

Finding a location for restaurant in Chicago

Jenny J. Housley

June 29, 2020

1. Introduction:

1.1 Background

It is hard to open a new restaurant, especially one that caters to a specific community. Selecting the correct location is a critical step in the initial set up of the restaurant, in order to place it on the path of success. It is of great importance to select a location in order to reach the community the restaurant wants to serve.

1.2 Problem

A client seeks to open a cyclist friendly restaurant, with a emphasis on Mediterranean cuisine in a Chicago neighborhood. The main question, which neighborhood would be best for the restaurant.

The objective of this capstone is to locate the best neighborhood for this niche restaurant. Using Foursquare API, I will be able to determine locations with access to city bike racks, other restaurants and shops within a half mile radius of the bike rack.

1.3 Interest

Others wanting to cater to the growing cyclist community will be interested in finding potential neighborhoods in which business opportunities would be easier to establish.

2. Data acquisition and cleaning

2.1 Data Sources

Using the website Kaggle, a data file was found with locations of bike racks in Chicago. This file included community areas and latitude and longitude of the bike racks. The latitude and longitude of the bike racks were then used in an explore call to FourSquare to get the ten closest venues in a half mile radius of the bike rack.

2.2 Data Cleaning

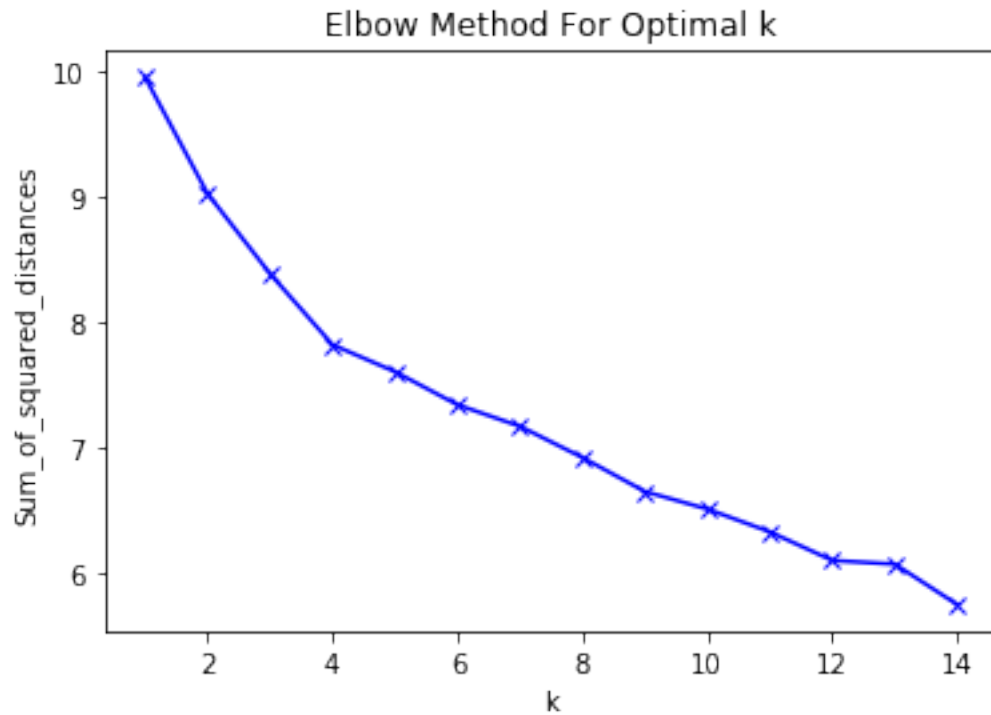
The data from the csv file obtained from Kaggle was very well organized and no cleaning was needed. It was determined that all bike racks had coordinates and a corresponding community area assigned to it. There were extra columns that needed to be removed. The average of the bike racks latitude and longitude was used for the coordinates for each community area. Once the explore function was called using the FourSquare API, it was found that a total of 165 unique venue categorizes were located between the 77 community areas. Once the data was encoded and prepped for k-means clustering, the number of clusters need to be determined.

3. Predictive Modeling

K-Means clustering was used to group the community areas into similar clusters. All the venues found were used as the features in determining the clusters.

