

# Utilizing Yelp Cost Estimates for Estimating Neighborhood Affluence

DSI 10 Atlanta

Holly Bok, Houston Bova, Raven Hinson, Lisa  
Hwang, Julian Lafaye, Owen Swetenburg

# Problem Statement

“This tool will estimate the affluence of a neighborhood based on Yelp data.

This tool will expect to get, as an input, a list of Yelp data corresponding to zip codes or names of neighborhoods and will estimate the median income.

While traditional methods typically estimate wealth of a locality based on demographic characteristics (e.g. unemployment rate), the novelty of this approach is in its use of big data related to commercial activity and cost of product and services as an indicator for affluency.”

TL;DR

Can we predict median household  
income based on Yelp data?

# About Yelp

Yelp is a website that allows users to submit ratings and reviews for local businesses

While many reviews tend to be of restaurants, many other businesses reviewed include boutiques, grocery stores, doctors, mechanics, and more

Yelp was founded in July 2004 in San Francisco

At the end of 2019, there were 205 million cumulative reviews posted

The number of reviews makes Yelp a promising resource for neighborhood data

# Yelp Pricing System

The average amount spent by each customer at a business is designed by one (\$) to four (\$\$\$\$) dollar signs

Yelp pricing guidelines are not easily located on the site

However, Yelp users are invited to vote for a business's \$ assignment on the app

## Vortex Bar & Grill

✓ Claimed

★ ★ ★ ★ ★

1256 reviews

Details

\$\$

American (Traditional), Bars

Edit

★ Write a Review

📷 Add Photo

🔗 Share

🔖 Save

No Service 1:16 PM 100%

Back

Duck Donuts

HELP IMPROVE YELP ⓘ

What's the price per person?

\$  
Under \$10

\$\$  
\$11 - \$30

\$\$\$  
\$31 - \$60

\$\$\$  
Over \$61

PHOTOS

Food  
14 Photos

Inside  
11 Photos

Menu  
5 Photos

See All 38

SHARE THIS BUSINESS


Messages

Copy Link

More

# Yelp and Atlanta

Decided to focus on Yelp businesses in Atlanta due to our familiarity with the city and its neighborhoods



tacos, cheap dinner, Max's

Atlanta, GA

Q


For Businesses

Write a Review

Log In

Sign Up

All Results




**1. Poor Calvin's**

(404) 254-4051  
510 Piedmont Ave NE  
Downtown

★★★★☆ 3116

\$\$ · Asian Fusion, Southern, Comfort Food

"Ohh. My. God. This place is the bomb. The food. The atmosphere. The decor. The service. The FOOOOOD. So good! We started with apps. I had the deviled egg..." [more](#)



**2. Atlanta Breakfast Club**

(470) 428-3825  
249 Ivan Allen Jr Blvd

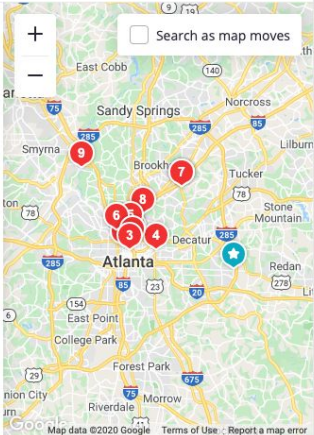
★★★★☆ 3277

\$\$ · Breakfast & Brunch, American (Traditional)

"This place has over 3,000 reviews for a reason, a very good and delicious reason. Great service, awesome food, the coffee, the omelette, the burgers, the peach..." [more](#)

Start Order

Offers takeout and delivery



+

-

Search as map moves

# justfly.com ▼  
(800) 717-5015  
[Visit Website](#)

**Cheap Atlanta to George - Book Our Cheapest Flight Today**  
Impressive Cheap Flights on 500+ Airlines.

# About Statistical Atlas

Located at [statisticalatlas.com](https://statisticalatlas.com)

Offers data from the US Census Bureau, particularly the 2010 Census and the 2012-2016 American Community Survey

 **Statistical Atlas** Demographic Reports For The United States

Detailed demographic information presented in charts and maps across all geographical levels of the United States, ranging from the entire nation down to the individual block.

Topics include race, ethnicity, ancestry, income, employment, educational attainment, and more.

