Risk Analysis of the Causes of Food Deserts

Lillian Wang, Audrey Hong, Jaden Chen, Aaroh Tak

Modeling the Future Challenge 2023-2024

Table of Contents

Executive Summary

A food desert is defined as an area with limited access to affordable and nutritious food. As of 2021, food deserts pose a significant challenge to over 3 million residents in Illinois [1]. Research indicates that these areas can exacerbate health issues such as obesity and diabetes [2], with potentially disproportionate impacts across different racial groups [3]. This study seeks to predict the evolution of food deserts in Cook County (the most populous county of Illinois) and the overall state of Illinois, pinpoint their spatial distributions, analyze critical contributing factors, and propose risk mitigation strategies based on these factors.

First, to characterize the food desert risk, we employed several statistical forecasting techniques to project the food desert rate up to 2027 for both Cook County and Illinois. Additionally, utilizing geospatial mapping, we analyzed the evolution and pinpointed the locations of food deserts in Cook County for 2010, 2015, and 2019. Concurrently, through a comprehensive literature review, we identified nine pertinent data sources that provide insight into the factors related to food deserts. By conducting multivariable regression analyses, including ridge regression and random forest, combined with correlation analyses between the food desert rates and these data sources, we identified unemployment, poverty, income, and traffic as the primary actionable factors. Subsequently, we utilized Monte Carlo simulations to predict the Illinois food desert rate in 2023 and quantified risk mitigation for all four actionable factors. We also addressed the chronic disease management costs from poor diets for those living in food deserts and conducted a full cost-benefit analysis of the Supplemental Nutrition Assistance Program (SNAP) used to mitigate the effects of Illinois food deserts, estimating $32 million in annual benefits.

Our analysis yielded four key recommendations aimed at reducing the incidence and impact of limited access to nutritious food. The first is to increase access to supermarkets, particularly in low-income areas, through incentivizing development and improving transportation infrastructure to ensure accessibility. We also propose expanding the SNAP program [38] and advocating for other incentives to promote healthier dietary choices and limit fast food consumption. Moreover, providing unemployment benefits to those struggling with the cost of living would enable them to take advantage of healthy food options in newly developed grocery stores. Lastly, enhancing access to farmers' markets is suggested as a practical and cost-effective solution, supported by zoning policies and development subsidies. All of these recommendations aim to address the underlying issues contributing to food deserts and promote healthier food environments within the community.

1. Introduction

Most would think that accessing fresh and nutritious food is a simple chore in the modern era. And yet, almost one in every six Americans struggle to put fresh and nutritious food on their tables daily **[4]**. The United States is facing its worst hunger crisis in generations **[6]**. Food bank lines stretch for miles, millions of children go to bed hungry every night, and countless families are forced to choose between buying groceries and paying rent. A report by the U.S. Department of Agriculture (USDA) found that between 2021 and 2022, the number of people living in food-insecure households jumped from 33.8 to 44.2 million, with a 45% increase in children experiencing food insecurity **[5]**.

Addressing this issue involves understanding the terms food insecurity and food deserts. The USDA defines food insecurity as the lack of access to sufficient food due to distance or affordability. A food desert is a geographic area that meets the USDA's criteria for food insecurity **[49],** typically defined as residing more than one mile away from supermarkets in urban areas or over ten miles in rural regions **[37]**. We analyzed the available food insecurity data to gain insight into the causes of food deserts. We define the food desert rate in our paper to be equivalent to the food insecurity rate.

In Illinois, food desert rates have persisted at 10% in recent years, affecting communities all across the state. This issue is particularly acute in the city of Chicago, where over 500,000 residents struggle with limited access to fresh food or produce **[16]**. Without these necessities, diet diseases such as obesity and diabetes thrive **[2]**. From complicated origins such as redlining and zoning laws **[9]**, food deserts are now attributed to differing factors of unemployment rate, population, poverty rate, and transportation from which large cities like Chicago are especially affected **[8]**. There is also significant disparity across different racial lines: 19.1% of Black households and 15.6% of Latinx households experienced food insecurity, whereas the rate for White households was 7.9% **[4]**.

The public must understand the underlying factors that contribute to this food injustice, and take the necessary actions to mitigate the effects of this in the future. Thus, this paper aims to predict the factors that contribute to the prevalence of food deserts in Illinois and Cook County (which encompasses the city of Chicago), as it affected 3.3 million residents in Illinois in 2021 **[1]**. Based on these factors, we propose four main recommendations to reduce the incidence of food deserts and mitigate the adverse effects of limited access to nutritious food.

2. Data Methodology

2.1 Collecting Data

The data for our study is derived from the following nine sources: Feeding America food insecurity data, personal income data from the Bureau of Economic Analysis, poverty data from the U.S. Census, population data from the U.S. Census Bureau, crime data from the Chicago Police Department, weather data from the National Centers for Environmental Information, traffic data from the Illinois Department of Transportation, unemployment data from the Chicago Police Department, and geospatial data from the U.S. Census Bureau. These sources are crucial for elucidating the relationship between food deserts and their driving factors, as well as for forecasting trends to predict the future occurrence of food deserts.

1) ***Feeding America Food Insecurity Data* [10]**

   Rationale: This paper is focused primarily on food deserts, which are defined as geographic areas that meet the USDA's definition of food insecurity (low access and low income) **[49]**. We obtained food insecurity data from Feeding America, a nationwide charity focused on ending hunger and has been widely credited for being credible and transparent.

   Variables: This dataset provides us with the food insecurity rate at county and state levels in the U.S. by year from 2009 to 2021. Because the food insecurity rate is only provided by year, we will have to create annualized data for our other analysis variables.

   Objective: We will study food insecurity data together with other variables listed below, building a multivariable model to predict the prevalence of food deserts in Illinois overall.

2) ***Bureau of Economic Analysis Personal Income Data* [11]**

   Rationale: A USDA report suggests that low-income areas are more likely to become food deserts **[49].**

   Variables: This yearly dataset provides the average personal income (defined as the mean income **[17]** calculated for every man, woman, and child in a particular group including those living in group quarters) for the populations of each county in the USA from 2000 to 2022. We aggregated it to the state level.

