

A Boston Brewery: What Location is Most Ideal?

Applied Data Science Capstone (IBM/Coursera)

By Carina DeBarcelos

Note: For the purposes of this project, Country Roads Brewery and DeBarcelos Contractors are fictional and are not affiliated with any other companies with the same name.

Introduction

i. Background

Boston, Massachusetts is one of the most populous cities in the United States of America, containing nearly 700,000 residents across 26 official neighborhoods. Boston is also a central hub for various higher education institutions, including Boston University, Emerson College, Northeastern University, Suffolk University, and UMass Boston, as well as multiple community colleges. While these institutions are dispersed across the city, students, recent graduates, and other young professionals who live in Boston live in a wide range of neighborhoods. Living in newly developed neighborhoods such as the Seaport District in South Boston expose young adults to a different mixture of venues compared to West Roxbury, one of the furthest boroughs from Downtown Boston.

ii. Problem

Country Roads Brewery, based in Western Massachusetts, is looking to expand its operations into a location in Boston. CRB has emerged as one of the highest-rated breweries in New England for its eclectic mixture of beers, and its flagship location in South Deerfield has attracted customers from many states. It also holds a great reputation with young adults after holding promotions at restaurants near college towns like Amherst. The owners of CRB have a great knowledge of the cities and towns in Western Massachusetts, but are unfamiliar with Boston's neighborhoods outside of the downtown area. As a result, they hired DeBarcelos Contractors to help them determine an ideal location for their second brewery. The owners prefer a neighborhood that lacks breweries, but is close to a university area that attracts students with residential housing and student centers.

With this, we ask the following question: **When controlling for university proximity and proximity to other breweries, what neighborhood would be most ideal for Country Roads Brewery to open a location in Boston?**

iii. Interest

While this project will help Country Roads Brewery, DeBarcelos Contractors hopes that the project will attract future clients, such as other emerging or trending restaurants, to strategically determine locations that fit their customer demographics.

Data

i. Data sources

There are a few factors that will influence the answer to our business problem, such as:

- the number of existing breweries in each Boston neighborhood; and
- the number of university venues and their proximity in each neighborhood (such as classrooms, residential halls, and student centers).

The neighborhood data will be officially obtained via the [City of Boston's Analytics Team](#) in order to define the 26 neighborhoods in Boston, and their coordinates will be obtained using **Google Maps API geocoding**.

Further, the project will use the **Foursquare API** to explore and visualize location and venue data within the neighborhoods of Boston, using the following endpoints:

- the **College and University endpoint**, which will help identify higher education location data and the various venue categories these institutions hold; and
- the **Brewery endpoint** to determine how many breweries are located in each neighborhood.

ii. Data cleaning and methodology

Through the identified data sources, I will scrape, extract, and clean the necessary data for this project, and will further analyze the neighborhoods via k-means clustering. The following libraries were also imported to conduct the analysis:

- Pandas,
- Numpy,
- Json and Json_normalize,
- Nominatim,
- Google Geocoder
- Matplotlib,
- KMeans, and
- Folium.

The Pandas dataframe containing data of the 26 Boston neighborhoods had to be cleaned by dropping columns that were not relevant to the analysis, such as shapefile sizes and object

IDs. Further, the dataframe was appended with neighborhood coordinates that were obtained via the Google Geocoder library. Folium was used to visualize neighborhoods, as well as the clusters created to observe neighborhoods with similar higher education and brewery venues.

In order to explore higher education and brewery venues, the Foursquare API was used to look up to 100 venues in each neighborhood, within 500 meters of coordinates. The venues, as well as their latitude and longitude, were stored into respective dataframes. Some neighborhoods had much more than 100 venues in certain searches (such as higher education venues in the Chinatown neighborhood), but could not be explored due to limits with using the Foursquare API.

For both the higher education and brewery venues, further analysis was conducted through one hot encoding in order to determine how many venues are in each category for each neighborhood. This will help with predicting what cluster each neighborhood best “fits” in. The one hot encoding was done in separate dataframes and were cleaned up to neatly represent each neighborhood in 26 consolidated columns. Each neighborhood had a printed report indicating the 10 most common college and university categories, as well as brewery categories.

