IBM Applied Data Science Capstone(Battle of Neighborhoods Houston Texas)

Eirik Kviten

Introduction

- Houston, Texas is one of the fastest growing cities in the United States

- One of the biggest problems facing the city is that of food deserts

- Individuals living in areas classified as food deserts are shown to have an increased risk of developing diabetes, cardiovascular disease, and being obese

- While bad for public health, these food deserts locations are relevant to business as they are areas that are under served or lack competition

- Good for business owners looking to build a new grocery store.

Business Problem

- Find what areas could be considered food deserts

- Use this data to further answer the question what would be the best area for a business owner or property developer to open a new grocery store that it has the least competition leading to a higher possibility of a successful store.

Interested Parties

- Public Health officials

- People looking to move to Houston and wanting to live in a super neighborhood with easy access to grocery stores

- Business owners and project developers looking to build a new grocery store that has the least competition possibly leading to a more successful store.

Data

- Data Required
  - List of Super Neighborhoods in Houston
  - Coordinates of super neighborhoods
  - Frequency of Grocery Stores in each Neighborhood
- Source of Data
  - Wikipedia page showing Houston Super Neighborhoods
  - https://en.m.wikipedia.org/wiki/List_of_Houston_neighborhoods
  - Geocoder for coordinates
 - Foursquare API for grocery store data
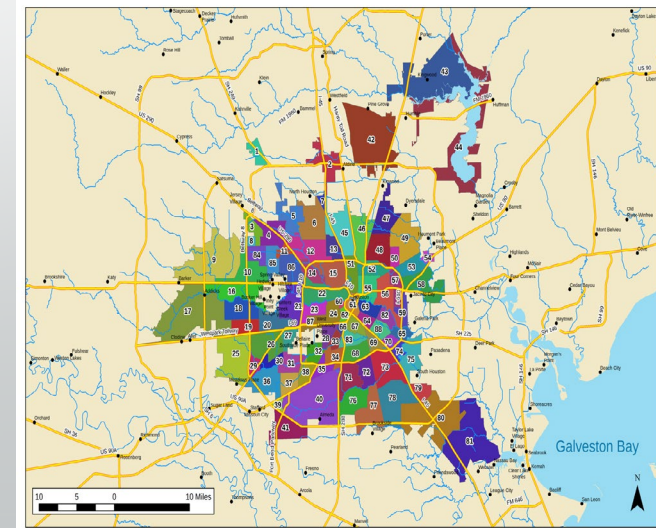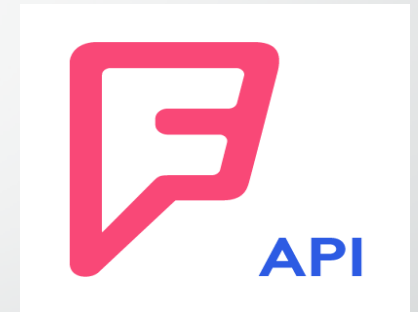
Methodology

- Web scrape Wikipedia page for list of super neighborhoods in Houston, Texas

- Use geocoder to get latitude and longitude coordinates of neighborhoods

- Check neighborhood locations for accuracy

- Use Foursquare API to find frequency of grocery stores within each neighborhood

- Cluster data using k-means clustering and visualize data using Folium

- Further analyze the best cluster to open a new grocery store
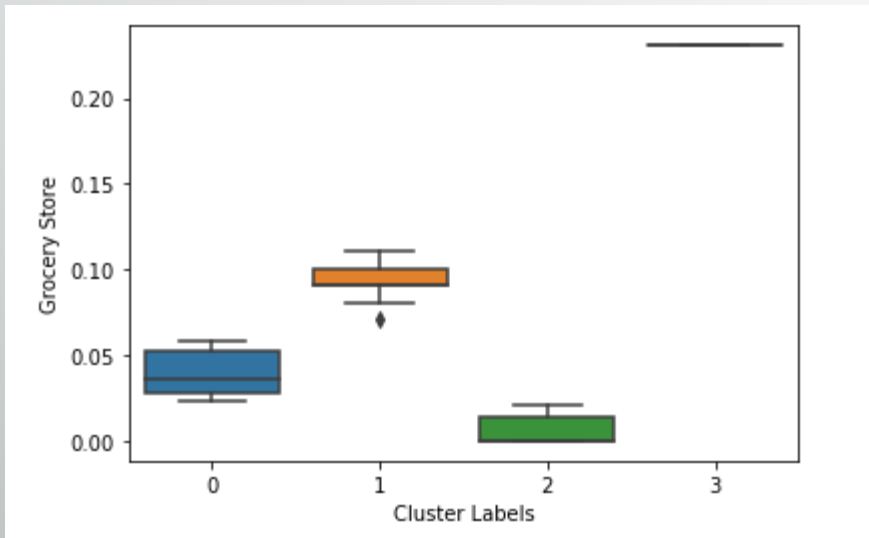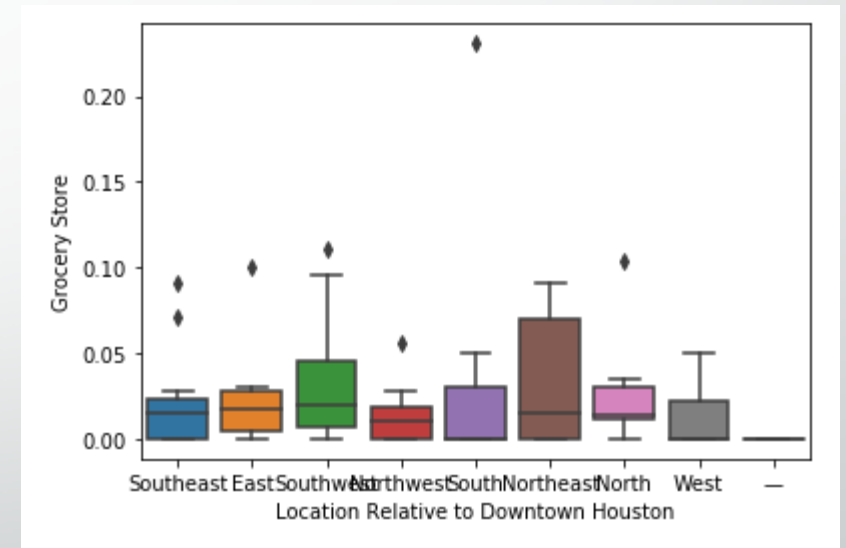
Results

