Best Place to Establish a Coffee Shop

# Introduction

## Background

The establishment of a coffee shop is a very critical decision. The location of a coffee shop is one of the most important considerations that affects the future success of the business.

## Business Problem

Business problem and My interest

Within a 20 Mile Radius of Winston Salem, North Carolina

## Methodology

### Business Survey and Input

The Business owner was interviewed to determine aspects which would be important considerations for the location of a new coffee shop.

The business owner provided various stories about past businesses. The main consideration for a location is foot traffic.

### Decision on how to address business input

It was decided to base the analysis on competition (other coffee shops in the area), population (as a rough indicator of foot traffic) and income (to determine which markets might have the most expendible income).

Also, the following articles mention the importance of location and income for finding a location for a coffee shop

<https://foodtruckempire.com/coffee/location-analysis/>

<https://www.shopkeep.com/blog/what-to-consider-when-choosing-a-coffee-shop-location#step-1>

<https://www.ncausa.org/Industry-Resources/Market-Research>

### Exploratory data analysis

Method:

1. Gather data on competing coffee shops in the area
   1. Obtain list of Zipcodes within 20 miles of center of Winston Salem NC
   2. Get Longitute and Latitude of each of the zipcodes
   3. Call Foursquare API to pull closest coffee shops to each of these zipcodes
   4. Remove duplicate coffee shop IDs
2. Obtain ZCTA boundaries for the area
   1. Load GeoJSON for all ZCTAs in North Carolina
   2. Filter by the list of Zipcodes within 20 mile radius (note: ZCTAs are approximately the same as zipcodes)
   3. Create GeoJSON file for ZCTAs in the area
3. Choropleth map of coffee shops
   1. Map each ZCTA with color according to number of coffee shops per ZCTA
   2. add circles at Longitude and Latitude for each coffee shop venue found from Foursquare
4. create data for additional maps
   1. Add count of Starbucks, Dunkin' and other coffee shops per ZCTA
   2. add Census population data per ZCTA
   3. add Census Household count and mean income data per ZCTA
5. Plot these maps
6. plot comparison of various features to see if and how they correlate

### Inferential Statistical testing

* 1. The data for income (per household), population, and total coffee shops per ZCTA was compared to Starbucks locations per ZCTA was reviewed to see if there was any correlation. It was considered that if a correlation was found, then the same criteria could be applied to use Starbuck’s formula to place a new coffee shop. No prevalent correlation was found with the data at hand.

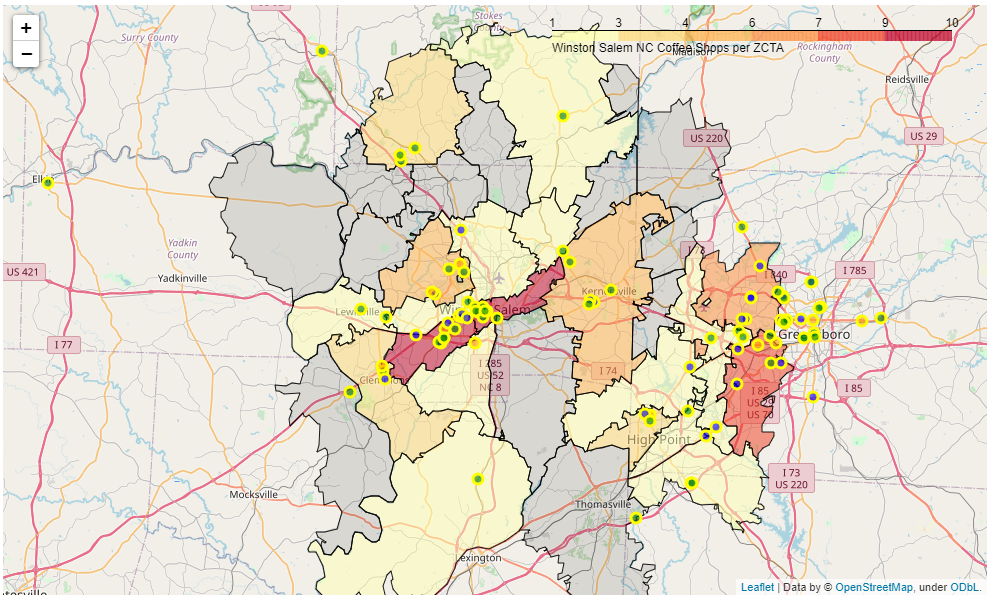
### Challenges Encountered

More time was spent on this capstone project than probably most students invest. There were many challenges. To name a few:

* It took a while to figure out that zipcode data was not available and I needed to use ZCTA data
* Data for Winston Salem was difficult to find: scope was expanded to the county level where data was difficult to find after searching so many places, scope was then expanded to the whole state which also proved a challenge. The university was contacted several times and finally the census bureau. There were many conversations with the census bureau to understand how to get the data that I needed. It was a very small subset of a massive amount of data features.
* GeoJSON for ZCTAs was difficult to find. Initially data was found that wasted countless hours only to find there was some type of corruption in the data rather than a python code problem. This had me spend countless hours on trying to debug code when it was later found to be a data issue.