3.1 Determining the Number of Clusters

In order to use k-means clusters an initial decision on how many clusters to use is needed. In order to determine the number of clusters, the graph of square error was used to find the K Elbow Visualizer.

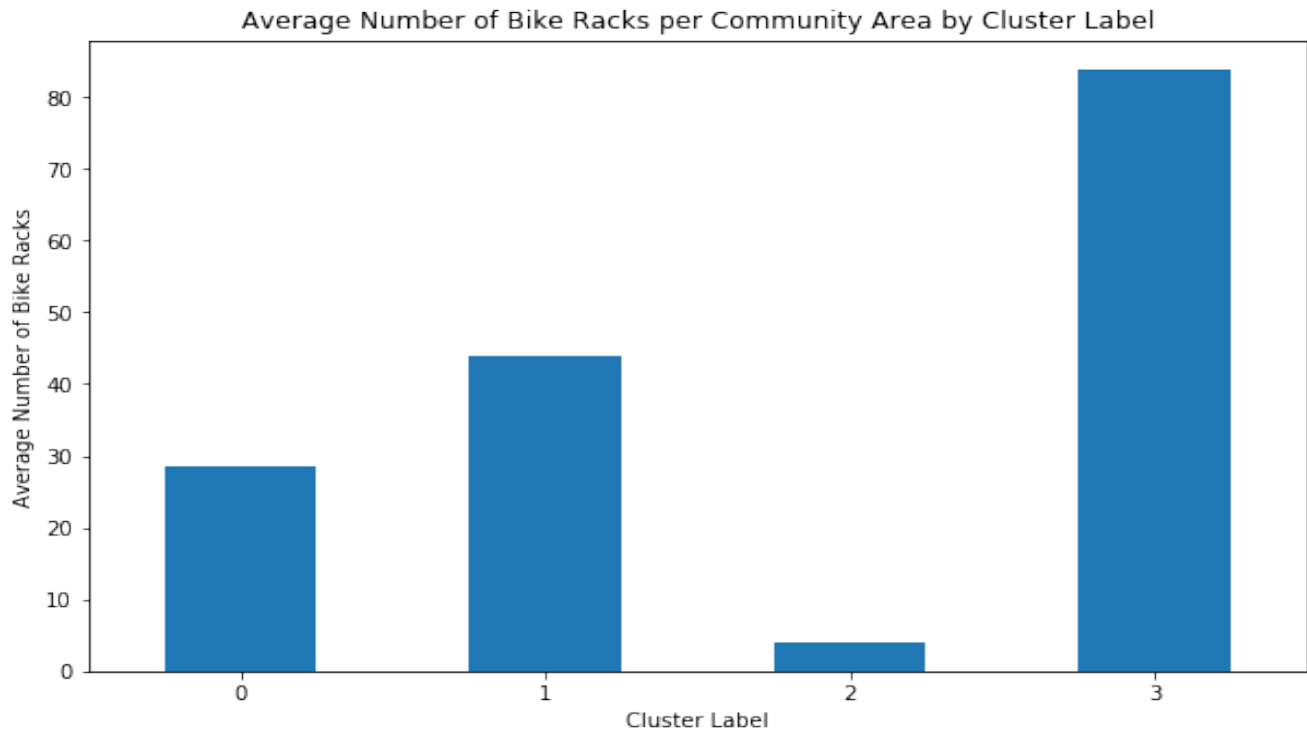


After studying the graph, it was decided to cluster our features into 4 clusters.

