

# Battle of Neighborhoods Report

Annie Helms

June 4, 2020

## 1. Introduction/Business Problem

For this project, a successful restaurant in Berkeley, CA is interested in opening a new restaurant in Chicago, IL. They believe that most of their success in Berkeley is due to the fact that their neighborhood, Gourmet Ghetto, is located near a university and contains many different types of venues. They would like to find the neighborhood in Chicago that is most similar to their current location. In order to do so, location data from Foursquare will be pulled and used to compare the neighborhoods according to their most popular venues. Additionally, the list will be ordered by distance from the University of Illinois at Chicago. The final deliverable of this report will be an ordered list of the neighborhoods in Chicago that are most similar to the restaurant's current location in the Gourmet Ghetto, Berkeley, CA.

## 2. Data

In order to address the business problem above, Foursquare location data will be used to obtain the top 100 venues within 500 meters of each neighborhood in Chicago and of the Gourmet Ghetto. To obtain a list of the neighborhoods in Chicago, a .csv file will be pulled from [www.data.cityofchicago.org](http://www.data.cityofchicago.org). In order to plot the venues onto maps of each city, coordinate data will also be obtained for each neighborhood with the *geopy* library in Python. Lastly, this same library will be used to obtain the coordinates for the University of Illinois at Chicago.

## 3. Methodology

### 3.1 Gather list of Chicago neighborhoods

The first step in the process was to pull a list of Chicago neighborhoods from the city of Chicago's website. This list was cleaned and the *geopy* library was used to fetch latitudes and longitudes from each neighborhood. Coordinate data was unable to be obtained for three neighborhoods, so these rows were dropped from the data frame. Lastly, some neighborhoods returned coordinate data that was erroneous, so neighborhoods with outlying coordinates (beyond 1 degree of latitude and longitude from the rest of the neighborhoods) were also dropped from the data set.

### 3.2 Map out Chicago neighborhoods

The resulting list of 90 neighborhoods were plotted on a map of Chicago using the *folium* library. Each neighborhood was superimposed on the map and labeled.

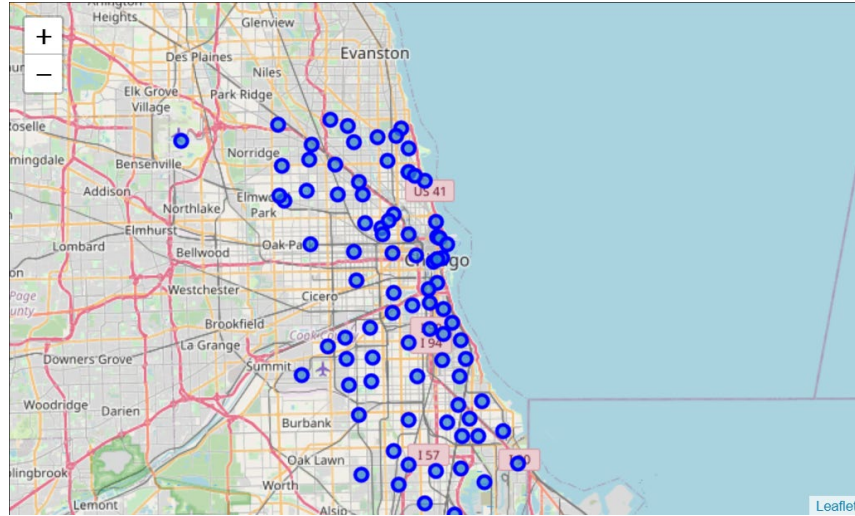


Figure 1. Map of the 90 Chicago neighborhoods.

### 3.3 Add Gourmet Ghetto to neighborhood list

Next, the coordinates for the Gourmet Ghetto neighborhood in Berkeley, CA were fetched using the *geopy* library. The neighborhood name and the corresponding latitude and longitude values were added as a new row into the Chicago neighborhood data frame.

### 3.4 Fetch location data from Foursquare

The coordinate data from this list were then submitted to an API call to Foursquare, in order to obtain a maximum of 100 venues within 500 meters of each coordinate set. The call resulted in location data from only 89 neighborhoods, so the two Chicago neighborhoods without venue information were dropped from the set and the neighborhood data frame was merged with the venue data frame. Using one-hot encoding, the merged data frame was transformed to include 285 columns representing 286 unique types of venue. The data frame was grouped by neighborhood and the means across each venue type column were obtained for each neighborhood.

### 3.5 Perform k-means clustering

Following the “elbow bend method” the ideal  $k$  value was obtained for the k-means clustering method. For this set of data,  $k$  was set to 10. This machine learning technique resulted in an array which labeled each neighborhood in the data frame with its corresponding cluster. This array was added to the data frame as a new column.

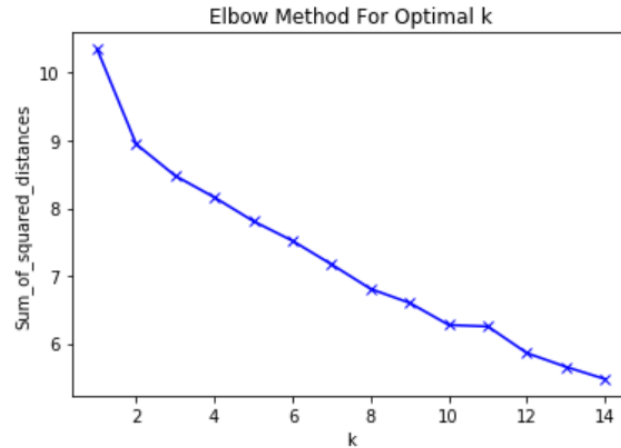


Figure 2. The “elbow curve” indicating the ideal k value to use.

