

FindMyCoffee



Data Science Project 1 - Northwestern University
Katelyn Burke, Xuan He, Prathyusha Challa

Project Summary



- **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.

Picking locations—An art and a science

- After the global economic crisis that hit in 2008, business owners are trying to be smarter about where they open their next enterprise. Many enterprises today are using technologies like ‘GIS’ to determine where to build new outlets.

What is FindMyCoffee?



- 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.
- FindMyCoffee helps hopeful coffee shop owners determine which area of Chicago to open their shop
- Predicts coffee shop success based on neighborhood population (visibility), and competition of other successful businesses

Questions



- 1) **Which neighborhood in Chicago currently has the most coffee-shop competition?**
- 2) **Does the population of an area in Chicago impact the success of a coffee shop?**
- 3) **How can coffee-shop owners predict where to open their next location/new shop in Chicago?**

FindMyCoffee Definitions



- **‘Popular’ Coffee Shop:** More than 50 Yelp Reviews with an Average Rating of 4 stars or higher (out of 5 total stars possible)
- **‘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

Data Used in FindMyCoffee



Yelp API Fusion API

- Figured out how to return more than 50 business data points from the Fusion API
- Determining which data we wanted to include from the JSON

Data from the US Census

“US Zip Codes from 2013 Government Data” - CSV by Eric Hurst

“US Census Population by Zip Code” - CSV by The Splitwise Blog

-Read in CSV files and merged DataFrames

Data Clean-Up and Exploration



- Organized JSON data we needed from the Yelp API into a data frame
- Defined ‘great/popular’ coffee shop (more than 50 reviews, 4 or more average stars) and created a data frame
- Merged Yelp data frame with the Population data frame on Zip Code
- Used GMaps to create Heat Maps - looked for trends between where the ‘great’ coffee shops are located and the population of the area

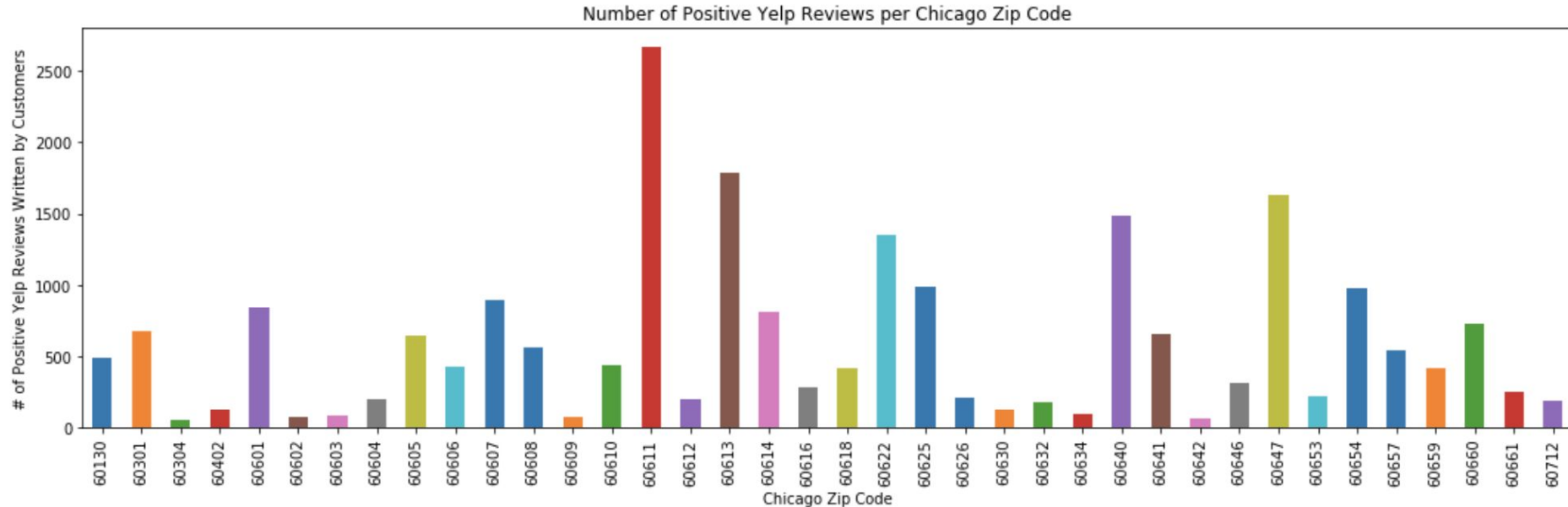


Data Analysis Findings and Conclusions

Illustrates the Coffee Shop Competition by Chicago Neighborhood

Top 3 Zip Codes with the highest number of Yelp reviews:

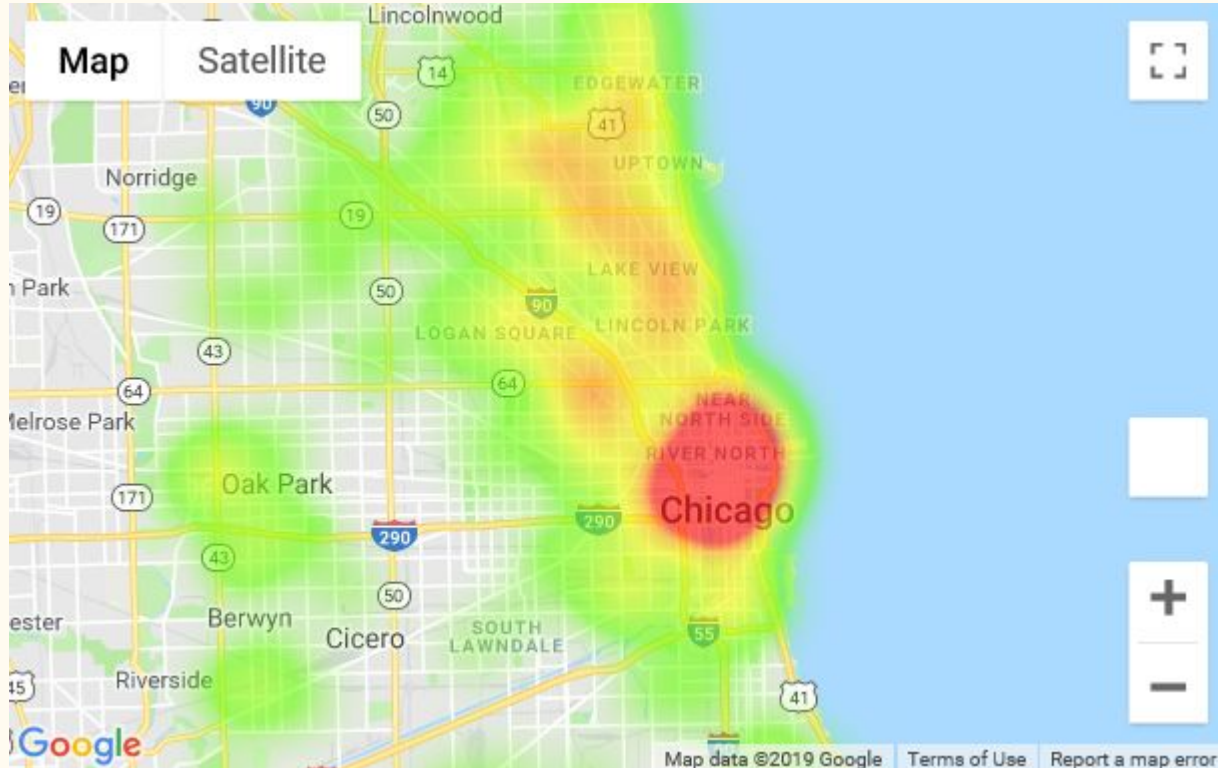
- 60611- Streeterville - 2,666 positive reviews
- 60613- Lakeview East - 1,791 positive reviews
- 60647- Logan Square - 1,630 positive reviews



