Introduction

Houston Texas is the largest city in the state comprising of 88 super neighborhoods with an estimated population of 2,325,502. Houston is commonly known for its food, job availability, night life, and NASA. With this popularity Houston is classified as one of the fastest growing cities in the nation. With this ever-increasing population comes a growing demand to address the needs of the residents that make Houston home. One of the largest problems facing the city is that of food deserts. Food deserts according to the CDC are "areas that lack access to affordable fruits, vegetables, whole grains, low-fat milk, and other foods that make up the full range of a healthy diet." Individuals living in areas classified as food deserts are shown to have an increased risk of developing diabetes, cardiovascular disease, and being obese. While food deserts can be present in areas typically associated as having low income, they can be present in any area that causes an individual to not have easy access to an available grocery store, and overall, about 23.5 million Americans and a reported 250,000 individuals living in Houston, TX lack easy access to grocery stores according to Rice University.

While living in food deserts are shown to negatively impact ones health, these areas would present lower competition for new grocery stores and allow for a perfect opportunity for business owners to open new grocery stores that not only help alleviate public health concerns, but to also serve the demand for fresh food for individuals that live in impacted areas leading to a greater possibility for a successful store.

Problem

The objective of this project is to find what neighborhoods could possibly be labeled as food deserts and to select the best locations to open a new grocery store that not only alleviates the problem of food deserts but also has the best probability of success due to lack of competition in the city of Houston, Tx.

Interest

This project is useful to a variety of groups, such as public health officials in finding possible food deserts within the Houston community, people looking to move into a neighborhood that has easy access to a grocery store, and finally business owners or developers looking to open new grocery store in an area with lower competition in Houston, Tx. This project is relevant as food deserts are not just a public health issue, but are an untapped resource for business and developers who are looking at finding an area to create a store with the highest probability of success and in areas with low competition.

Source of Data

- The data for the project comes from the page (http://www.houstontx.gov/planning/Demographics/demograph_docs/income_avgs.htm) which contains the list of neighborhoods within Houston as well as their location relative to downtown, Texas totaling to 88 super neighborhoods.
- Geocoder

- Foursquare API

Methodology

The Wikipedia webpage along with the information contained in the table will be scraped using BeautifulSoup and geographical coordinates such as latitude and longitude data will be found for each neighborhood using a geocoder. Further data for each neighborhood will be collected using the Foursquare API. The Foursquare API will allow for the gathering of venue data for each neighborhood such as the frequency and occurrence of groceries stores each neighborhood. Once the frequency of grocery stores are found in each Houston super neighborhood using foursquare, the data will be combined with the location data of the neighborhoods. After combining the two data sets we cl and will be clustered and mapped using the merged data set. Lastly each cluster will be examined to determine the best location to place a new grocery store.

Further data analysis on the clusters will be done using bar graphs to determine the best clusters to open a new grocery store. In addition the best location relative to down town Houston will be

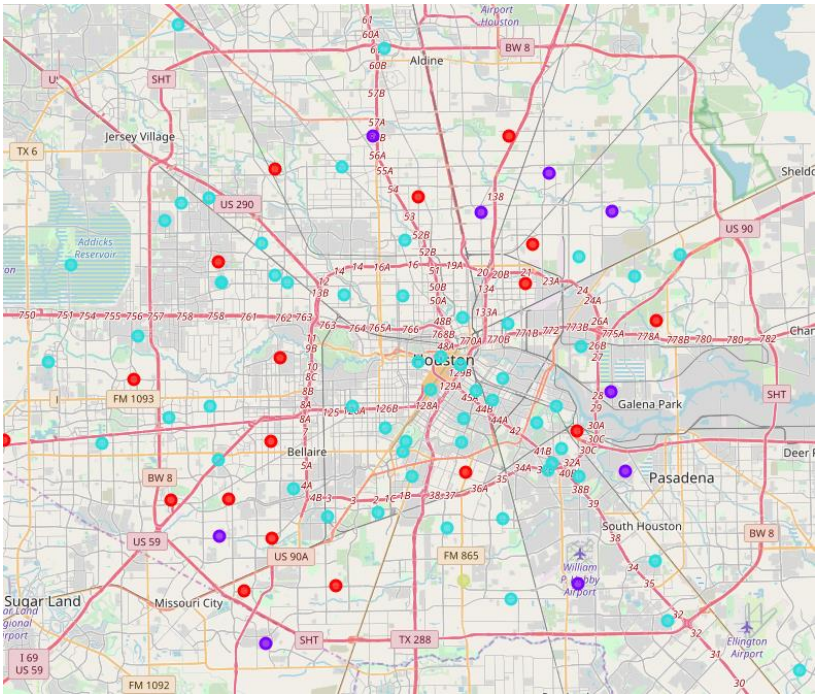
Results

Houston averages 0.024497 grocery stores overall within its 88 super neighborhoods

