

Analyzing Quality Grocery Options in Wake County, NC

Applied Data Science Capstone Project

Jon Puryear

Business Problem

- Wake County, NC is one of the fastest growing areas in the United States. However, there has been concern about access to grocery options in some parts of the county. Some parts of the county have been called a "Food Desert". A Food Desert is an urban area where it is difficult to buy affordable food or good-quality fresh food.
- This analysis looks at the Grocery Options in Wake County to determine if a Food Desert does exist and where a good place would be to open a grocery store that had the potential to both make money and provide food services to an under-served community



Data Acquisition



The Data used in this analysis has been sourced directly from various gold data sources

- Venue information from Foursquare has been used to determine the Quantity and Quality of Grocery Options for a given Zip Code
- US IRS Tax Return Data has been used to determine mean income and number of households by Zip Code.
- Wake County Zip Codes has been used to determine the scope of Zip Codes to include in this analysis
- Geospatial Data for Wake County Zip Codes has been sourced from the US Census Database.

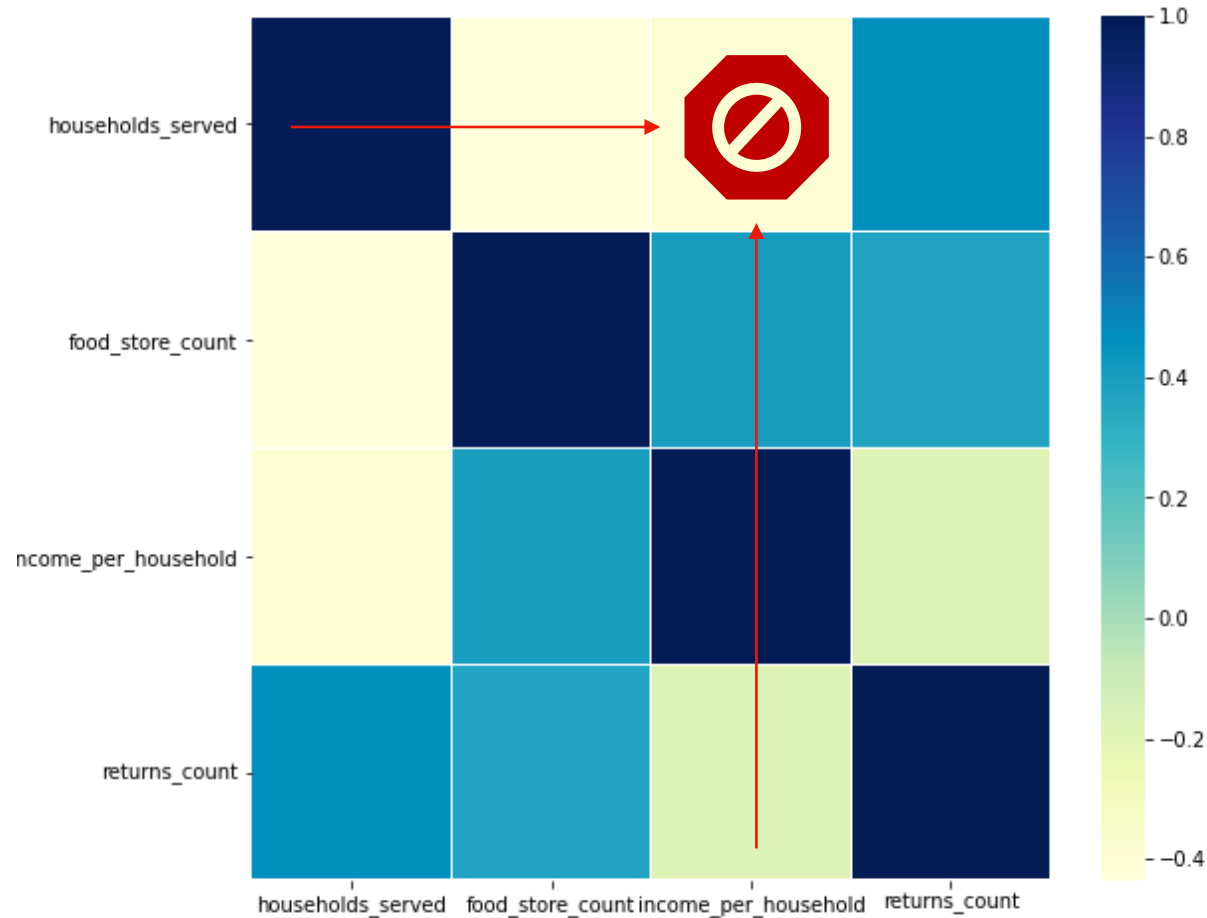
Data Cleaning

Data collection began with downloading the source datasets from Wake County Open Data, IRS, and the US Census

