

# Capstone Project - The Battle of Neighborhood: From Cambridge, MA to Manhattan, NY

By Arturo Avila-Lares

## 1. Introduction/Business Problem

Due to the COVID-19 pandemic, a lot of people around the world have been forced to work from home. However, seeing as a lot of companies have determined that having employees work remotely is a feasible long-term option, a lot of people have moved from more expensive cities like New York—while saving in rent and maintaining their same jobs. Due to the decreased demand for real state, rent prices have reduced substantially so for a lot of people now would be a good opportunity to move to a city that they couldn't afford pre-pandemic. According to Forbes, rent in New York City has dropped to 10-year lows. Rent in Manhattan has dropped 12.7%, which is more than the 10% drop that happened during the 2008 recession. This decrease has been propelled by an overall increase of 37,000 rentals.

For the purpose of this project, we are going to look at my particular case. I live in the city Cambridge, MA, which is adjacent to the city of Boston—more specifically, I live in a neighborhood called The Port (Zip Code 02139). Given that rent has decreased substantially in Manhattan and is now comparable to rent in the Cambridge/Boston area, I want to explore the idea of moving to a similar neighborhood in Manhattan. For that, I will be doing a clustering analysis with neighborhoods in Boston and Cambridge to understand more about the type of neighborhood I live in. Then, I will proceed to do a clustering analysis with neighborhoods in Manhattan (my preferred borough in New York) to find a similar neighborhood to the one I currently live in and could potentially move to.

## 2. Target Audience

Although this project will be focused on people wanting to move from Boston to New York, it could also serve for people who are looking to move to a different city or different neighborhoods in the same city. It could be for people, like me, who want to live in a similar neighborhood they currently live in or for people who are seeking to live in a neighborhood that has a specific set of venues. Additionally, this project is especially helpful because it can condensate everything a neighborhood has to offer, which is something that would be very time consuming to do by browsing the web. Thus, this is something that would be useful for both people who know the city they want to move to and for people who don't know much about it. Lastly, not only would this allow for people to find similar neighborhoods to potentially move to, but it could also help the realize that it might not be worth moving elsewhere because another neighborhood or city might not offer what they are looking.

## 3. Data Required

- JSON file containing data about all neighborhoods in all boroughs of New York City. From this data source, I will be extracting the name of each neighborhood along their

respective latitude and longitude and will be filtering the data to only the Manhattan borough, which is the one I would like to move to.

- CSV containing list of neighborhoods in the Cambridge and Boston area with their respective zip codes.
- JSON file containing latitude and longitude coordinates for all zip codes in the United States. This data will be merged with the list of neighborhoods in Cambridge and Boston to obtain all the necessary information to extract data for each neighborhood in these cities.
- Foursquare API to extract venues for neighborhoods in Manhattan and neighborhoods in Cambridge and Boston. This would allow me to do a K-means clustering analysis for the neighborhoods in both locations.

#### 4. Methodology

Since the purpose of this project is to find a neighborhood in Manhattan, NY to move to a similar to the one, I currently live in Cambridge, MA, I will be exploring first the neighborhoods in Cambridge, MA, which is called “The Port” and has Zip Code 02139. First thing to point out is that Cambridge, MA is adjacent to Boston, MA so for this project, it makes sense to explore neighborhoods in Cambridge and Boston together—given how these are considerably smaller individually than Manhattan, NY.

To understand more about my neighborhood The Port, I will be performing a K-means clustering analysis with the rest of the neighborhoods in Boston/Cambridge area. To perform this analysis, I will be leveraging the Foursquare API to extract the venues of each of these neighborhoods. However, in order to extract them, I need to know the location (latitude and longitude) for each of these neighborhoods. Since I couldn’t find one data source that had all this information, I had to use two data sources and merge them, which I was able to download from the web:

- I. CSV containing a list of neighborhoods in Boston/Cambridge.
- II. JSON file containing latitude and longitude for each zip code in the US.

Having those two data sources merged, I will be proceeding to use the coordinates to extract the venues for each of the neighborhoods listed as well as their corresponding venue category. The venue category allows me to group these venues and understand what type of venue is most common in each of these neighborhoods. In order to figure out the most common venue categories, I transformed the data using the one hot technique, which converts each of the categories into individual columns and assigns a 1 if the venue in the row falls in that category or 0 if it doesn’t. That way, I am able to group venue categories and determine which venues on average are most common on each of these neighborhoods. Having done this, I will be proceeding to run a clustering analysis with these neighborhoods, trying to optimize the number of clusters, and draw some conclusions about the neighborhood I live in The Port.

However, seeing as there are so many venues per neighborhood, I will be focusing on the 10 most common ones to perform the analysis.

Having completed this clustering analysis, I will be performing a similar analysis with neighborhoods in Manhattan, NY, which is the borough I would like to live in NY, and, similarly to the first analysis, this would give me an idea of the neighborhoods I will be looking at in Manhattan. The one difference of this process is that I have JSON file that contains all the info I need to extract data with the Foursquare API.

Now that I have an idea of the neighborhoods I'm dealing with in both areas, I will be proceeding with running a clustering analysis with all the neighborhoods in both areas at the same time. For that, I will be concatenating both data sources to run the clustering analysis for all neighborhoods. Having done that, I will be able to determine which neighborhoods in NY fall in the same cluster as The Port in Cambridge. This will give me an idea of which neighborhoods in Manhattan are similar to the Port.

As a final step I will be running a K-means clustering analysis exclusively using the neighborhoods in Manhattan that fall in the same cluster as The Port, plus The Port. That way, I will be able to narrow down even more the list of potential neighborhoods to live in Manhattan.

## 5. Results

After running the K-means clustering analysis in the Cambridge and Boston area, The Port was assigned to Cluster #X. This cluster is characterized for one

## 6. Discussion

## 7. Conclusion

I think this analysis was very insightful to understand what are the characteristics of a neigh

