**Selecting a New Rock-climbing Gym Location in Denver**

Shea Dougherty

May 1, 2021

**1 Introduction**

**1.1 Background**

The summertime weather in CO is wonderful. Sunny warm days and crisp cool nights. The mountains and foothills all over the Front Range have their rock faces covered with climbers eagerly searching for the next foothold. But the winters can be long and cold. A solution to the bitter winds and frost preventing an intrepid climber from perching on their favorite peak is to put the rock ledge inside a big warm building.

**1.2 Problem**

Indeed, several indoor rock-climbing gyms can currently be found scattered along the Front Range. But none of them are your rock-climbing gym. We want to show you exactly where your dream gym should be opened. And what neighborhood site will provide the highest return on your investment. Nothing is a sure thing, and all ventures carry risk. We are living in a time when these risks can be calculated and mitigated by powerful technologies and machine learning algorithms. We are going to provide you with an array of possible locations that have the maximum potential of financial success and longevity.

**1.3 Interest**

Deciding on a new location is a long and complicated process that carries a substantial risk of making the wrong choice. Our findings can be an extremely useful guide to help make this difficult process easier.

**2 Data acquisition and cleaning**

**2.1 Data sources**

To facilitate this process, we will be using FourSquare’s vast collection of venue location data. The actions and comments of thousands of users are collected and stored for access by our algorithms. Countless actual reactions to myriad venues, events, and experiences can provide a backdrop of information for us to base our calculations on. Every currently operating venue in the city will be provided for an exact count of gyms, health clubs, yoga studios, and especially other climbing gyms in the vicinities. The likes and preferences of the clients that patronize these venues are also collected and will allow us to provide actionable insights into not only the locations, but also the clientele in those locations and their personal preferences. This information will allow you to open your climbing gym in a location that suits your style, and also allow you to operate in a way that will cater to your local clients. We also use Denver.org as a source for the neighborhood divisions and names.

**2.2 Data cleaning**

Our primary dataset consists of the FourSquare API venue location .json that we transposed into a data frame for easier manipulation. We sourced the geolocations for each neighborhood from the google.maps API into a csv file. This data frame was merged with our primary data frame to provide an exact geocoordinate for each city and its corresponding venue proximity data.