Least competitive neighborhood: Oak Park

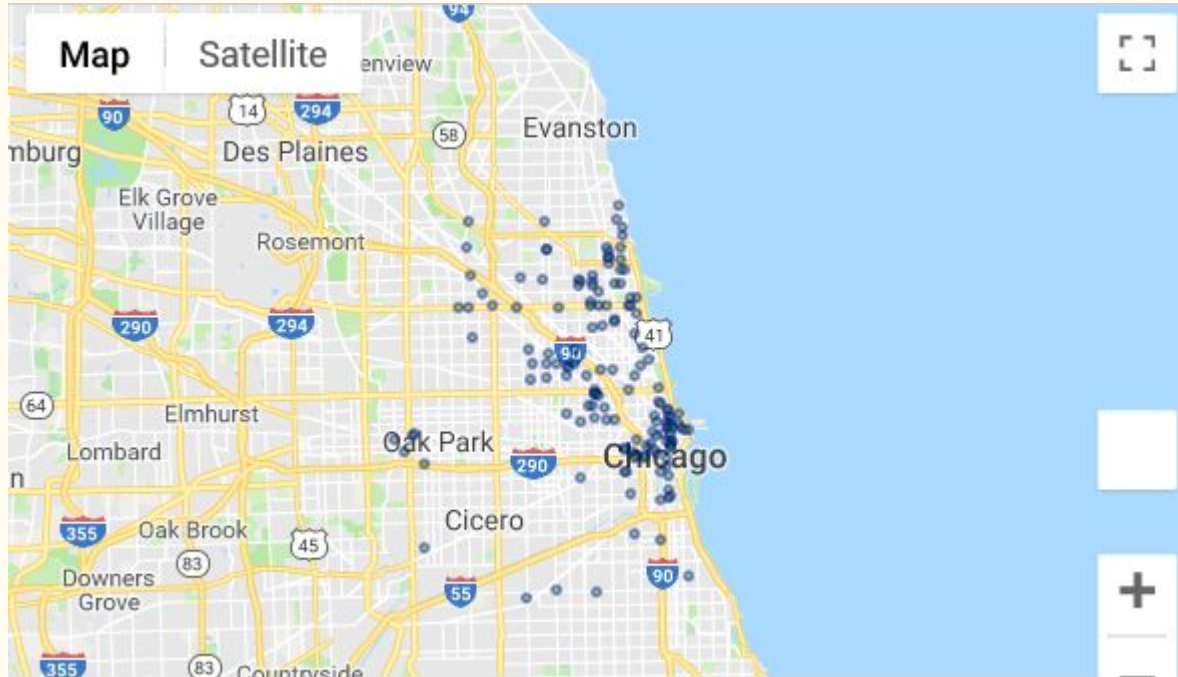
Heatmap - The Number of Coffee Shops in Chicago

- Access to 945 records in total from Yelp API
- Based on 482 of them with Rating > 3.5