4. Exploratory Data Analysis

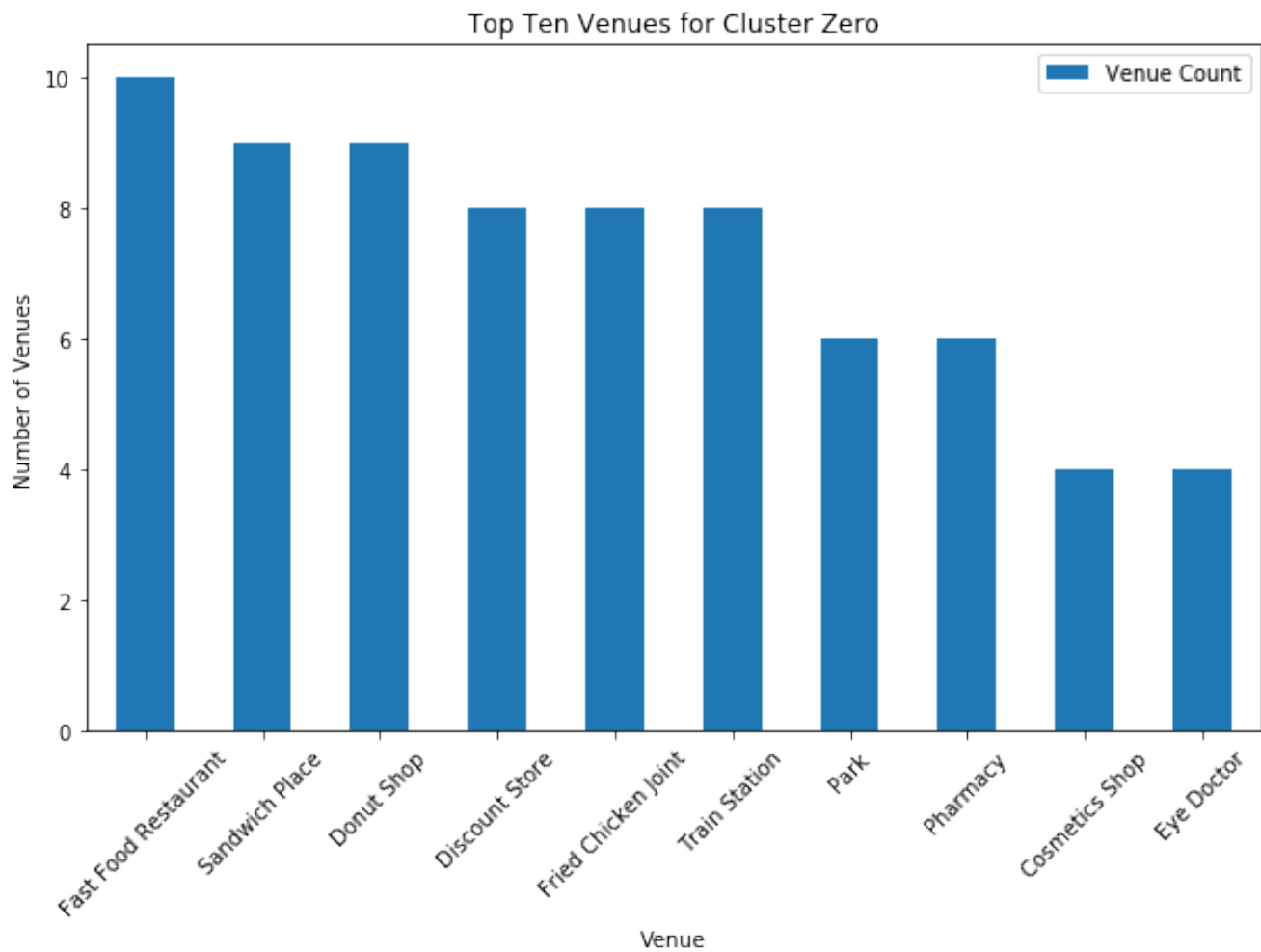
I decided to explore the data after I clustered the community areas to compare and contrast the the different venues for each cluster. Once I computed the clusters, I was able to find valuable information such as average number of bike racks per community area and the top ten venues for each cluster.

