**Where Denver Demands Yoga**

Parker Servello

1. **Introduction**
   1. **Background**

My friend has confessed to me that he has a dream in which he owns and operates a yoga studio. This project will be the first step in making that dream become a reality. Since we will have went to Denver, Colorado the week after I wrote this, we will work under the assumption that he fell in love with it and wants to settle down there.

* 1. **Problem**

We need to determine the locations in Denver with the high demand for a yoga studio to ensure my friend’s business succeeds. We will analyze other venues that are typically associated with yoga studios and pinpoint the neighborhoods with such venues that have a relative yoga studio deficit. Ideally, there will be at least one neighborhood with the associated venues that does not currently have a yoga studio.

* 1. **Interest**

Real estate professionals have preached “location, location, location,” for as long as their professions have existed. This project presents a simple solution to find that location. Any company planning to open a brick and mortar style business in a new city will want to consider the output of this process when selecting potential neighborhoods.

**2. Data acquisition and cleaning**

**2.1 Data sources**

The list of neighborhoods and the corresponding coordinates were extracted from the “list of neighborhoods in Denver” Wikipedia page along with the links contained in the page. Information relevant to the venues surrounding the neighborhood coordinates came from the foursquare API.

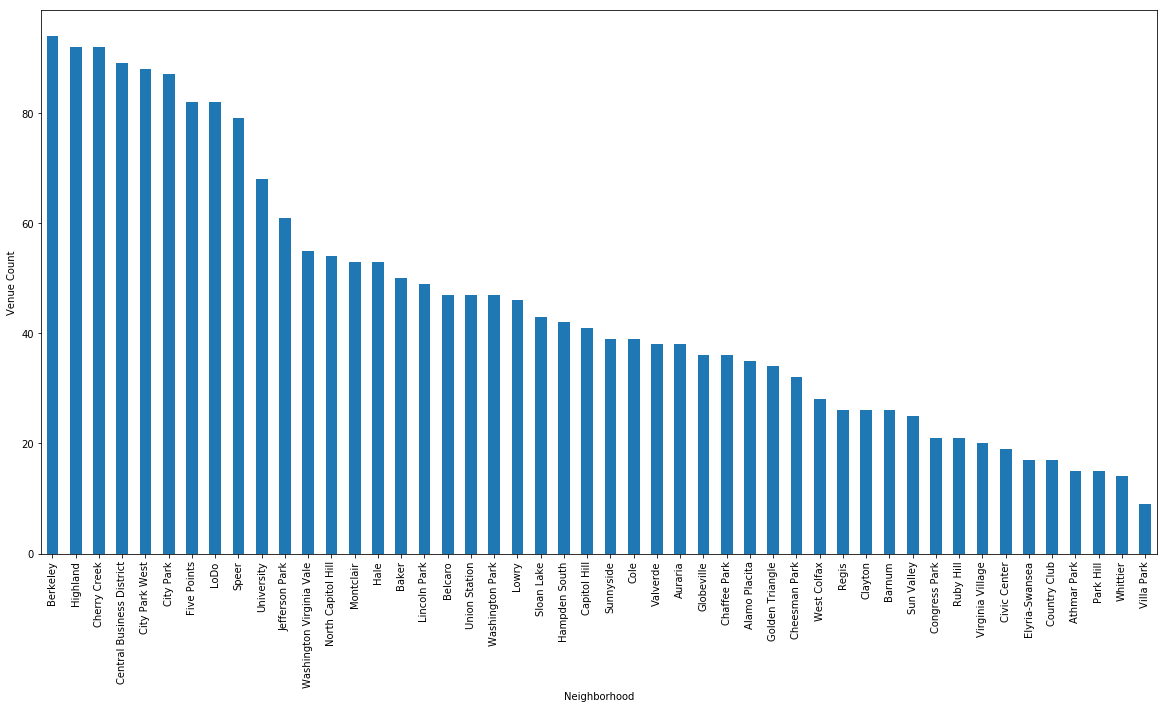
**2.2 Data cleaning**

Neighborhoods without Wikipedia pages, and the airport were removed from the dataset. For each neighborhood, the closest 100 venues within 1,000 meters were selected using the computed Euclidean distance. In the foursquare API, each venue had an associated category. The number of each venue category per neighborhood was computed and standardized to perform the clustering.