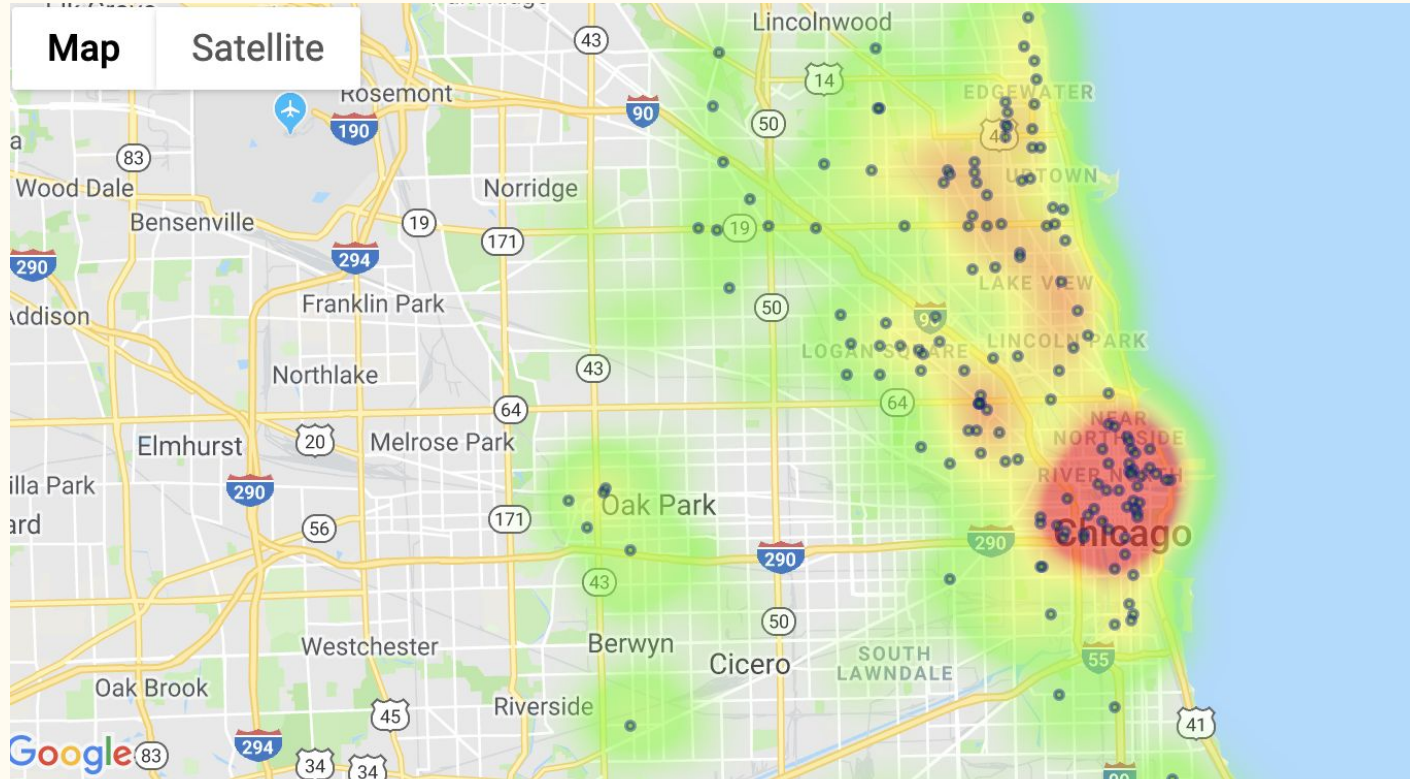


Where to find a good Coffee Shop

- Overall rating > 4.0
- Number of Reviews > 50
- 149 records within Chicago



- Illustrates the concentration of ‘popular/great’ coffee shops in Chicago
- Each point on the Heat Map is a positively rated/reviewed coffee shop location