These venues were then placed into a dataframe to prepare for K-Means clustering, which would then be spun off into a new dataframe containing the cluster labels and the top 10 venue categories in each neighborhood. Two K-cluster groups were created – one for the higher education venues, and one combining the higher education venues with the brewery venues. Both groups had a total of 5 clusters, and the dataframes were cleaned to remove null values. Null values represented neighborhoods that did not share a common thread with other neighborhoods to be clustered. Then, the clusters were visualized through Folium, and each cluster was identified by color and label markers.

Analysis

i. Preliminary data analysis

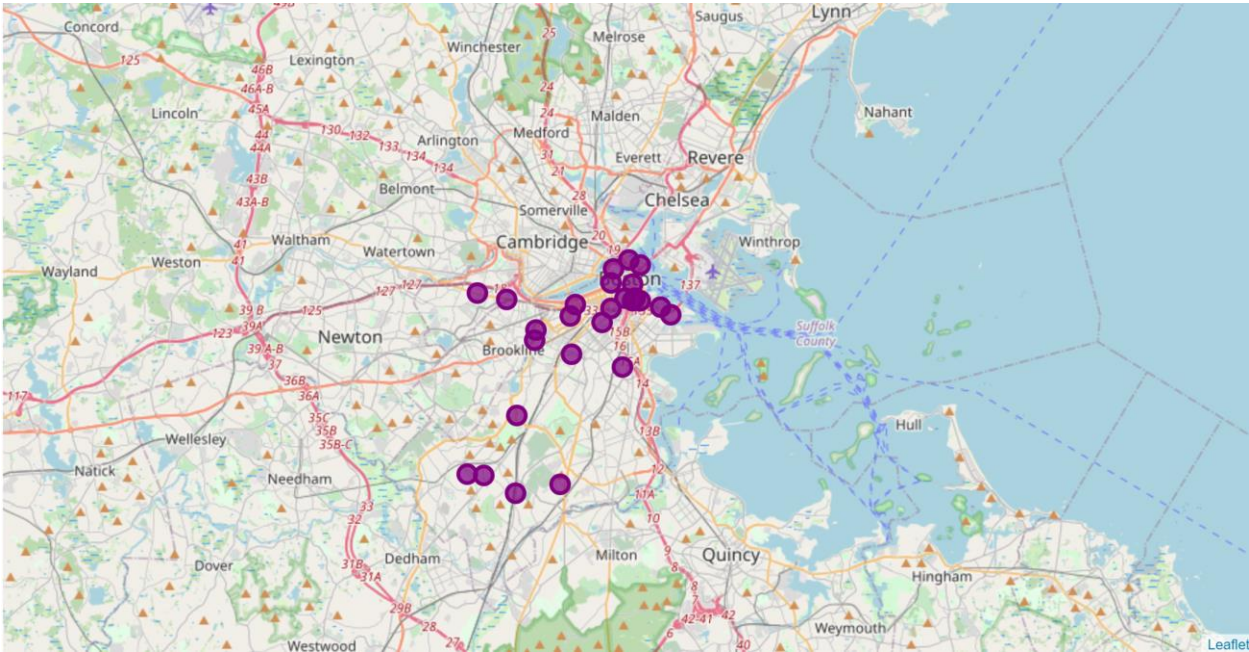
Boston officially has 26 neighborhoods identified by the City’s analytics team. These neighborhoods include:

1. Allston	14. Jamaica Plain
2. Back Bay	15. Leather District
3. Bay Village	16. Longwood
4. Beacon Hill	17. Mattapan
5. Brighton	18. Mission Hill
6. Charlestown	19. North End
7. Chinatown	20. Roslindale
8. Dorchester	21. Roxbury
9. Downtown	22. South Boston
10. East Boston	23. South Boston Waterfront
11. Fenway	24. South End

