Relocation Project

Stephen Suydam

November 23, 2020

Introduction: Someone living in Chicago has received a job offer from a company in Seattle. They do not have time to explore all the neighborhoods in Seattle and would like to live in a similar location with reference to amenities and establishments to the area they currently live. They enjoy being able to walk to the shops, gyms, parks, and public transit, so proximity is important. In order to have a similar neighborhood feel, the shops and venues need to be present within a 1 kilometer radius. This has been found to be an acceptable walking distance to patronize these establishments. The aim of this study is to find the closest matched neighborhood in Seattle to that of the authors neighborhood in Chicago.

Method: All code was written on Watson Studio[1] using a Jupyter Notebook or on a local instance of Jupyter Lab and Python 3.7.

Data Collection: This problem will require collecting data from several sources. The neighborhoods in Chicago, the neighborhoods in Seattle, and the respective venues were all scraped from online sources and post processed. The City of Chicago website[2] provided each of the neighborhoods and the outline of their borders for Chicago. Data for the neighborhoods in Seattle was taken from Wikipedia[3] . Venue data including name, category, and location was provided by FourSquare[4].

Data Analysis: Data were cleaned to establish the latitude and longitude of the neighborhood center. Most of these neighborhoods, in both Chicago and Seattle, are square or rectangular, and a radius was used to determine nearby venues. Most of the neighborhoods were encompassed by the 1km radius. The radius sometimes reached into other neighborhoods, which is considered acceptable because that would still be within walking distance of the person moving.

The Chicago data were given as a series of latitude and longitude points that provided the perimeter of the neighborhood. The latitude and longitude data were averaged together to determine the center of each of the neighborhoods. The Seattle neighborhood data needed additional cleaning as the Wikipedia page was not comprehensive and some of the neighborhoods were not included in the PiPy Geocoder

library[5] which was used to retrieve the Seattle data. Each of the neighborhoods from each of the cities is marked in Figure 1.
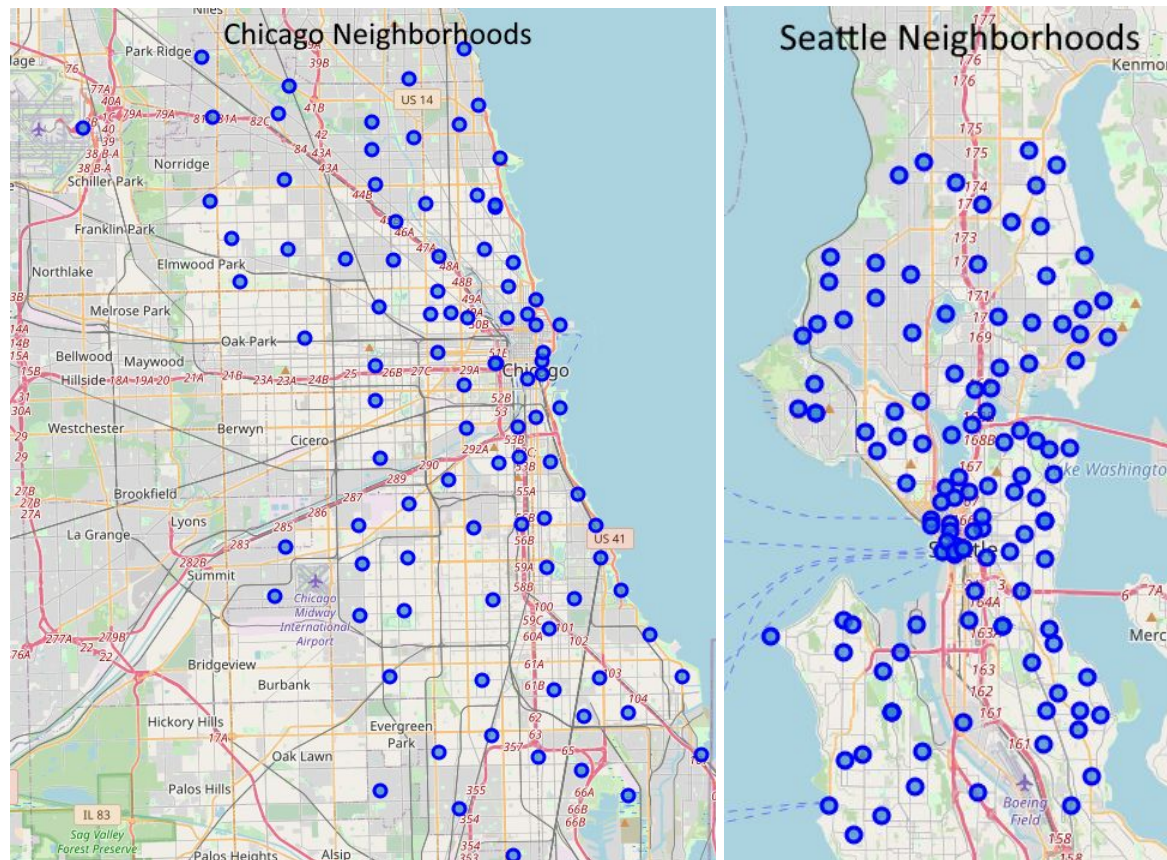
Figure 1. The neighborhoods of Chicago (left) and Seattle (right) marked as the blue dots. There is a concentration of blue dots in the center of Seattle compared to a more uniform distribution in Chicago. This implies there could be more overlap of walkable range for Seattle neighborhoods that are more centrally located.