   Objective: We will analyze this data to determine exactly if the average personal income amount of the population of an area is related to food deserts.

3) ***US Census Poverty Data* [12]**

   Rationale: Similar to how low-income areas are more likely to become food deserts, areas with high poverty rates have a similar risk. We chose to analyze both personal income and poverty data as factors because they may have slightly different correlations to food deserts.

   Variables: This yearly dataset contains the total number of people in poverty and the poverty rate at the county level during 2009 - 2022. We aggregated it to the state level.

   Objective: We will analyze if this poverty data correlates to the personal income data, and determine its relationship to the prevalence of food deserts.

4) ***US Census Bureau Population Data* [13]**

   Rationale: The USDA predicts that population growth will cause the demand for food to increase rapidly **[28]**. Additionally, areas with a low population may lack grocery stores and other food options due to the lack of abundant customers.

   Variables: This dataset provides yearly population counts per county from 2010 to 2022, including race population counts. We aggregated it to the state level.

   Objective: We will analyze this data to determine exactly if the population of an area is related to the prevalence of food deserts.

5) ***CPD Crime Data* [14]**

Rationale: According to a study published by the National Library of Medicine, high crime rates can increase food deserts by discouraging customers from traveling on foot or using public transportation due to safety concerns **[35]**. This can reduce the sales of grocery stores, making it more difficult for the store to stock perishable, healthy food items.

Variables: This monthly dataset provides all reported crimes in Illinois from 2001 to 2024, as well as the type of crime, location, and date. Our multivariable analysis aggregates it into the state and year level.

Objective: We will analyze this data to determine exactly if the frequency of crime is related to the prevalence of food deserts.

6) ***NCEI Weather Data (Temperature and Precipitation)*** **[15]**

Rationale: Weather conditions have a strong impact on human behavior, especially criminal activity and transportation, so it is necessary to take this factor into account **[23]**. Also, precipitation can hinder access to grocery stores because of rain, sleet, freezing rain, snow, or hail **[16]**.

Variables: This dataset provides the monthly average temperature and amount of precipitation at the county level in Illinois from 1895 to 2024. To prepare the variables for the study we utilize annual precipitation and temperature data for the state.

Objective: We will analyze this data to determine if changes in weather conditions are related to food desert prevalence. They are independent variables for our multivariable regression models.

7) ***Illinois Department of Transportation Traffic Data*** **[18]**

Rationale: Having convenient and accessible methods of transportation has been found to improve one's ability to purchase healthy food, with approximately 20% of U.S. households having transportation barriers that limit their ability to buy food **[39].** Traffic congestion can be roughly estimated by how many cars are on the road thus we use the vehicle miles in the county to estimate congestion and ability to buy food.

Variables: This yearly dataset provides the average vehicle miles of traffic at the county level in 1997-2022. We aggregated it to the state level.

Objective: We will analyze this data to determine if vehicle miles traveled are related to food deserts.

8) ***CPD Unemployment Data*** **[19]**

Rationale: Unemployment rates are key indicators of economic health. Understanding trends in unemployment can provide insights into the overall economic stability of a region, which is important for food desert study.

Variables: This Illinois unemployment monthly rate is at the state level during the timeframe 2000-2023. We aggregated it to the yearly level.

Objective: We will analyze this data to determine if the unemployment rate is related to food deserts.

9) ***Illinois Geospatial Data*** **[20]**

Rationale: When studying food deserts over time in the context of Cook County and Illinois, it helps to better understand where the food desert occurs and its time frame **[21]**.

Variables: Illinois geospatial data, such as census tract boundaries, urban or rural census tracts, and location of supermarkets and farmers markets in the years 2010, 2015, and 2019.

Objective: Utilize geospatial data and the food desert definition to identify locations of food deserts for the years 2010, 2015, and 2019. This approach enables us to observe the evolution of food desert areas over time as well.

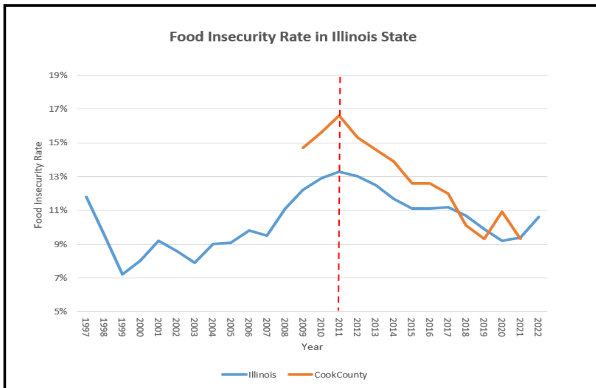
2.2 Exploratory Data Analysis (EDA)

Our initial look at the data revealed that some modifications to the data were necessary:

- For data sources that we could find monthly but not yearly data, we performed data aggregation by taking the average, namely for our weather temperature and precipitation data.
- For data sources available only at the county level rather than for the entire state of Illinois, we aggregated the data by summing data for all counties, as we did with the traffic vehicle volume data.
- For Cook County thefts, we refined the raw crime data to exclusively include theft incidents, encompassing both general thefts and specifically those targeting grocery stores.
- For all of our data in general, if only one data point was missing, we did an imputation with an average to preserve its distribution.
- For the dataset used in the multivariable regression model, we performed min-max normalization to keep a consistent scale, improve model performance, and have better interpretability.

3. Data Characterization

3.1 Food Insecurity Rates in Illinois & Cook County Over Time

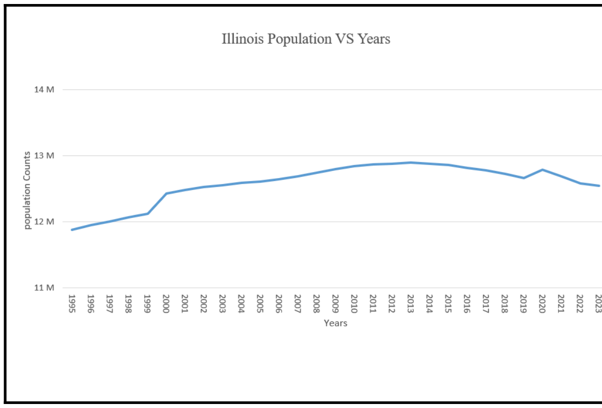


<u>Observation</u>: The food insecurity rates for both Illinois and Cook County peaked in 2011. Since then, Illinois' overall food insecurity rate (1997-2022) has exhibited a notable downward trend ($R^2 = 0.83$), while Cook County's rate (2009-2021) has shown an even stronger downward trend ($R^2 = 0.93$), possibly influenced by missing 2022 data. The correlation of the food desert rate between Cook County and Illinois during 2009-2021 is 0.94.