12. Harbor Islands	25. West End
13. Hyde Park	26. West Roxbury

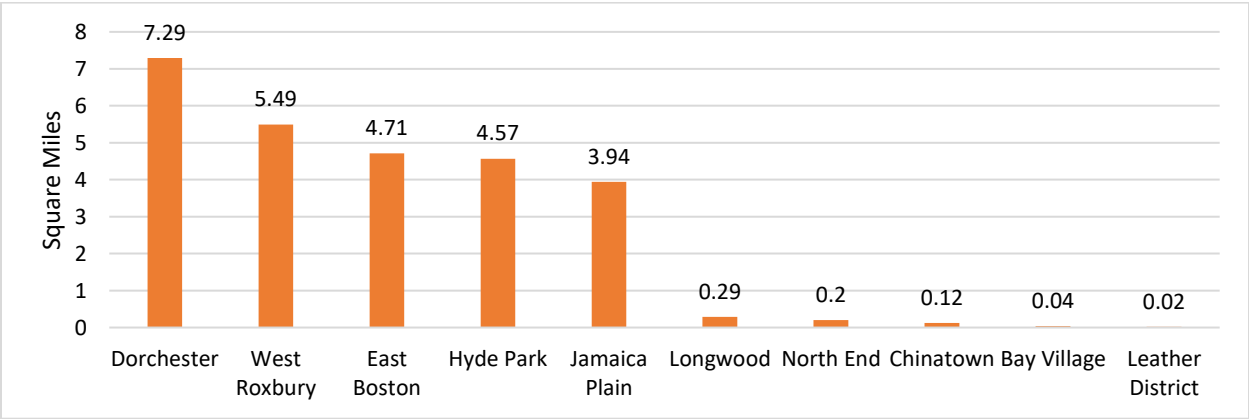
Boston’s neighborhoods cover a large swath of land, as identified in Figure 1:

Figure 1 – Visual Map of Boston Neighborhoods



In order to get a good sense of Boston’s size as a city, as well as the size of individual neighborhoods, I identified the five largest neighborhoods, as well as the smallest, as shown in Figure 2.

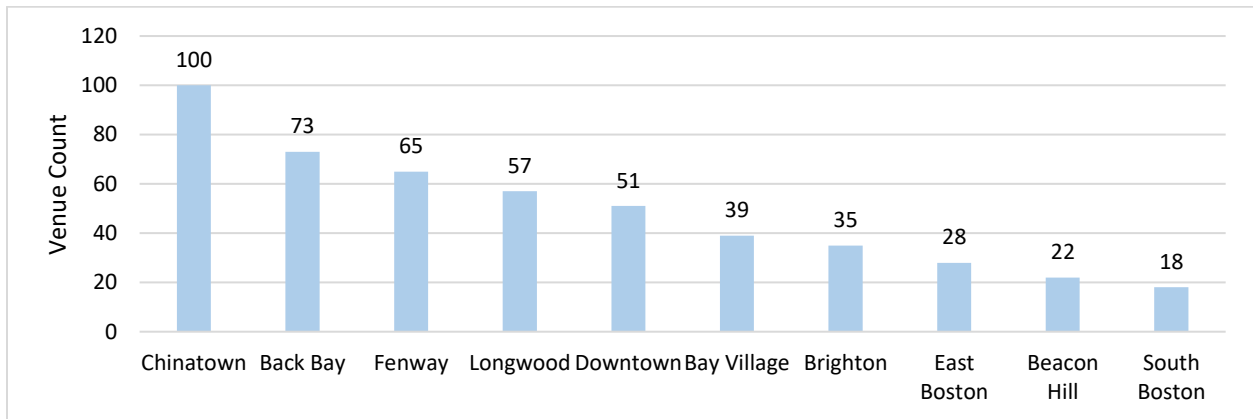
Figure 2 – Largest and Smallest Neighborhoods in Boston, Massachusetts by Square Mileage



ii. Neighborhood observations and clustering: colleges and universities

At initial observation, at least 561 higher education venues were identified by the Foursquare API's College and University endpoint across Boston's 26 neighborhoods. According to Figure 3, Chinatown, Back Bay, Fenway Longwood, and Downtown have the highest count of higher education venues, representing 346 of the 561 venues in the city. Neighborhoods with the lowest count of higher education venues were Jamaica Plain, Mattapan, and Roslindale – all who had just 1 venue each.

