**Data Science Capstone**

Stephen C. Brown

**Problem:**

The experience of people in Chicago, Illinois in the United States varies wildly from neighborhood to neighborhood. I'd like to explore the relationship between the most common venues in a neighborhood and its key economic and sociological indicators.

Specifically

1. Is there a relationship between number of grocery stores and income level
2. how many grocery stores are there in the highest and lowest income areas per-capita

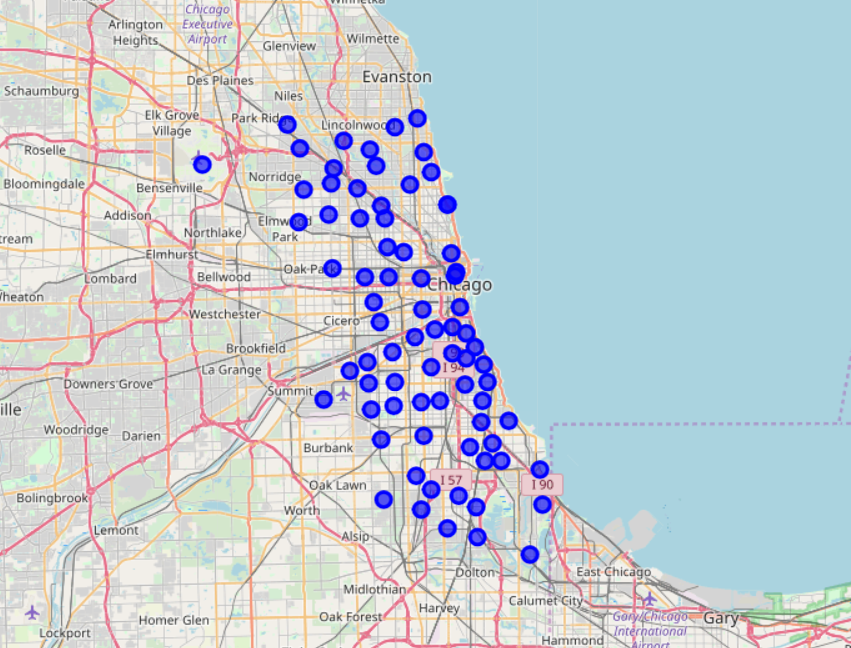
**Data:**

Foursquare API: foursquare.com These data are fundamental to the analysis. The response to the API calls are in the form of JSON files

Census Data - Selected socioeconomic indicators in Chicago, 2008 – 2012: <https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2> These data contain multiple measures of socioeconic health, including education, housing and income. I will be leveraging the income data primarily.

Spreadsheet: 2010 Census Data Summarized :<https://datahub.cmap.illinois.gov/dataset/2010-census-data-summarized-to-chicago-community-areas/resource/b30b47bf-bb0d-46b6-853b-47270fb7f626?inner_span=True> This contains the total poulation of each of the community areas related to my analysis

These data will allow us to separate out locations and see which venues are most prevalent. I look forward to evaluating the data and gaining deeper understanding of a premier global city. The venue data will come from the Foursquare API. The income information will be from the socioeconomic indicators and the population totals will come from the summarized census data.

 Here’s a visualization of each neighborhood I surveyed.

**Methodology**

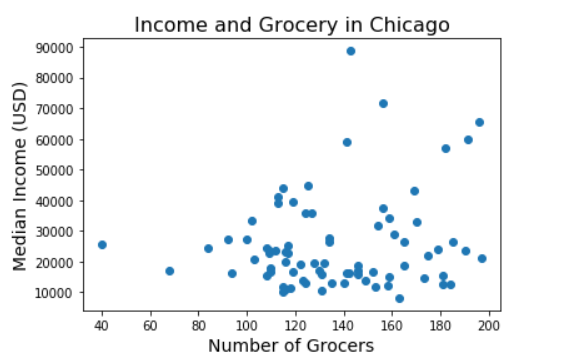
The analysis was conducted by taking the location of each neighborhood in the Census data set and exploring the surrounding 1-mile radius. I searched for grocery stores in each radius and compared it to the median income of each neighborhood as shown by the socioeconomic data.

The number of grocery stores was determined by calling the Foursquare API to return any location that matches the Grocery Store category or explicitly has "Grocer" in the name. These data were to a new dataframe along with the income and location information.

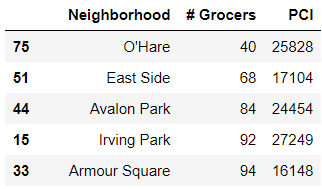
The resulting number of grocers was charted against the median income to determine the relationship. I determined the magnitude of the relationship with a simple scatter plot and calculation of correlation between the number of groceries and the income level.

**Results**

I found that there **was** a relationship between the number of grocery stores and the income level of each neighborhood. The relationship was limited but apparent. The correlation coefficient was 0.17 indicating that there is a weak positive relationship.



Top 5 fewest Grocer Neighborhoods



**Discussion**

These data showed a limited relationship between income and access to grocery stores. That is positive and indicates that the city of Chicago is doing alright in one measure of equity. There are limits to my analysis. I was not able to ascertain the quality of the grocery stores which leaves open the possibility that there are differences between high income and low income grocery stores.

I would recommend that Chicago continue to work on reducing food access inequality. These data are promising but it’s far from perfect for the residents of the city.

**Conclusion**

There is a relationship between income and grocery stores. While it is not a strong relationship it shows that income can predict the prevalence of grocery stores. Future research will incorporate more detailed information of each grocery store and more precise location data.