Figure 1: Food Insecurity Rate vs. Years (Illinois & Cook County, IL)

3.2 Population by Year & by County

<u>Illinois Population Over Time</u>



Observation: Illinois' population grew steadily from 1996 to 2000, reaching a peak of 12.48 million. Then there was a slight decline until around 2005, followed by stability with minor fluctuations. By 2010, the population had modestly increased again but saw a slight decrease by 2023. Overall, Illinois's population showed a pattern of growth, stabilization, and some fluctuations during this period.

Figure 2: Illinois Population Counts Over Years

County Population Distribution in Illinois

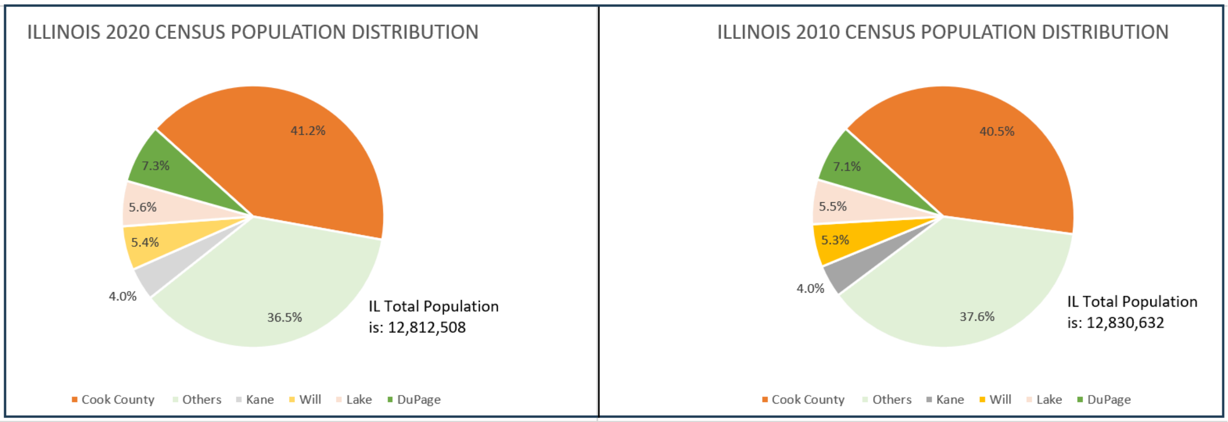


Figure 3: Illinois Census Population Distribution Among Major Counties

Observation: In the 2010 census, Cook County contained 40.5% of the Illinois' population. In the 2020 census, Cook County remains 41.2% of Illinois' population. Being the most populated county in Illinois, Cook County results should be important for all of Illinois. Cook County's percentage increased because Illinois' total population slightly decreased from 12.83 million to 12.81 million between 2010 and 2020, Cook County's population increased from 5.19 million to 5.28 million.

3.3 Personal Income by Year

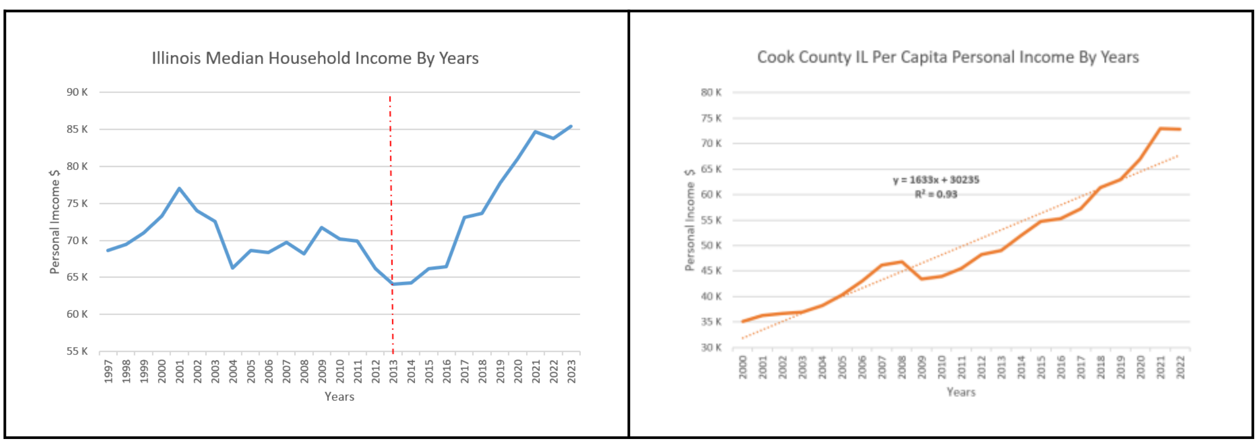


Figure 4: Income Data vs. Years (Illinois & Cook County IL)

Observation: The left graph shows Illinois' median household income amount from 1997 - 2023, we can see it dips in 2013. From 2000 to 2023 the median household income amount is increasing (the trend line $R^2 = 0.24$). The right graph shows Cook County residents' average personal income during 2000 - 2022. As time goes on, the Cook County personal income increases, which has a strong positive trend ($R^2 = 0.93$).