Figure 3 – Boston Neighborhoods with Highest Higher Education Venue Frequency



Altogether, the 561 venues represented 52 unique categories under the College and University endpoint, as seen in Figure 4. Scrutiny was given to venues that were labeled under categories belonging to classrooms and general buildings, but we also paid attention to venues that may be prime for student interaction such as student centers, recreational centers and gyms, college quads, and residential halls.

Figure 4 – College and University Endpoint Categories Captured in Boston Neighborhoods

College Classroom	General College & University	Student Center	College Auditorium
College Administrative Building	College Library	College Science Building	Medical School
College Stadium	Medical Lab	College Rec Center	Athletics & Sports
University	College Academic Building	College Arts Building	College Residence Hall
College Lab	College Quad	College Cafeteria	College Gym
College Bookstore	College & University	Community College	Language School
Trade School	School	Office	Library
Law School	TV Station	Dance Studio	College Theater

Performing Arts Venue	College Communications Building	Gym	Fraternity House
General Travel	Track	Music Venue	College Technology Building
Social Club	Bookstore	Speakeasy	Theater
Coworking Space	College Baseball Diamond	College Track	Pool
College Hockey Rink	Health & Beauty Service	Entertainment Service	College Engineering Building

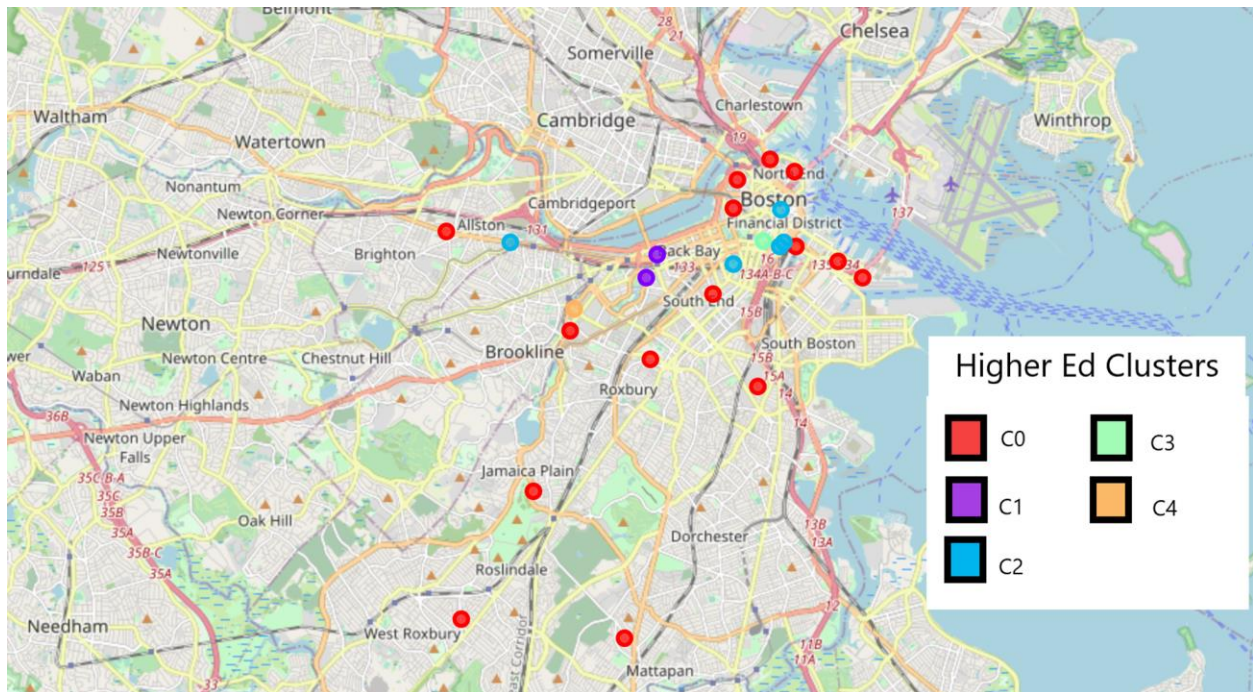
After some preliminary one hot encoding, we found that there were some neighborhoods that had a large frequency of academic buildings, residential halls, *and* student centers. These neighborhoods included:

- Back Bay
- Chinatown,
- Downtown,
- Fenway, and
- Longwood.