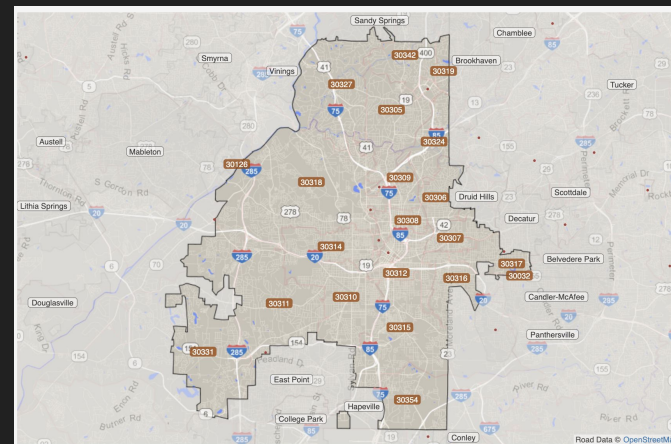
<https://statisticalatlas.com/about>

# Atlanta Zip Codes

## Statistical Atlas includes zip code information for cities

# There are 30 zip codes for Atlanta

One zip code may contain several neighborhoods



State: [Georgia](#)

Counties: [Cobb](#), [DeKalb](#), [Fulton](#)

Metro Area: [Atlanta Area](#)

ZIP Codes: [30030](#), [30032](#), [30126](#), [30303](#), [30305](#), [30306](#), [30307](#), [30308](#), [30309](#), [30310](#), [30311](#), [30312](#), [30313](#), [30314](#), [30315](#), [30316](#), [30317](#), [30318](#), [30319](#), [30324](#), [30326](#), [30327](#), [30331](#), [30332](#), [30334](#), [30336](#), [30342](#), [30344](#), [30354](#), [30363](#)

Unified School Districts: [Atlanta](#), [Cobb County](#), [Fulton County](#)

Atlanta, Georgia (City)

456,378

189,343

Overview

Population

Age and Sex

Race and Ethnicity

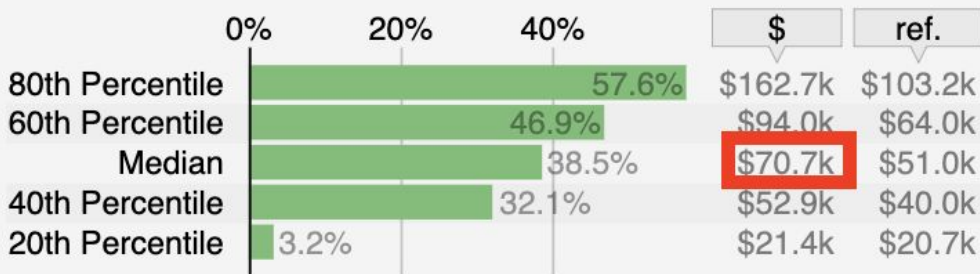


# Median Household Income

## Relative Household Income Percentiles #2

Selected measures of household income in ZIP Code 30030, as a percentage more or less than Georgia at large.