- Neighborhoods were clustered into one of 4 groups

  - Cluster 0: Neighborhoods with moderate frequency of grocery stores

  - Cluster 1: Neighborhoods with high frequency of grocery stores

  - Cluster 2: Neighborhoods with low frequency of grocery stores

  - Cluster 3: Neighborhoods with very high frequency of grocery stores
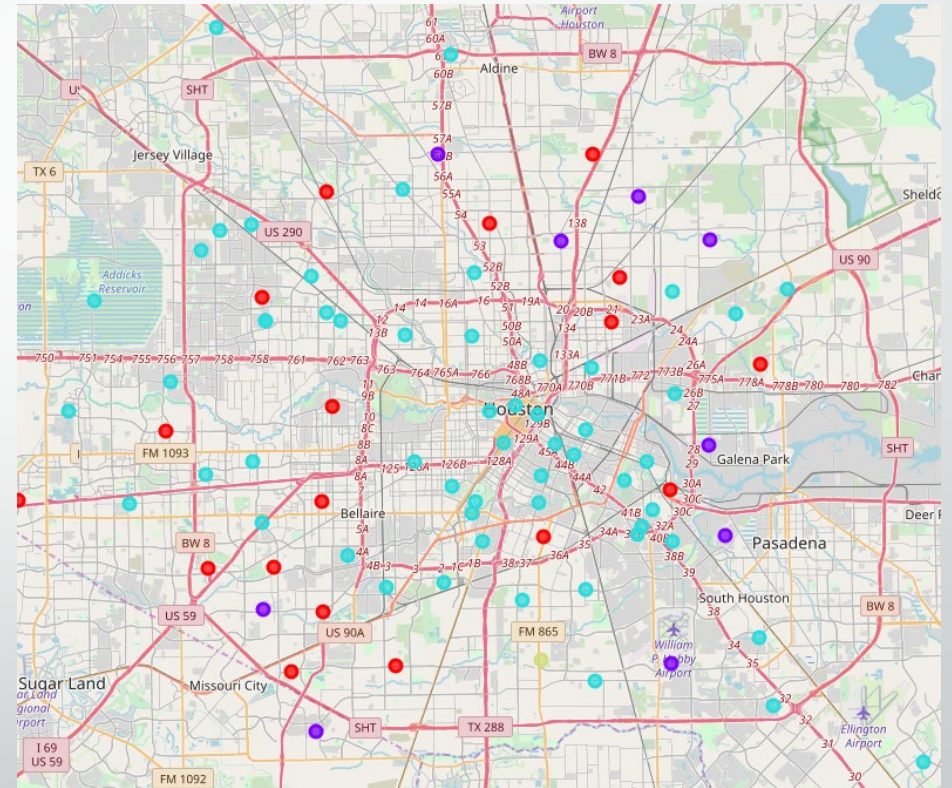
Tables of Results

Discussion/Conclusion

- Houston overall has a low frequency of grocery stores within its super neighborhoods, having a mean frequency of grocery stores being 0.24497. For clusters, cluster 2 has the lower overall frequency of grocery stores by comparison to other clusters. With cluster 3 having the highest frequency of grocery stores by comparison making it a possible outlier cluster super neighborhood in Houston Texas.

- In terms of the best location to set up a new grocery store, cluster 2 would be the best due to the low frequency and competition present, and the best overall location would be either Central Houston or areas within the loop. The worst location would be the Northeast of central Houston or areas within cluster 1.