Choosing A Neighborhood in Chicago

Samantha Goodman

March 6, 2020

1. Introduction

1.1 Background

Chicago is a large city in the midwest, and attracts many new residents each year. It is often described as a “city of neighborhoods” where each area of Chicago has its own unique feeling and characteristics. Choosing a neighborhood when moving to such a varied and large city can be a daunting task, especially if you have spent little time in the city. Realtors and apartment agents can be little help, as they make decisions about what neighborhoods you might like with very little information (either about you, or both you and the neighborhoods themselves). Going out and exploring the neighborhoods yourself can be time consuming and much information (such as prices for apartments) is only available online and not in person.

In addition, Chicago is connected by an extensive metro and bus network, run by the CTA (Chicago Transit Authority). This is important to many new residents, as 42% of renters do not own a car.[1]

1.2 Problem

Given information about a new resident (price range for apartments, preferences for types of businesses in the neighborhood), can we use data about Chicago to match them to a neighborhood? One unique aspect of this project is that instead of recommending a neighborhood, my analysis recommends a train stop. The reasons for this are twofold: many neighborhoods have overlapping areas or many names. Data sources might label a neighborhood according to their own designations that do not coordinate with other data sources. There is no central authority on neighborhood names and many are disputed. The second reason is that distance to a train stop (and the train stops distance from downtown) can greatly affect how desirable an apartment, business or neighborhood is. Chicago can have extreme weather and without a car or parking lot, even a distance of 500m can be significant.

---

[1] https://www.chicagomag.com/city-life/April-2015/no-cars-no-problems/

1.3 Interest

Anyone who is considering moving to Chicago, or has an interest in the neighborhoods would find this interesting. Also, people who already live in Chicago and are considering moving or even buying property would find this analysis useful.

1.4 Given Scenario

For this analysis, the following scenario was used: a young student from University of Illinois at Chicago (UIC), who is looking to live alone, doesn't have a car, and has an active nightlife. I translated these limitations into the following requirements:

- A cheap 1 bedroom or studio apartment near the train (within 600m)
- Near the Blue Line (the only train line that stops at UIC)
- Many bars also within 600m

2. Data Acquisition and Cleaning

2.1 Data Sources

Basic location data about CTA rail stations can be found on the Chicago Data Portal here. Additional data sets that will use the rail stations as a reference (based on longitude and latitude) are Foursquare data for types of businesses near each stop, and scraped apartment listings from Craigslist. This gave me the types of businesses near each CTA rail station, and the cost of an apartment near that train stop.

2.2 Data Cleaning

The three sources of data (City of Chicago, Craigslist, Foursquare) all had to be obtained and cleaned.

The City of Chicago data was the easiest to work with. A .csv file was downloaded and needed only minimal cleaning for this project. Each train stop was listed twice (EB and WB, for instance) and since they had the same location data, only one was used. The Blue Line was the only train line analyzed, so other train lines were eliminated. The location data had to be split and converted into floats to use later. Each stop needed a zip code (for Craigslist searching) so Nominatim geocoder was used to get the zip code given the coordinates.

Scraping Craigslist was initially very straightforward, but after examining the data I found that much was left out. Each apartment listing scraped from the search page had a neighborhood, but no location or zip code or anything usable. However, the link to each individual listing did have location data. I wrote a function that scraped that location data, while scraping each search result page.

Another issue was that when the entire city of Chicago was searched, the data scraped was limited - only a few hundred unique apartments seemed to be available. I found that when scraping by zip code, other apartments that were not listed before were then in the data. I developed a workaround to scrape each zip code that had a train stop (so not the entire city) and then remove duplicates.

Duplicate apartment listings were a problem to be overcome. Many landlords will relist an apartment automatically using Craigslist, and this will create multiple listings but the same URL. Other times, they copied and pasted their previous listing, so it had the same data, but a new URL. Other times the same pictures and data was present, but different prices - in this case it was unclear if they were listing different units in the same building at different prices, or relisting and changing the price (usually lowering it). Those cases I removed if the listing description was the same, but the prices were different.

Other listings that were removed were those without a price (uncommon) and those with a too low price - less than $300 (usually spam, or mis-classified roommate ads) and those that were higher than $5000 (mis-classified apartments).

The Foursquare data obtained was not too difficult to work with, once I found the right queries to the API. For instance, if I used the ‘explore’ option, it returned the top 100 establishments, many of which would be bars, but not often. Also, some less popular bars would not be in the top 100 and would be omitted. If I entered a search for ‘bar’ I got many businesses with ‘bar’ in the name, that were not in fact bars. The solution I found was to use the ‘explore’ option, but set the ‘section’ to ‘drinks’. This returned all bars, and many establishments that were similar. I then filtered them based on the category. If the category contained Bar|Speakeasy|Lounge|Nightclub|Pub|Gastropub|Brewery then I considered it a bar.