The venue information was reported by FourSquare. Each venue was categorized by neighborhood which was determined by the proximity of the venue to the center of each neighborhood, which could allow a venue to be listed in two separate neighborhoods. Also reported was the venue's types and its latitude and longitude. The categories of 'Art Gallery', 'Cosmetics Shop', 'Dance Studio', 'Dog Run', 'Pet Store', 'Hotel', and 'Yoga Studio' were dropped from the analysis because they were not of interest to the author.

After separating the venue categories into dummy variables, each neighborhood in Seattle was compared to the author's neighborhood in Chicago, West Loop. This similarity was performed as a cosine similarity which reported a value from 0 to 1, with 1 being an exact match. An additional filtering step was performed to ensure the closest

match to West Loop, by clustering the venue data across neighborhoods including the West Loop.

Clustering: K-Means clustering was used on the neighborhood venue data of Seattle and the West Loop. 5 clusters were used for the K-Means model. Once the clustering was performed, the neighborhoods in Seattle which were included in the West Loop cluster were selected. These data were used to determine the neighborhood with the largest similarity in Seattle compared to West Loop, Chicago.

Results: Miller Park in Seattle presented the most similar neighborhood to West Loop, Chicago (Figure 2) with a value of 0.813. The next closest were Pike/Pine and Stevens with values of 0.812 and 0.804, respectively. Venue overlap of West Loop and Miller Park were plotted on Figure 3 to show which were the similarity between neighborhoods.
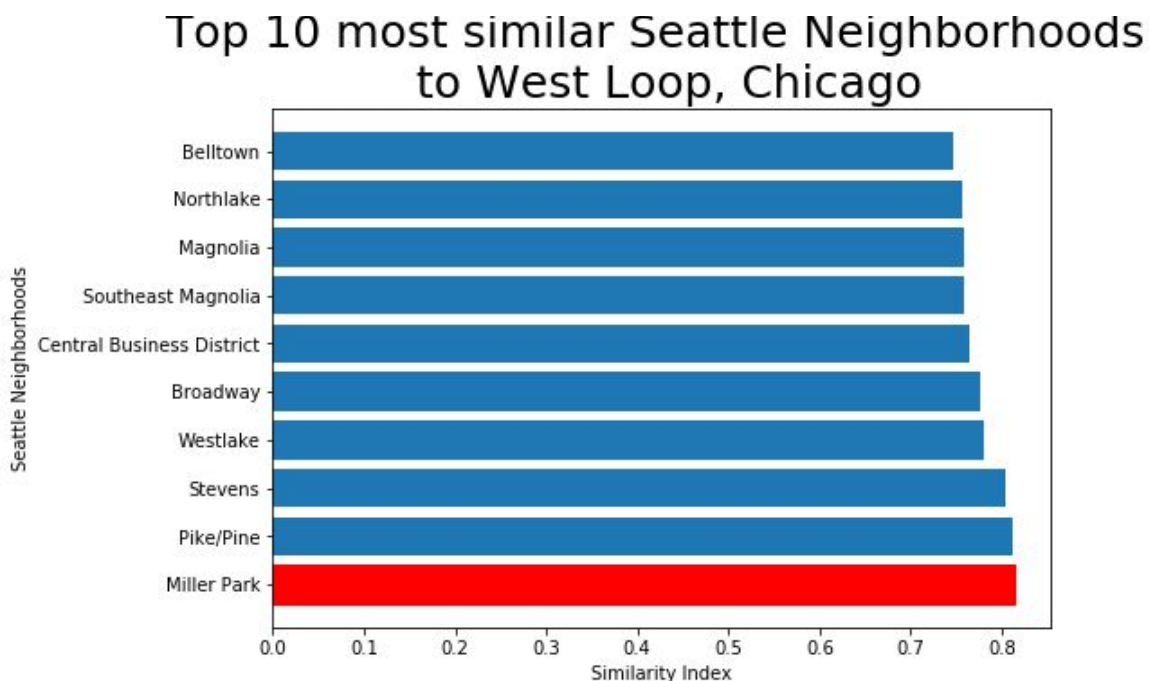
Figure 2. The top 10 most similar neighborhoods to West Loop, Chicago. The maximum similarities were >0.80 from a cosine similarity score.
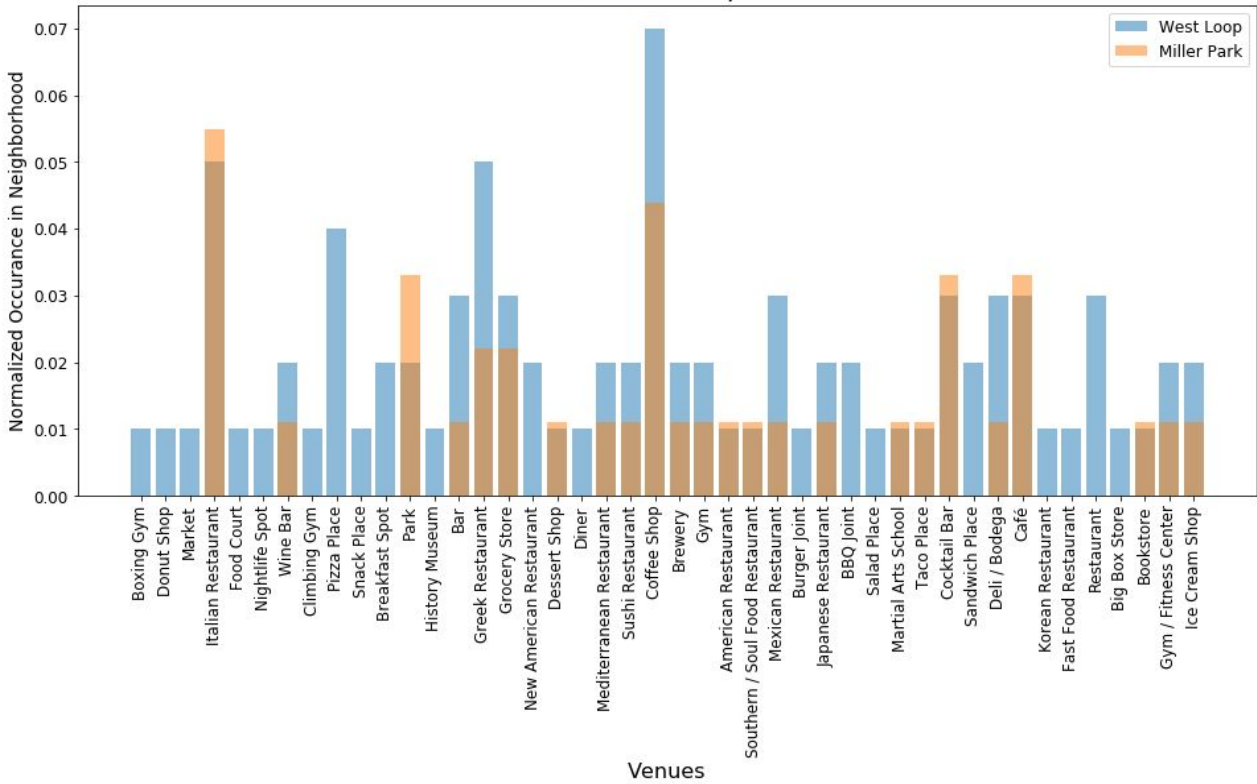
Figure 3. The venues which occurred in both Miller Park and West Loop. Key matches for the author were the number of parks, gyms, restaurants, coffee shops and grocery stores similar to their current neighborhood.

