

# **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 located in Central Cambridge. 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 Central 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 the Boston/Cambridge area (**Clustering No. 1**). 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 of 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 one hot encoding, 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 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 (**Clustering No. 2**).

Now that I have an idea of the neighborhoods I'm dealing with in both areas, I will proceed to run 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 (**Clustering No. 3**).

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 (**Clustering No. 4**).

## **5. Results**

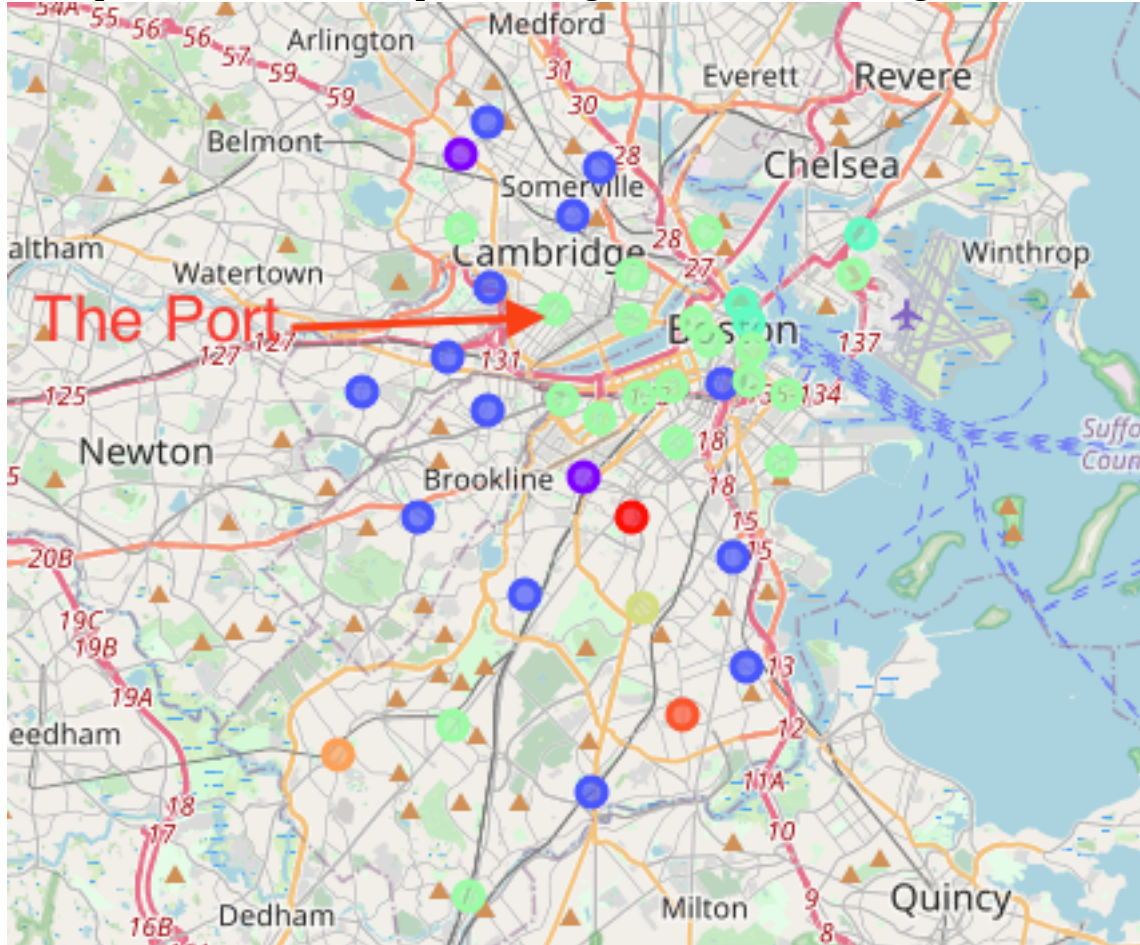
### Clustering No.1: Neighborhoods in the Cambridge/Boston area

After trying different numbers of clusters, I deemed 10 clusters to be the optimal number of clusters for the K-means clustering analysis in the Cambridge and Boston. The Port was placed in Cluster 7, which is characterized by having a lot of restaurants (from a wide array of cuisines) and fitness centers/gyms. It's a cluster that is likely to often receive out-of-town visitors as evidenced by the number of hotels and car rentals in the area, and the wide array of restaurants. A number of these neighborhoods have a considerable number of business centers, which may indicate the presence of offices, which can also be a reason for the number of restaurants, especially the many coffee shops.

