

Exploring location attributes of successful breweries in the United States

1. Introduction/Business Problem

A beer connoisseur would like to visit breweries across United States to supplement his research on attributes that make a brewery commercially successful. The end goal is to assess whether there are business opportunities in the beer brewing industry and to identify drivers that increase the likelihood of establishing a successful brewery. There are numerous attributes that draw more customers to a venue or a product. Besides obvious and intrinsic factors such as quality, uniqueness and taste, extrinsic factors such as market demand, location, and brand play important roles in determining the viability of a brewery business. Some of the external factors can be studied using data. Population density is a simple proxy for exploring the impact of market demand, and the influence of location can be studied by exploring the places around successful breweries. These two aspects can be studied concurrently by exploring the locations around successful breweries in big populated cities across the united states.

The beer enthusiast intends to use his beer palate to reveal the qualitative aspects of a good product and use data analysis to quantitatively assess what makes a brewery thrive. The task at hand involves the following steps: (1) select top N big cities in the country; (2) identify highly rated breweries in each city; (3) explore the location of breweries and identify commonalities among them using clustering;(4) interpret the results and assess business viability in light of this assessment. The insights gleaned from studying successful breweries will supplement the development of a business plan to establish a new brewery.

2. Data Requirements

Data required corresponding to each step discussed above:

- A list of cities with their population density/size along with their coordinates – The data points are widely available, albeit not well organized. A free tier on Simplemaps.com provides access to a list of world cities with their coordinates and population. <https://simplemaps.com/data/us-cities>
- Foursquare API can be leveraged to identify top 'liked' breweries in each city selected in step 1.
- Foursquare API can be leveraged to explore the location/vicinity of selected breweries.
- Vicinity attributes of each brewery across the country can be grouped and analyzed.

3. Methodology and model selection:

The following assumptions were made to simplify the analysis.

- Breweries in most populated cities are better than breweries not located close to the city.
- The popularity of the brewery was based on # of 'likes' on foursquare.

3.1 Exploratory Analysis:

The list world cities along with populations data was downloaded from <https://simplemaps.com/data/us-cities> and processed using spreadsheets. Cities in US were filters and sorted based their populations. As mentioned above, the underlying assumption is that there is more reliable information about breweries in highly populated cities since these breweries are rated by a large population. To keep the analysis simple, top 10 cities shown in Table 1. were selected for our analysis.

Table 1. Top 10 cities in US based on population
(Population data as of September 2019)

	City	Latitude	Longitude
0	New York	40.6943	-73.9249
1	Los Angeles	34.1139	-118.4068
2	Chicago	41.8373	-87.6862
3	Miami	25.7839	-80.2102
4	Dallas	32.7936	-96.7662
5	Philadelphia	40.0077	-75.1339
6	Houston	29.7869	-95.3905
7	Washington	38.9047	-77.0163
8	Atlanta	33.7627	-84.4225
9	Boston	42.3188	-71.0846

The locations of these cities were plotted as shown in Figure 1. to create a base map.



Figure 1. Top cities selected for analysis – base map
(Created using Folium package)

A list of breweries in each of the 10 cities was retrieved using Foursquare API. Foursquare provides the category of venues. Each category can be called individually by specifying the category ID in the API calls. To simplify the analysis, a maximum of 10 breweries within 10,000 meters of each city center were

selected for the analysis. The locations (latitude and longitude) were retrieved and plotted. The map can be zoomed-in in the jupyter notebook to view the distribution of breweries around the city. As an example, Figure 2. shows the distribution of selected breweries around New York city.