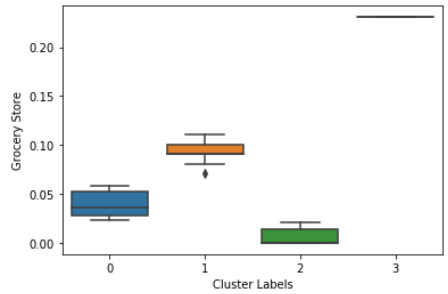
The super neighborhoods were split into a total of 4 clusters

- Cluster 0 (red): Neighborhoods with moderate frequency of grocery stores
- Cluster 1 (blue): Neighborhoods with high frequency of grocery stores
- Cluster 2 (light blue): Neighborhoods with low frequency of grocery stores
- Cluster 3 (green): Neighborhoods with very high frequency of grocery stores

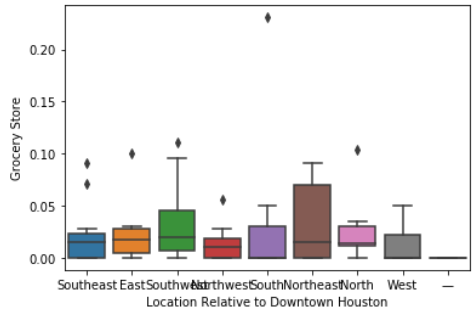
Map. Clustered map of grocery store frequency



Graph1. Box plot of frequency of grocery stores within each clustered super neighborhood in Houston, Texas



Graph2. Box plot of frequency of grocery stores and location relative to downtown Houston Texas.



Discussion

The analysis of the results shows that Houston overall has a low frequency of grocery stores within its super neighborhoods, having a mean frequency of grocery stores being 0.024497. In terms of the clustered data, cluster 2 has the lowest overall frequency of grocery stores by comparison to other groups. Cluster 3 is shown to have the highest frequency of grocery stores by a large margin when compared to other clusters, making it a possible outlier cluster group for grocery stores in super neighborhood within Houston, Texas. In further looking at the frequency of grocery stores based on location relative to downtown Houston, areas located to the Northeast, and Southwest have the largest frequency of grocery stores. The lowest frequency of grocery stores relative to downtown Houston is central Houston showing 0 grocery stores in the area.

Overall the best location to set up a new grocery store would be within cluster 2, as it has the lowest frequency of grocery stores and therefore the lowest amount of competitors present. The overall best location in cluster 2 would be areas in either central Houston or areas within the loop. The worst location would be the Northeast or Southwest as they would be closer to competitors or any areas within cluster 1 or 3.

Conclusion

The goal of this project was to analyze the 88 super neighborhoods that are located within Houston Texas in order to see possible food desert locations as well as determine the best location to open a new grocery store that would have the least competition leading to a possible more successful store. Neighborhood data was collected using a web scrape of Houston neighborhoods located on Wikipedia web page. This data was geolocated and then mapped to check for accuracy. The frequency of grocery stores were then found for each of the 88 super neighborhoods using the Foursquare API. The data sets were then merged and clustered in order to better analyze each super neighborhood. Overall the conclusion of the project is that Houston has a low frequency of grocery stores and a large amount of food deserts, and the best place to open a new grocery store would be in central Houston cluster 2. This project was made using the Foursquare API Sandbox Tier Developer Account, this limits the number of API calls and therefor limits the depth of our search and the results returned. Before a final decision should be concerning the projects results, further consideration should be done that looks at areas other than frequency or competition data before setting up a new store. Other areas such as income level and population density within the defined neighborhoods should also be considered to best determine the optimal location for a grocery store.