1. Introduction

Finding a location for a new restaurant can be difficult. There are many factors that influence this decision, including the cost of rent, restaurant competition or saturation, the number of potential nearby patrons, and nearby amenities. The many factors can be daunting for the new restaurant owner to ponder, but given one key requirement, local data can aid in narrowing down the potential areas.

Here we consider a restaurant owner whose cuisine complements nicely with flavorful beers. She would like to place her business near 'good' local breweries in Downtown Denver with high traffic so she also might get overflow business. The problem here is to find a way to define 'good' with available data.

2. Data

Foursquare API results are leveraged to search for breweries within a certain radius of a location in Downtown Denver. After making the identification of these breweries, additional information from Foursquare, including ratings, the number of ratings, the number of pictures taken, the number of Foursquare 'tips', and the number of 'listings' are pulled from the database.

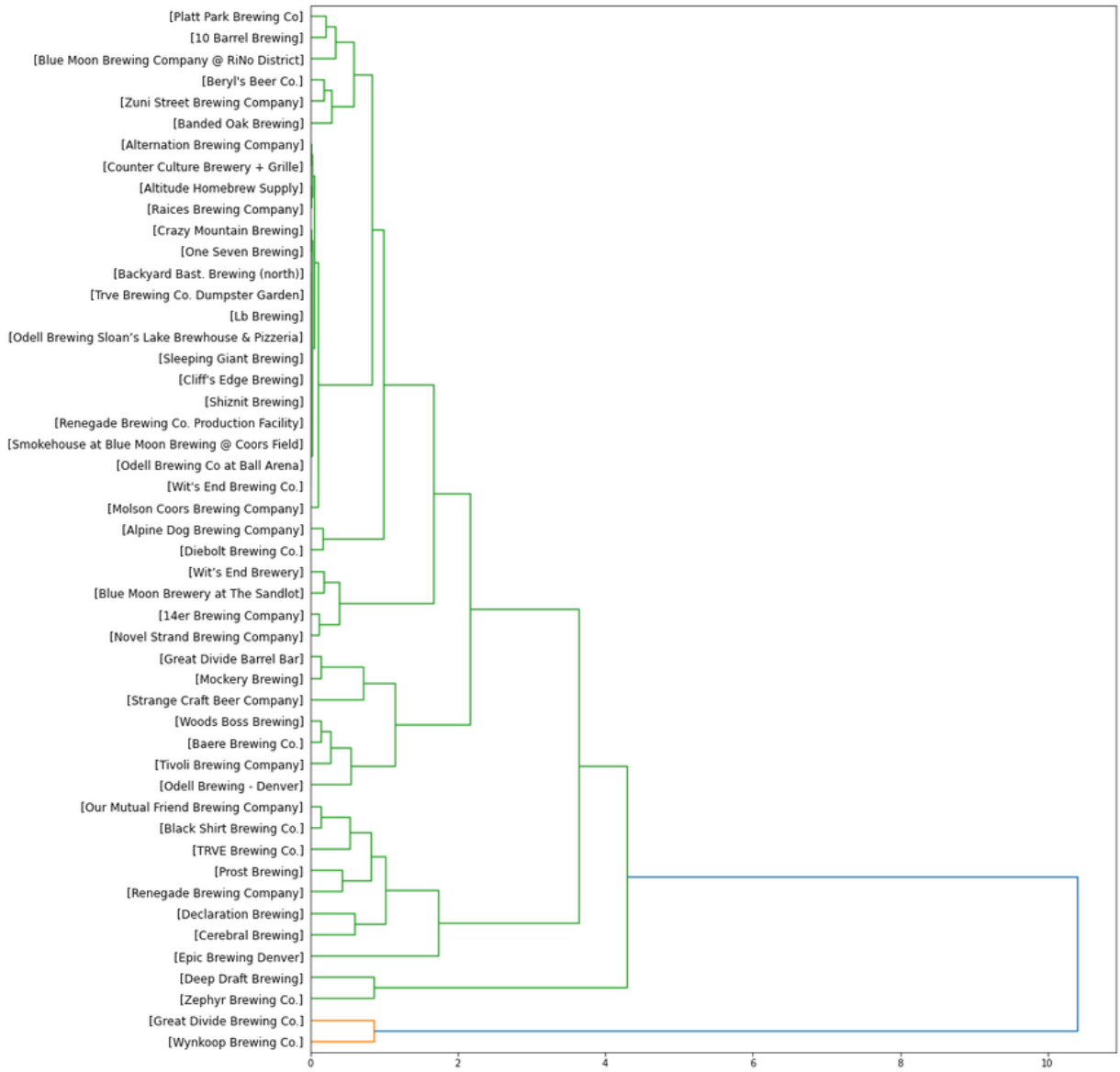
As with all datasets, a bit of cleaning must be applied in order to use the data in statistical methods. One of the main elements that is needed for the analysis is 'ratings', however not all breweries returned in the API call contain that information. In order to retain this, the mean of available ratings (8.226 out of a scale of 10.) was substituted for missing ratings. The number of ratings for these cases was set to zero so that a rating count wouldn't influence all of the results.