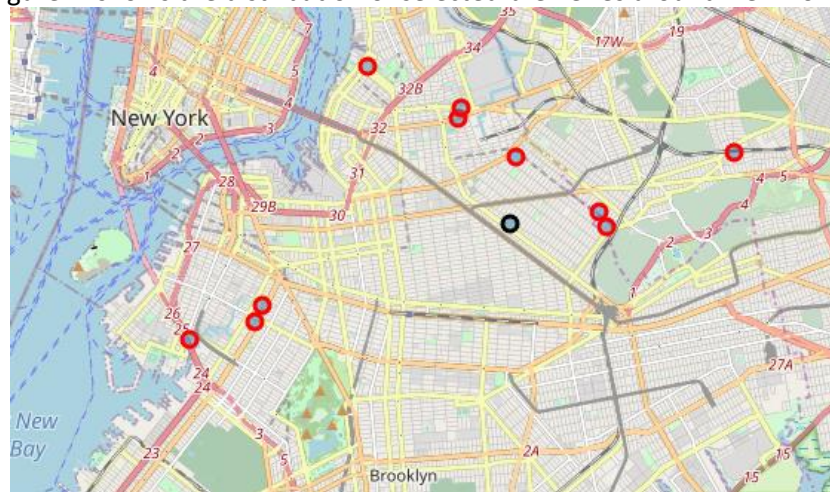


Figure 2. Location of breweries in New York city

(Created using Folium package)

The number of 'likes' on Foursquare was used as a proxy for the popularity/quality of the brewery and their product. A separate API call was made to retrieve the number of 'likes' for each brewery. 'Likes' is arguably the best available measure to rate the breweries. Higher the number of 'likes' greater the popularity of the brewery, and all things equal, better the business prospects.

The 'likes' data was further explored for each city. It is important to note that the 'likes' were analyzed on an absolute basis, without taking into account the city population or other aspects such as tourism that typically affect the number of people visiting a brewery, and hence the number of 'likes'. The distribution of 'likes' was plotted for each city as shown in Figure 3. New York, Boston, Chicago, and Miami have some of the highest rated bars in our dataset.

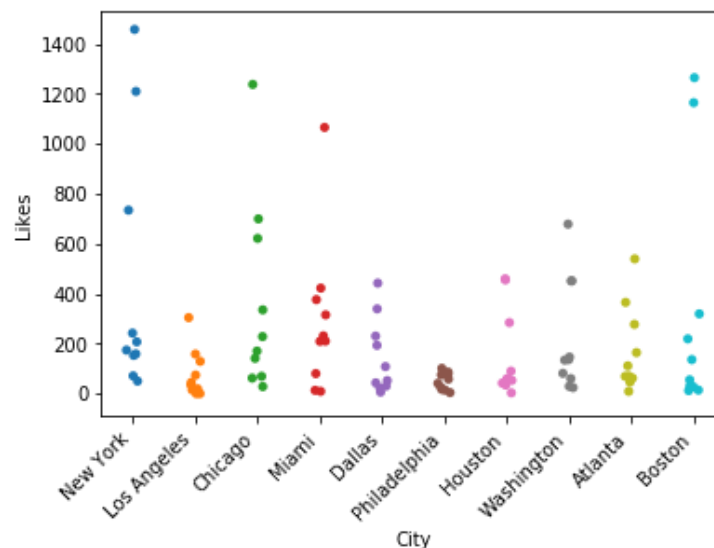


Figure 3. Distribution of number of 'likes' in each city (10,000-meter radius)

The insight from Figure 3. Gently guides the analysis towards exploring to a greater the degree the breweries that are highly rated and are in cities with more highly rated breweries.

One of the main goals of this analysis is to supplement existing research with new information or pattern based on establishments in the vicinity of top breweries. In other words, we are seeking the answer to these questions – what lies in the vicinity of top breweries? Is there any commonality in the type of venues around top breweries?

Foursquare provides information about venues around a requested location. The venues around each brewery in our dataset was retrieved using the API. Other information retrieved with the venue information includes the location and the venue category. The number of venues was limited to 50 and the search radius was limited to 1000 meters. The partial table below shows the information retrieved from Foursquare API. The venue category is our feature of interest. The feature captures the type of establishments around a brewery. The dataset consisted of at least 6 venues for each brewery.