*Scope: households in Georgia and ZIP Code 30030*



\$ given measure of household income in ZIP Code 30030

ref. given measure of household income in Georgia

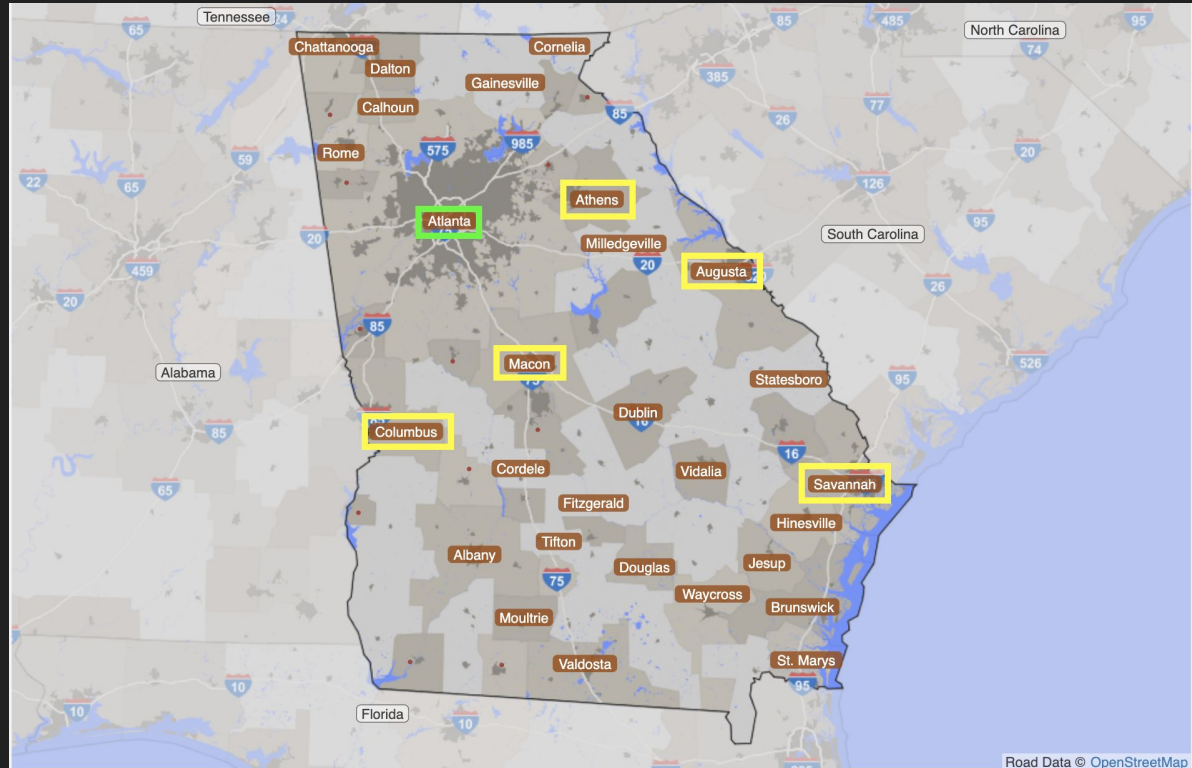
For each zip code, Statistical Atlas shows several graphs for household income

Median household income by zip code was obtained

# Adding Additional Cities

For more data and modeling purposes, we added the following cities in Georgia:

- Athens
- Augusta
- Columbus
- Macon
- Savannah



# Data Collection

In order to help predict the affluence of a neighborhood, we scraped data from Yelp using an API.

Using the website [statisticalatlas.com](https://statisticalatlas.com) and BeautifulSoup, we scraped the median income of each zip code in Atlanta, Athens, Augusta, Columbus, Macon, and Savannah.

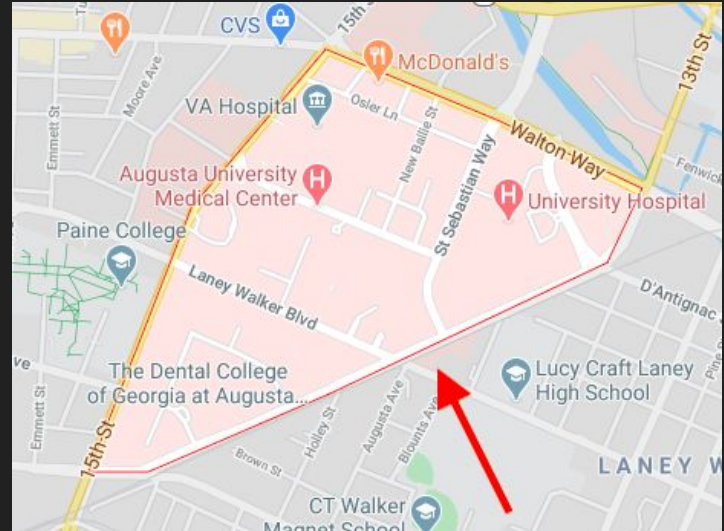
Nine zip codes presented problems due to page formatting and lack of income information. 78 zip codes are used in our model.