2.3 Feature Selection

For the CTA dataset, I only used the Blue Line data, and the location of each train stop.

The Craigslist search only returned 1BR and studios, and from that data all I used was the price of the apartment and its location. I used distance from the geopy library to calculate and tabulate the train stops that each apartment was close to.
I used pandas .cut to classify each apartment by the rental price as Cheap, Medium or Luxury.

When the final data was obtained and cleaned, I created a dataframe where each row was a train stop and the total number of apartments available, the number of each classification (Cheap, Medium, Luxury) and the number of bars in the area.

3. Exploratory Data Analysis

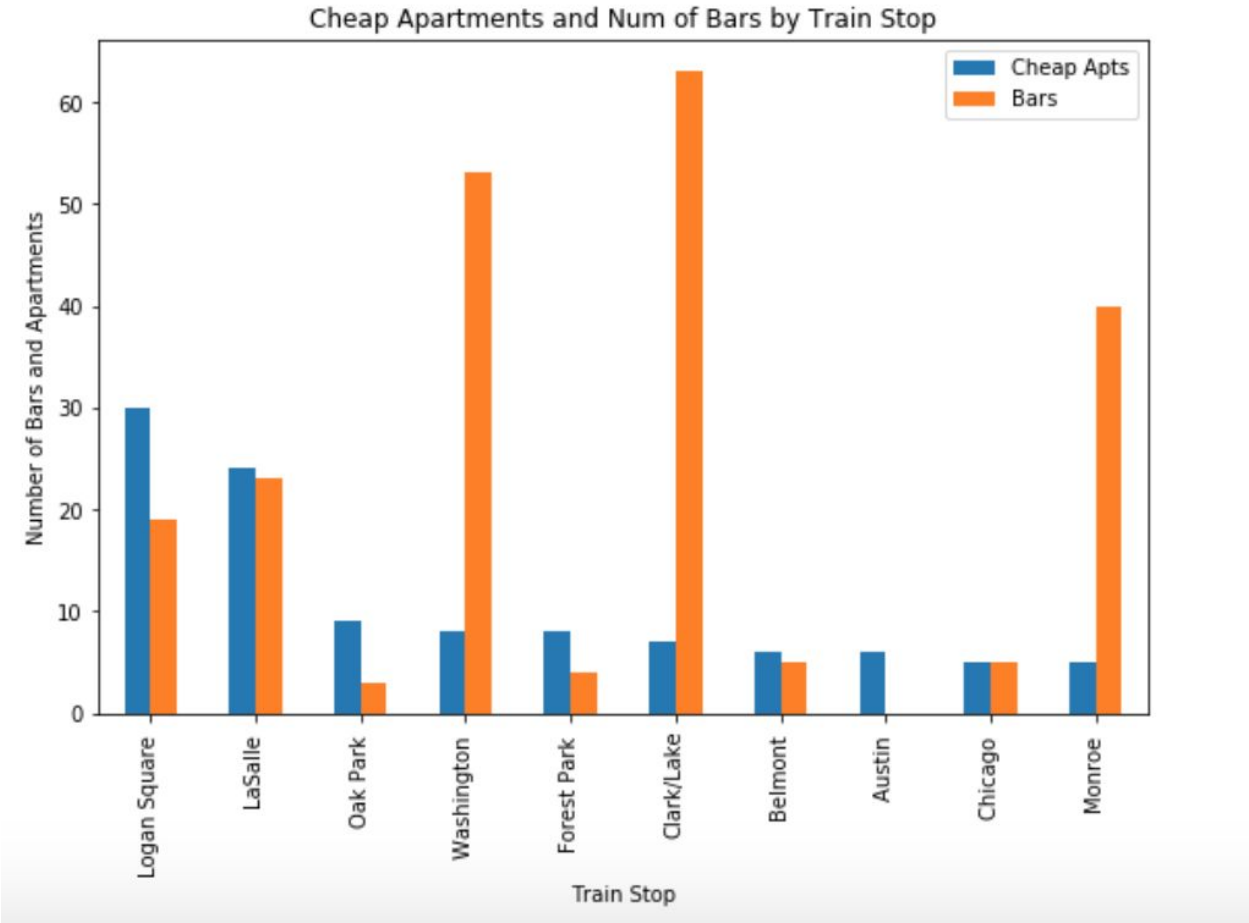
3.1 Sort based on counts

Since the goal was pretty straightforward (find the train stop with the most cheap apartments and the most bars), once I had the data tabulated, I only had to sort it based on number of cheap apartments and then compare the average rent (to get an idea how expensive the bars and restaurants might be) and the number of bars nearby.