3.4 Poverty Rate Over Time (in Years)

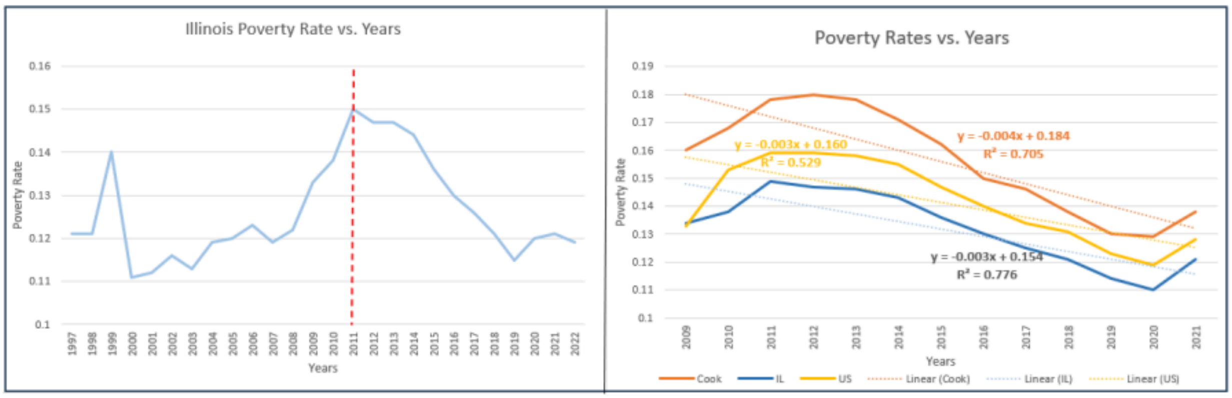


Figure 5: Poverty Rate vs. Years (Cook County IL & Illinois & US)

Observation: Illinois's poverty rate peaked in 2011. The average poverty rate during 1997-2022 in Illinois is 12.6%. The general trend for poverty rates in Cook County, Illinois and the US are all generally decreasing from 2011 with a slight increase near 2020 and 2021 which the COVID-19 pandemic may explain.

3.5 Traffic Data Over Time

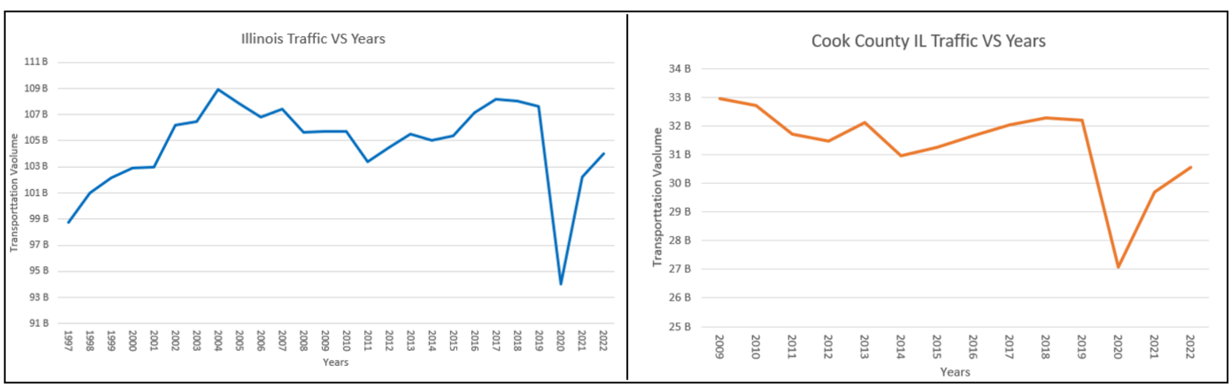
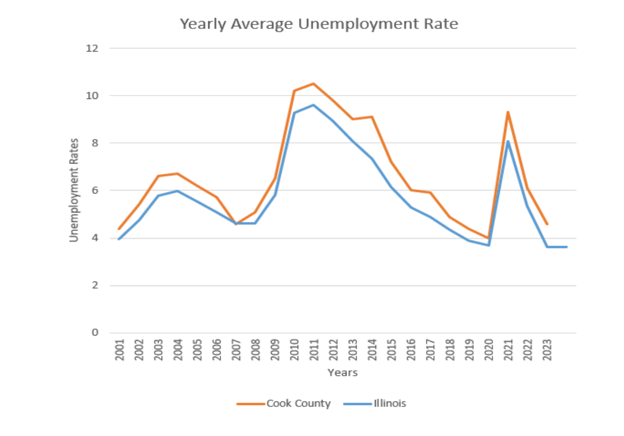


Figure 6: Traffic Data vs. Years (Illinois & Cook County IL)

Observation: The left graph is Illinois traffic data for 1997 - 2022, and the right graph is Cook County traffic data for 2009 - 2022. There might be a downward trend in traffic volume, but the COVID impact is very obvious in both graphs. Traffic volume dips around 2020 which is explained by the COVID pandemic as people were quarantined at home.

3.6 Illinois Unemployment Rate Over Time



Observation: Unemployment showed 3 spikes during 2000-2023. Unemployment spikes in 2003 and 2009 correspond to market crashes as shown in the SPY ETF price history provided in Appendix A. The unemployment spikes around 2020 can be explained by the COVID-19 pandemic. Cook County has always had higher unemployment rates compared to the state average due to various factors such as economic disparities, city issues, and industry composition [22].

Figure 7: Unemployment Rate vs. Years (Illinois & Cook County)

3.7 Total Crime Indices Over Time

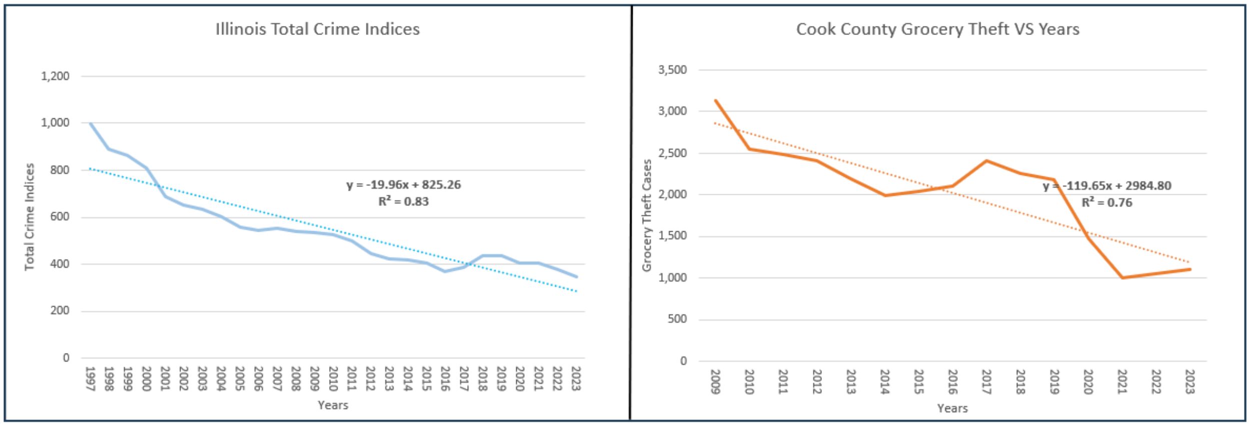


Figure 8: Crime Cases vs. Years (Illinois & Cook County IL)

