Identifying Potential Locations for New Craft Breweries in Denver, Colorado

Stephen Shirley

1/12/2021

1. Introduction and Business Problem

According to the Denver Microbrew Tour, the City of Denver has 92 brewery venues as of early 2020. In a city spanning a modest area of 401 km^2, there exists approximately one brewery for every 4 km^2. Home to North America's annual Great American Beer Festival and listed as the number one destination for craft beer tourism on Wikipedia, it appears Denver's local economy is willing to accommodate for such a high density of craft beer venues.

This project leverages venue geolocation and a machine learning clustering algorithm to help small business entrepreneurs identify which neighborhoods in Denver, Colorado have the greatest potential for being the home of the city's next craft brewery.

2. Data

The data used for this project are a list of 51 neighborhoods in Denver, Colorado, obtained by scraping the "Neighborhoods in Denver" category from Wikipedia using the BeautifulSoup python package. Once the list of neighborhoods was read into a data table, each neighborhoods' geographic information was georeferenced using ESRI's ArcGIS geocoder. Then, using Foursquare's API, a search request returned the top 100 venues within 4 km of each neighborhood's centroid. Note that a radius of 4 km was chosen based on the 4 km^2 density of Denver breweries. The dataset returned by Foursquare containing the top 100 venues was then sorted by neighborhood and venue type, and lastly the top 10 most occurring venues per neighborhood were selected. A total of 33 neighborhoods out of 51 total had a brewery listed as one of the top 10 most occurring venues.
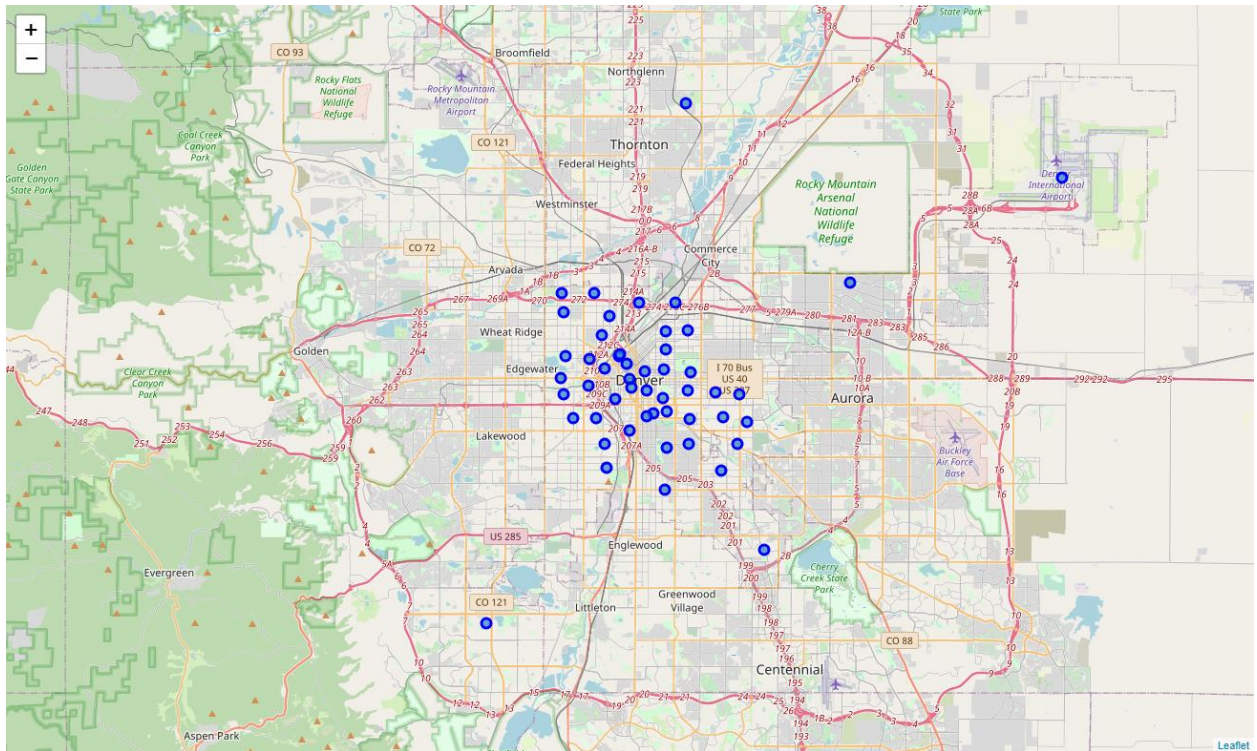
Map 1. Dots representing the 51 neighborhoods in Denver, CO.
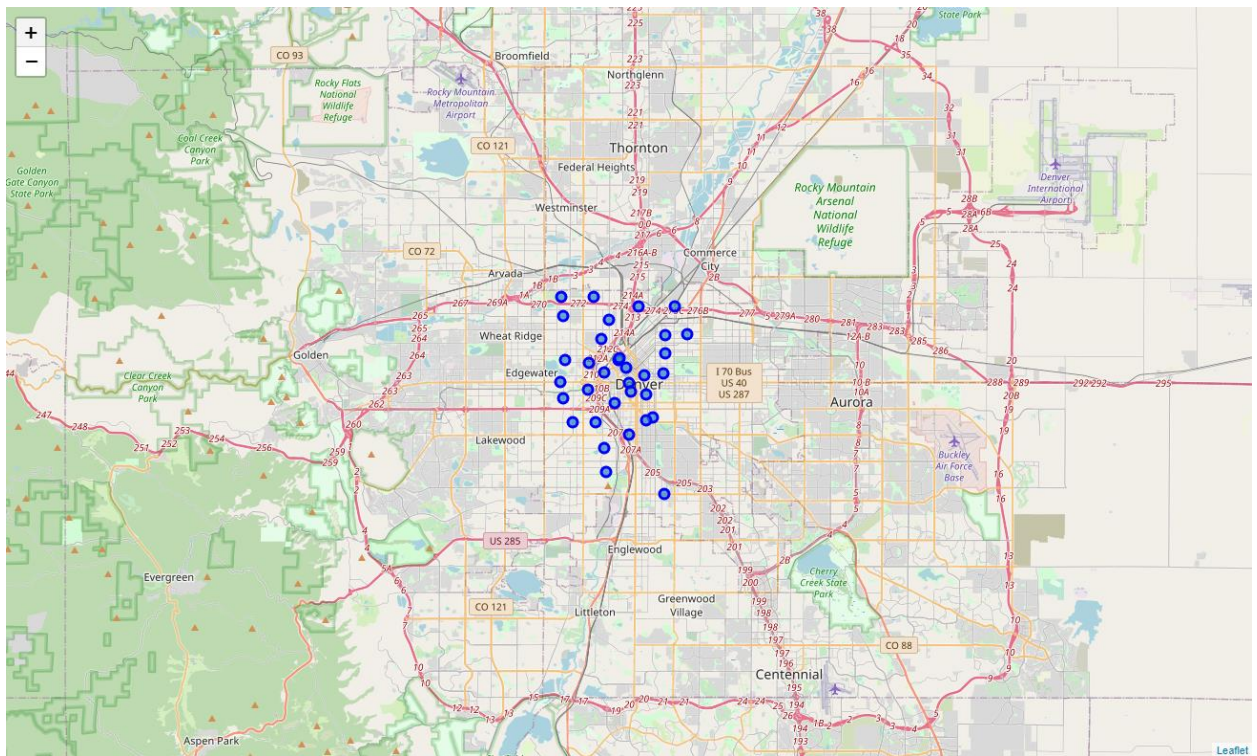
Map 2. Dots representing the 33 neighborhoods in Denver, CO with a brewery listed as one of the top 10 venues based on occurrence.

