Predicting the Emergence of Food Deserts

Nigele X McCoy (nmccoy9@gatech.edu), Anthony Philip Lee (alee657@gatech.edu)

Problem Statement

Food deserts, defined as geographic areas with limited access to affordable and nutritious food, have become a significant concern in both urban and rural communities. According to the USDA's Economic Research Service, over 6,500 food desert tracts were identified in the United States based on Census data and information about grocery/supermarket locations. These tracts represent areas where residents have minimal access to fresh, healthy food options. Specifically, 23.5 million people live in low-income areas more than a mile from a supermarket or large grocery store in urban areas, and over 10 miles in rural areas, further highlighting the disparities in food accessibility. [1]

The issue of food deserts is of critical importance due to its profound effects on public health, economic stability, and overall quality of life. The FAO defines food security as a state in which all individuals have access to sufficient, safe, and nutritious food to lead active and healthy lives. Food deserts undermine this goal by restricting access to essential resources, often leading to poorer dietary outcomes and increased rates of diet-related diseases such as obesity and diabetes. In times of economic and supply chain disruptions, ensuring communities have the necessary resources for food security is essential for resilience and well-being. [2]

Furthermore, addressing food deserts aligns with global initiatives, such as the United Nations' Sustainable Development Goal 2, which aims to eradicate hunger by 2030. With the UN projecting that over 600 million people worldwide will still face hunger by 2030, understanding and mitigating the development of food deserts is essential in this broader fight against hunger. [3]

Different studies have previously looked at characteristics and predictors of food deserts. This research has been instrumental in helping shape government policy. These studies have looked at demographic, socio-economic and mobility/transportation factors that are significant predictors of food deserts. However, most of these studies have not incorporated crime rate data into their analysis and may not be able to capture the recent trends of higher level of petty thefts and vandalisms that have driven business away from potentially high risk food desert areas.

This projects seeks to address two key questions: **What demographic, socio-economic, and crime rate factors contribute to the formation of food deserts?** and **How can we predict their emergence?** By leveraging data from the USDA, census reports, and crime statistics, this study will fit classification and clustering models to identify the key drivers of food deserts and potentially forecast their development at the county level.

Related Literature

One of the most relevant studies in this space is a sanctioned study by the Department of Agriculture – *Characteristics and Influential Factors of Food Deserts* by Dutko, Ver Ploeg and Farrigan. In this study, they evaluated US Census data and 2006 data on locations of supermarket. They found that the there are different factors that are common and different between designated urban and rural areas. For both urban and rural areas, they found that minority population, poverty rates and region of the country are significant predictors of food deserts. In rural areas, additional factors like percent of vacant housing and the inverse of unemployment rates. Furthermore, they found different set of predictors for highly dense and less dense urban areas. [4]

Dutko, et. al. used two approaches to their analysis. First, they creatively used Descriptive Analysis to look at the data and the difference between the categories. Second, they used a logit multivariate regression to identify significant predictors for food deserts. They had set up different models that covered the different cases and used a Chow test to determine if the factors are significant.

This study is different from the cited paper by evaluating the impact of crime to predicting food deserts. That is answering the question – is the presence of high crime rates correlated to the emergence of food deserts? In addition, this study will use different classification and clustering models to see if we can identify the same or a different set of significant predictors.

Data Sources

In this study, we utilize several key data sources to investigate and predict the emergence of food deserts by county in the United States. These data sets provide insights into food access, crime rates, and environmental factors that contribute to food insecurity.

1. **Food Access Research Atlas**
   This dataset offers a comprehensive overview of food access indicators for low-income and other census tracts across the U.S. It measures supermarket accessibility through different metrics, helping to assess the availability of affordable and nutritious food in various regions. This data is vital for identifying areas where food access is particularly limited.
   Source: USDA Economic Research Service