Discussion: It was found that Miller Park is the neighborhood in Seattle most similar to West Loop Chicago. There was a distinct similarity in restaurants, parks, cafes, cocktail bars and grocery stores which help create this likeness. The restaurants, grocery stores, parks and cafes were the venues of most interest to the author. It was unfortunate that the neighborhood of Miller Park did not have a public rail line, though a bus system may exist that was not included on the list of venues in FourSquare.

It is to be noted that upon further review of the data, the top three neighborhoods are located next to each other, with Miller Park being in the center. As stated previously, the concentration of neighborhoods in certain areas of Seattle could cause some amenities and venues to be located in another neighborhood, though it is within the walkable 1km set as the standard. This is acceptable to the author, though something to note. This data should be investigated more carefully before a relocation as not to rule out the adjacent neighborhoods which may have the more key venues located within their borders.

Limitations: Data for Seattle neighborhoods was found through Wikipedia which is a crowd sourced data depository and therefore may contain inaccuracies. The data coming from FourSquare is not ideal for this project. A better dataset would have more generalized locations (restaurants vs "food" which contained cafes, markets and high end dining) and less redundancies (gyms and gyms/fitness centers).

Conclusion: When moving from the West Loop in Chicago, the most similar neighborhoods in Seattle are Miller Park, Pike/Pine and Stevens, which are located within the same general area of the city of Seattle.

Future Directions: This project can be taken in a more generalizable format by allowing users to select the neighborhood they live in Chicago to be compared to those in Seattle. This can go further by targeting major cities across the country and creating a model that helps users find neighborhoods that mirror that of the neighborhood they currently live. Additional updates to this code would include cleaning the FourSquare data further into more useful categories for this particular application. Finally, a check box could be added for users to determine what the user's most important venue categories are in order to create a weighting system.

References:
1. https://cloud.ibm.com/developer/watson/
2. https://www.chicago.gov/
3. https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Seattle
4. https://foursquare.com/

5. https://pypi.org/project/geocoder/