Table 2. Partial dataset showing venues around selected breweries

	City	Brewery	Brewery Latitude	Brewery Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	New York	Kings County Brewers Collective	40.705974	-73.923487	Local Honey	40.704902	-73.924325	Nail Salon
2	New York	Kings County Brewers Collective	40.705974	-73.923487	House of Yes	40.706777	-73.923704	Nightclub
3	New York	Kings County Brewers Collective	40.705974	-73.923487	Arepera Guacuco Restaurant	40.704775	-73.924500	Arepa Restaurant
4	New York	Kings County Brewers Collective	40.705974	-73.923487	The Sampler BK	40.705627	-73.922485	Bar
5	New York	Kings County Brewers Collective	40.705974	-73.923487	Théâtre XIV	40.706121	-73.923448	Theater

Each brewery had multiple venues of same or different category in its vicinity. To determine the most common venue category for each brewery, the venue category was grouped for each brewery using one hot coding. One hot coding allows us to analyze categorical variables and apply numerical algorithms to categorical variables. The coded variables were grouped and averaged to determine the normalized frequency of each venue category for each brewery.

```

----Max Lager's Wood-Fired Grill & Brewery.
                                venue freq
0                                Hotel 0.12
1                                Hotel Bar 0.10
2                                Coffee Shop 0.08
3                                Bar 0.06
4                                Rental Car Location 0.04
5                                Eastern European Restaurant 0.02
6                                Shipping Store 0.02

```

Figure 4. Normalized frequency of each venue category for each brewery

The dataset was sorted to list the top 15 most common venues. The number of topmost common venues did not affect the analysis. (Analysis was carried out with 10 and 20 top venues)

The resulting dataset was grouped based on the most common venue. Figure 5. shows the number of times a venue appears in the most common venue position.