I knew a few of these as a someone who has been living in the area for 8 years, but it's interesting to see the analysis confirming what I already knew. Plus, my old neighborhood 'Prudential' is in the same cluster, which in a way validates the analysis. I'm someone who likes to eat out often, goes to the gym, and likes living close to my place of work, and all of these are reflected in the cluster. Additionally, looking at this cluster can help me figure out what other neighborhoods in the area are similar to The Port in case I decided to just move neighborhoods and not cities.

Below, I've included a map depicting the clusters in the area. Cluster 7 is marked with a light green color. Also, there is a table including a sample of the neighborhoods in Cluster 7. As someone who has lived in the area, most of these neighborhoods are attractive to me and would consider moving there based on their similarities with my neighborhood.

**Map of the 10 clusters comprised of neighborhoods in Cambridge and Boston**



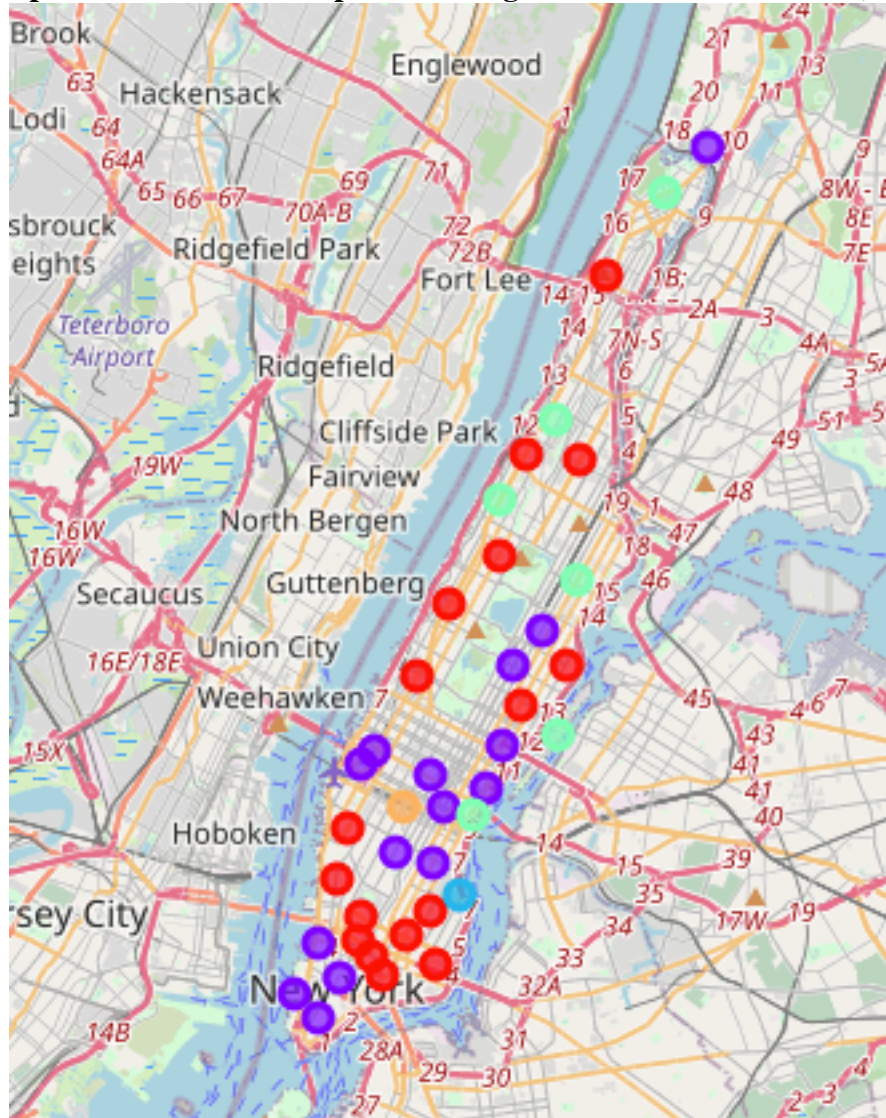
**Sample of Cluster 7, which includes The Port**