Observation: The left graph is for total crime indices in Illinois during 1997 - 2023, which is good for a relative comparison of the total number of crimes over the years. The graph on the right is for the number of grocery theft cases in Cook County during 2009 - 2023. Both illustrate a general downward trend over time.

3.8 Weather Data Over Time

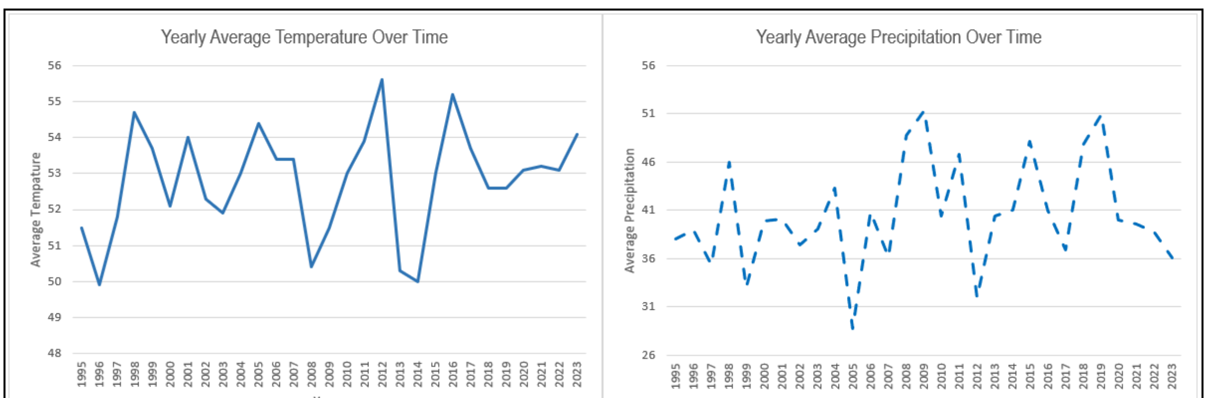
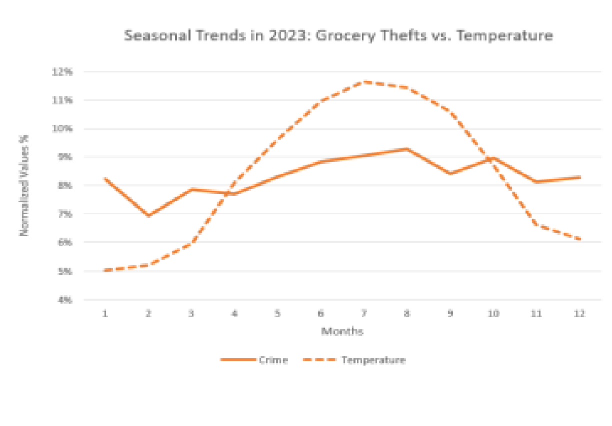


Figure 9: Illinois State Yearly Average Temperature (Left) and Precipitation (Right)

Observation: Figure 9 shows the yearly temperature and precipitation for Illinois from 1995 to 2023 (Appendix B has Cook County average monthly weather graphs). Precipitation levels vary throughout the years, with some years experiencing higher precipitation than others. The correlation between temperature and precipitation is 0.49, a weak correlation, so we used both variables in the later multivariable analysis. The correlation is positive which indicates higher temperatures are associated with increased precipitation tendency.



Using Cook County data, we find that the correlation between grocery theft and weather temperature over 12 months for 2023 is 0.75. This suggests that incidents of theft tended to rise during warmer weather conditions. Previous literature supports that there is an increase in crime rates during warmer weather **[23]**. Moreover, our observation of peak occurrences of grocery theft aligning with the summer season provides further confirmation of this correlation.

Figure 10: Cook County IL Monthly Grocery Thefts & Temperature in 2023

4. Mathematics Methodology

Our main goal is to identify the key factors contributing to food deserts in Illinois and develop effective strategies to address them. To achieve this, we performed 5 types of analyses:

1. Geospatial analysis for the years 2010, 2015, and 2019 to virtually identify the locations of food deserts in Cook County, IL
2. Statistical forecasting (ARIMA, Prophet, and ETS) to predict food desert prevalence until 2027
3. Pearson correlation analysis to access the relationships between different factors and the prevalence of food deserts, including examining pairwise correlations among these factors
4. Multivariable regression models: symbolic regression, ridge regression, and random forest are built to identify the factors that predict food deserts. We identify the best model based on accuracy metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE), along with the considerations for model interpretability and result understanding
5. Monte Carlo simulations to quantify risk mitigation

4.1 Assumptions

To patch missing information, we made the following assumptions:

I. The food insecurity data is used to identify food deserts.

Reasoning: The food desert datasets identified in Illinois span only three years and are at the census tract level. Annual food insecurity data is available for Illinois with 26 data points. We leverage this information to identify food deserts, as they are defined as geographic areas that meet the USDA's criteria for food insecurity.

II. The census tract population is assumed constant for geospatial data 2010 - 2019.

Reasoning: The food access geospatial data for 2010, 2015, and 2019 relies on the census population data in 2010. It assumes that the population distribution observed in 2010 remains consistent from 2010 to 2019.

III. *The entirety of Cook County can be considered as an urban region.*

Reasoning: In the 2010 Cook County census food access data, the number of rural tracts was very limited. This suggests the necessity of an assumption that Cook County is primarily an urban area.