There were other neighborhoods that may be worth exploring. For example, the neighborhoods of Brighton and the North End had a notable frequency of residential halls. On the other hand, Beacon Hill, Charlestown, East Boston, the Leather District, Mission Hill, and South Boston had a considerable frequency of student centers.

We were able to see more common threads between the neighborhoods once they were clustered and visualized, as seen in Figure 5. The neighborhoods of Hyde Park and Roslindale are not included in the clusters as they were not able to find common threads with the other clusters. Therefore, we can determine that these neighborhoods will not be ideal locations based on Country Roads Brewing's interests.

Figure 5 – Higher Education Venue Clusters in Boston Neighborhoods



The first cluster (visualized as C0) includes the neighborhoods of Roslindale, Jamaica Plain, Mission Hill, the Leather District, the North End, Roxbury, the South End, Charlestown, the West End, Beacon Hill, Mattapan, Dorchester, the South Boston Waterfront, Allston, and the Harbor Islands. This cluster consists of more than half of the city’s neighborhood, and their most common higher education venues consist of general buildings and some of the less common categories we are not looking for in our search. Most notably, this cluster does not appear to have as many venues for student interaction, such as residence halls and student centers.

The second cluster (visualized as C1) includes the neighborhoods of Back Bay and Fenway. There is more common thread seen between these two neighborhoods, as their most common venues are college arts and academic buildings. There are also student centers, residence halls, and student quads – areas where students may gather.

Figure 6 – Cluster C1 Common Higher Education Venues

	Neighborhood	Cluster Labels	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
10	Back Bay	1	College Arts Building	College Academic Building	College Administrative Building	Fraternity House	Student Center	General College & University	University	College Residence Hall	College Cafeteria	College Library
16	Fenway	1	College Academic Building	College Arts Building	College Residence Hall	Student Center	General College & University	College Administrative Building	University	College Cafeteria	College Quad	College Library

The third cluster (visualized as C2) consists of Bay Village, East Boston, Downtown, Brighton, and South Boston. This group of neighborhoods has more frequency of general higher education buildings, as well as administrative facilities, but do not have as many areas to congregate for students, such as residence halls or student centers.

The fourth and fifth clusters (visualized as C3 and C4) are standalone neighborhoods – Chinatown and Longwood. These are interesting because they have areas for students to socialize but have an unusually high frequency of medical school buildings.

Figure 7 – Clusters C3 and C4 Common Higher Education Venues

	Neighborhood	Cluster Labels	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
6	Chinatown	3	General College & University	College Lab	College Academic Building	College Administrative Building	College Residence Hall	Student Center	Medical School	College Library	College Classroom	Law School

	Neighborhood	Cluster Labels	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
3	Longwood	4	College Residence Hall	College Cafeteria	Medical School	College Gym	College Academic Building	General College & University	College Classroom	College Quad	College Administrative Building	College Arts Building

Before looking at the presence of breweries in these neighborhoods, we can see that Clusters C1, C3, and C4 might be some of the most ideal locations by just looking at higher education venue attributes. We still have more analysis to go, so it is too early to determine a finite location for a new Country Roads Brewery location.