1. Extraneous data was removed, and the IRS and US Census information was filtered and joined to the Wake County Dataset
2. Each record in this combined dataset was passed through a custom python function which collected information from the foursquare API and appended that information into the combined dataset. The dataset was then filtered again to only include food stores such as Grocery Stores and Supermarkets
3. OneHot encoding was then used to convert the categorical Venue data into numeric data for ease of measurement
4. This final dataset was used to calculate the Key Indicators used in the Analysis (**Income per Household** and **Households Served per Food Store**)



Data Visualization and Analysis

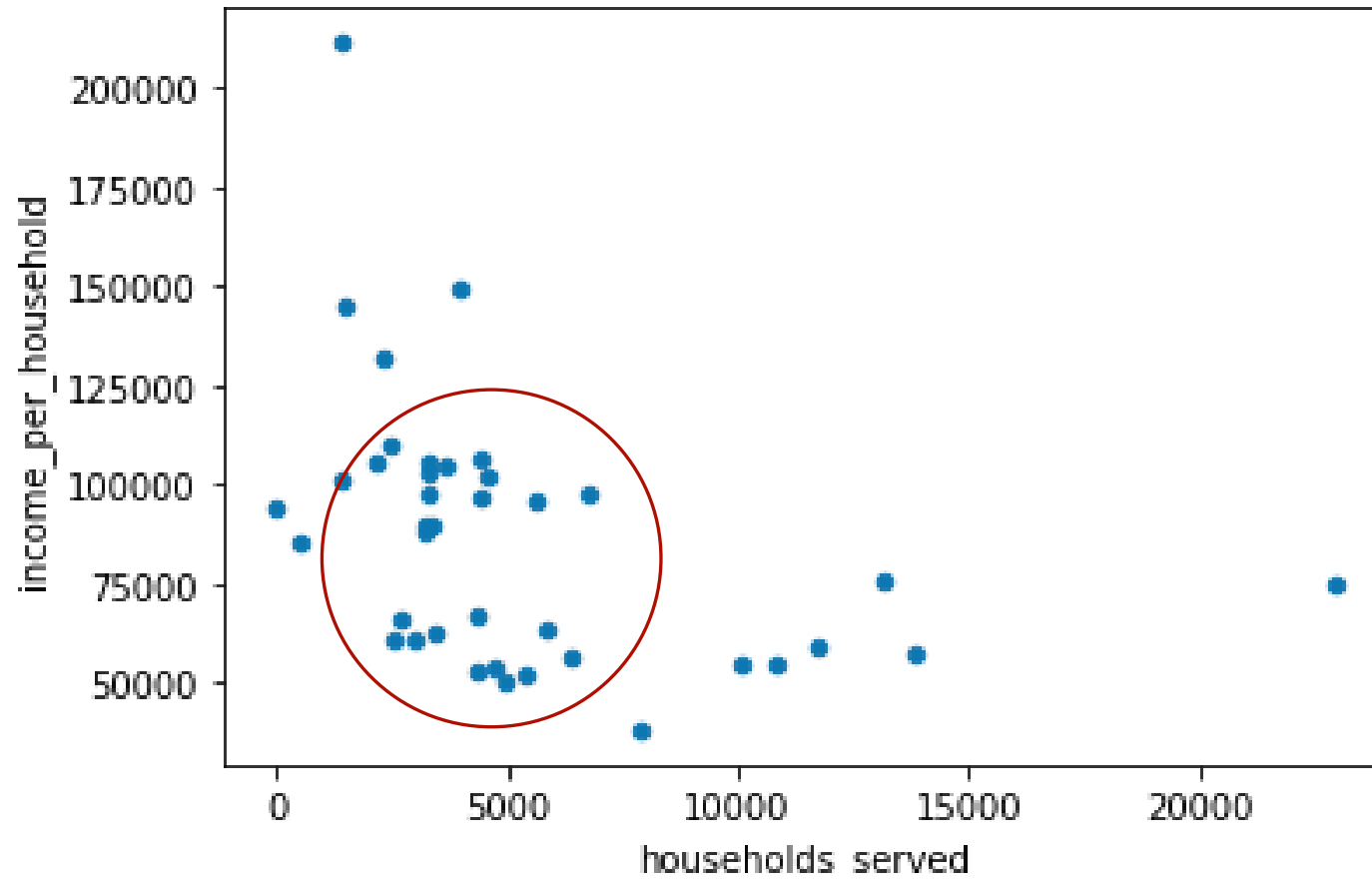


This heat map shows the correlation between the individual Key indicators

One of the questions in the initial business problem was to determine if Quality Of Food Stores and Income was related