# Zip codes with no income information:

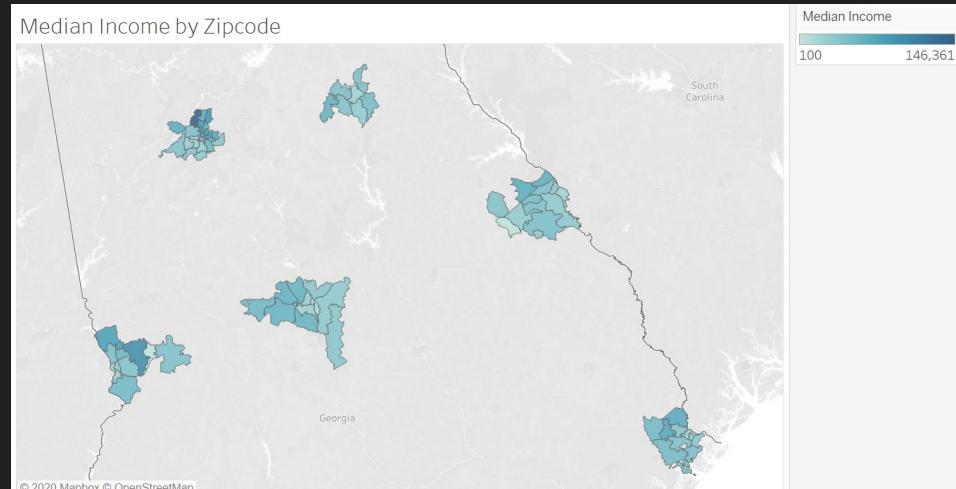
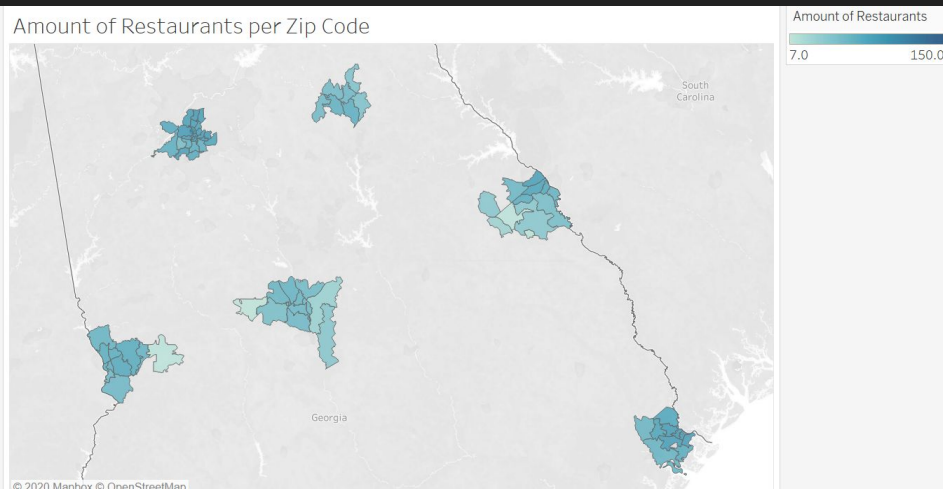
31207 in Macon