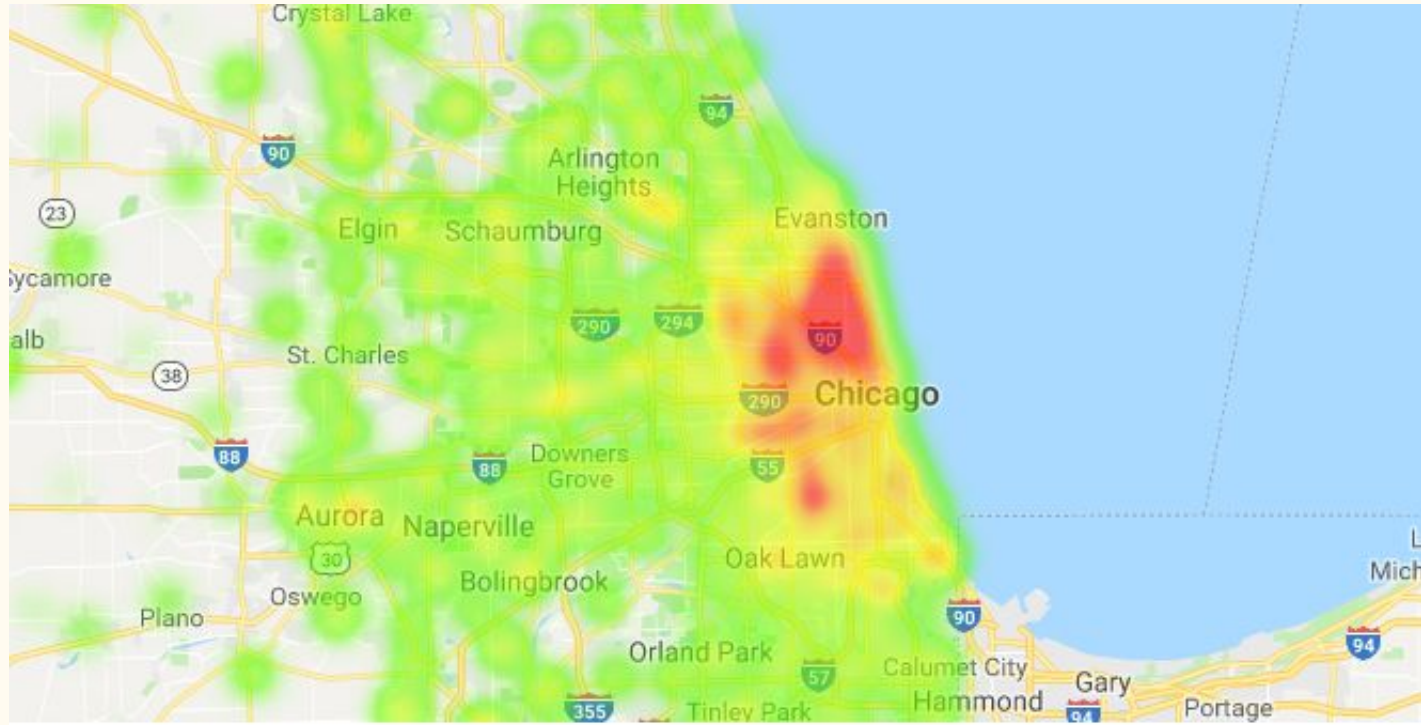


Which neighborhood in Chicago currently has the most coffee shop competition?

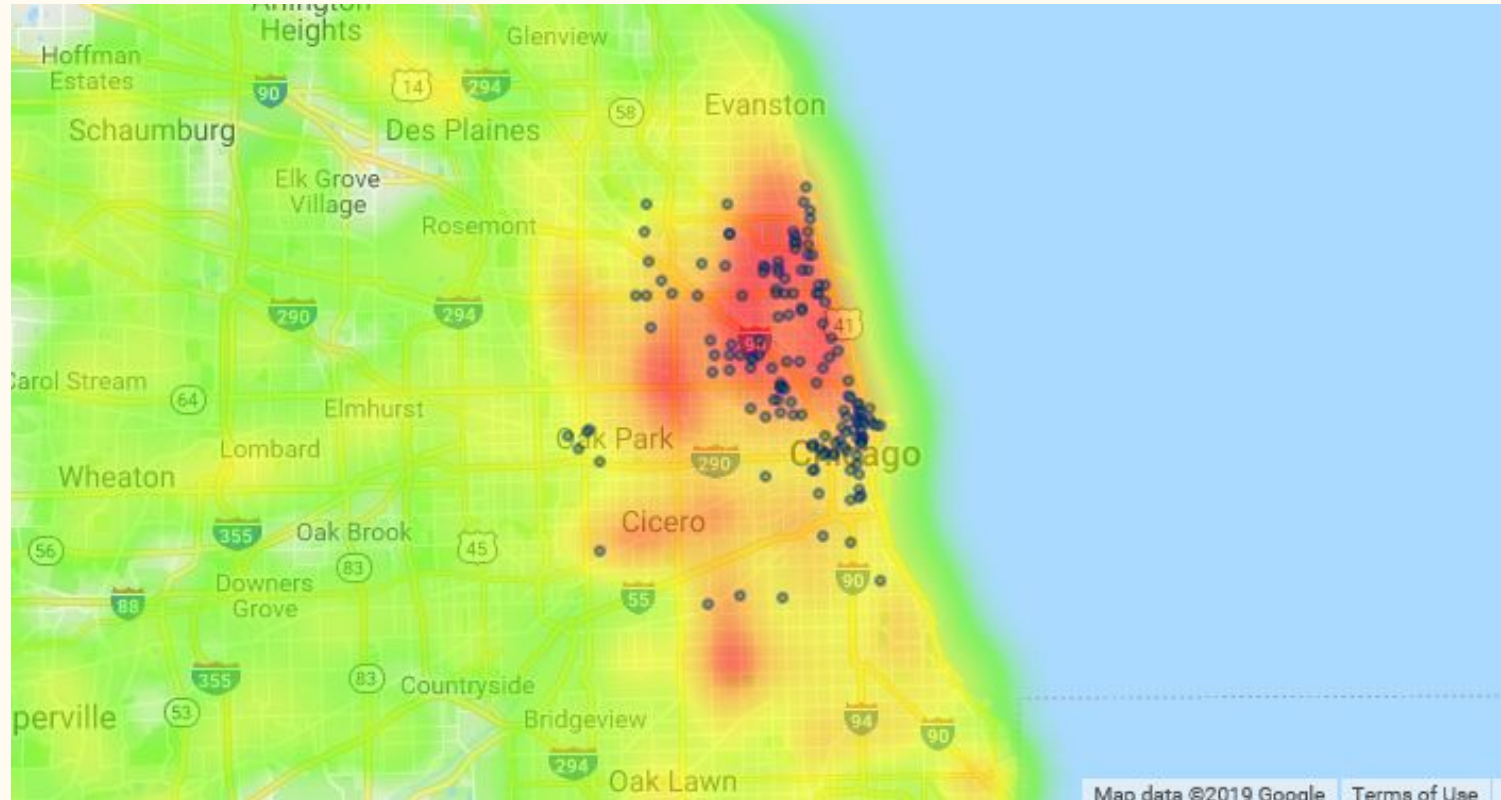
- The Loop has a high concentration of popular coffee shops
- Streeterville (near the Loop), Logan Square and Lakeview East (north of the Loop) have the highest number of positively rated coffee shops and the most businesses competing against one another
- Several of the popular coffee shops are open along major streets in Chicago (higher visibility - more foot/car/bike traffic)