## Visualization of the Data

### Coffee shops per ZCTA (below)

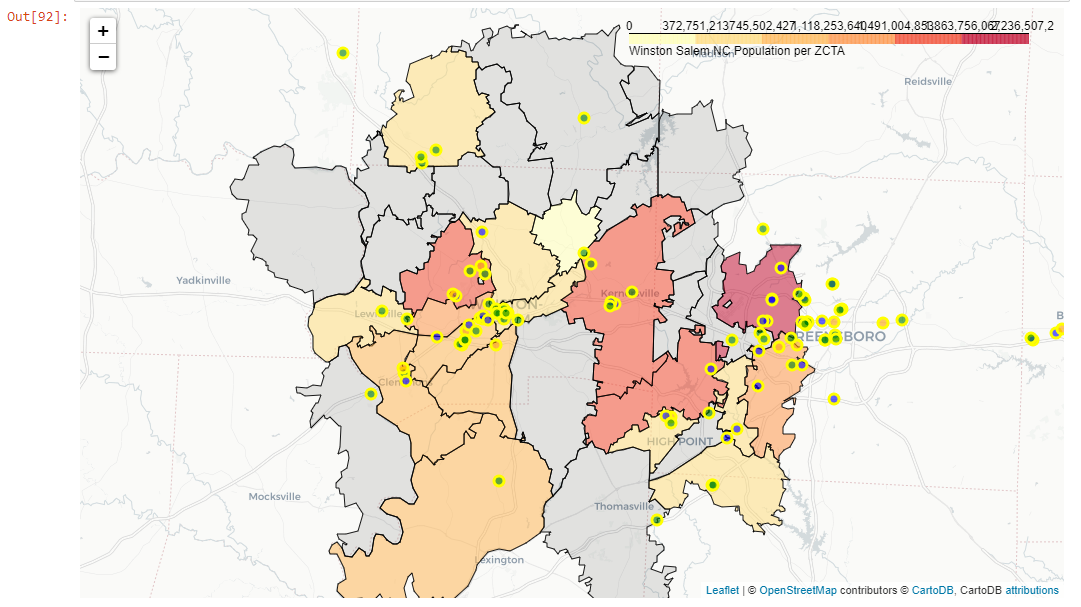


Circle Color: Green: Starbucks, Orange: Dunkin’, Grey: Other

ZCTA boundary area colors: darker for more coffee shops per ZCTA (grey = 0 coffee shops in the ZCTA)

Note: zipcodes of Coffee Shops approximately align to ZCTA boundaries

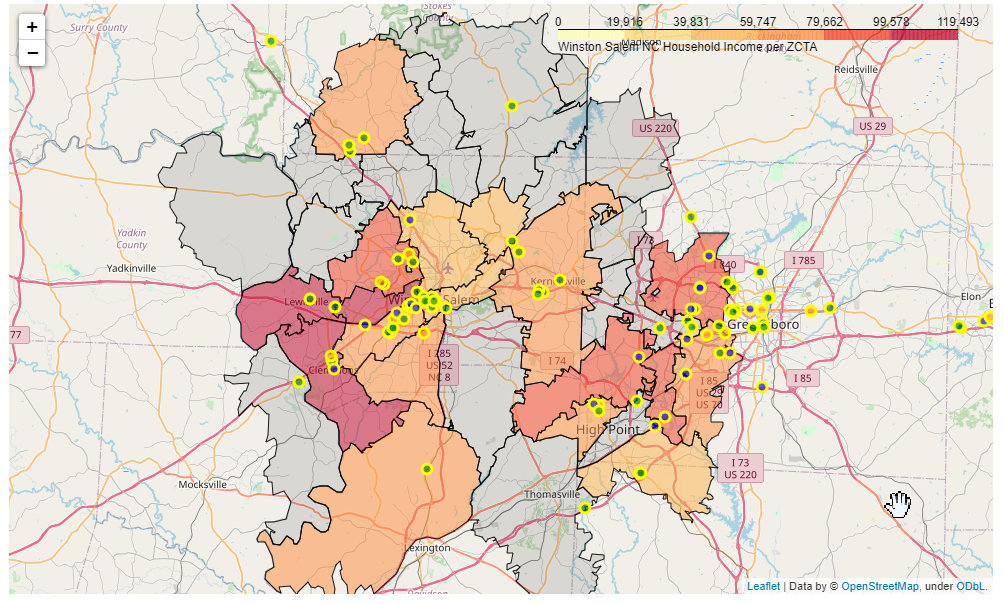
### Population per ZCTA (Below)