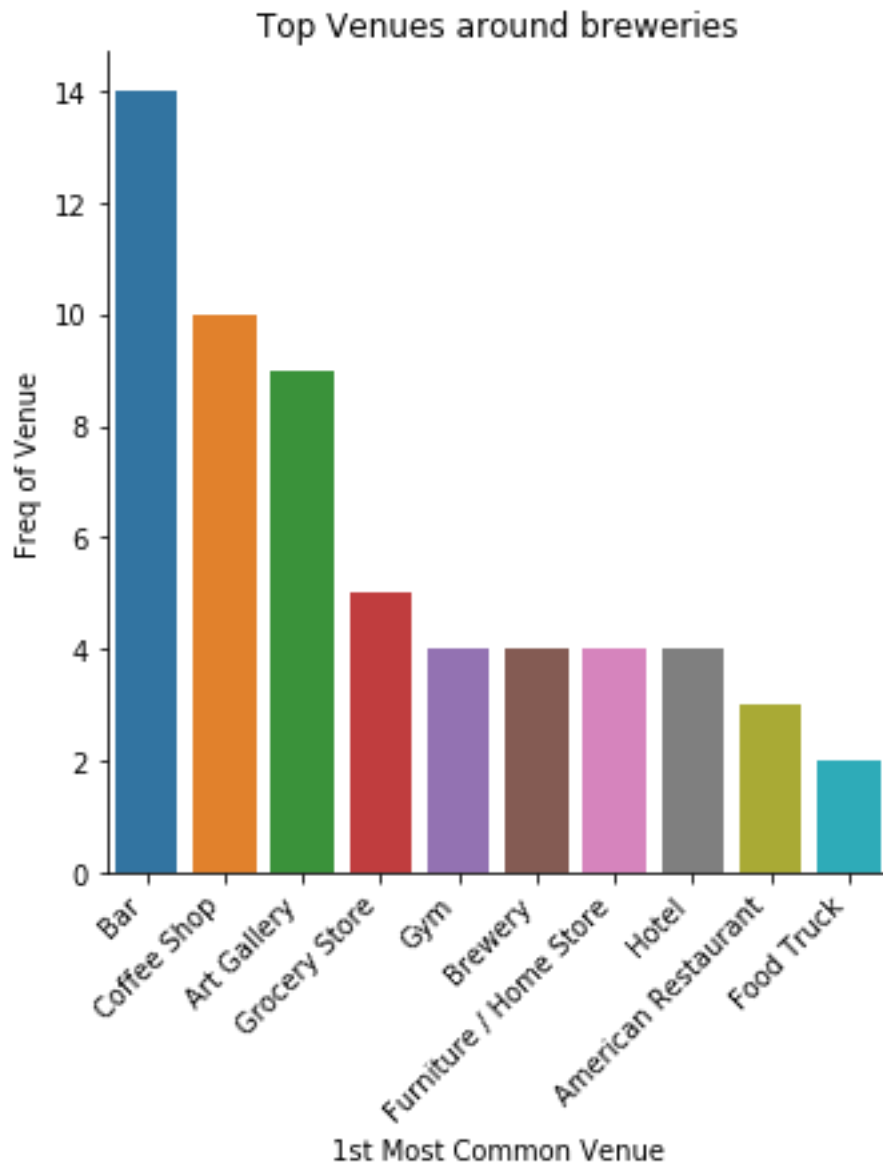


Figure 5. Top venues around breweries

Clearly, the top venues around breweries are bars. This fact almost seems obvious but is substantiated by data. Other venues are not so obvious. Coffee shops and art gallery suggest the neighborhoods usually attract crowds are situated near city landmarks.

3.2 Model selection

With information about venues close to top breweries, the next step is to identify similarities among breweries based on the venue category feature. This project used clustering to explore the similarities in our selected list of breweries based on venues. Cluster analysis is the task of grouping a set of objects that are more similar to each other than to those in other groups.

K-means clustering from scikit-learn package was utilized to perform clustering. The algorithm was run with different number of clusters ranging from 4 through 8. The optimal number of clusters was determined heuristically using elbow method. The value of k at the elbow was 5. Thus, 5 clusters were used to carryout the clustering analysis and interpret the results.

4. Results and discussion:

Figure 6. summarizes the number of breweries within each cluster. Majority of breweries were grouped under cluster 1 and cluster 4, with cluster 4 having the maximum number of breweries.

Cluster Labels	
0.0	2
1.0	26
2.0	10
3.0	1
4.0	60

Figure 6. Number of breweries in each cluster

The distribution of 'likes' across the clusters is shown in Figure 7. Cluster 2 has the highest average 'likes' with a relatively narrower range of 'likes'. Cluster 4 has breweries with high number of 'likes' but the average is slightly below cluster 2, and the spread is much higher compared to cluster 2. The average rating of Cluster 1 is lower than Cluster 2 despite having more breweries in its group. Cluster 0 and Cluster 3 do not have enough data to draw insights. The dissimilarities in these cluster's location is discussed later in this section.