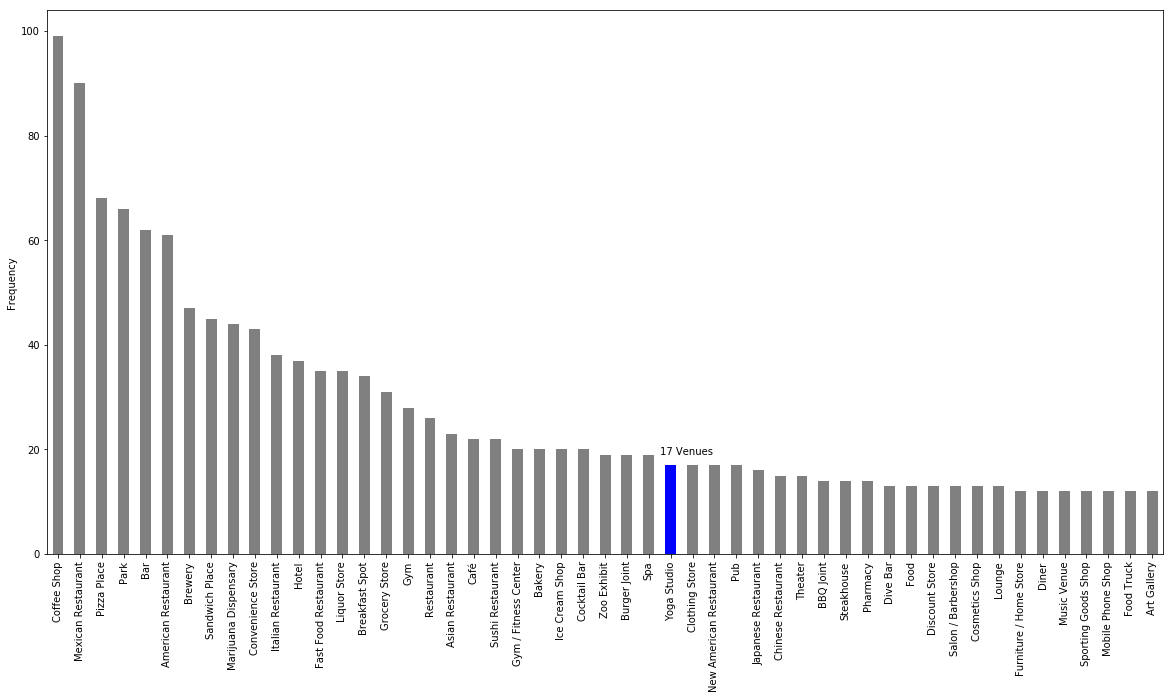
**3. Exploratory Data Analysis**

**3.1 Summary statistics**

Neighborhoods – There are 48 neighborhoods left after the cleaning. The mean number of venues is 45 with a standard deviation of 24.

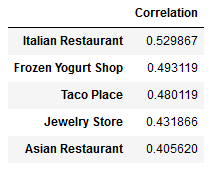
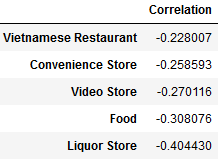


Venues - The most popular category is coffee shop with 99 instances. There are 283 categories and 2,167 total venues. In the graph below, only the categories with at least a dozen venues were included. We also see that there are only 17 yoga studios in Denver.



**3.2 Correlations**

These were the venue categories that were correlated to yoga studios. We’ll check these as a reference for later to validate our clustering. For now, they should still be interesting to observe. For example, some of the results almost seem stereotypical – yogis eat frozen yogurt and drink less liquor. Apparently, they like Italian and tacos as well.

**4. Modeling**

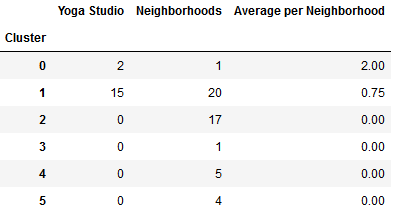
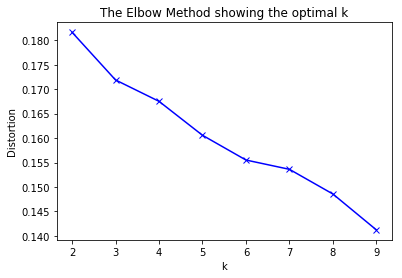
**4.1 Preprocessing**

To start, the venue category counts were converted to proportions of their respective neighborhoods so that neighborhood size would not have an impact.

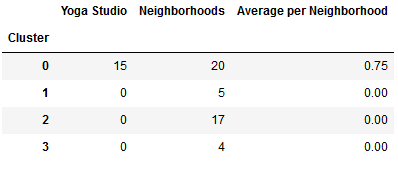
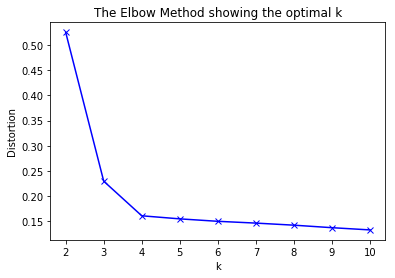
We wanted the yoga studios to influence the results, but we saw that there were only 17 of them. To remedy this, we doubled the proportion for yoga studios. Though I don’t go into further detail about it, this did have a noticeable impact on the formation of the clusters.

**4.2 Clustering**

The first clustering did not yield a definite value for k as indicated by the distortion curve. Furthermore, I settled for six clusters and observed the results. As we see, all the yoga studios were clustered together as desired, but the clusters are not evenly split.