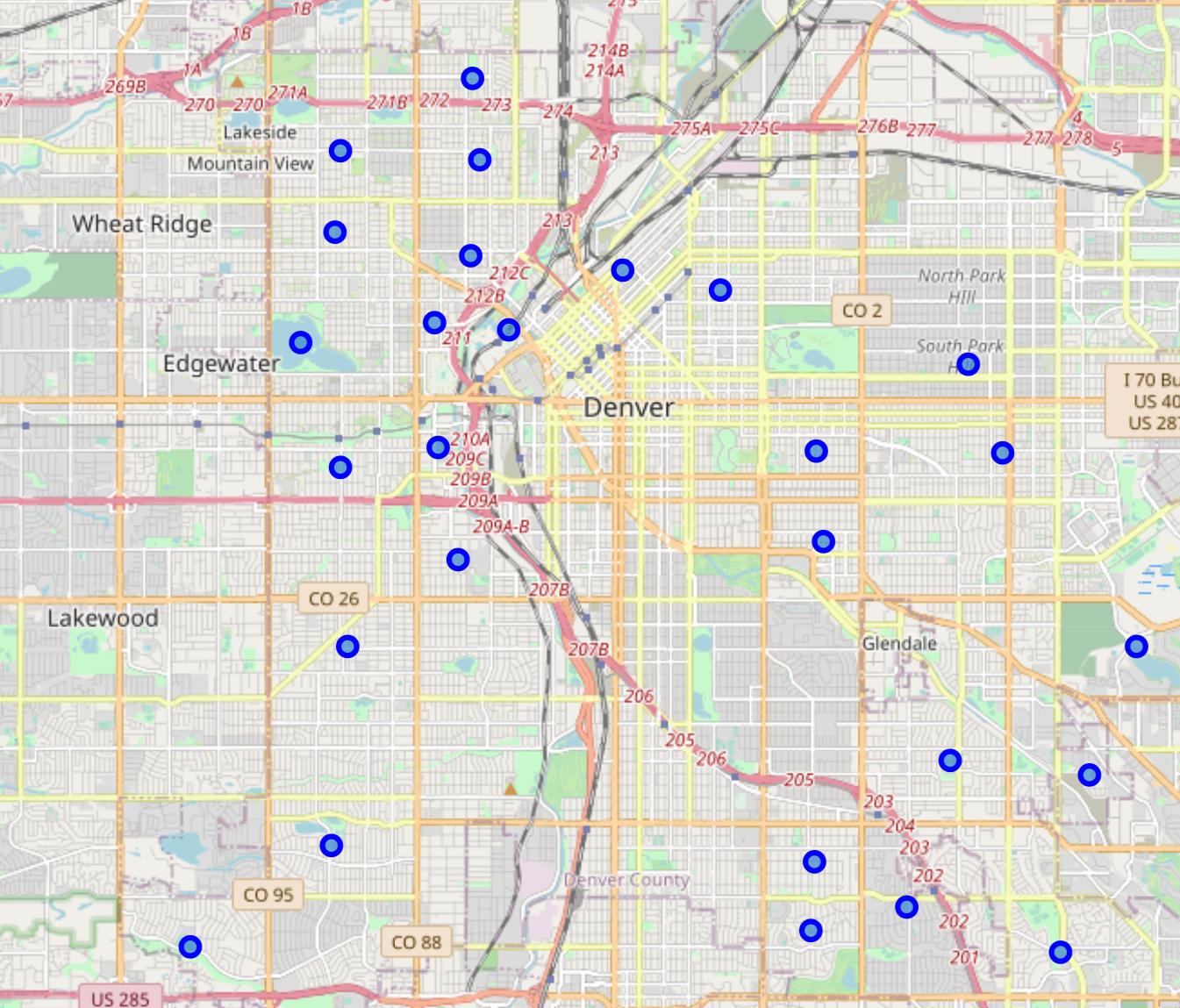
**2.3 Feature selection**

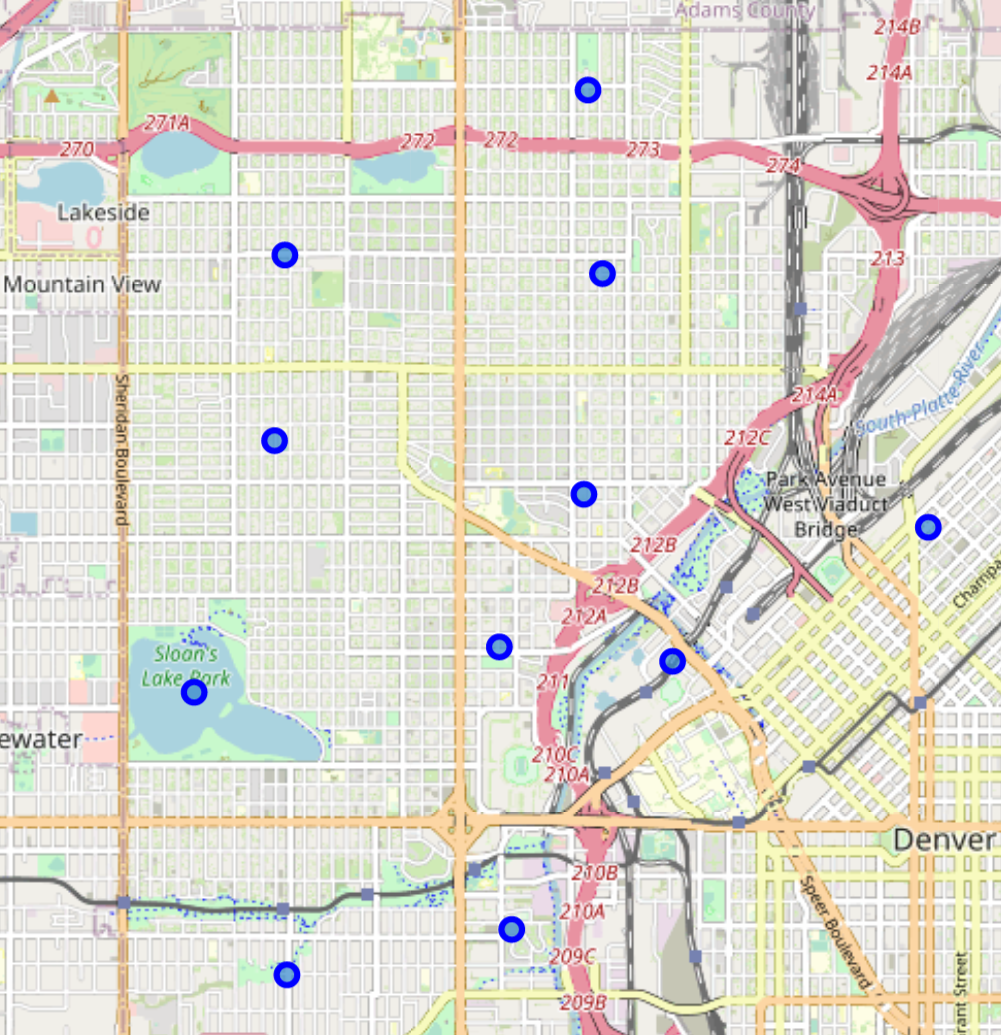
Our initial dataset consisted of 18 boroughs and 261 neighborhoods. Some of these are in such close proximity to one another it was immediately evident that many of said instances could be combined. These combinations reduced our dataset considerably. Based on specifications in building codes and available real estate the dataset was even more substantially truncated. Our final dataset consisted of 6 boroughs with 27 neighborhoods from which we based our analysis and conclusions from.