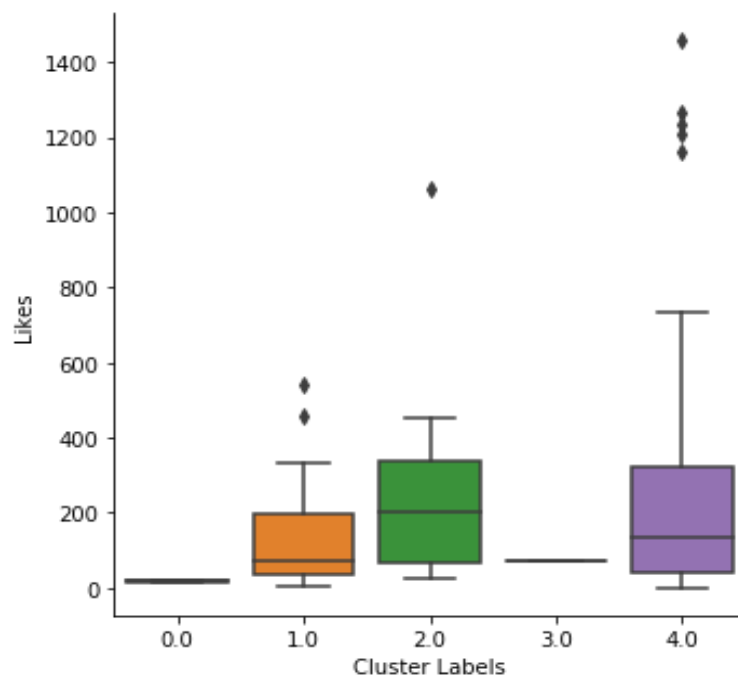


Figure 7. Distribution of 'likes' across clusters

With 'likes' as an indicator of brewery success, the breweries in Clusters 2 and Cluster 4 are more likely to attract more customers. Figure 7. shows a bar plot with venue distribution for top 20 breweries based on 'likes'.

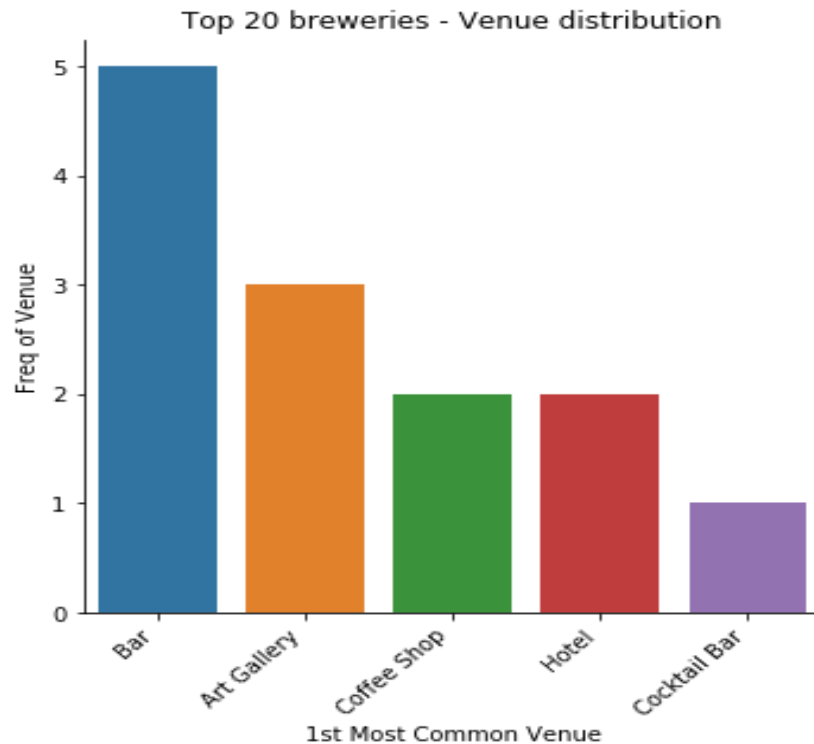


Figure 7. Top venues around top 20 breweries

Comparing Figure 5. and Figure 7, the most common venues in the vicinity of top 20 breweries are similar to the rest of the breweries. Bars, art galleries and coffee shops are the most common venues. However, venues like grocery store or gym do not appear around top 20 breweries. This suggests that highly ranked breweries are typically surrounded by these venues, which may be impacting the brewery operations and business.

Before exploring the venues around these clusters, the distribution of breweries and clusters across cities would provide some insight about the impact of cities on the number of 'likes' of breweries. Figure 8. shows a swarm plot of cities, likes and clusters.

