

Directions:

Imagine you're a business owner in Chicago looking to open a new location. Any kind of business will do.

In the form of writing, potentially supplemented by sketches (computer-drawn or hand-drawn) and links, we want to see your response to these questions:

- What questions could you potentially explore/answer with this data?
 - Ideally, what other data would you gather or combine to learn even more?
 - How would you want to see data presented, to make it actionable by you or others?
-

Starting Point:

For the purposes of this challenge, my business, Chicago Guitar Emporium (CGE), is a boutique guitar and accessory store (amplifiers, pedals) that caters to all buyers but targets customers who are looking for harder to find guitars and pedals and specialty and rare, guitar-related items.

Overall Process:

The data was downloaded and imported into a Python Jupyter notebook. After assessing the state of the data (dtypes, nulls, etc.), I performed some exploratory data analysis to get an overall feel for the data set. The thinking that guided my data analysis here was to identify areas where ridership was relatively high and increasing over time. Ultimately, given time, I would be able to compare stops with increasing traffic in growing neighborhoods that have a buffer between any competing businesses.

Potential Questions:

The initial data set pertaining to CTA ridership could be analyzed, by itself, in a number of different ways. Time is a key element in this data set as we have data over 15.5 years. Obviously, high-traffic locations are desirable with respect for space rental costs. Ideally, I would want to target a growing neighborhood where space and property could currently be acquired more cheaply, establish the new location and grow the business along with the neighborhood. Second, the daytype column interesting feature of this data set. From this, we could learn how pedestrian traffic changes based on the day of the week. Ideally, for our new business, we would want to be near el stops that have relatively high traffic on both weekdays and weekends. This would maximize exposure in total, but also increase the number of weekend shoppers.

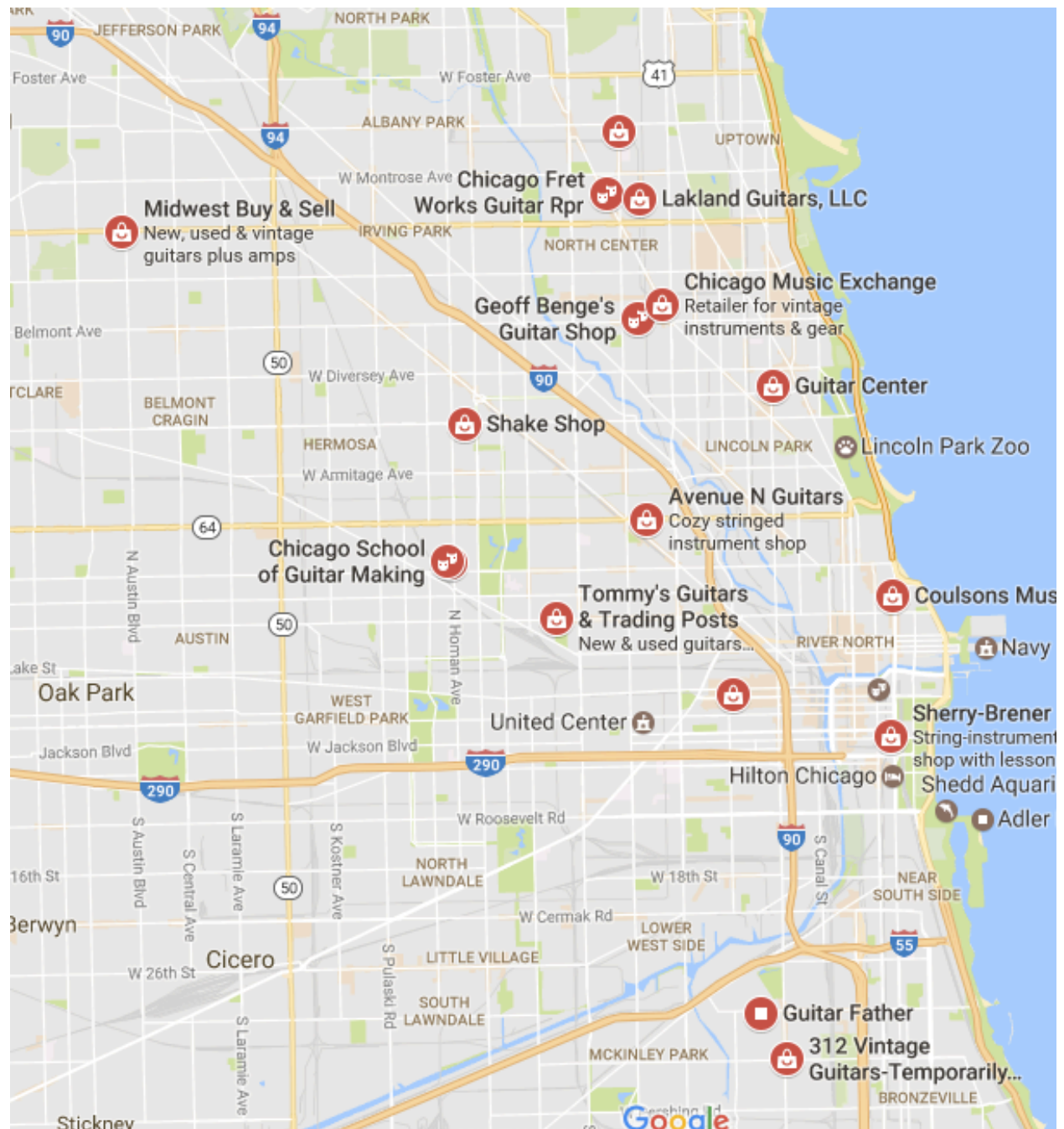
Other Potentially Informative Data Sets:

This data could be paired with the Chicago Community Area Data. Link below, the file is also in the GitHub repository.

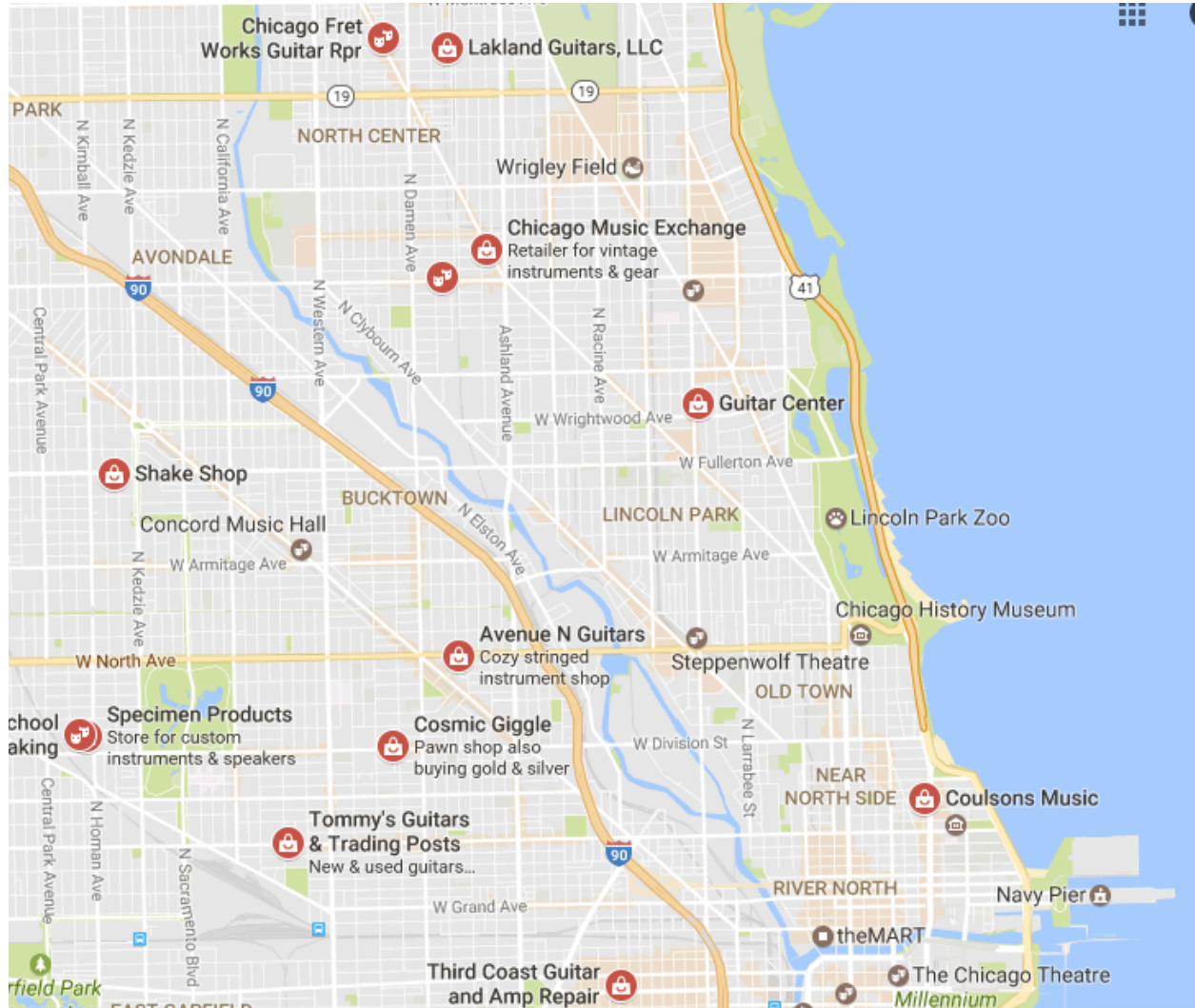
(Link: <https://datahub.cmap.illinois.gov/id/dataset/community-data-snapshots-raw-data/resource/8c4e096e-c90c-4bef-9cf1-9028d094296e>)