Heatmap - population of Chicago based on Zip Code

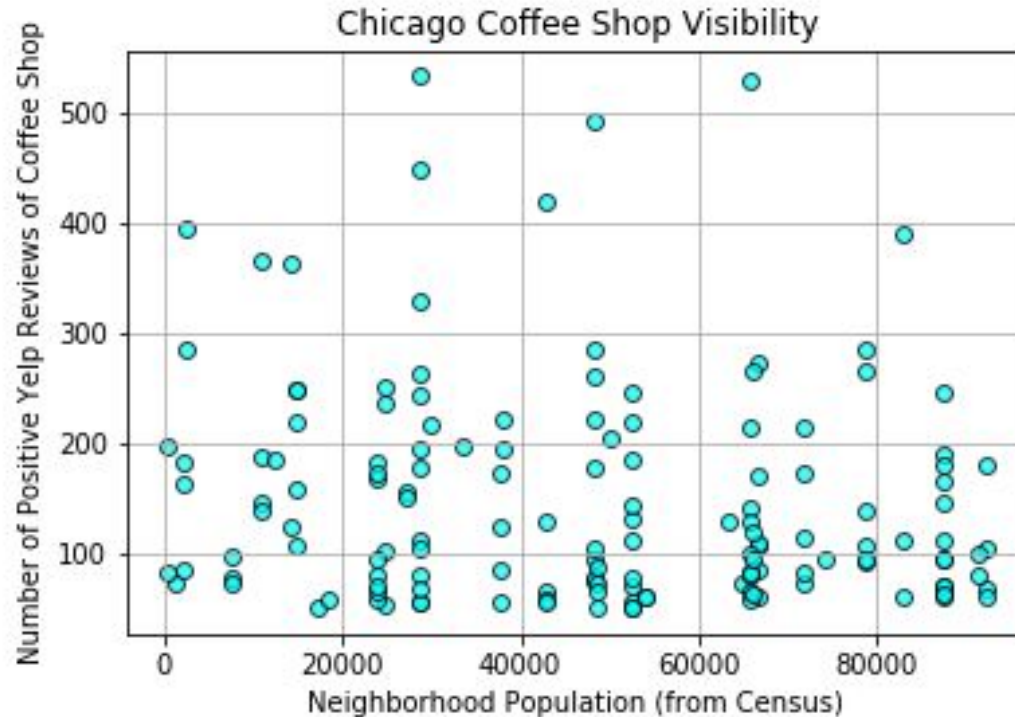
- Chicago Zip Codes range from 60007 - 60827
- Data is from Census