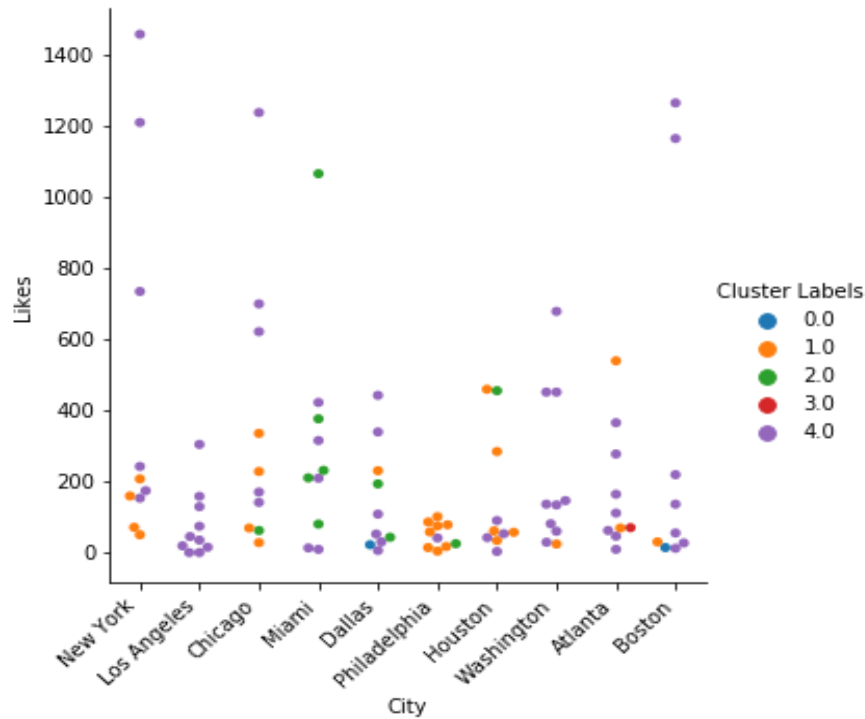


Figure 8. Distribution of 'likes' and clusters in each city

Each city has breweries from different clusters. The cities do not seem to have any significant impact on the distribution of 'likes'. The number of 'likes' is presumably a function of population, but the plot above shows no correlation between the population and the number of likes. Other factors such as demographics and tourism could affect the rating of breweries.

The distribution of clusters in two cities – Los Angeles and Philadelphia, are distinct. All breweries selected in Los Angeles belong to cluster 4, the most common cluster. Philadelphia has disproportionately more cluster 1 breweries. Figure 9. shows a map of clusters in Philadelphia. Cluster 1 is the most common cluster Table 3. shows a partial list of breweries along with their respective most common places.

Table 3. Breweries in Philadelphia with most common venues

	Brewery	Likes	Cluster Labels	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
50	Philadelphia Brewing Company	78	1.0	Philadelphia	Bar	Café	Pizza Place	Athletics & Sports
51	Fermentery Form	17	1.0	Philadelphia	Art Gallery	Café	Food Truck	Brewery
52	Evil Genius Beer Company	86	1.0	Philadelphia	Dive Bar	Café	Indian Restaurant	Taco Place
53	Love City Brewing	58	1.0	Philadelphia	Gastropub	Park	Art Gallery	Food Truck
54	Crime & Punishment Brewing Co.	75	1.0	Philadelphia	Bus Station	Pizza Place	Café	Coffee Shop