3. Methodology

To identify neighborhoods estimated to be a good potential host for a new brewery site the machine learning algorithm k-means clustering from the scikit-learn python package was used. Given a user defined number of clusters to group the dataset, k-means clustering attempts to group input variables by mean distance from data centroids to minimize inertia. In order to avoid overfitting or underfitting of the data by grouping the dataset into too few or too many clusters, k-means clustering was run multiple times using a minimum of 3 clusters and a maximum of 20 clusters. Data for each run of k-means iteration was then output to a table in order to identify which cluster contained the most breweries, the total number of neighborhoods per cluster, and the number of neighborhoods. This information helped to identify the most desired outcome of the k-means model run.

N Clusters	Mode Cluster	Total Neighborhoods	Neighborhoods Without Breweries
3	2	18	0
4	3	13	1
5	4	12	1
6	1	12	0
7	2	11	0
8	4	10	0
9	3	6	0
10	4	7	0
11	2	7	0
12	2	7	0
13	2	7	0
14	2	7	0
15	2	7	0
16	0	7	0
17	3	6	0
18	3	6	0
19	0	7	0
20	2	5	0

Table 1. Output generated by k-means clustering using a minimum of 3 clusters and a maximum of 20 clusters.

4. Results

The k-means clustering algorithm with clusters ranging between a minimum of 3 and maximum number of 20 returned between 18 and 5 total neighborhoods, respectively, and each being the maximum or minimum number of total neighborhoods returned. K-means clustering outputs using 4 and 5 clusters where the only outputs to return a neighborhood that did not have

brewery venue as a top ten most occurring venue. Given that the k-means run with 4 clusters contained 13 neighborhoods, one more than the run with 5 clusters, this output was chosen as the desired model outcome given the larger sample size. The one neighborhood identified as no having a brewery as a top ten most occurring venue was Denver's Central Park neighborhood. It is important to note that the one neighborhood without a top ten most common brewery venue in the k-means clusters = 5 output was not Central Park, but instead the Montbello neighborhood.