	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	Downtown Boston	Business Service	Rental Car Location	Hotel	American Restaurant	Coffee Shop	Italian Restaurant	Food Truck	French Restaurant	Sandwich Place	Bakery
1	Beacon Hill	Italian Restaurant	American Restaurant	Steakhouse	Restaurant	Historic Site	Pizza Place	Gym / Fitness Center	Coffee Shop	New American Restaurant	French Restaurant
3	Financial District / Wharves	Boat or Ferry	Harbor / Marina	Hotel	Historic Site	Aquarium	Seafood Restaurant	Salad Place	Park	American Restaurant	Coffee Shop
31	Mid Cambridge	American Restaurant	Italian Restaurant	Hotel	Grocery Store	Historic Site	Park	Pizza Place	Bakery	Restaurant	Flea Market
32	The Port, Cambridge	Vegetarian / Vegan Restaurant	Coffee Shop	Cocktail Bar	Bar	Yoga Studio	New American Restaurant	Supermarket	Gift Shop	Martial Arts School	Massage Studio

## Clustering No.2: Neighborhoods in Manhattan, NY

For this K-means clustering analysis, I decided to follow the example given in the lab and did 5 clusters. Below, I included a map depicting the 5 clusters in Manhattan. If I were to move here, this analysis would be useful if I ever wanted to move to another similar neighborhood in the area.

### **Map of the 5 clusters comprised of neighborhoods in Manhattan, NY**



## Clustering No.3: Neighborhoods in both Cambridge/Boston area and Manhattan, NY

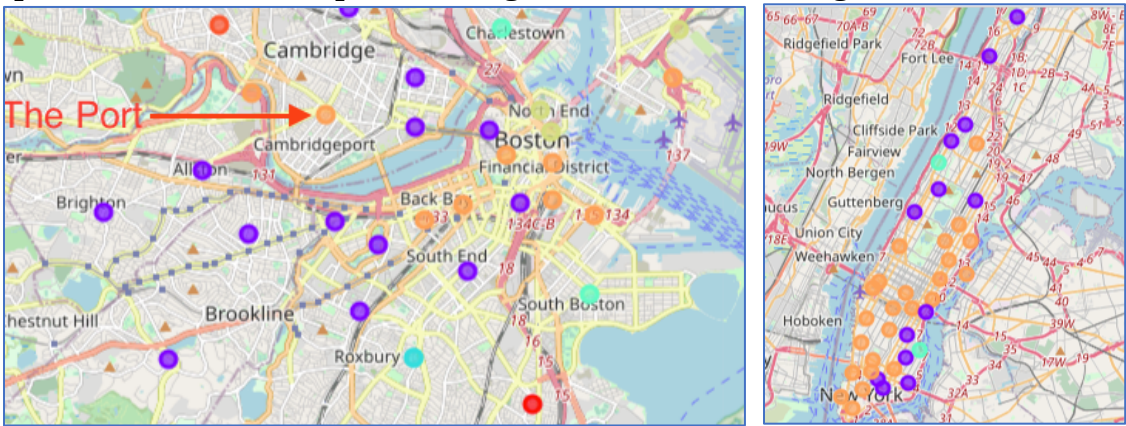
I performed a K-means clustering analysis using neighborhoods in both locations and 10 clusters. The Port is part of Cluster 9 along with neighborhoods in both Cambridge/Boston and Manhattan. Similar to my analysis of cluster 2 for only the Boston/Cambridge area, this cluster is also categorized by having lots of restaurants, shops, and fitness centers/gyms. Thus, these are



neighborhoods that often receive visitors, where people work, and where people live to be in close proximity to gyms and a wide variety of restaurants.

Below I have two pictures of the map where the clustering was done for both locations. Cluster 9 is marked with an orange color. However, seeing as this cluster has a considerable number of members, I did an additional clustering analysis to narrow down this list.

Map of the 10 clusters comprised of neighborhoods in Cambridge/Boston & Manhattan



Sample of Cluster 9, which includes The Port

	Neighborhood	State	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	Downtown Boston	MA	Business Service	Rental Car Location	Coffee Shop	American Restaurant	Hotel	Food Truck	Italian Restaurant	Bakery	Sandwich Place	French Restaurant
1	Beacon Hill	MA	American Restaurant	Gym / Fitness Center	Steakhouse	New American Restaurant	Historic Site	Pizza Place	Italian Restaurant	Restaurant	Coffee Shop	Falafel Restaurant
3	Financial District / Wharves	MA	Boat or Ferry	Harbor / Marina	Hotel	Aquarium	Historic Site	Salad Place	Seafood Restaurant	Park	Bakery	Coffee Shop
5	Downtown Boston	MA	Business Service	Rental Car Location	Coffee Shop	American Restaurant	Hotel	Food Truck	Italian Restaurant	Bakery	Sandwich Place	French Restaurant
32	The Port, Cambridge	MA	Vegetarian / Vegan Restaurant	Coffee Shop	Bar	Cocktail Bar	Plaza	Massage Studio	Mediterranean Restaurant	New American Restaurant	Nightclub	Playground

Clustering No.4: The Port Cambridge and Manhattan neighborhoods in same cluster.