IV. *The frequency and location of grocery store thefts provide an accurate indication of the locations of grocery stores in Illinois.*

Reasoning: We were unable to obtain geospatial data for grocery stores for all years. As an alternative approach, we utilized data on grocery store thefts. Although grocery stores can experience varying degrees of theft, the locations of theft can still be used as an indicator of where grocery stores are.

V. *There exists a high correlation between the total crime rate and incidents of grocery theft.*

Reasoning: We need this assumption to analyze the crime index for Cook County, IL. We don't have a total crime index for Cook County, we do have it for Illinois State. Figure 8 shows these two have downward trends. Our model factors use total crime rates. Cook County's total crime rate is represented by grocery theft data.

VI. *For data sources providing monthly but not yearly data, we aggregated the data by calculating the average. This assumes that the yearly data effectively captures insights from the monthly data.*

Reasoning: Food desert data lacks seasonality. So, the fluctuations between months do not have to be considered.

4.2 Model Development

4.2.1 Geospatial Visualization

Geospatial analysis, a widely used analytical technique leveraging Geographic Information Systems (GIS) data, holds particular significance in the study of food deserts. Unlike conventional analyses relying solely on data points, Geospatial analysis uses spatial data to provide a more accurate representation of real-world phenomena. Utilizing a list of stores selling fresh food, we can employ the following measure of food access **[24]** based on proximity to supermarkets.

We define food deserts as low-income census tracts where a significant portion of the population (at least 500 people or 33%) is over 1 mile (urban) or 10 miles (rural) from the nearest supermarket **[37]**. Around 19 million people, or 6.2% of the U.S. population, belong in this area. To determine the locations of food deserts, we:

1. Draw Cook County census tract boundaries
2. Separate urban and rural census tracts
3. Mark the location of supermarkets and farmers' markets
4. Calculate the distance based on the USDA definition
    - Low Income: The poverty level in a census tract is ≥ 20%
    - Low access: At least 33% of the census tract is >1 mile away from a fresh food source in urban areas or >10 miles away for rural areas
5. Find the proportion of low-income census tracts that are too far away from sources of fresh food, resulting in a possible food desert

4.2.2 Statistical Forecasting

We also employed three distinct statistical forecasting methodologies to predict the future food desert rate in Illinois and Cook County. These methodologies are:

1. Exponential Smoothing (ETS) **[42]**
   - Strengths: Simple to understand and implement, can handle time series data with various patterns
   - Weaknesses: Cannot detect complex trend and seasonality patterns, performance deteriorates when predicting too far into the future
2. Facebook Prophet algorithm **[43]**
   - Strengths: Automatically can detect trends and multiple seasonality patterns in data, can handle missing data easily
   - Weaknesses: More difficult to understand and control underlying mechanics
3. ARIMA (Auto-Regressive Integrated Moving Average) **[42]**
   - Strengths: Can capture both short-term correlation and long-term trends, can accurately predict far into the future
   - Weaknesses: Lacks flexibility by assuming linearity in data, requires selection of parameters, like the number of the lagged observations, the degree of differencing, etc.

Utilizing multiple models provides a deeper insight into analyzing the trends in data. Each of these models has unique strengths and weaknesses that contributed to our understanding of our topic, and in the end, we chose to use the ARIMA model to characterize the time series trends.

4.2.3 Multivariable Regression Modeling

Similarly, we employed five different multivariable modeling techniques to determine the relative importance of various factors, including crime, traffic, unemployment, total population, poverty rate, income, weather temperature, and precipitation. These methodologies are pairwise correlation, symbolic regression, ridge regression, random forest, and Monte Carlo. We determine the strengths of all of these model methods and decide which one is the most optimal for characterizing our risk and predicting the growth of food deserts.

- **Pairwise Correlation Analysis**
  - **Purpose**: Provides a measure of the linear relationship between each pair of variables. Here we have continuous variables so Pearson correlation **[40]** is used
  - **Strengths**: Simple and straightforward, easy to understand
  - **Weaknesses**: Limited to linear relationships, susceptible to outliers, doesn't capture interactions between multiple variables simultaneously
- **Symbolic Regression** **[41]**
  - **Purpose**: Helps explain complex models by generating interpretable equations, aiding in understanding how input variables contribute to the output
  - **Strengths**: Can capture complex interactions that traditional models might miss, easy interpretation of results

- **Weaknesses**: With risk of overfitting, particularly when the search space is not constrained effectively or when the data contains noise or irrelevant features

- **Ridge Regression [45]**
  - **Purpose**: Produces a linear equation that predicts the target variable based on multiple predictors. The relative importance of features can be determined by coefficients
  - **Strengths**: Good at handling many highly correlated independent variables, prevents overfitting and improves generalizability, has more stable parameter estimates
  - **Weaknesses**: Can produce biased estimates
- **Random Forest [44]**
  - **Purpose**: Provides a method to assess relative feature importance by measuring how much each feature contributes to decreasing impurity when making splits in decision trees
  - **Strengths**: Can capture complex nonlinear relationships between variables, results are from multiple decision trees which reduces overfitting
  - **Weaknesses**: Not suitable for extrapolating beyond the data set; challenging to interpret
- **Monte Carlo [27]**
  - **Purpose**: Provides estimates of uncertainty around model predictions and parameter estimates, making it useful for decision-making
  - **Strengths**: Flexible, can handle complex and nonlinear relationships
  - **Weaknesses**: Difficult to use and achieve convergence, especially for complex problem

The results of all of these models were used together to gain a comprehensive understanding of how our factors contribute to food deserts.

5. Risk Analysis

5.1 Risk Characterization

Before we begin our analysis of how our various factors impact food deserts, we must characterize the risk of food deserts in the future based on analysis of historical data. We will attempt to characterize food deserts by their location, severity, and growth.

5.1.1 Geospatial Visualization

We used the USDA Food Desert data as well as Illinois geo-mapping data to generate an original model of where food deserts are located for the years 2010, 2015, and 2019 for Cook County. We can see that the number of food deserts is decreasing over the three years.