- Illustrates the population density of Zip Codes in Chicago
- Each point on the Heat Map is a positively rated/reviewed coffee shop location



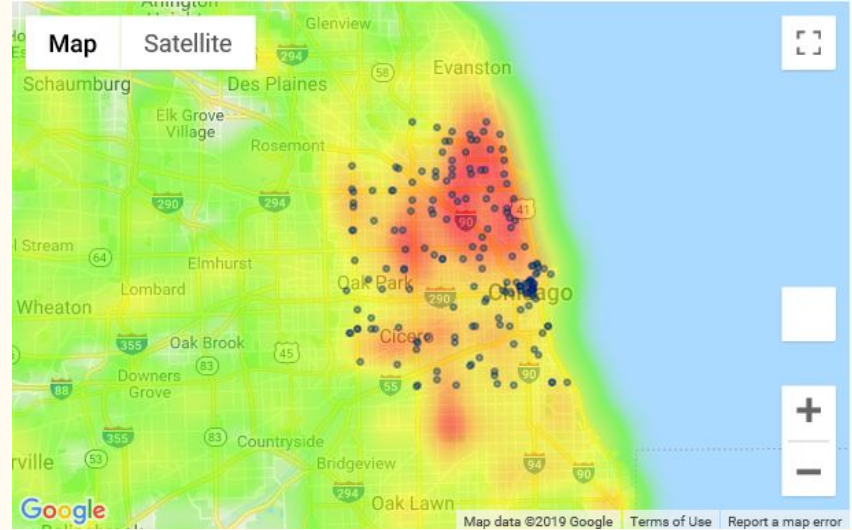
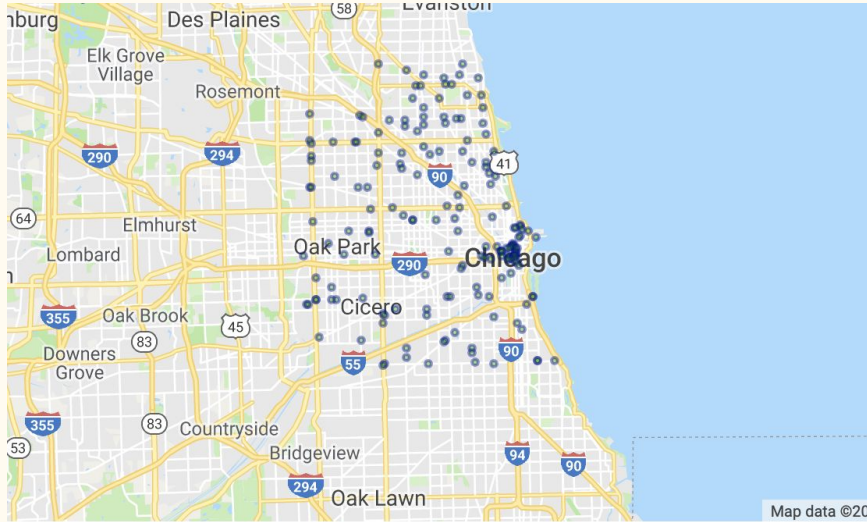
- Shows the neighborhood population count vs. the number of positive reviews written (each point represents a zip code)
- Does not show any super significant patterns
- Neighborhoods under 20,000 people have a lower average number of reviews



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

- Higher populated areas of the city yield a higher number of popular coffee shops
- Opening a coffee shop near a major road/intersection where there is a higher population (see heat map) will increase visibility and most likely business
- There are more successful coffee shops open in Chicago areas with a higher population (as seen on the heat map)

- Illustrates where the Poor Performing Coffee Shops are located
- Each point on the Heat Map is a coffee shop that has an average rating of 2 or fewer stars (out of 5 possible)



-Compares the ‘poor performing’ large chain coffee shops with the total number of locations