I performed the final K-means clustering analysis using only the Manhattan neighborhoods, included in Cluster 9 from the analysis before, and The Port. The analysis consisted of 5 clusters. The Port is part of Cluster 3, which indicates that the other neighborhoods in this cluster are neighborhoods where I could potentially move to. I will discuss more about this in the discussion section.

Below, I have two pictures of the map where the clustering was done for both locations. Cluster 3 is marked with a light blue color.

## Map of the 10 clusters comprised of neighborhoods in Cambridge/Boston & Manhattan



**Cluster 3 Table**

	Neighborhood	State	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	The Port, Cambridge	MA	Vegetarian / Vegan Restaurant	Bar	Cocktail Bar	Coffee Shop	Yoga Studio	Dance Studio	Bookstore	Breakfast Spot	Shanghai Restaurant	Sandwich Place
2	Central Harlem	NY	African Restaurant	Chinese Restaurant	Bar	French Restaurant	American Restaurant	Seafood Restaurant	Cosmetics Shop	Boutique	Gym	Library
7	Lincoln Square	NY	Plaza	Café	Theater	Concert Hall	Performing Arts Venue	Gym / Fitness Center	Italian Restaurant	Wine Shop	French Restaurant	Indie Movie Theater
9	Midtown	NY	Hotel	Clothing Store	Theater	Sporting Goods Shop	Sandwich Place	American Restaurant	Steakhouse	Coffee Shop	Spa	Bakery
11	Chelsea	NY	Art Gallery	Coffee Shop	Bakery	American Restaurant	Ice Cream Shop	Hotel	Seafood Restaurant	Sushi Restaurant	Cosmetics Shop	Bar
13	Tribeca	NY	Park	American Restaurant	Italian Restaurant	Coffee Shop	Spa	Café	Wine Bar	Poke Place	Skate Park	Playground
16	Battery Park City	NY	Coffee Shop	Park	Hotel	Gym	Clothing Store	Boat or Ferry	Memorial Site	Burger Joint	Food Court	Gourmet Shop
20	Civic Center	NY	Coffee Shop	Cocktail Bar	American Restaurant	Hotel	Spa	Gym / Fitness Center	Yoga Studio	French Restaurant	Park	Bakery

## 6. Discussion

Looking at the 4<sup>th</sup> K-means clustering analysis, we can see that The Port is in Cluster 3, which means the other neighborhoods in this clusters are deemed as similar neighborhoods to The Port and, hence, potential neighborhood I could move to. Given that I know these two cities, I am able to see how these are actually really good recommendations, which is validating to see.

As we have seen throughout this project, The Port is a neighborhood with a wide range of restaurants, gyms/fitness centers, hotels, and car rentals. All of these are common trends along the other neighborhoods in this cluster and confirm my analysis. These neighborhoods are good for people, like me, who like going out to eat, work out, live close to their place of work, and have places nearby where friends or family can stay when visiting. Thus, based on this project,

I could move to any of the other 7 neighborhoods in this cluster, like Midtown and Tribeca, and they would offer similar venues as the ones I currently have in my neighborhood The Port.

Ideally, I would like to expand this analysis to include other features such as median rent, crime rate, etc. and come up with a better choice of neighborhood. However, this project in its current state is flexible if someone wanted to do look at venues to find similar neighborhoods. It could be used if someone wanted to move within the same city or move to a different city that New York within the US. Given that I have all the coordinates for all zip codes in the US, the only data source missing would be the list of neighborhoods to be able to run this analysis.

## **7. Conclusion**

This project allowed me to explore different facets of what it means to be a Data Scientist. I had to extract and wrangle data and run an unsupervised algorithm like K-means clustering, but, more importantly, it gave me an idea of how to use Data Science to solve real-life scenarios. I'm looking to use this experience to embark in other Data Science project and have the opportunity to work with stakeholders to solve business problems through data.

Through this project, I was able to find neighborhoods in Manhattan, NY that are similar to my current neighborhood The Port in Cambridge, MA where I could potentially move to and take advantage of the decrease in rent prices in Manhattan due to COVID-19.