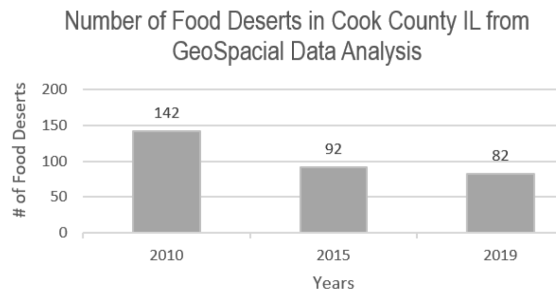


Figure 11: Number of Food Deserts in Cook County IL

The location of food deserts in Cook County IL has changed over the years 2010, 2015, and 2019, as shown in the maps below as an area of approximately 45 miles wide and 45 miles high which includes Chicago, Cook County, and most of the Chicago metropolitan area.

Location of Food Deserts in 2010, 2015, and 2019 Using Geo-mapping Methodology		

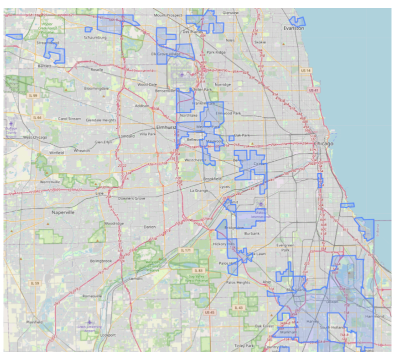
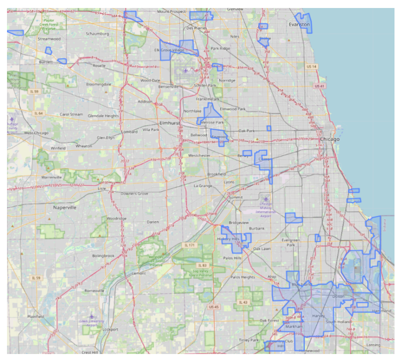
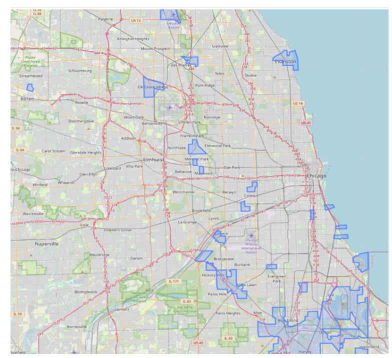

		

Figure 12: Cook County Food Desert Locations Identified by Geospatial Analysis in 2010 Number of food deserts: 142	*Figure 13: Cook County Food Desert Locations Identified by Geospatial Analysis in 2015* Number of food deserts: 92	*Figure 14: Cook County Food Desert Locations Identified by Geospatial Analysis in 2019* Number of food deserts: 82

We can also see the general areas where the food deserts are still prevalent remain nearly the same, including regions such as South Shore, West Englewood, North Lawndale, Austin, and Roseland. Since food deserts have consistent locations, this indicates where resources should be allocated as there are clear patterns of where food deserts exist. It should be noted that although the data shows a decreasing trend, since the most recent is from 2019, it does not capture the significant leap in food insecurity in the recent years of 2020 and 2021. Therefore, this data may not depict the most accurate severity of the situation.

Figure 15 below shows Illinois food deserts locations in 2010. We can see that food deserts occur throughout the state. The color scale indicates the percent of the census district population that lives in a food desert. Yellow is the highest percentage, indicating that over 90% of the population lives in food deserts. Green indicates that over 50% of the population lives in food deserts.

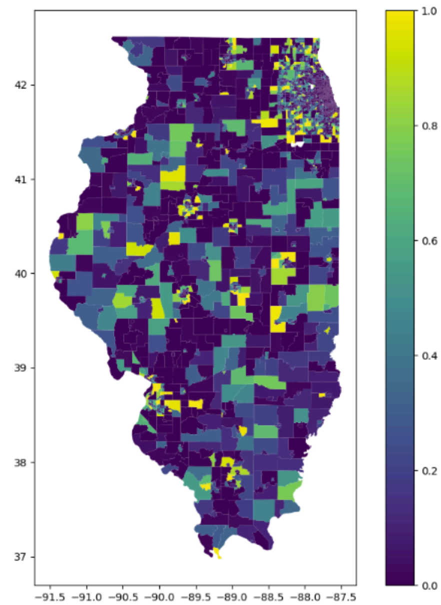


Figure 15: Illinois Food Desert Locations Identified by Geospatial Analysis in 2010

5.1.2 Statistical Forecasting

As shown in Figure 1, for the overall state of Illinois, we have 26 years of historical food desert data (1997-2022). In contrast, we only have 13 years of Cook County historical data (2009-2021). The correlation of the food desert rate between Cook County and Illinois during 2009 - 2021 is a high 0.94, meaning that Cook County reliably reflects the food desert rate in the overall Illinois area.

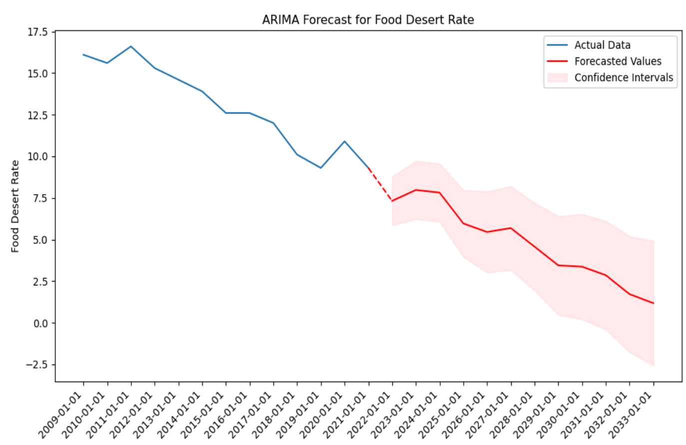


Figure 16: ARIMA Forecasting based on Cook County 13-Year Historical Data

When examining the 13-year history of Cook County, a downward trend is evident in the food desert rate. Figure 16 shows ARIMA forecast results with confidence intervals (Appendix G has its Python code). Fitting a trendline in Figure 16 including the forecasted data points could reveal that in approximately 30 - 35 years all food deserts in Cook County would disappear. The longer history of Illinois state food desert data showcases fluctuations with both lows and highs, indicating a slightly increasing or flat overall trend (Figure 17). We aim to leverage the more extensive 26-year Illinois data to enhance the forecast of the shorter 13-year Cook County data.