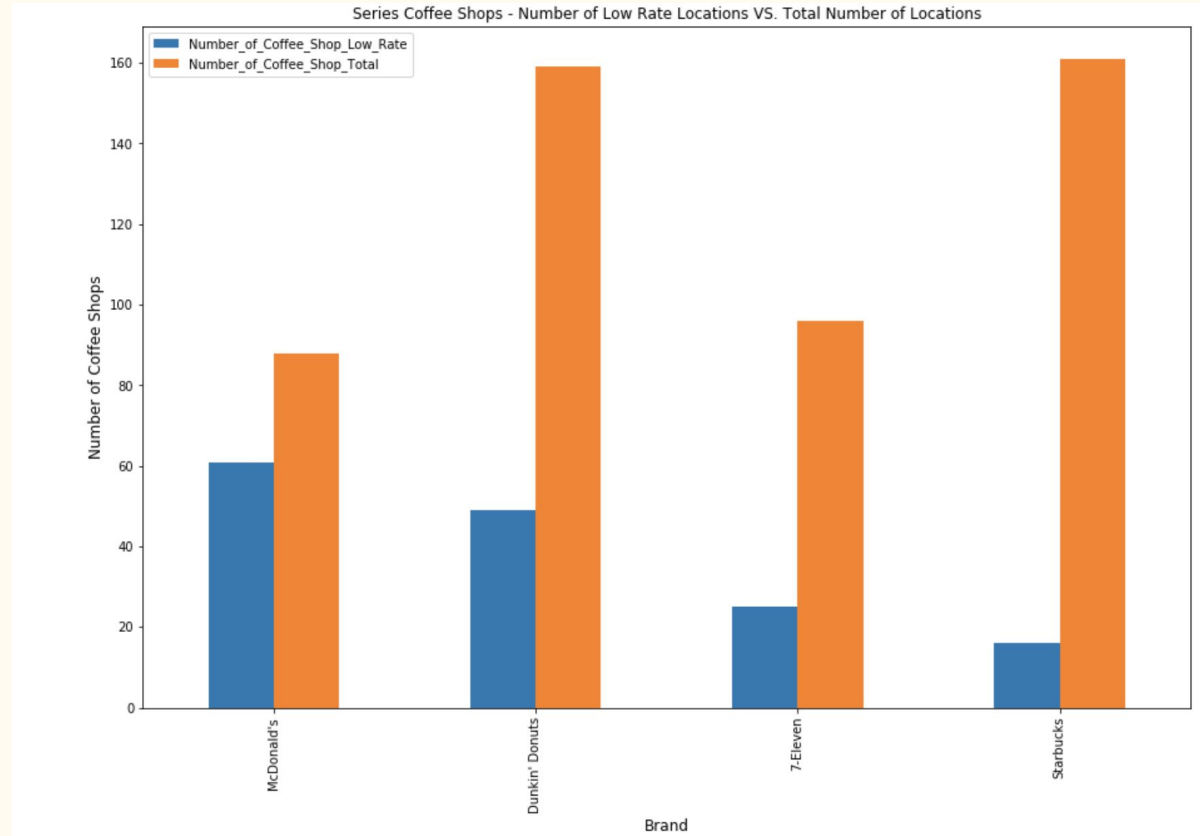
69% of McDonalds

31% of Dunkin’ Donuts

26% of 7-Eleven

10% of Starbucks

locations in Chicago are
considered ‘poor-performing’
(average of 2 or fewer stars on
Yelp)



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

- Opening a shop near a 'large chain' coffee shop (i.e. Dunkin Donuts, McDonalds, 7-Eleven) will most likely not be detrimental to business - could actually help increase business
- A coffee shop located in a more populous area increases the visibility of the coffee shop and will most likely increase business
- Opening a 'large chain' location of a coffee shop is more likely to be poor performing than opening a 'non-chain' coffee shop

Post-Mortem



- How do the rental prices in various areas of Chicago impact the opening of successful coffee shops? (Utilize Zillow)
- Include average prices of rent for businesses in areas of Chicago
- Daily limit of Yelp Fusion API Calls (5,000 per day)
- Difficult to find current population by Zip Code data (we wanted to represent the areas/neighborhoods of Chicago by Zip Code)

Time for Question & Answer!