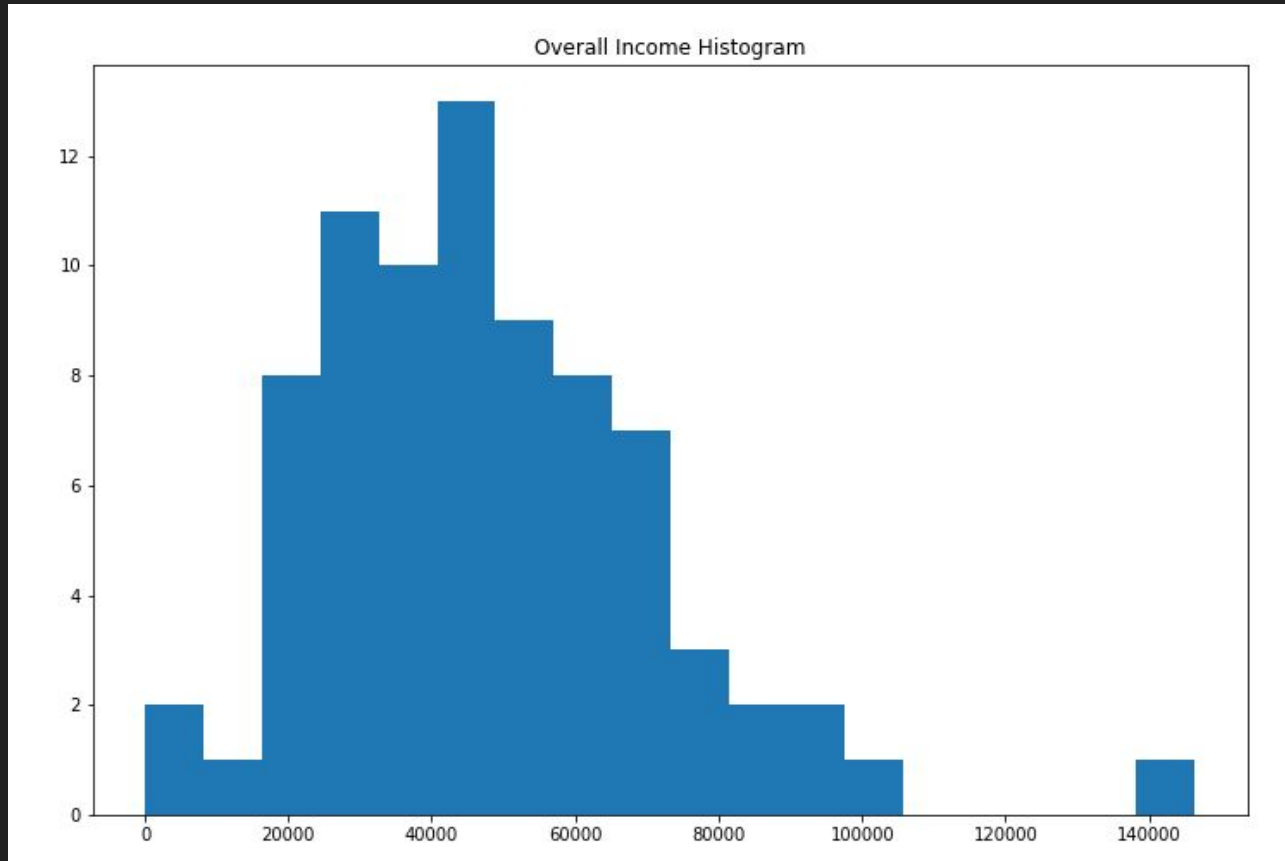
30912 in Augusta



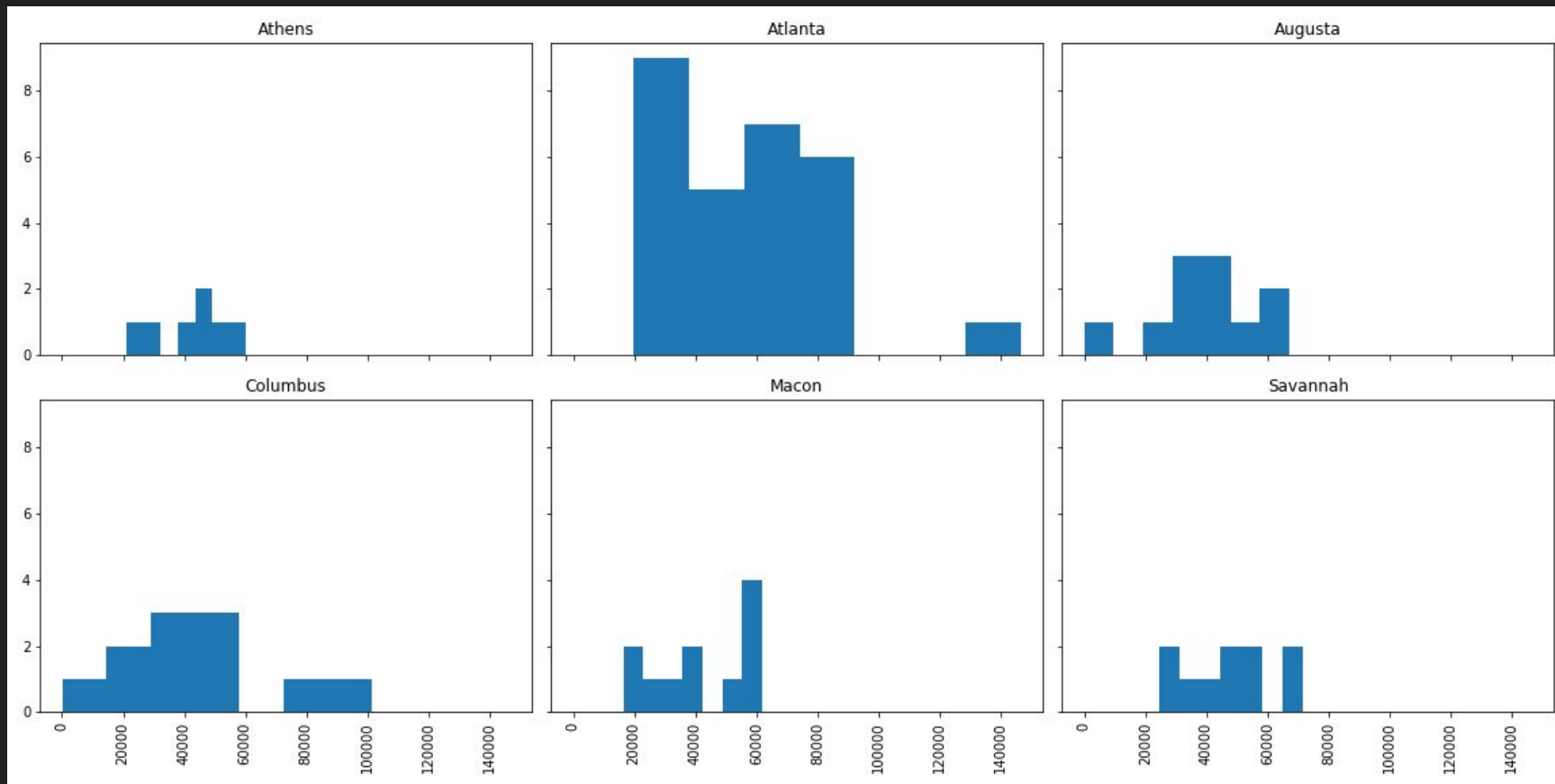
# Comparing Amount of Restaurants and Median Income by Zip Code