4.1 Calculation of average number of Bike Racks

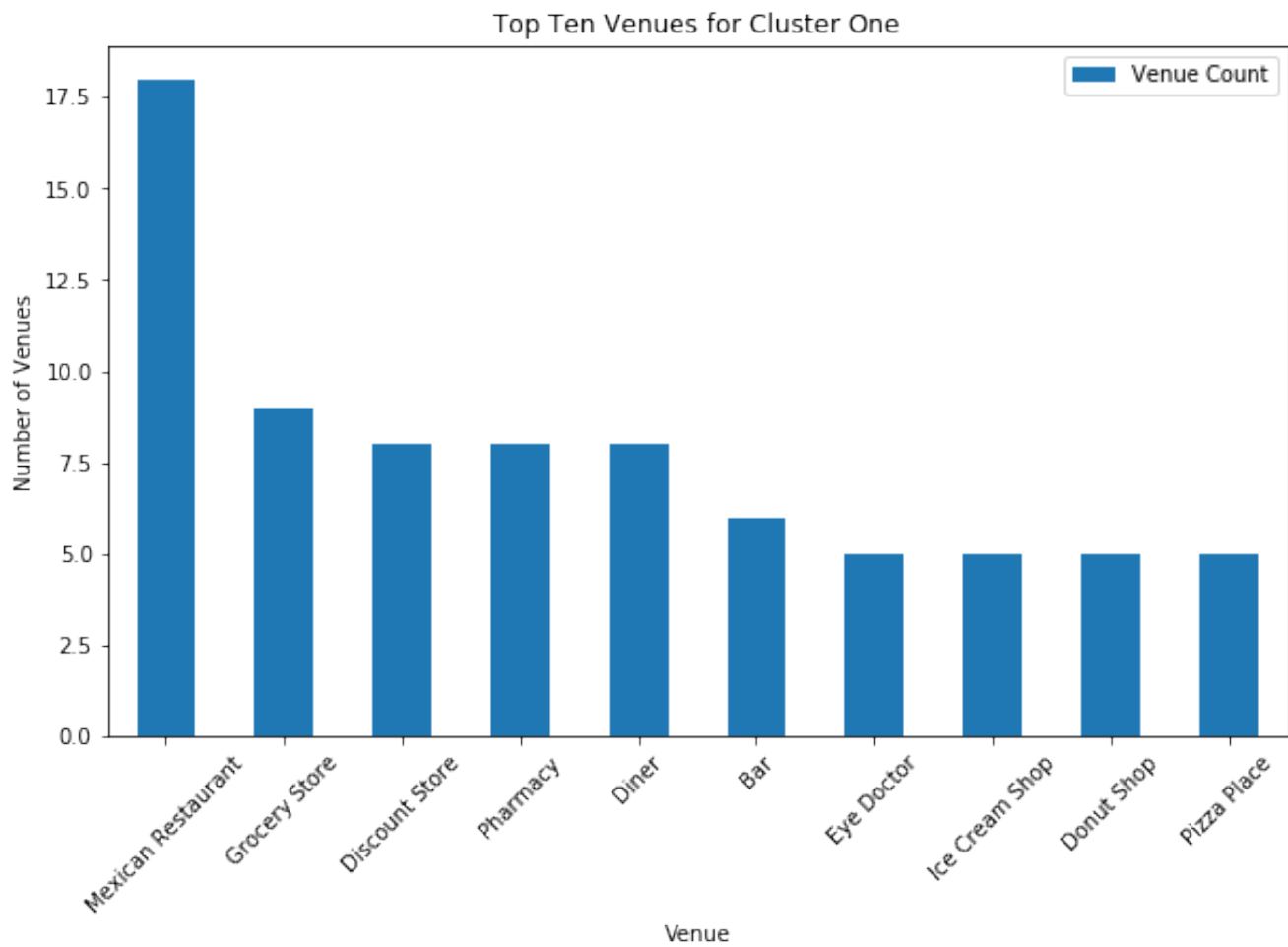


4.2 Top Venues by Cluster