	Train Stop	Avg Rent	Num Apts	Cheap	Med	Luxury	Bars	Longitude	Latitude
0	Logan Square	1322	31	30	1	0	19	-87.708541	41.929728
1	LaSalle	1882	101	24	74	3	23	-87.631722	41.875568
2	Oak Park	1330	9	9	0	0	3	-87.791602	41.872108
3	Washington	2015	54	8	34	12	53	-87.629440	41.883164
4	Forest Park	1006	8	8	0	0	4	-87.817318	41.874257

3.2 Ranking based on cheap apartments

I created a bar graph to easily see how each train stop compared. Many had lots of bars, but Logan Square had the highest number of cheap apartments.



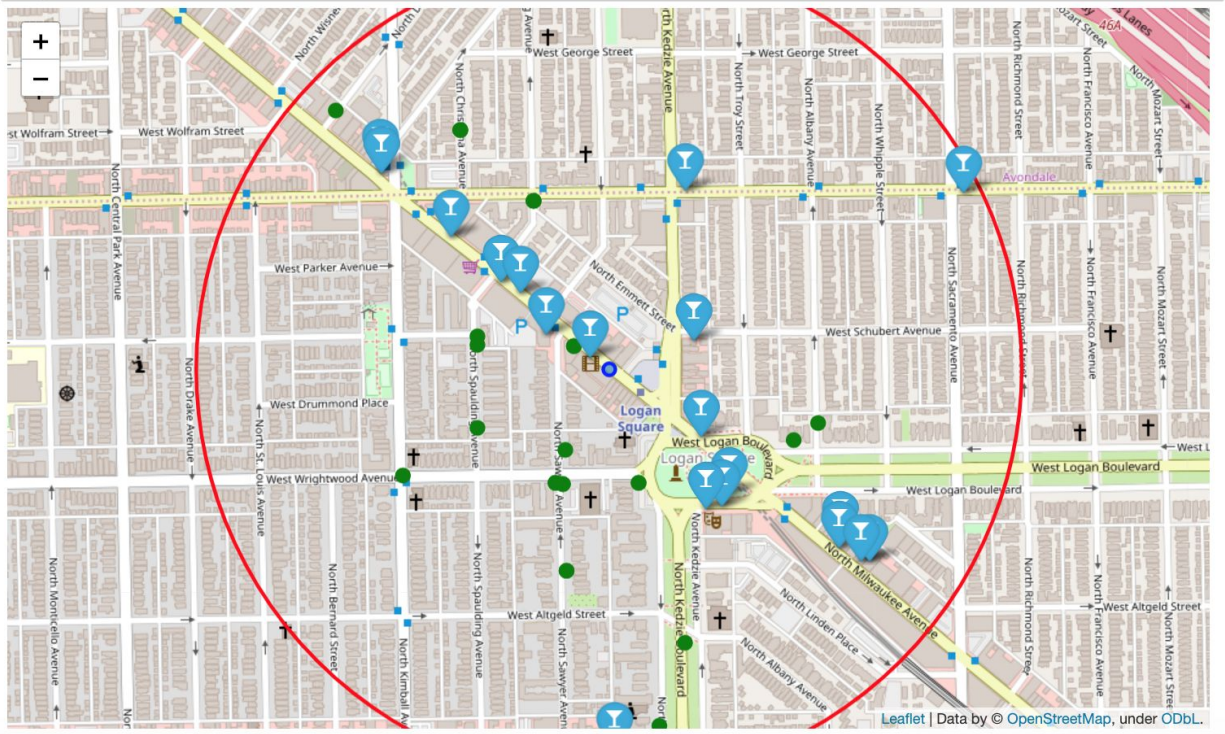
3.3 Comparison of top two candidates

The two train stops, Logan Square and LaSalle had a similar number of cheap apartments and bars, but I noticed that the average rent was much higher for LaSalle. I created two box plots to further visualize the spread of overall rent for the two stops.