### 3.6 Create list of Chicago neighborhoods most similar to the Gourmet Ghetto

The resulting data frame indicated that the Gourmet Ghetto neighborhood belonged to Cluster 0. Using a Boolean mask, the merged neighborhood-venue data frame was subset so that only neighborhoods in Cluster 0 remained.

### 3.7 Determine distances from the university

Finally, the coordinates for the University of Illinois at Chicago were obtained from the *geopy* library. A new data frame was created that only contained the list of Chicago neighborhoods belonging to Cluster 0 and their corresponding latitudes and longitudes. Next, using the calculation for Euclidean distance, the distance between two vectors, a new column was added to the data frame that indicated each neighborhood’s distance from the university. The data frame was then sorted by this value.

## 4. Results

Using the methodology above, a total of 62 Chicago neighborhoods were grouped in the same cluster as the Gourmet Ghetto. Figure 2 shows the neighborhoods according to their cluster labels, where neighborhoods in red are most similar to the Gourmet Ghetto.

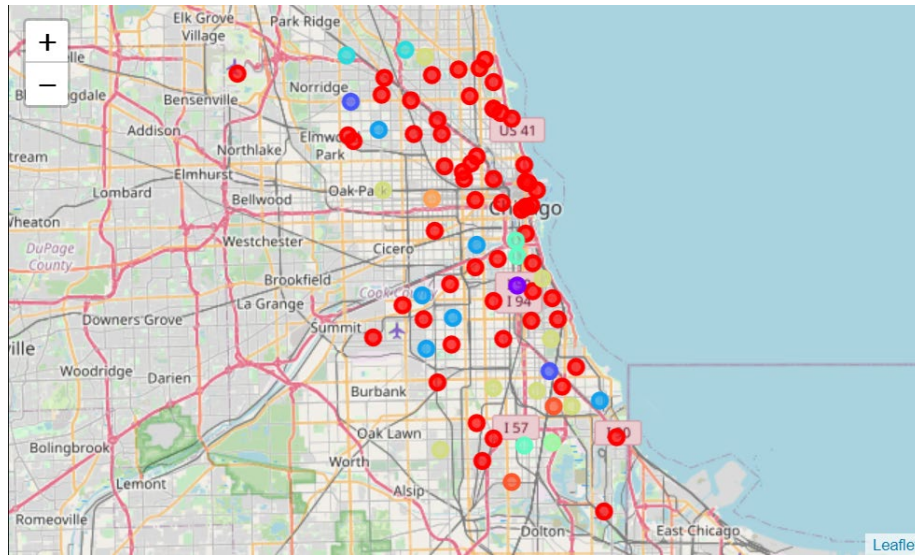


Figure 3. Map of 88 Chicago neighborhoods, color-coded by cluster label. Neighborhoods labeled with red were found to be most similar to the Gourmet Ghetto.

Once the distance of each neighborhood from the University of Illinois at Chicago had been calculated, the list of 62 neighborhoods was re-sorted by distance, from smallest to greatest. The first six neighborhoods from this data frame are shown below, where Greektown is the neighborhood that is both similar to the Gourmet Ghetto and closest to the university.

	Neighborhood	Distance from uni	Latitude	Longitude
0	Greektown	0.009595	41.878445	-87.646932
1	Printers Row	0.020240	41.873787	-87.628900
2	East Village	0.025020	41.875562	-87.624421
3	Loop	0.025020	41.875562	-87.624421
4	Little Village	0.025020	41.875562	-87.624421

Table 1. Chicago neighborhoods most similar to the Gourmet Ghetto, sorted by distance from the university.

## 5. Discussion

In order to hypothesize why these neighborhoods pattern similarly to the Gourmet Ghetto, a data frame containing the top ten venues in each neighborhood was generated. According to the list, it appears that Cluster 0 neighborhoods all have a wide range of venues that cater to young professionals, such as restaurants, bars, and a variety of shops. Perhaps this feature set, in conjunction with the close proximity to the University of California, Berkeley, is why the client's

restaurant in the Gourmet Ghetto has been so successful. The client was wise to look for the same features when deciding where to start a new restaurant in Chicago.

Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
Beverly	41.718153	-87.671767	0	Cosmetics Shop	Boutique	Farmers Market	Martial Arts Dojo	Burger Joint	Italian Restaurant	Bakery
Washington Heights	41.706423	-87.656160	0	Caribbean Restaurant	Cosmetics Shop	American Restaurant	Salon / Barbershop	Park	Yoga Studio	Dog Run
Hyde Park	41.794225	-87.592562	0	Sandwich Place	Bookstore	Park	Train Station	Dive Bar	Taco Place	Shopping Mall
Bucktown	41.912499	-87.672467	0	Bookstore	Park	Bar	Pub	Pizza Place	Mexican Restaurant	BBQ Joint
Wrigleyville	41.947022	-87.656477	0	Bar	Sports Bar	General Entertainment	Sandwich Place	Pizza Place	Restaurant	Dive Bar

Table 2. The top seven most common venues in a selection of neighborhoods in Cluster 0.

## 6. Conclusion

To conclude the report, the client has received the list of the 62 neighborhoods in Chicago that are most similar to the Gourmet Ghetto in Berkeley, according to the type of and density of venues. Additionally, the list is sorted according to distance from the University of Illinois at Chicago, where Greektown is the neighborhood that is both similar to the Gourmet Ghetto and closest to the university.