

**FindMyCoffee Analysis**  
Northwestern University  
Data Science Bootcamp Project

By Katelyn Burke, Prathyusha Challa, & Xuan He

**Project Title:** FindMyCoffee : Using GIS in Chicago (Geographic Information Systems)

**Abstract:**

After the global economic crisis that hit in 2008, business owners are trying to be smarter about where they open their next enterprise. Selecting a profitable location is most important endeavor a business owner can invest in. This analysis uses Geographic Information Systems ‘GIS’ to determine where to build new coffee shop in Chicago area. Geographic information Systems, or GIS for short, can be thought of a system that combines maps and information in one easy-to-read venue, using an array of software applications.

**Hypothesis:** If the coffee shop location has a higher population and more visibility, then opening a coffee shop in that area will result in higher success and more business.

**Overview of FindMyCoffee:**

In this project, we will be working with the Yelp API and U.S. Census data sets to determine where the next coffee shops should open in Chicago. It’s important for shop owners to be informed before choosing a location to open their business. Success of a business depends on several factors including the visibility (foot, car, and bike traffic) of the location, surrounding competitors, and consumer demographics. In this project, we work to understand the relationships between neighborhood population (by zip code) and the number of popular coffee shops in the neighborhood. We analyze the factors that promote business success in order to determine where future coffee shops should open in Chicago.

**Definitions:**

- **‘Popular’ Coffee Shop:** More than 50 Yelp Reviews with an Average Rating of 4 stars or higher (out of 5 total stars possible)
- **‘Low/Poor Performing’ Coffee Shop:** Average Rating of 2 or fewer stars on Yelp (out of 5 total stars possible)
- **Chicago Neighborhood:** based on Zip Code

**Data Preparation:**

The project required a number of data sets to conduct the site selection analysis. For the first data set, we used Yelp Fusion API and read all the current coffee shop information from Yelp for Chicago location. This helped us in the analysis process to determine how many stores are currently located with in each neighbourhood.

To determine the second data set we considered Chicago population from Census board by Zip codes and combined it with Zip codes related latitudes and longitudes. This helped us to understand the location vicinity and demographics.

## **Data Analysis Process:**

The Project extent was on Chicago metro area. We took advantage for YELP Fusion free API and extracted 1000 results and divided the ratings into good/bad ratings and number of reviews. We compared the store locations based on Zip codes and used Google earth latitudes and longitudes to map each location.

## **Which neighborhood(s) in Chicago currently have the most coffee-shop competition?**

**Plots:** Bar Chart- Number of positive Yelp Reviews per Zip Code, Heat Map - Where the Highest Reviewed/Rated 'Great' Coffee Shops are Located

The Chicago neighborhoods with the most coffee-shop competition are Streeterville (zip code 60611), Lakeview East (60613), and Logan Square (60647). These areas of the city have the highest concentration of 'popular' coffee shops - average rating of 4 stars or higher and more than 50 reviews. Based on the data from the heat map we created, we can conclude that the Loop area has the most popular coffee shops and is the most concentrated area of coffee-shops. It could be challenging for a new business owner to open a coffee shop in the Loop and find success right away because of the competition that exists. By looking at the heat map, we can also see that there is coffee-shop competition along major streets in Chicago. These streets include N. Milwaukee Ave., W. Division St., N. Armitage Ave., Michigan Ave., and W. Irving Park Rd. We can conclude that these coffee-shops are frequently visited and reviewed because of their high visibility. There is more foot, car, and bike traffic near these major streets rather than smaller side streets and less-seen areas of the city.

## **Does the population of an area in Chicago impact the success of a coffee shop?**

**Plots:** Scatter Plot- Neighborhood Population vs. Number of Positive Yelp Reviews, Heat Map- Great Coffee Shops and Chicago Neighborhood Populations

We determined that the higher populated areas of the city yield a higher number of popular coffee shops. The heat map shows us that the majority of the high-rated coffee shops (by Yelp) fall in red and orange areas of the map. The neighborhoods with the most coffee shop competition are the highly-populated areas that many Chicagoans choose to reside in. The scatter plot we created shows the neighborhood population count vs. the number of positive reviews written on Yelp in that neighborhood. We did not see any significant patterns on the scatter plot. We did notice that the majority of 'popular' coffee shops have approximately 100-200 reviews. We also noticed from the plot, that neighborhoods with higher populations have more coffee shops to review. The plot also shows neighborhoods with populations from 20,000 to 80,000 that had over 500 Yelp reviews of their coffee shops. Neighborhoods with populations under 20,000, did not have more than 400 reviews of their coffee shops.

## **How can coffee-shop owners predict where to open their next location/new shop in Chicago?**

**Plots:** GMap - 'Poor Performing' Coffee Shops, Bar Chart - Number of 'low rated' coffee shops vs. the total number of coffee shops

Aspiring coffee-shop owners can predict where to open a location based upon the visibility of the area. We have come to the conclusion that coffee shops open close to major roads and in areas of a higher population are more successful. Opening a coffee shop near a chain coffee shop (ie Dunkin Donuts, McDonalds, 7-Eleven) will most likely not be detrimental to business or cause more competition. Many of the chain coffee shop locations have

the lowest ratings according to the Yelp data we pulled from the Yelp Fusion API. We consider 'poor performing' coffee shops to have 2 or fewer average stars on Yelp (out of 5 total stars possible). According to the Yelp data, 69% of McDonald's locations in Chicago are rated as poor performing. 31% of Dunkin' Donuts locations in Chicago are rated as poor performing. 26% of 7-Eleven locations in Chicago are rated as poor performing. 10% of Starbucks locations in Chicago are rated as poor performing. Opening a new shop close to a chain could actually increase business for the shop owner.

### **Conclusion:**

Overall, we can conclude that a coffee shop located in a more populous area increases the visibility of the coffee shop and will most likely increase business. We can also conclude that opening a 'large chain' location of a coffee shop is more likely to be poor performing than opening a 'non-chain' coffee shop.

### **Resources Used:**

- Yelp API Fusion API
- Google Maps Platform
- "US Zip Codes from 2013 Government Data" - CSV by Eric Hurst - Data from the US Census
- "US Census Population by Zip Code" - CSV by The Splitwise Blog - Data from the US Census