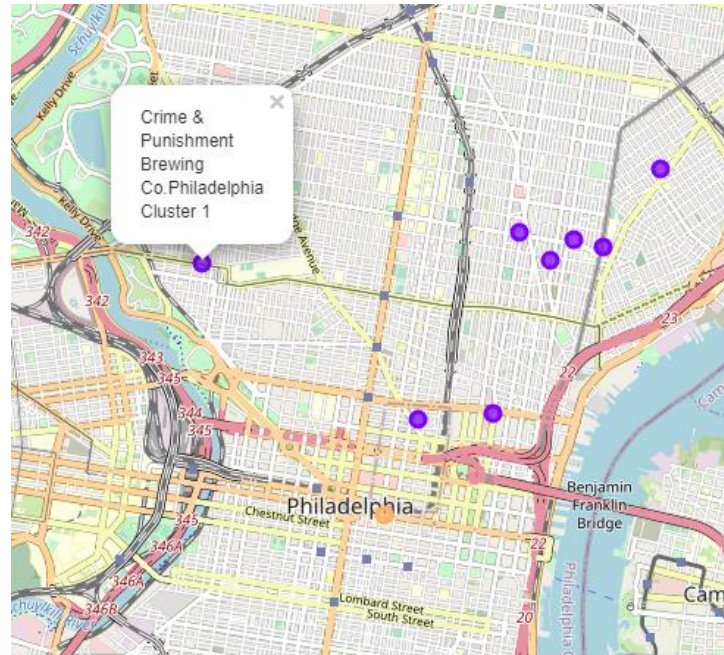


Figure 9. Map of Philadelphia with brewery cluster

Comparing the common venues of breweries in Philadelphia to that of breweries in Los Angeles show in Table 4., it can be inferred that the most common venues are different but not as much to support any underlying pattern. This suggests that other less common (4th, 5th.and so on) venues around the breweries may also be impacting the status of breweries.

Table 4. Breweries in Los Angeles with most common venues

	Brewery	Likes	Cluster Labels	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
10	Karl Strauss Brewing Company	129	4.0	Los Angeles	Theme Park Ride / Attraction	Theme Park	Souvenir Shop
11	MacLeod Ale Brewing Co.	74	4.0	Los Angeles	Rental Car Location	Food & Drink Shop	American Restaurant
12	Broxton Brew	15	4.0	Los Angeles	Coffee Shop	Pizza Place	Mediterranean Restaurant
13	The Stalking Horse	35	4.0	Los Angeles	American Restaurant	Sandwich Place	Diner
14	6th & La Brea	19	4.0	Los Angeles	Coffee Shop	Yoga Studio	French Restaurant
15	The Surly Goat	158	4.0	Los Angeles	Coffee Shop	Lounge	Art Gallery
16	BJ's Restaurant & Brewhouse	45	4.0	Los Angeles	Coffee Shop	Pizza Place	Grocery Store
17	The Village Idiot	304	4.0	Los Angeles	Café	Clothing Store	Boutique

The bottom 20 breweries based on 'likes' have the venues distribution shown in Figure 10. The venues around these breweries are less similar to each other indicating that these breweries are not located in dense commercial areas such as downtown, where it is more likely to have more venues of same category.