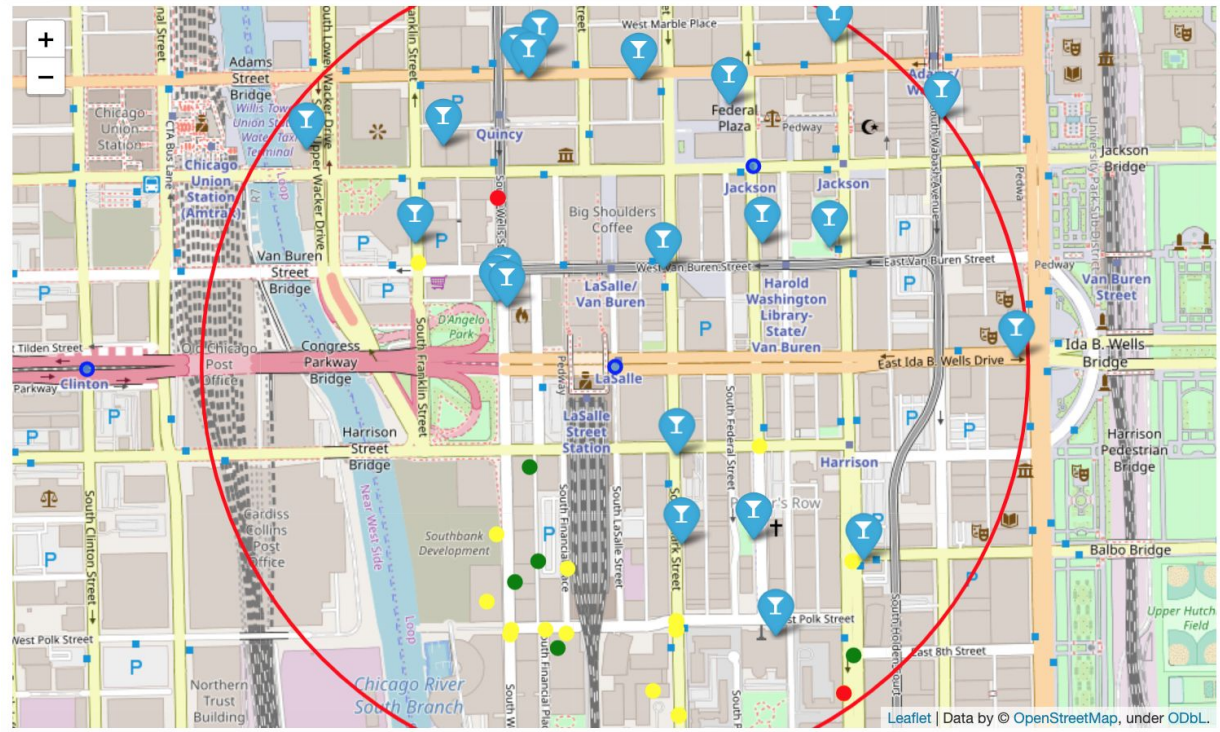
Although LaSalle had almost as many apartments that were labeled “Cheap” the overall rent was much higher and the cheap apartments only represented the lower end of the available apartments. From this fact I concluded that it would be more expensive to live in that neighborhood (price of drinks, price of food and restaurants) when compared to a neighborhood like Logan Square where the overall rents are lower.

Maps were generated for the user to compare the density and types of bars and apartments available near each of the two train stops:

Logan Square:



LaSalle:



4. Conclusion

The results of this analysis is that for a young, price conscious apartment seeker, Logan Square is the train stop that they should search for an apartment near. There are the highest number of cheap one bedroom and studio apartments and has many bars within walking distance. While it's true that one person only needs one apartment, with many cheap apartments to choose from, one is more likely to find one that is acceptable. In addition, it can sometimes be difficult to let an apartment, even if it is available. Landlords and building managers have many applicants to choose from and can be choosy (or discriminatory).

5. Limitations

Due to the fact that our apartment data was limited to one bedroom and studio apartments, some areas were overrepresented due to common architectural designs in the neighborhoods. Logan Square, the train stop recommended by this analysis, is located in a neighborhood that has many courtyard buildings all comprised of one bedroom or studio apartments. Nearby train stops (California, for instance) most likely

have many cheap apartments, but in that neighborhood 2 and 3 bedroom apartments are more common.

Another way that our data could be limited is that our source of apartment listings (Craigslist) is an online listing platform favored by middle-to-high rent apartments. In some neighborhoods, it is much more common to have a physical sign in the window and the owner's phone number to call for more information. Additionally, some high end apartments are only let through their own website portals or agents, and are not listed on Craigslist.

6. Future Directions

This project could easily be adapted to find a train stop or apartment for other scenarios or requirements. For instance - if our apartment seeker wished to move in with a roommate or partner, they could look for a two bedroom apartment instead and this would change the results significantly. In addition, one might be more comfortable taking the bus or walking further and could increase the radius from 600 meters. If someone was looking for a quieter living situation, they could look for apartments near parks, or libraries, or even away from bars. There are many possibilities with only a few changes.