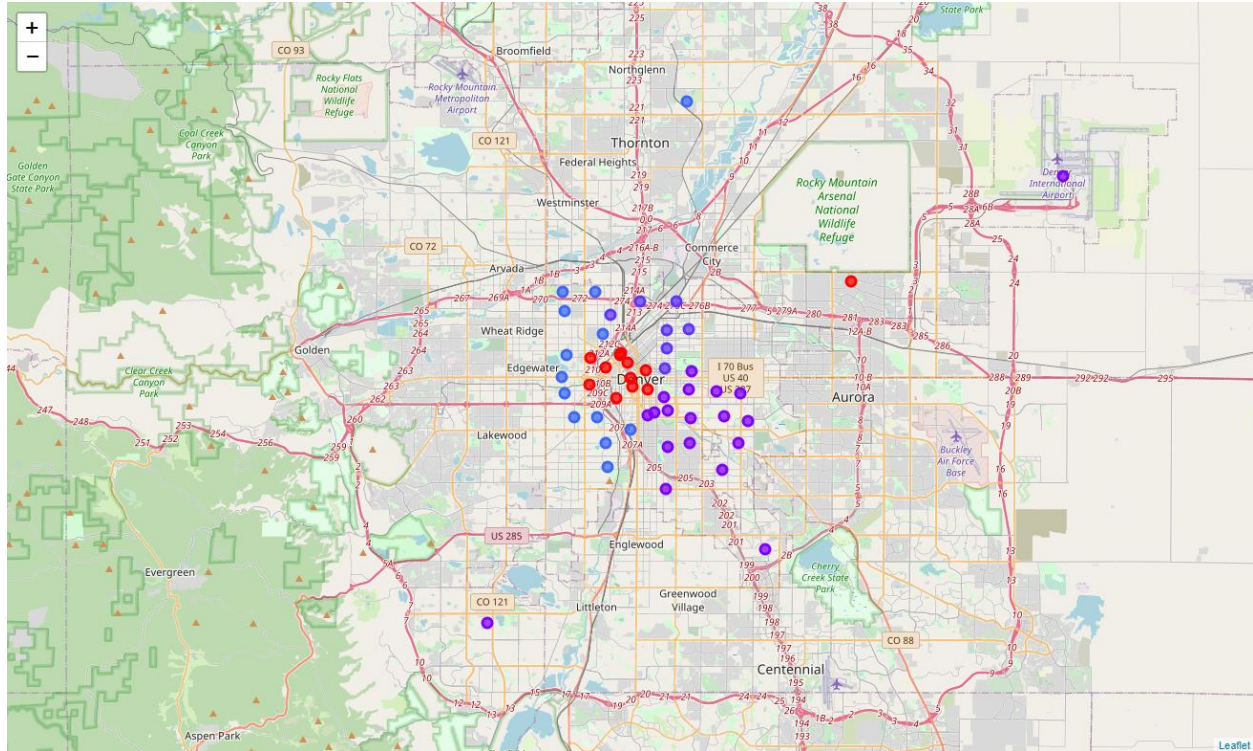
Map 3. K-means clusters = 4 outputs. Note the four clusters, the blue cluster represents the chosen brewery cluster.
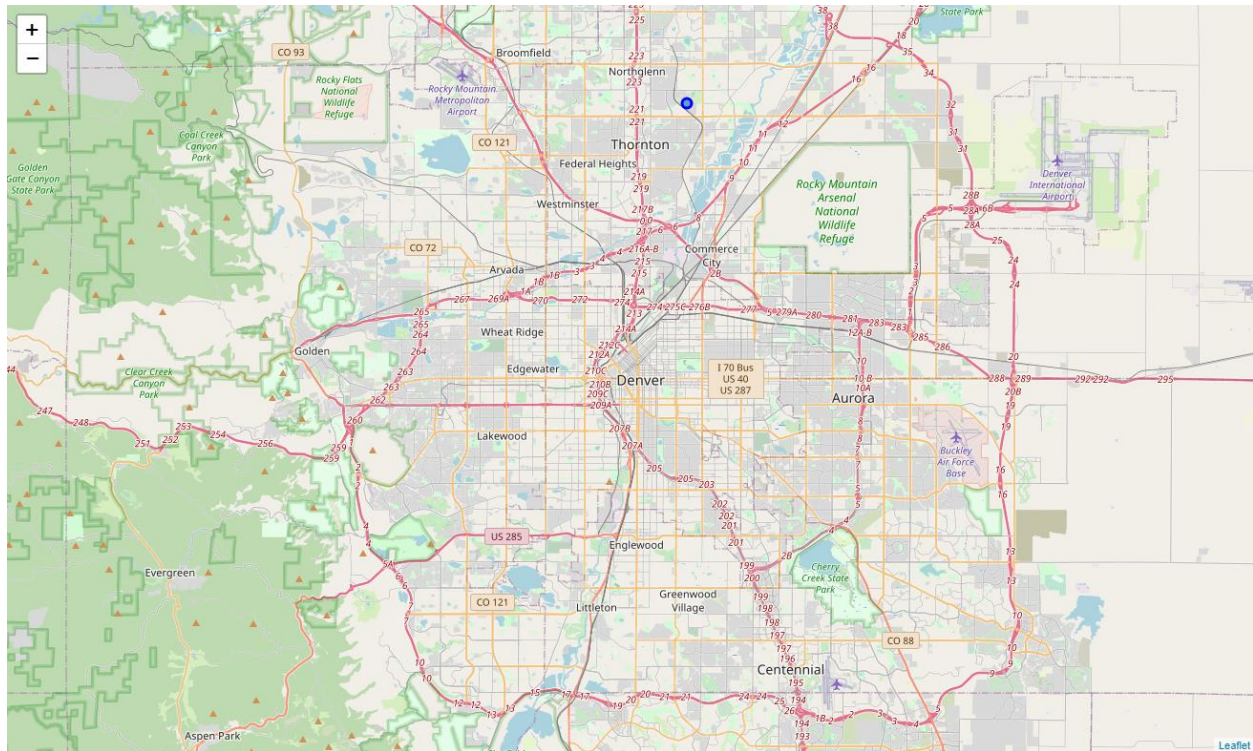
Map 4. The lone blue dot representing Denver's Central Park neighborhood. Estimated to be the optimum neighborhood for a new brewery.

5. Discussion

This assignment felt pressuring in the sense that k-means clustering was the appropriate model to use for the situation and that Foursquare venue occurrence was an acceptable variable identifying neighborhood characteristics, however, if given more freedom a classification model and no grouping of venues could potentially be a better and more reliable predictor. Additionally, more research into the problem could be given, and sources from said research cited to bolster the legitimacy of the project. There is also a potential for the chosen results for this project to have an equally successful chance as the k-means clusters = 5 outcome, and it is equally important to note that on a given day the results from the Foursquare search could change, completely throwing off the results in this report.

6. Conclusion

Using a k-means clustering machine learning algorithm this project identified, using the optimum model run, the Central Park neighborhood in Denver, Colorado as the most suitable neighborhood for stakeholders to open a new brewery based on its similarity to other

neighborhoods with brewery being listed in the top 10 most occurring venues. The Montbello neighborhood could also be a potential site but holds less confidence. It is important to note that the results of this model change frequently based on search results from the Foursquare API. It is recommended that future projects should search for other, possibly more reliable, means of neighborhood classification in order to identify potential venue sites.