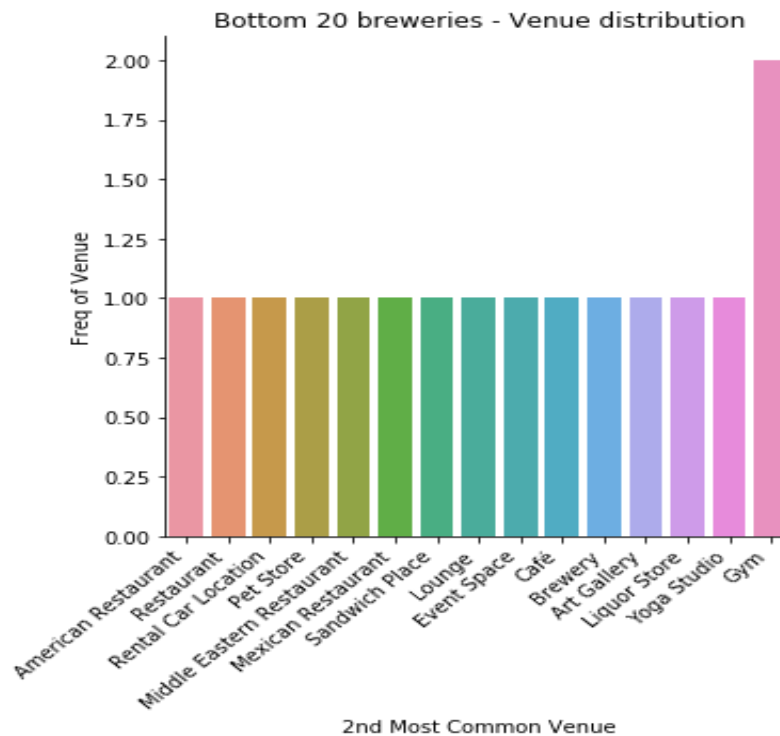


Figure 10. Top venues around bottom 20 breweries

Highly rated breweries were concentrated in Net Your and Boston. Boston, the least populated among the selected cities, has breweries grouped under cluster 4, like Los Angeles.

Cluster 2 shows an interesting pattern with noticeable similarities even across cities. The clustering analysis was able to identify similarities and group these breweries together. The most common venue for most of Cluster 2 breweries is Art Gallery. This is a useful and actionable insight. Table 5. shows a partial table with breweries and most common venues. Cluster 2 breweries are mostly concentrated in Miami, which suggests the popularity of the city might be influencing the rating of breweries.

Table 5. Cluster 2 breweries with most common venues concentrated in Miami

	Latitude	Cluster Labels	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
21	41.8373	2.0	Chicago	Art Gallery	History Museum	Food Truck
30	25.7839	2.0	Miami	Bar	Art Gallery	Ice Cream Shop
31	25.7839	2.0	Miami	Art Gallery	Bar	Ice Cream Shop
33	25.7839	2.0	Miami	Art Gallery	Bar	Ice Cream Shop
36	25.7839	2.0	Miami	Art Gallery	Ice Cream Shop	Brewery
39	25.7839	2.0	Miami	Art Gallery	Bar	Ice Cream Shop
45	32.7936	2.0	Dallas	Art Gallery	Furniture / Home Store	Yoga Studio
49	32.7936	2.0	Dallas	Art Gallery	Antique Shop	Fast Food Restaurant

Cluster 0 and Cluster 3 have the least number of breweries. As shown in Table 6, the venues around these breweries are less conventional. The ratings of the breweries are quite low indicating that these breweries are not located in areas that are quite different from the areas around highly rated breweries.

Table 6. Cluster 0 and Cluster 3

	Brewery	Likes	Cluster Labels	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
48	Four Corners Brewing Company	22	0.0	Dallas	Food Court	Food Truck	Museum
95	Backlash Beer Co.	14	0.0	Boston	Museum	Rental Car Location	BBQ Joint
	Brewery	Likes	Cluster Labels	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
86	Scofflaw Brewing Co.	70	3.0	Atlanta	Gym	Miscellaneous Shop	Juice Bar

5. Conclusion:

A set of breweries in top ten most populated cities were analyzed to supplement the research on the impact of venues in the vicinity of breweries on the popularity and success of breweries. 'Likes' data provided by Foursquare was used as an indicator of popularity and business profitability of breweries. Clustering analysis was carried to glean insights into the similarities among top rated bars. The analysis shows breweries with venues such as Art Galleries and bars receive higher rating. The analysis did not show any bearing of city population on brewery rating. Qualitatively, breweries in cities that attract more tourists had higher number of 'likes'. This could be due to the total number of people visiting the breweries. The most common venues in the vicinity of top breweries were bars, cafés, art galleries, coffee shops, and restaurants. The common venues in the vicinity of breweries with lower rating were sparse. These venues include Gyms, yoga studio and other non-food and drink venues.

6. Future work:

The analysis can be easily scaled to include more cities to gain better insight about features of cities that influence the rating. Other variables such as demographics, income, and age can be included in the analysis. Since the above analysis hinges on one source – Foursquare, more sources can be combined to expand the study of location features that impact a brewery business.

Acknowledgements:

Data sources: Foursquare, Simplemaps.com

Python Packages: Folium, scikit-learn, seaborn, matplotlib, requests