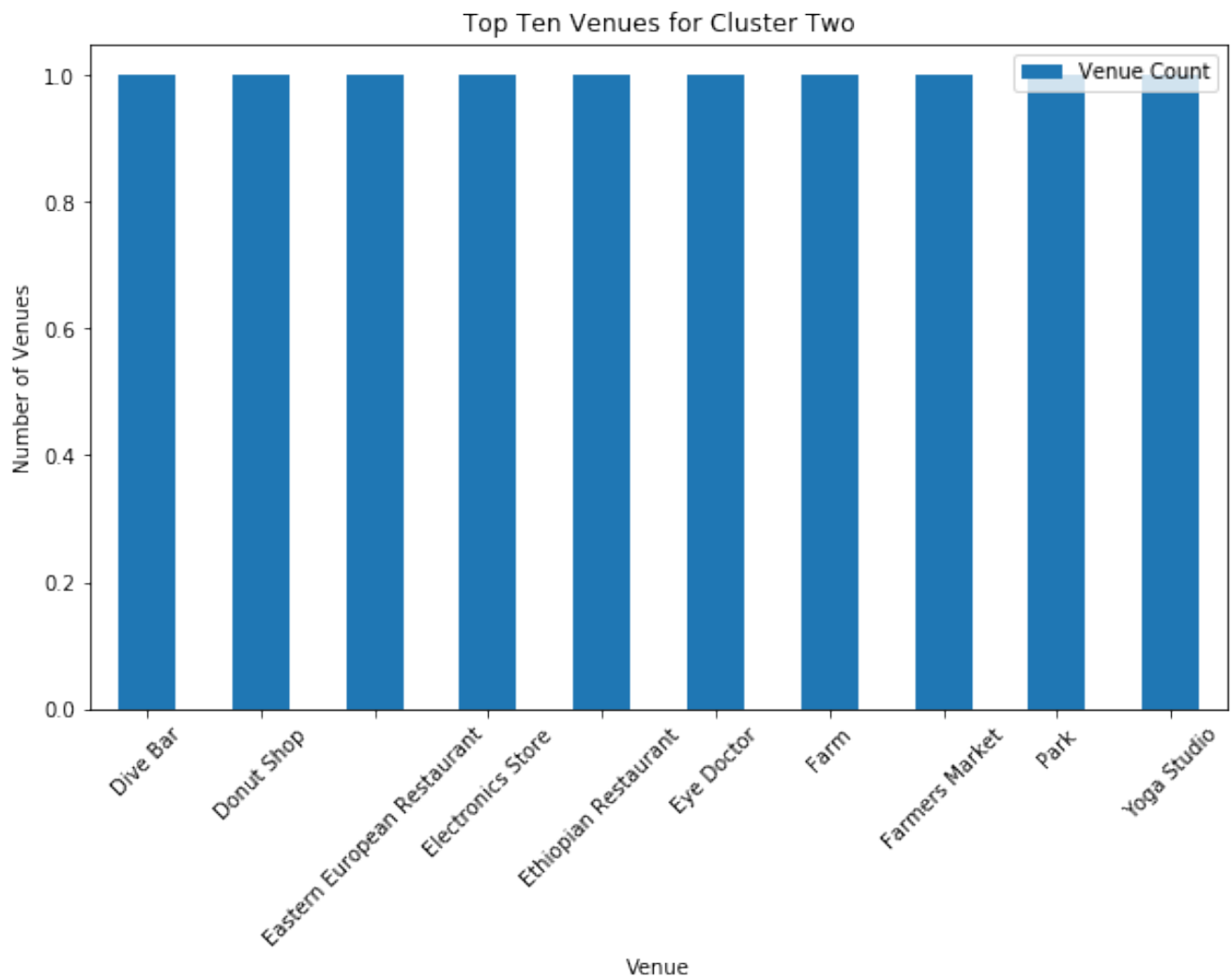
The following graphs are visualizations of the top ten venues with the highest frequency through out the cluster.