Based on this analysis, the answer is that they are not related

Data Visualization and Analysis



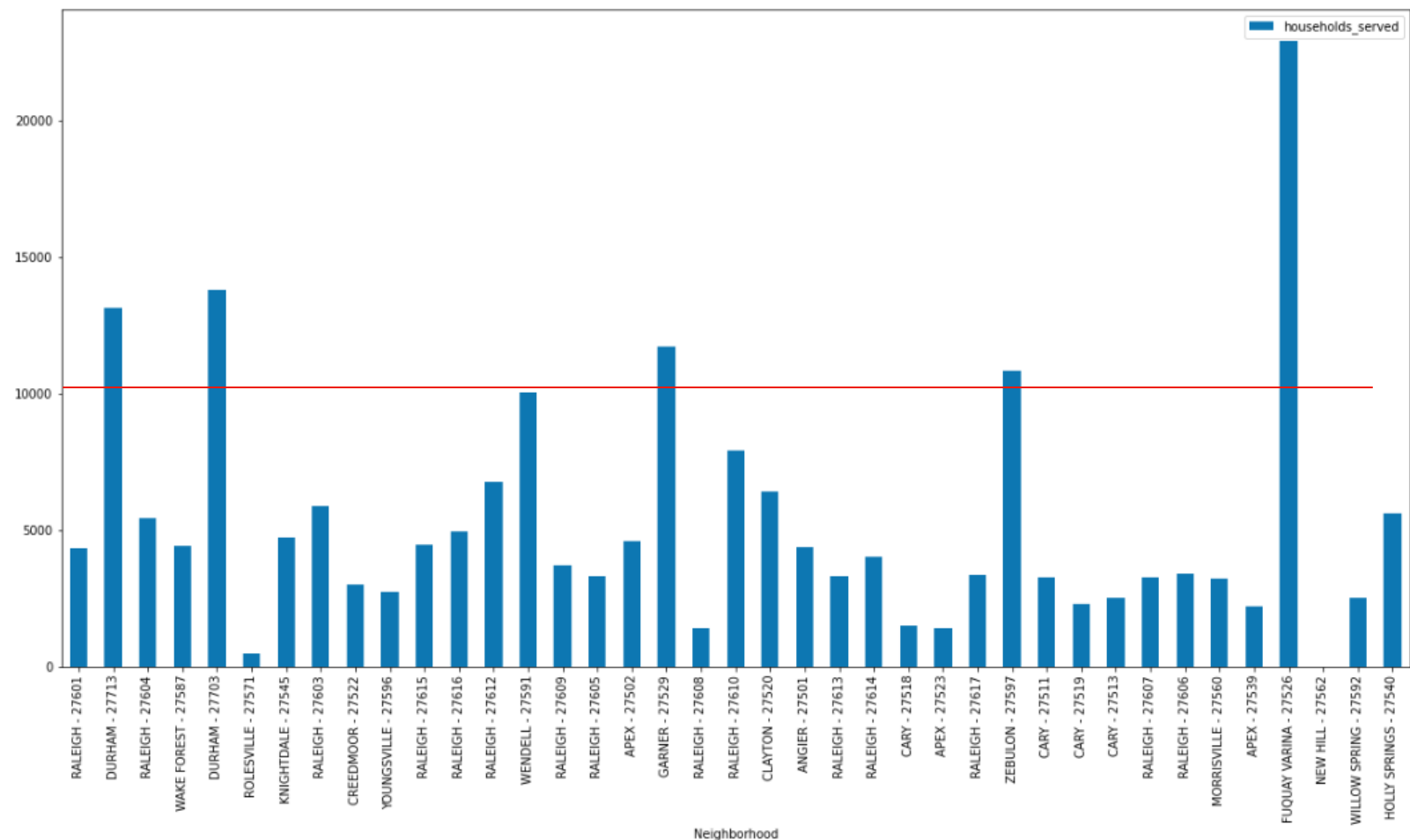
This scatterplot shows the Average Number of Households Served per Food Store versus the Average Income per Household for a given Zip Code

The Majority of the Data is clustered around the 5,000 households per Food Store Mark

This indicates two things

- 1) Income is not a factor in the number of Quality Food Stores
- 2) There are several outliers above 10,000 households per Food Store that could be advantageous to investigate the possibility of expanding food store offerings