Note color scale increases to the right with higher populations.

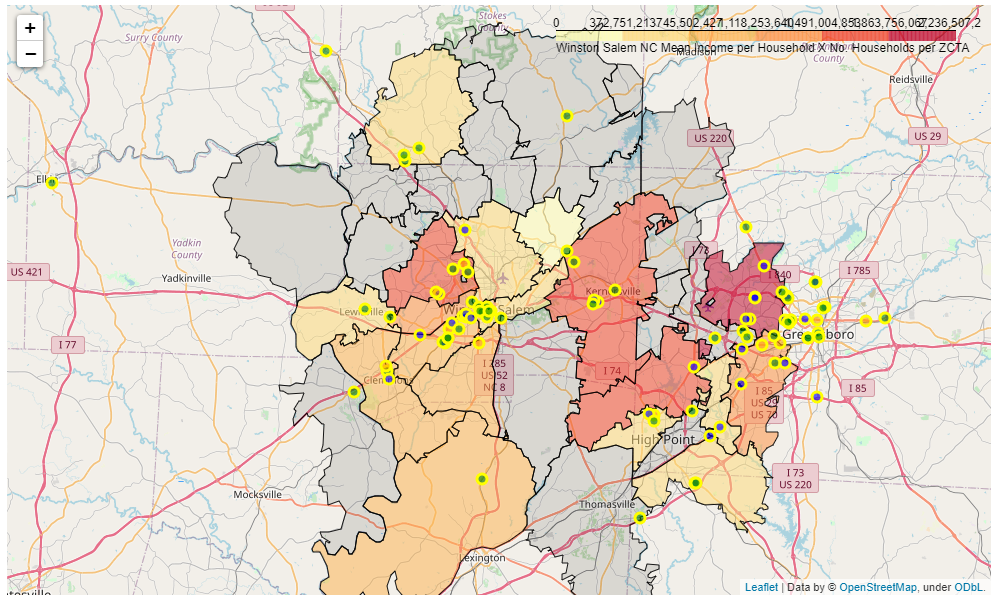
Note: Census Bureau only returns a subset of zipcodes (even requesting less ZCTAs did not provide data for the missing ZCTAs). Time did not permit asking the Census Bureau for additional input on this.

### Household Mean Income per ZCTA (below)

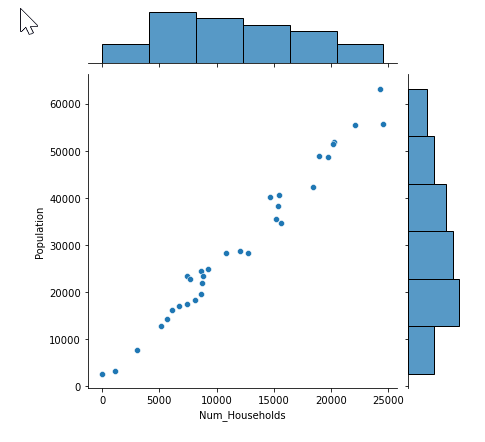


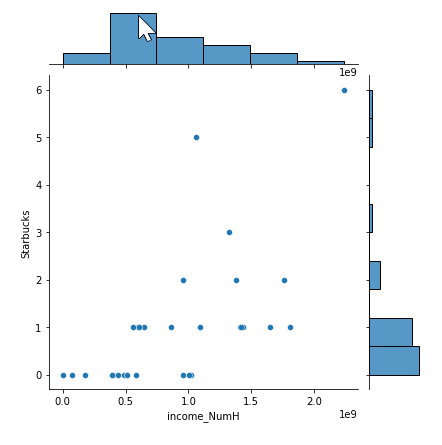
### Multiplying Mean Household Income by number of Households (below)