**3 Methodology**

**3.1 Exploratory data analysis**

We use frequency distribution to analyze the volume of existing venues for each neighborhood that are similar in nature to an indoor rock-climbing gym. We can graph the quantity of every venue type within a set proximity of our selected neighborhood center. This can be visualized with histograms to gain an understanding of the most common venue categories in our area of interest.

A Folium map shows that our neighborhoods are disbursed uniformly around the Denver urban center.

****

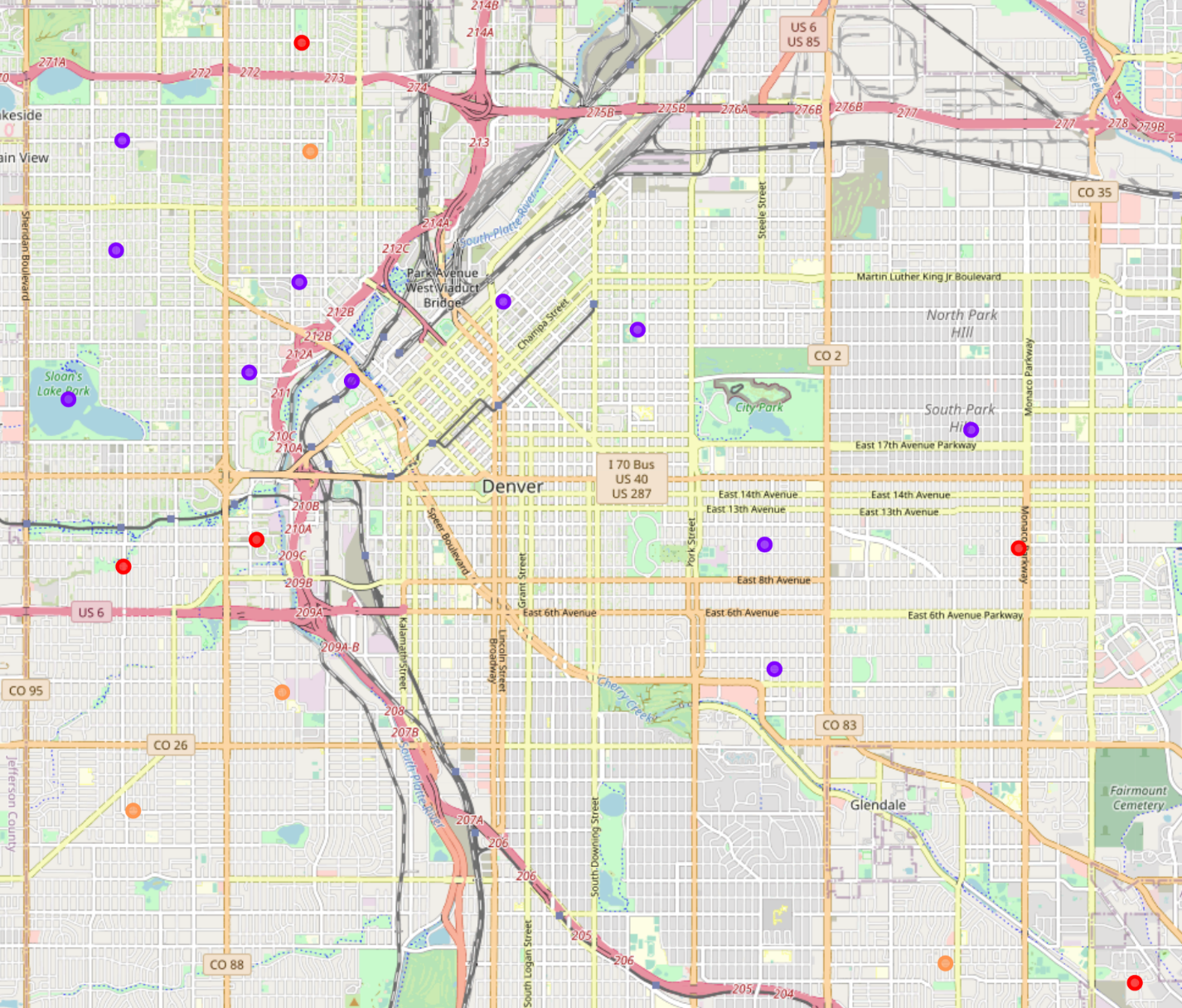
A particularly dense area of prospective neighborhoods resides in the northwest quadrant.