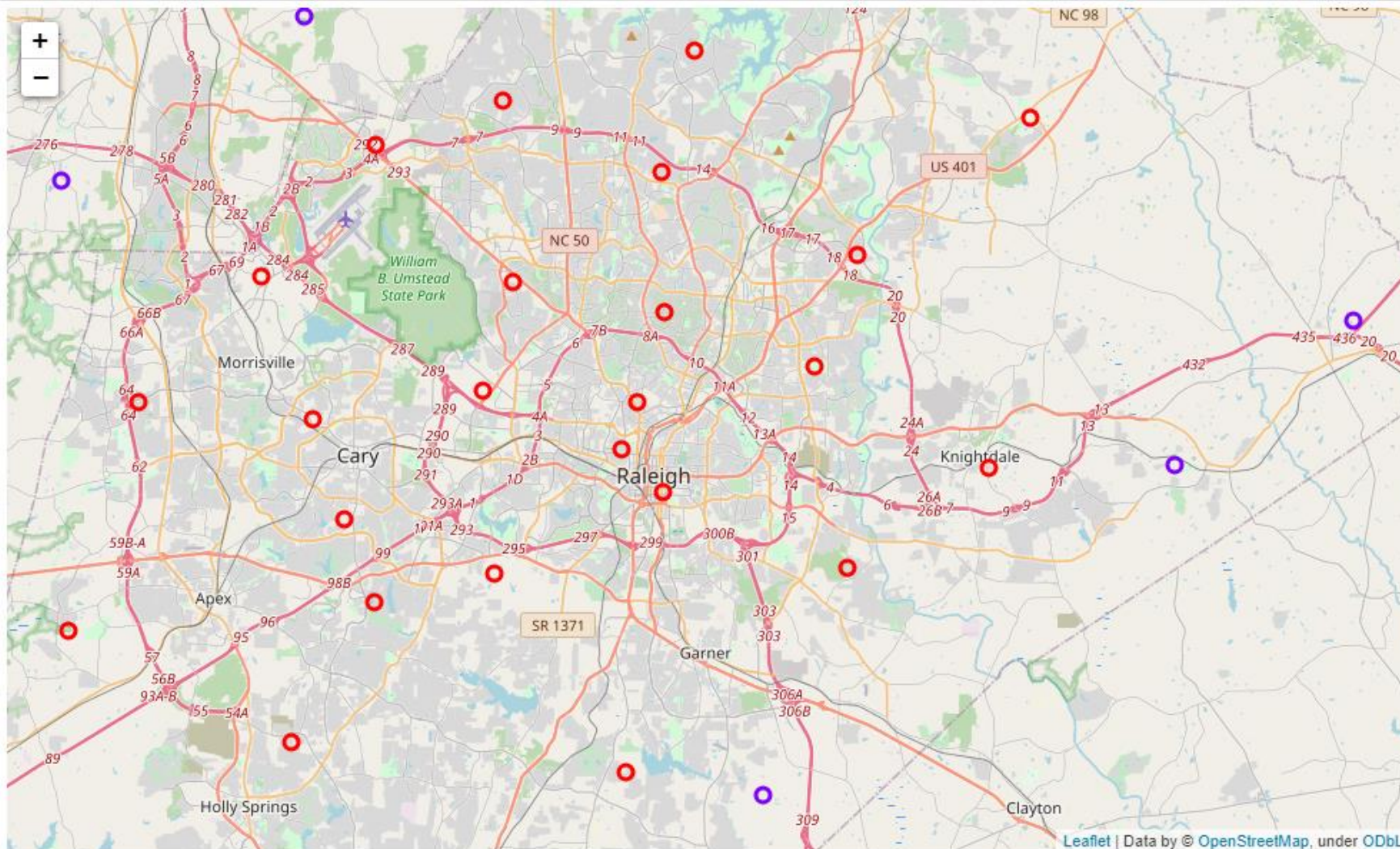
Data Visualization and Analysis



This Bar Chart shows that there are several outlying areas above the 10,000 households per Food Store mark

This chart shows which areas they are. The Red line shows the Threshold to determine if a given Area was under served by Food Store Options

Data Visualization and Analysis



This Map shows the Center Points of the Zip Codes that have been Analyzed

The Red Circles indicate that the area is well served by Food Options,

The Purple Circles indicate that the area is Under Served

As the map clearly shows the purple areas are in the rural outlying areas of Wake County

Conclusions

Based on our analysis of the data, the questions in our initial business problem can now be answered

- A food desert does appear to exist in certain zip codes where the population is “underserved” by quality food store choices.
- This food Desert does not appear to be tied to average income, but to geography. The underserved zip codes appear to be in the Rural areas of the county
- Any of the Zip codes that are under-served would warrant further investigation into provisioning additional food stores by an enterprising company willing to invest in providing high-quality food supplies to a rural environment