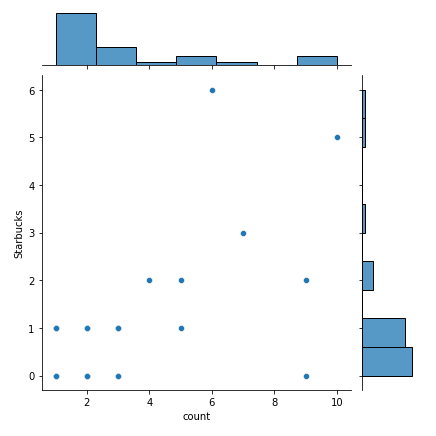
Note: Although this product of Mean Household Income per household times the number of Households does not provide an accurate number, it might be a helpful rough indicator of desirable ZCTAs to investigate.

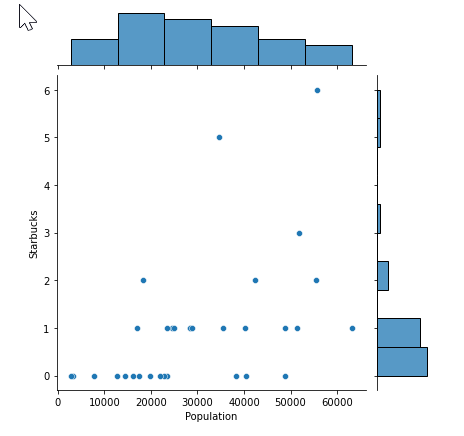


### Attempts to Find Correlations









### Machine Learnings

There were time limits on this study. If the study is extended in time, the following Machine Learning could be valuable: Using multiple features, determine the correlation of those features to the number of national coffee shops in the ZCTA. The features per ZCTA to analyze would be population, , number of other coffee shop businesses located in the area, foot-traffic for nearby retail businesses, retail rental costs, and income. This analysis might be able to determine what features where used by successful national businesses to decide where to locate their coffee shops. The results could then be used to determine where their analysis would recommend the best location for a coffee shop.

## Results

To determine a good location for a coffee shop in the Winston Salem North Carolina area, the following three aspects were considered:

1. Competition
2. Population
3. Income

Considering these criteria, the following Zip Code Tabulation Areas should be considered when looking for a location:

27265: This ZCTA has lower comptetion and higher population and income

The following ZCTAs are not as desirable but should be considered: 27106, 27284, 27410. It is worth noting that there is at least one Starbucks in each of these areas and 27410 has 6 Starbucks outlets while 27265 only has one.

## Discussion

While the three criteria analyzed are not the only considerations when locating a coffee shop, they may be helpful considerations to narrow the areas to review when considering a location. Many other criteria need to be considered.

## Conclusion

This report provides helpful information to a business owner who is looking to locate a new coffee shop in the Winston Salem, North Carolina area. The information can be used to help narrow the areas to consider when looking at locations. Of course, other factors should also be considered (such as rental fees, culture of the neighborhood to see if it is the type of place that people would like to gather at a coffee shop, other shops nearby to generate foot traffic and other considerations.

## Future Studies

In the future, providing more time is allowed, it would be helpful to study the following additional data for consideration:

* Future data studies for consideration:
  + Viewing retail rental fees and commitment timeframes
  + Analyzing foot traffic in the area by looking at Foursquare visits to nearby venues
  + Analyzing neighborhood cultures for receptivity to a Coffee Shop
  + Dividing areas into smaller areas for a more concise study
  + Where other national businesses put their locations and how their locations correlate to existing data (for example: population, other businesses, rental costs, and income). The resulting determination of what those national businesses most likely used in determining their locations might then be applied to the data to determine the best locations for a coffee shop.

### Recognition and Words of Thanks

I would like to thank the following organizations for their assistance to me while working on this report

* The U.S. Census Bureau – several people answered emails and provided help in obtaining the census data and answering questions
* Websites listed in the notebook that provided help in obtaining data and finding ways to work on the data
* Coursera for providing the data science, python and machine learning courses
* IBM for providing the Coursera courses in a learning program to me and others in the cohort
* ADP for providing the time, offering the cohort and courses to me as well as some of my fellow chatbot team members who aided with overcoming some issues I ran into with python (you know who you are)
* My family for being patient with me while spending so much time on this program and this capstone project
* God for giving me the life, time, and ability to understand and apply this learning

This report prepared 2021-01-27 By Henry Will IV

Twitter: henrywill

Linkedin: http://linkedin.com/in/henrywill4

If you want to connect, please provide a personalized reason