**3.2 Inferential statistical testing**

We use a descriptive model to show the relationships of the neighborhoods and clusters. A mean distributions of venue categories helps to illustrate the primary characteristics of each neighborhood.

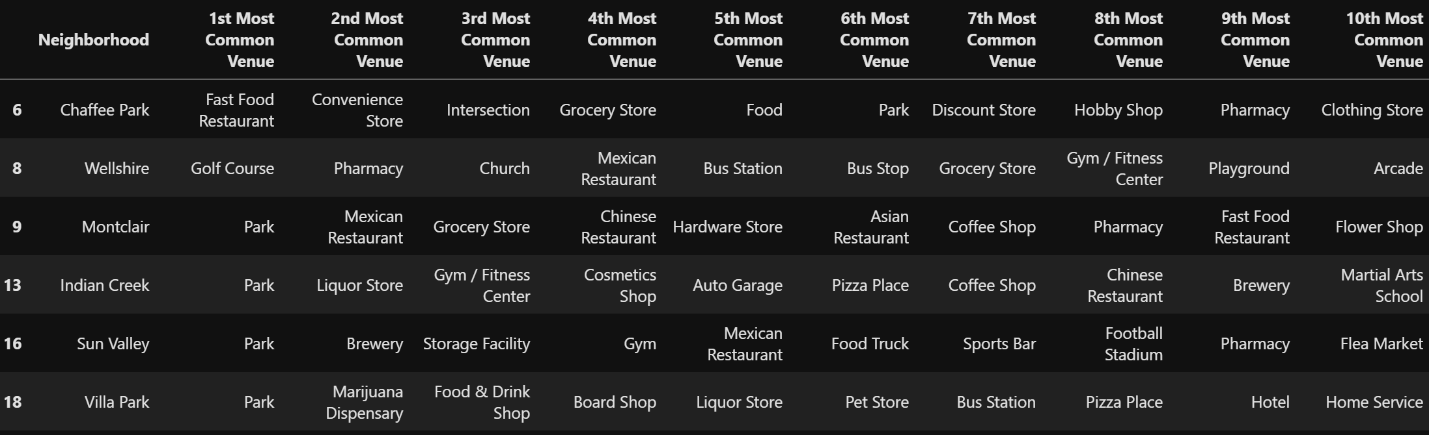
**3.3 Machine learnings**

We utilized k-means clustering to group or neighborhoods into similar categories. Once grouped into clusters we are able to obtain a generalization of the primary characteristics of each cluster.



**4 Results**

The strong similarities among neighborhoods of many instances led them to be clustered into two large cluster groups.



The first cluster consisting of Chaffee Park, Wellshire, Montclair, Indian Creek, Sun Valley, and Villa Park are all very heavily populated with Parks. Four of the six have Park as their most common venue. Wellshire replaces Park with Golf Course, which in many respects is extremely similar to a park. Gyms and coffee shops make a large proportion of the most common venues.



Cluster two consists of fourteen neighborhoods and is the largest cluster. The dominant venue in the most common category is Coffee Shop, followed by Bar/Brewery. Cosmetic Shop, Pizza Place, and Yoga Studio make a significant showing in this cluster’s most common venues.





The remaining three clusters have one neighborhood each. This suggests a very differentiated character to each of these neighborhoods that causes them to be resistant to categorization with other cluster labels. Bear Valley has Park for the most common venue, followed by Video Store, Pool, and Music Store. Cluster four contains Harvey Park with a most common venue of Construction & Landscaping followed by Recreation Center, Lake, and Park. The most common venue for Windsor in cluster five is Gym. Golf Course, American Restaurant, and Lake comprise the remaining most common venues.

**5 Discussion**

Based on the venue frequency we can likely conclude clusters four and five are commercial districts with business or office park zoning. Cluster three data shows it is a residential area. Cluster one shows a strong indication for mostly single-family homes.

The most promising cluster appears to be Two. The most common venue categories seem to indicate a large population of younger professionals are even college students. A domain study will help determine the neighborhoods within this cluster than will provide the maximum compatibility with your decision.

**6 Conclusion**

While many neighborhoods show promising characteristics as potential locations for opening an indoor rock-climbing gym, several stand out as potential candidates. Our preliminary analysis indicates the Five Points neighborhood as the most efficacious location. Final surveys and polls will be conducted to further substantiate data and modeling processes.