iii. Neighborhood observations: breweries

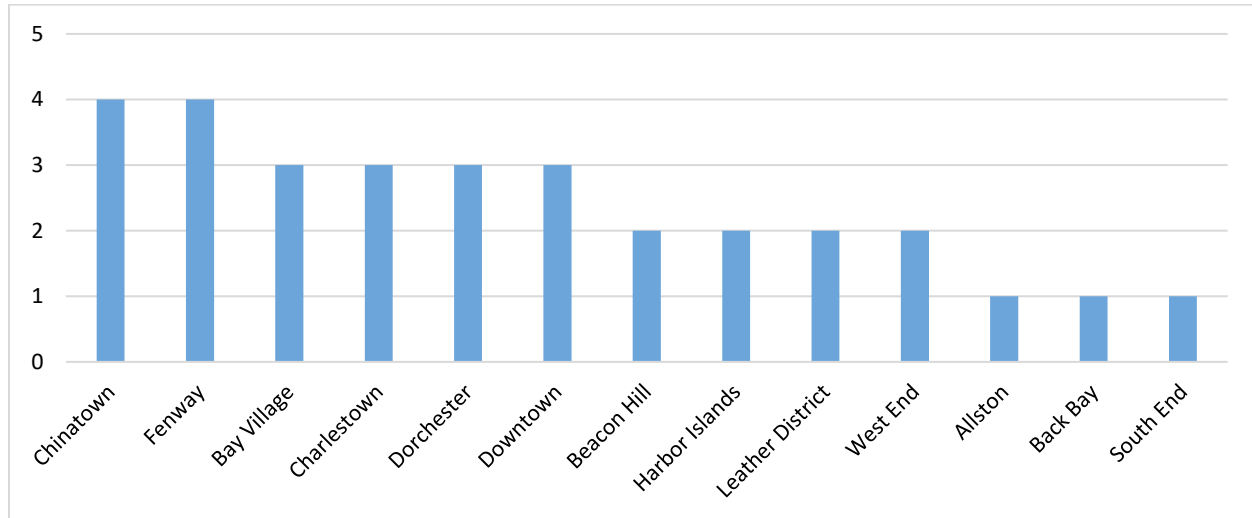
At initial observation, only 31 brewery venues were identified by the Foursquare API's Brewery endpoint across Boston's 26 neighborhoods. According to Figure 8, Chinatown, Fenway, Bay Village, Charlestown, Dorchester, and Downtown have the highest count of brewery venues, representing 20 of the 31 venues in the city. In fact, only 13 neighborhoods, or half of the neighborhoods in Boston, have a brewery. Altogether, the 31 venues represented 9 unique categories under the Brewery endpoint, including the following:

- Brewery,
- Pizza Place,
- Sandwich Place,
- Burger Joint,
- Bar,
- American Restaurant,
- Pub,
- Mexican Restaurant, and

- Gastropub.

These categories indicate that where there are true breweries in the mix, some of the venues in this endpoint have venues that have brewery-like qualities and serve as other purposes. Interestingly enough, the neighborhoods of Allston, Beacon Hill, the South End, and the West End do not have a brewery venue listed as “Brewery” as its venue category.

Figure 8 – Boston Neighborhoods with Highest Brewery Venue Frequency



iv. Clustering based on higher education and brewery attributes

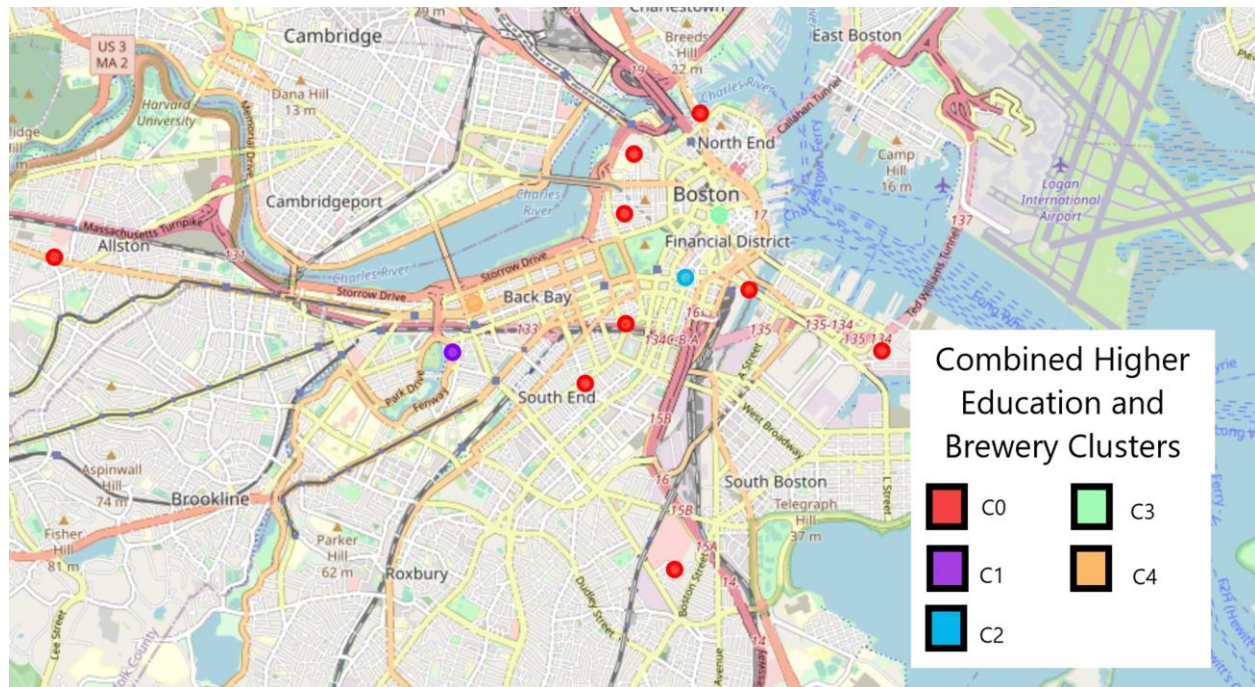
In order to establish new cluster groups, I will merge the two consolidated, one hot encoded dataframes that were created for higher education venues and brewery venues into a new dataframe. We were able to see more common threads between the neighborhoods once they were clustered and visualized, as seen in Figure 9. By doing this, the following neighborhoods were excluded from the new cluster analysis due to null values:

- Brighton,
- East Boston,
- Hyde Park,
- Jamaica Plain,
- Longwood,
- Mattapan,
- Mission Hill,
- North End
- Roslindale
- Roxbury,
- South Boston,

- South Boston Waterfront, and
- West Roxbury.

As a result, we can safely say that these neighborhoods might not be the most ideal locations to locate a new brewery based on Country Roads Brewing's expectations. This will also make it easier to narrow down locations as they represent nearly half of the city's neighborhoods.

Figure 9 – Higher Education and Brewery Venue Clusters in Boston Neighborhoods



The first cluster (visualized as C0), consists of Bay Village, the Leather District, the South End, Charlestown, the West End, Beacon Hill, Dorchester, Allston, and the Harbor Islands. In this cluster, we found that these neighborhoods have a brewery, or a brewery-like venue that are just as common as their college and university venues. Therefore, we can remove these neighborhoods for contention as a potential location for Country Roads Brewery.

The second cluster (visualized as C1) consists of the Fenway neighborhood, one of the communities we identified as ripe for student interaction. We can see that a brewery is among one of the most common venues in the neighborhood, so we can also eliminate this neighborhood for contention. Similarly, the fourth cluster (visualized as C3 and consists of the Downtown neighborhood) also has a brewery among its most common venues and can also be removed from contention.

That leaves us with the third cluster (C2) and the fifth cluster (C4) – which consist of the neighborhoods of Chinatown and Back Bay, respectively. These two neighborhoods do not have a brewery among their most common venues, as seen in Figure 10, but the two neighborhoods have differences. For example, Chinatown has centers for student interaction but appears to have

more medical school campuses. Meanwhile, Back Bay has a similar composition of venues compared to Fenway, but unlike Fenway breweries are not common in the neighborhood.

Figure 10 – Clusters C2 and C4 Common Higher Education and Brewery Venues

	Neighborhood	Cluster Labels	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
6	Chinatown	2	General College & University	College Lab	College Academic Building	College Administrative Building	College Residence Hall	Student Center	Medical School	College Classroom	College Library	Law School

	Neighborhood	Cluster Labels	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
10	Back Bay	4	College Arts Building	College Administrative Building	College Academic Building	Fraternity House	Student Center	General College & University	College Residence Hall	University	College Library	College Technology Building

Upon further individual analysis of the Chinatown and Back Bay neighborhoods, we uncovered a few key findings. In the Back Bay neighborhood, there are a number of colleges houses there representing 73 individual venues. These institutions include the Boston Architectural College, Berklee College of Music, the New England College of Optometry, the Boston Conservatory, and a Boston University sub-campus. We also found that there is one brewery – named Edge! – as well. External analysis reveals that Edge! Belongs to a guest hotel facility.