# Overall Income Distribution:



# Income Distribution by City:



# Modeling

Linear Regression to look for coefficient trends

After seeing trends, it was passed through a Lasso model to see if they were substantial enough to build a model off of.

Lasso was the best performing model overall

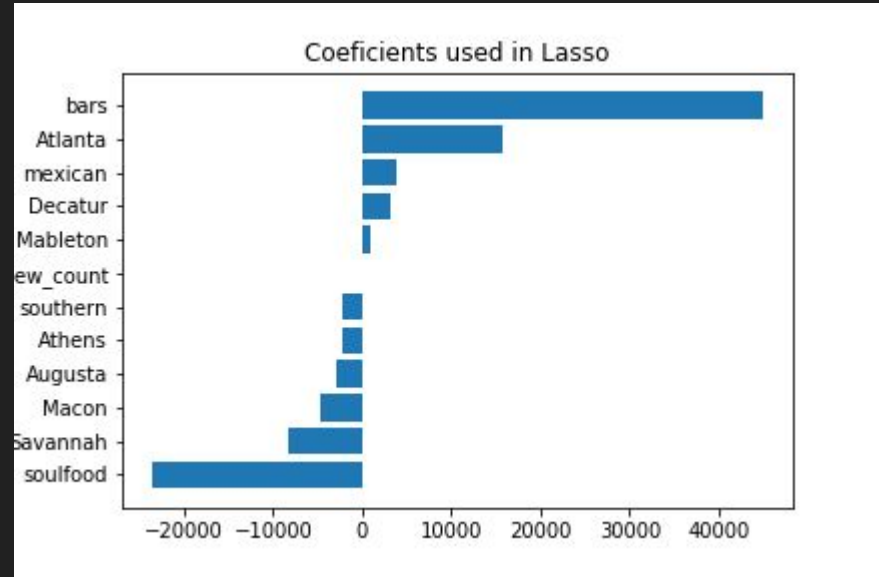
The result was  $R^2 = .33$ , which isn't the best but better than 0.

Tried SVR and 8-layer Neural Nets model on all the data- still didn't work



# Bar Graph of Categories

Graphs showing which categories were most telling of neighborhood affluence predicted using LASSO model against Data Collected



# Limitations

Lack of Yelp reviews for regions with fewer business such as rural areas

Quality of Yelp data (duplicates, missing price information)

Inconsistency across cities in the Yelp data

Not much of a correlation between restaurant price and affluence of a neighborhood

# Conclusions

Yelp data are not ideal for predicting neighborhood affluence

LASSO regularization might be the best option for this problem

Yelp data are geared towards apps and not data analysis

Interesting the best predictors where the categories of restaurants

# Next Steps/Future Directions

- Collect more data (entire US) to improve neural net performance
- Combine Yelp data with other datasets such as housing sales for better insights
- How does the size of a zip code (sq mi) factor in?
- How does housing price data in zip codes correlated with the affluence of an area of town?

Thanks!