This data set shows increases and decreases in population by neighborhood. It also shows the modes of transportation for how residents of a neighborhood commute. For example, 34% of Albany Park residents commute by public transit. This is valuable data if we are looking to expand to locations near el stops.

Additionally, mapping data on the locations of other guitar stores in Chicago would be useful.

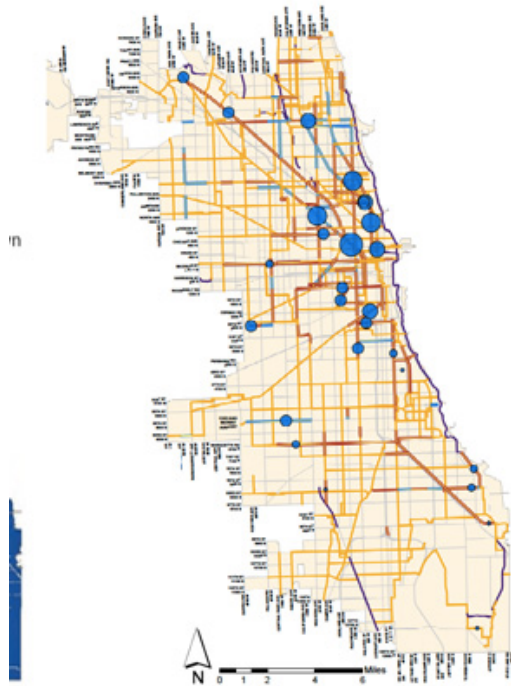


Here is a closeup view of guitar stores on the Northside:



Data Presentation:

In the Jupyter Notebook, I attempted to create at least one way that I would like to see the data presented. I created pivot tables that show aggregations of rider numbers based on location (station). This allows for analysis of high-traffic, medium-traffic, and other locations. Secondly, I would hope to see visualizations such as this (Note: This is just a sample, I plucked from the internet, this is not a reflection of this data set):



This type of visualization, realized through a combination of data sets, would allow for a different manner of viewing the growth of the city, and thus potential locations for the CGE.

Analysis:

I identified a list of 8 potential CTA stops near which locations could be scouted. This list is a starting point and certainly further data analysis would reveal other possibilities.

Rosemont (Blue Line)
 Morse (Red Line)
 Western/Milwaukee (Blue Line)
 Morgan-Lake (Green & Pink Lines)
 Cermak-Chinatown (Red Line)
 Logan Square (Blue Line)
 Damen/Milwaukee (Blue Line)
 California/Milwaukee (Blue Line)

Each of these locations has had an increase in ridership of at least 3 years since 2011. None of these locations reside in the expensive downtown area as those were eliminated from consideration due to their likely cost. These locations could be compared with the neighborhood data (from above) to identify locations in growing neighborhoods. Similarly, zones could be identified in those neighborhoods and/or near to CTA el stops that are the maximized distance from competing guitar stores.