3. Methodology

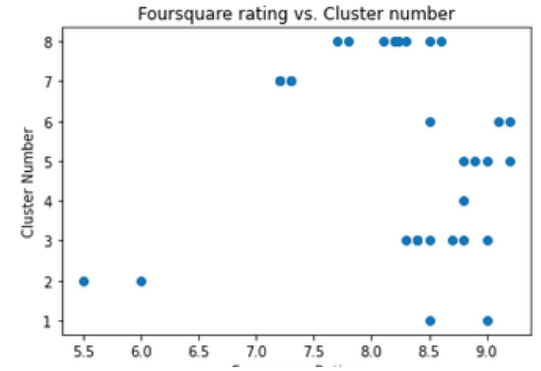
A starting location of latitude 39.731 N and longitude of 104.988 W was chosen as a central location. The search radius is 5000 m, with a search query of 'brewing'. Different lengths of radius were explored, with the final chosen due to the number of results returned by the API call. With those, the initial list included 49 breweries. Producing a map confirmed a variety of potential locations with nearby breweries. Once the data were cleaned, initial exploratory analysis included looking at the mean values of the pulled information. Because the goal is to find a location with the best breweries, the best course of action is a clustering method of the quantifiable information to find particular groups that share traits. From the Python library scipy, the fcluster method was used, with several tests of the criterion and number of clusters to end with the final analysis being a criterion of "maxclust" with the number of clusters of eight.

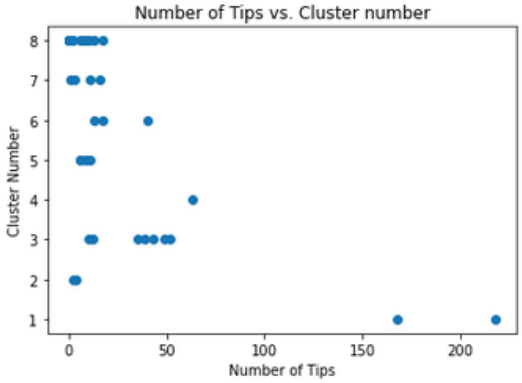
4. Results

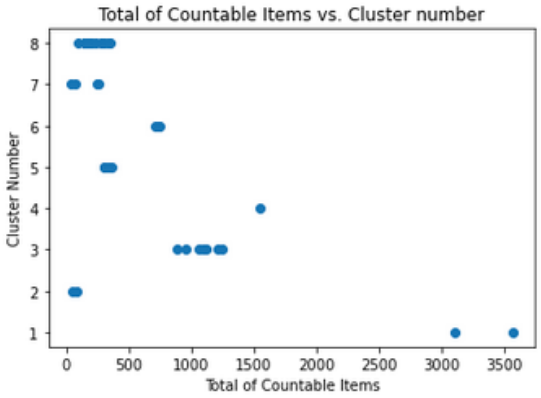
Once the clustering algorithm completed, this dendrogram shows how the clusters are arranged:



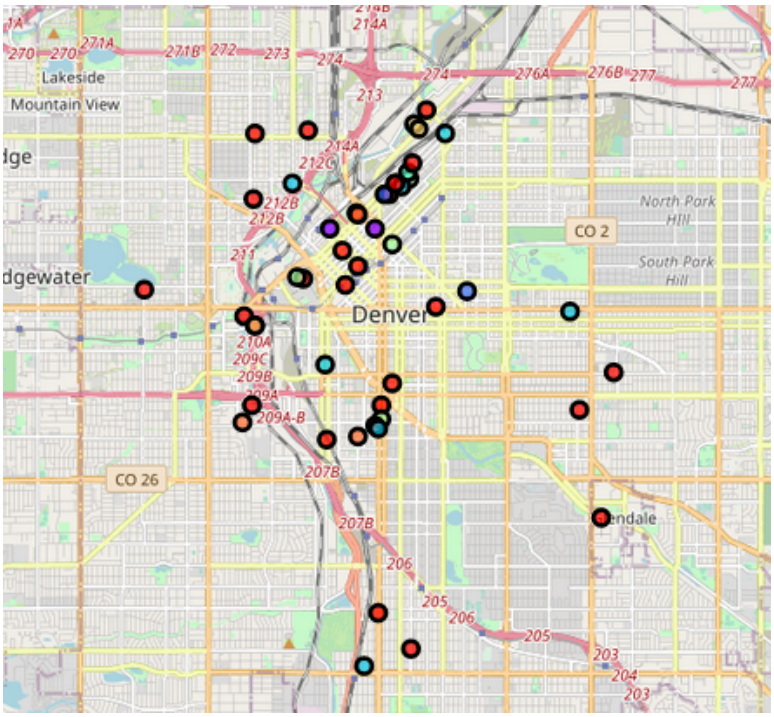
While the dendrogram gives a nice ordering of the clusters, more information is needed to relate how each cluster is related to the input parameters. Scatter plots assist in this analysis. Presented here are a few of these, including the Foursquare rating versus the cluster number and the total number of countable items (the number of tips, photos, ratings and lists) vs. the cluster number.





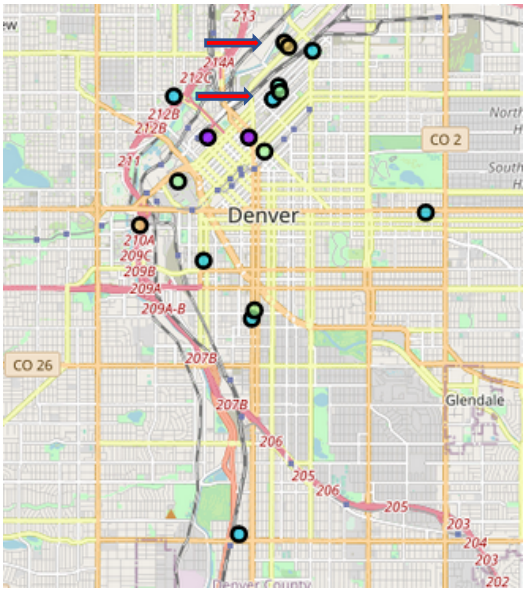


Here is a map of the breweries color-coded by cluster. In the discussion section, the cluster list will be refined for the final area selections.



5. Discussion

From the results, we can see clearly that based on ratings, clusters 2 and 7 should be eliminated from our consideration based on ratings and the countable items. Cluster 8 has too much spread in the ratings, plus few countable items. Cluster 5 is kept despite the relatively low number of countable items because of the higher ratings. With the three clusters being eliminated from consideration, we are left with a list of 16 breweries. The final map is:



There are two areas that immediately jump out with denser number of breweries. The restaurant owner could start a search for a suitable location near either of these.

6. Conclusion

The cluster method has assisted in reducing the possible locations for a restaurant by focusing on the ratings and traffic to each brewery. From the initial list of 49 breweries, the analysis eliminated all but 16. These on a map show two areas to start the hunt for a location to place a restaurant. These methods can be applied to other cities and qualifications.