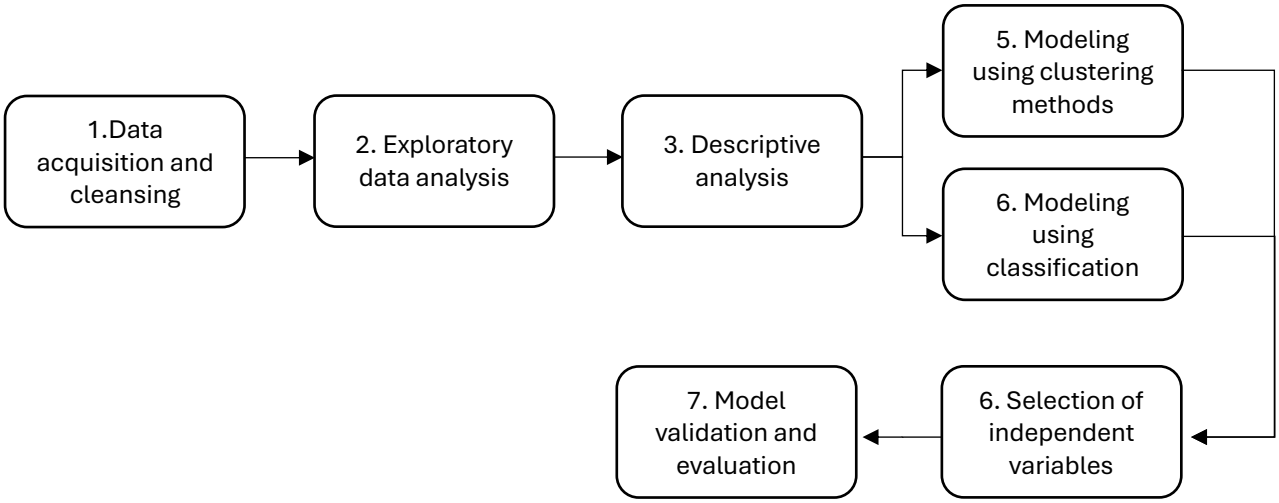
2. **United States Crime Rates by County**
   Crime rates at the county level are an important factor in understanding the broader context of food deserts, as high crime areas may deter grocery store development or contribute to diminished food access. This dataset provides crime statistics across U.S. counties, offering critical insight into how safety concerns interact with food availability.
   Source: United States Crime Rates by County

3. **Food Environment Atlas**
   This dataset highlights various food environment factors, such as the proximity of stores and restaurants, food prices, nutrition assistance programs, and community characteristics. These factors directly influence food choices and diet quality, which are essential for identifying the causes of food deserts and exploring potential policy

interventions.
Source: USDA Economic Research Service

By combining these data sources, we aim to develop a more comprehensive understanding of the factors that contribute to food deserts and use this information to predict their emergence and inform potential solutions.

Methodology



1. **Data acquisition and cleansing**
   Data sets will be collected from the Data Sources identified in the previous section. From here, we'd look to combine the data sets into a common set. This may include efforts to cleanse the data since we expect to be combining from multiple unrelated sources.

2. **Exploratory data analysis**
   At this stage, we will look to identify the summary and shape of our data set. This includes identify and handling missing data, scaling or normalizing the data, correlation analysis, identifying and handling outliers, and potentially introducing categorical variables for certain attributes. We will also evaluate the benefits of dimensionality reduction through principal component analysis.

3. **Descriptive Analysis**
   We will try to understand and interpret the summary statistics of the data set and determine if we can derive qualitative observations. This creates some foundational understanding that would help us interpret the results of the analytical models.

4. **Modeling using clustering methods**
   K-means and k-nearest neighbors (KNN) models will be used to identify food deserts across the data set.

5. **Modeling using classification methods**
   Linear regression and support vector machines (SVM) models will be used to identify significant predictors of food deserts.

6. **Selection of independent variables**
   For the results of both the clustering and classification models, we will apply various methods to select the significant predictors of food deserts. Once this is completed, the models will be retrained using this narrowed list of factors.

7. **Model validation and evaluation**
   With the final models available, an evaluation will be done to determine how each model performs with the task at hand and determine if any model performs significantly better than others in identify food deserts based on the dataset available.

Planned Evaluation Strategies

Selection of Predictor Independent Variables

For the clustering methods, we will look at doing principal component analysis and correlation analysis to identify significant factors for identifying food deserts. PCA will be applied to the dataset and the *explained variance ratio* will be analyzed to determine which factors capture the most important features of the data set. Correlation analysis will be applied using the Pearson correlation metric to identify and eliminate highly correlated variables.

For the classification methods, a Recursive Feature Elimination (RFE) approach will be taken to recursively eliminate unimportant features from the model. In addition, Lasso regularization will be used to penalizes the absolute size of the regression coefficients, forcing less important feature coefficients to become exactly zero, effectively performing feature selection.

Evaluation of Model Performance

To evaluate the performance of the different clustering models, an internal and external evaluation method will be taken. The internal method can use a Silhouette score and/or a Within-Cluster Sum of Squares (WCSS) to identify the effectiveness of each model.

To evaluate the performance of the different classification models, accuracy, precision and recall metrics will be used. In addition, an F1 score will be used to balance between precision and recall in anticipation that the dataset available will be highly imbalanced.

Bibliography

[1] USDA, "Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences," *USDA,* 2009.

[2] F. a. A. O. o. t. U. Nations, "Food Security," June 2006. [Online]. Available: https://www.fao.org/fileadmin/templates/faoitaly/documents/pdf/pdf_Food_Security_Cocept_Note.pdf. [Accessed October 2024].

[3] U. Nations, "Sustainable Development Goals," [Online]. Available: https://www.un.org/sustainabledevelopment/hunger/. [Accessed 16 October 2024].

[4] P. M. V. P. a. T. F. Dutko, "Characteristics and Influential Factors of Food Deserts," U.S. Department of Agriculture, Economic Research Service, August, 2012.