan
- Coffee shops and Italian restaurants are very prevalent in this cluster
- Neighborhoods have an average of 9 lines accessible through 6 to 8 nearby stations

Cluster 2: The following similarities were observed for neighborhoods in this cluster:

- The neighborhoods are located close to the center of the Manhattan island when considered from a west to east position
- Their central location means that several subway lines traverse through these neighborhood making them the most accessible from a public transit perspective:
    - Neighborhoods have an average of 15 lines accessible through 16 to 21 nearby stations
- The neighborhoods have a wide variety of most popular venues which makes them suitable for catering for a wide range of needs

Cluster 3: The following similarities were observed for neighborhoods in this cluster:

- The neighborhoods are generally residential neighborhoods on the eastern side of the Manhattan
- The neighborhoods have poor access to transportation with fewer stations and subway lines servicing these locations:
    - Neighborhoods have an average of 1 line accessible through 1 to 5 nearby stations
- Restaurants are less prevalent in this cluster relative to other clusters and are instead replaced by supermarkets, deli/bodegas, markets and other residential neighborhood businesses

Cluster 4: The following similarities were observed for neighborhoods in this cluster:

- The neighborhoods are generally residential neighborhoods surrounding Central Park
- The neighborhoods have limited access to transportation although it is not as poor as cluster 3:
    - Neighborhoods have an average of 4 line accessible through 2 to 6 nearby stations
- This is the cluster with the most neighborhoods

Cluster 5: The following similarities were observed for neighborhoods in this cluster:

- The neighborhoods are generally residential neighborhoods in Downtown Manhattan
- The neighborhoods have the second best access to subway transportation: neighborhoods have an average of 10 line accessible through 11 to 14 nearby stations
- This is the cluster with the least neighborhoods

5. Discussion

Our approach was to focus on the cluster into which the current office location falls. The idea was that neighborhoods clustered together with the present location would be viable candidates for a possible second office. The current office is located at 4 Times Square. This neighborhood is located in the Midtown neighborhood. The Midtown neighborhood was categorized in Cluster 2. We will therefore discuss Cluster 2 (cluster label 1) in more detail in arriving at a suitable recommendation for a second office location.

A review of Cluster 2 shows that 8 neighborhoods were categorized into this cluster. We omit the Midtown neighborhood as this is the neighborhood of the current office location and it would not be advisable to have 2 offices located in the same neighborhood. To ensure better coverage of the island it would be more prudent to locate the second office location in a different neighborhood; preferably one that is some distance for the current location to achieve the best possible area coverage. On this basis we omit Midtown South and Flatiron neighborhoods as they are too close to the current office location.

This leaves 5 neighborhoods that are located in Downtown Manhattan:

- Greenwich Village
- Little Italy
- SoHo
- Financial District
- Civic Center

Looking at the 5 remaining options we observe based on knowledge of Manhattan that 3 of the 5 remaining neighborhoods are primarily residential neighborhoods and would not be suitable for securing commercial real estate space. These neighborhoods are Greenwich Village, Little Italy and SoHo.

We are down to 2 possible candidates - Financial District and Civic Center.

As discussed earlier, subway transportation is an important consideration for making sure that the commutes of employees are as convenient as possible. We notice that the

Civic Center has better subway access to the Financial District when measured by number of stations and subway lines that are nearby. Based on knowledge of Manhattan we also observe that rentals in the Financial District will be much more expensive as this is the area were Wall Street is located and is highly sort after by financial services firm both locally and internationally.

6. Conclusion

After considering the neighborhoods that seem most similar to the current location we determine that the Civic Center neighborhood should be considered as the best option for locating a second office location. This determination was made based on the following factors:

- The most comparable subway transportation access to the current Midtown office
- The more economic neighborhood relative to other commercial district that are most similar to the Midtown location
- This neighborhood is in Downtown Manhattan which gives is good separation from the current Midtown location and therefore achieves good area coverage
- Similarity in common venues relative to the current office. For example 4 out of the top 10 venues are similar

In conclusion we recommend that the second location be located in the Civic Center neighborhood of Manhattan, New York City.