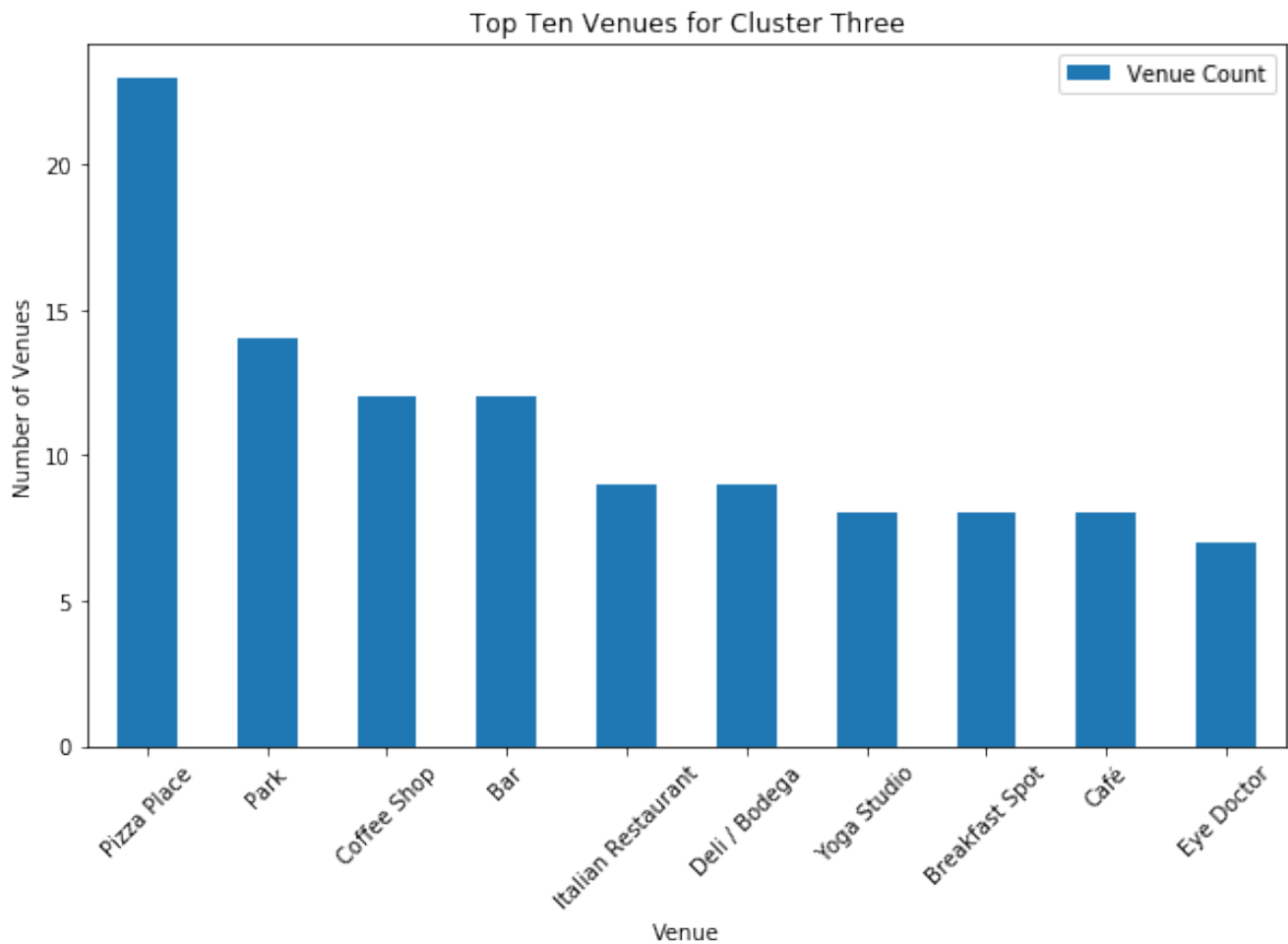
Cluster One seems to have potential for what I am looking for. Four of the top ten venues are restaurants along with a number of parks.



Cluster One also show potential with five of the top venues being restaurant.



While initially cluster two seems to have potential, only one community area has been grouped into this cluster and there are a limited number of bike racks. With this information, this cluster can be dropped from future discussions.



Cluster number three is the most exciting of finds. Seven out of the top venues are restaurants. Some other notable venues are the number of parks and yoga studios in the area.

5. Conclusions

In this study, I clustered community areas in Chicago by the top venues found using FourSquare's API explore call. Once I clustered the areas using K-Means with a cluster grouping of four, I was then able to determine the average number of bike racks per community area by cluster. Also the top venues per cluster were found in order to determine which would be best for opening a restaurant. This data will be very useful for an entrepreneur preparing to open a restaurant that caters to the growing cycling community of Chicago.

6. Suggested Community Areas

The following is a list of community areas in the Chicago area that would be a great location to opening a restaurant wanting to serve the cyclist community.

Armour Square	Ashburn	Beverly	Bridgeport	Brighton Park
Clearing	Dunning	Edgewater	Edison Park	Forest Glen
Garfield Ridge	Grand Boulevard	Greater Grand Crossing	Hegewisch	Humboldt Park
Hyde Park	Irving Park	Jefferson Park	Kenwood	Lake View
Lincoln Park	Lincoln Square	Loop	Mount Greenwood	Near North Side
Near South Side	Near West Side	North Center	North Park	Norwood Park
Oakland	Portage Park	Rogers Park	South Deering	South Shore
Uptown	West Ridge	West Town	Woodlawn	