Chinatown, on the other hand, houses Emerson College, New England College of Finance, Babson College, Tufts Medical School, New England Law, and Suffolk University Law School. These institutions represent much more than 100 venues under the Foursquare API. We also found that there are also four brewery-like venues in the neighborhood: Democracy Brewing Company, Rock Bottom, Boston Beer Company, and Back Deck.

Figure 11 – Brewery Venues in Back Bay and Chinatown

	Area	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
10	Back Bay	42.34999	-71.08765	Edge!	42.345795	-71.087747	Brewery

	Area	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
5	Chinatown	42.35251	-71.0609	Democracy Brewing Company	42.355073	-71.062135	Brewery
6	Chinatown	42.35251	-71.0609	Rock Bottom Restaurant & Brewery	42.351329	-71.065327	Brewery
7	Chinatown	42.35251	-71.0609	Boston beer company	42.352899	-71.064451	Brewery
8	Chinatown	42.35251	-71.0609	Back Deck	42.354448	-71.061898	Burger Joint

Findings and Discussion

Based on our analysis, we discovered there are a lot of higher education venues across Boston, in a wide swath of neighborhoods. That gave us the flexibility to go outside of the immediate downtown area, knowing that there were colleges and universities in some of the smaller neighborhoods. Conversely, there were less brewery venues than anticipated in Boston. This was surprising to our client because of the wide variation of breweries across New England and their inaccurate assumption that they were entering a crowded brewery market in the city.

It was no surprise that the neighborhoods with the highest frequency of higher education venues were in the most popular neighborhoods in the city, which have access to public transit, restaurants, historical places and various tourist attractions. Brewery locations were varied across the city, regardless of popularity. Clustering neighborhoods with both of these factors helped weed out neighborhoods that would not be ideal.

Filtering out neighborhoods was easier than expected - nearly half of the Boston neighborhoods were not included in these new clusters because they lacked brewery categories and/or higher education venue categories. We would not find it ideal to place Country Roads Brewing in a neighborhood that does not have many higher education venues that may posit student interaction. A caveat of this approach is that it may weed out neighborhoods that may simply not have *any brewery* but had plenty of higher education venues. This brought us to our final cluster groups, which were easily identifiable. The elimination of each cluster group was based on frequency of breweries, as well as lack of venue frequency that promote student interaction, such as student centers. The final two clusters - Back Bay and Chinatown - were known as having the most higher education venues, with Back Bay having the edge of having more venues with additional student interaction.

Ultimately, DeBarcelos Contractors will select **Back Bay** as the ideal neighborhood to locate the new facility for Country Roads Brewery, due to the lack of brewery venues in the area despite its large student population from neighboring colleges such as Berklee College of Music and the Boston Conservatory.

Conclusion

The purpose of this project was to identify an ideal location for Country Roads Brewery to open in Boston, when controlling for university proximity and proximity to other breweries. Through the Foursquare API, we were able to identify various higher education venues, as well as brewery-like facilities across Boston's 26 neighborhoods, and established k-clustering mechanisms to explore common threads across the neighborhoods. By merging one-hot encoded groups for higher education institutions and breweries, we were able to identify neighborhoods that were more likely to have a student presence, as well as those that may have a potential brewery following. We wanted to select an ideal location that the brewery industry hasn't fulfilled yet, but would attract that young adult crowd that made Country Roads Brewery so popular in Western Massachusetts.

We are comfortable with selecting **Back Bay** as the ideal neighborhood to place Country Roads Brewery's new location, as they fulfill our criteria of abundant higher education venues, as well as a brewery market that has not been fulfilled in this neighborhood. We hope that this exercise inspires other breweries and restaurants to reach out to us so we can continue strategizing potential markets for them.

Future Directions

It would be interesting to see how the results would differ if we focused on *all nightlife venues* and not just brewery-like venues. What implications would be there? Would the ideal location be in a neighborhood that lacks nightlife overall? Would the neighborhood already have multiple breweries? Back Bay is rich in other restaurants and may have some nightlife as well, and it could be possible it may have not been selected as the final venue if our analysis was approached differently.