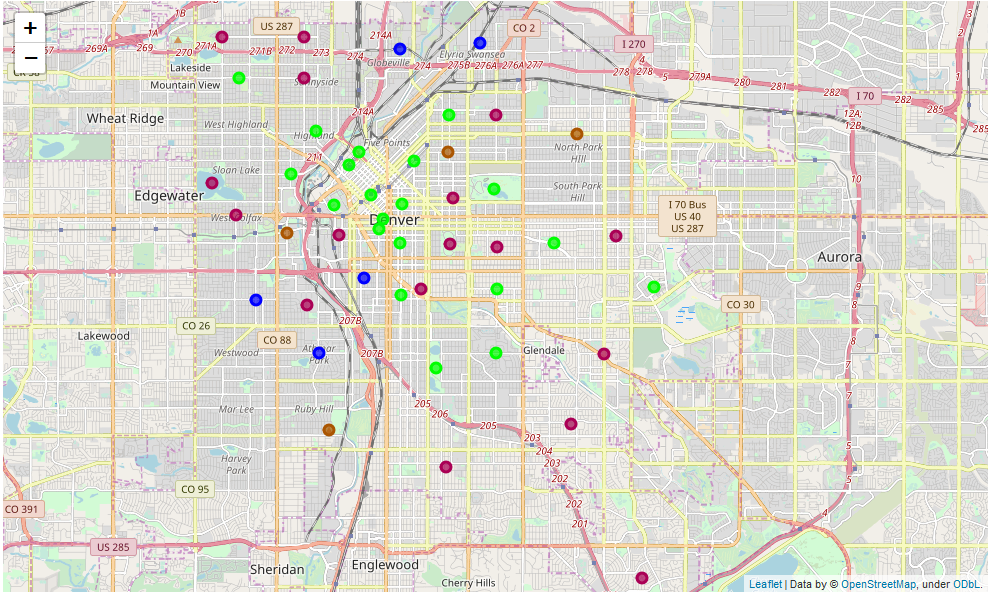


To clean up the results, I removed the neighborhoods whose excessive differences prompted individual clusters. The resulted in a more obvious selection for k.

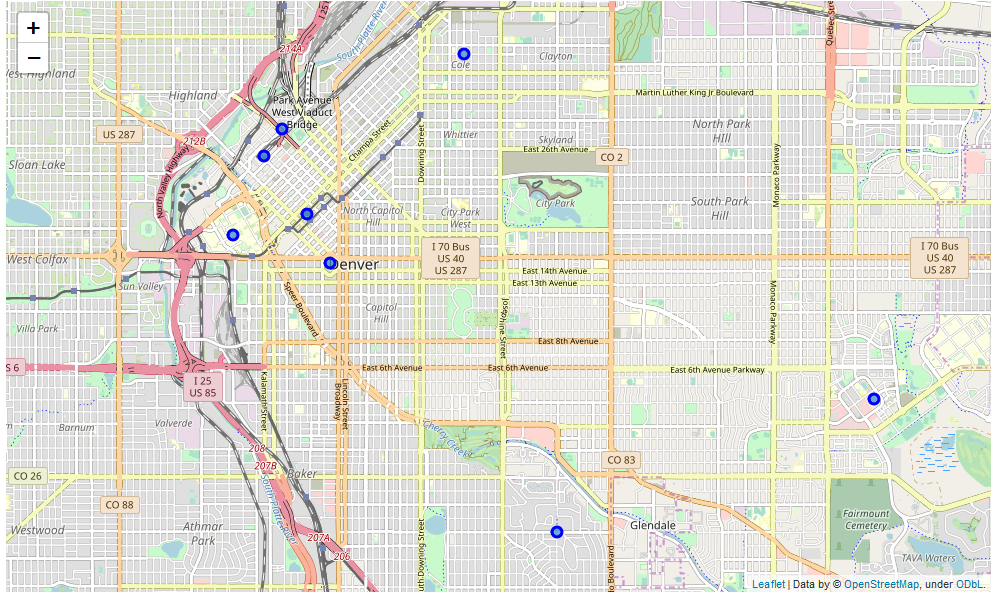


**5. Results**

Ideally, we would have clusters that are closer in size, but at least this clustering weeds out the neighborhoods that are not suitable for a yoga studio. The final clustering can be seen below.



The green circles represent cluster 0 which we see contains all the yoga studios. Let’s take a closer look at the neighborhoods within that cluster that don’t already have a yoga studio in them.



Most of the relevant neighborhoods are in the heart of Denver. This is likely due to the high rent and yoga studios are not suited to bring in the kind of revenue required for downtown real estate. Luckily, we see two locations that are not downtown.

**5. Conclusions**

There are not very many yoga studios compared to other venue types in Denver, so there may not be significant demand. Surveying would help to answer the question as to whether there is demand, but let’s work under the assumption that there is. We see that yoga studios have difficulty surviving downtown even though the downtown neighborhoods are like other neighborhoods that have yoga studios. There are only two neighborhoods, Belcaro and Lowry, that were clustered with the yoga neighborhoods that aren’t downtown as well. The analysis has reduced greatly reduced the number of potential neighborhoods. However, there still may be reasons that these neighborhoods lack yoga studios and we would want to investigate any indication that a yoga studio would not fit in these neighborhoods.