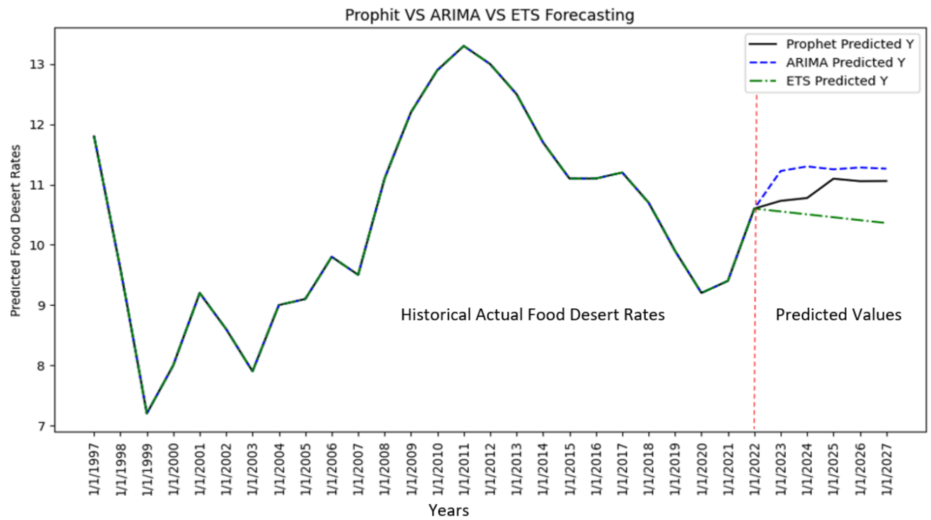


Figure 17: Forecasting based on IL 26-Year Historical Data

We compared forecasting results from the three different methods of ETS, Prophet, and ARIMA. The forecasts are based on 26-year Illinois history. Our ARIMA model setting is ARIMA(2,2,1), meaning that the number of autoregressive terms is 2, the number of differences is 2, and the number of moving average terms is 1.

Forecasted Food Desert Rates in Illinois			
Forecast Year	Prophet	ARIMA	ETS
2023	10.73%	11.22%	10.55%
2024	10.78%	11.30%	10.50%
2025	11.10%	11.25%	10.46%
2026	11.06%	11.28%	10.41%
2027	11.06%	11.26%	10.36%

Table 1: Forecasted Food Desert Rate in Illinois for 2023-2027

We preferred the ARIMA model to forecast future values of the food desert rate for two reasons. First, the model's autoregressive capabilities allow it to make both short-term and long-term predictions, surpassing the ability of the exponential smoothing method, which fails to detect complex trends. Second, the Prophet algorithm, which detects seasonality is not needed, because there is no seasonality involved with food desert yearly rates. Therefore, ARIMA stood out as the best methodology choice to forecast future food desert rates.

The forecasting results are all slightly different but all indicate that it will still take many years to drive the food insecurity rate down to 0%. Our goal is to identify risk mitigation strategies that will help force food deserts to disappear faster.

5.2 Multivariable Analysis

We will now determine how our various underlying factors contribute to the severity and growth of food deserts. Our model's target (dependent) variable is 26-year food desert rates in Illinois.

5.2.1 Pairwise Correlation Analysis

We have eight remaining factors aside from the food desert rate: weather temperature, precipitation, crime rate, traffic, unemployment rate, population, poverty rate, and income. These factors are potential variables in a multivariable model. After standardizing the data, the heat map below shows the results of the pairwise correlation analysis.

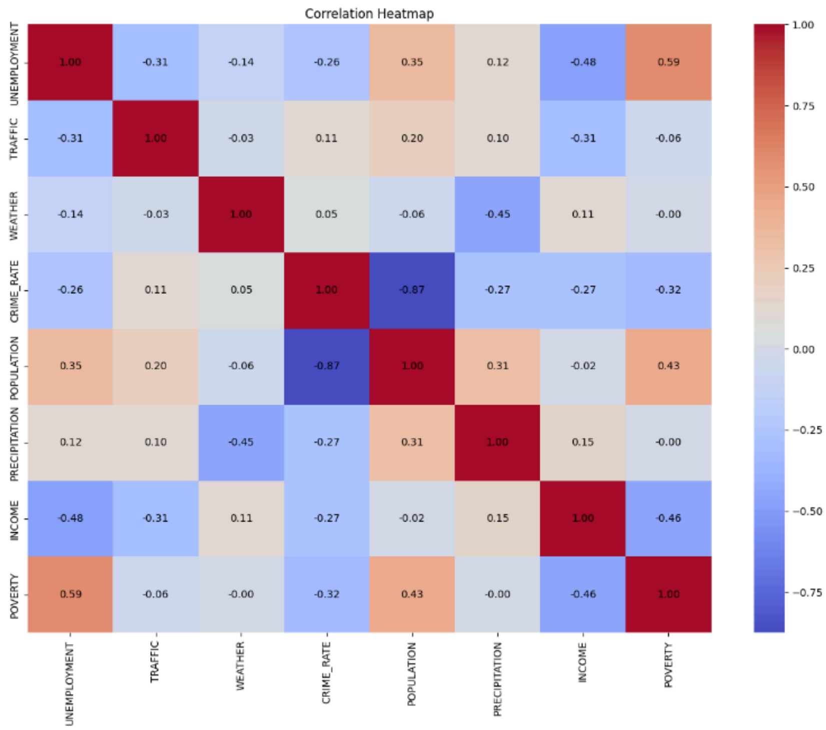


Figure 18: Pairwise Correlation Heatmap

When doing multivariable analysis, we must ensure that none of the factors are overly correlated with each other as this creates redundant data, making the results difficult and unclear to interpret. As discussed in section 2.1, we checked the correlation analysis between poverty rate and income data. The heat map above shows a correlation of -0.46; however, taking the absolute value, it is not the highest correlation and is lower than the correlation between income data and unemployment rate, indicating that poverty and income are not highly correlated.

Using 60% correlation as the cut-off line, we found one pair with too high correlation:

- Correlation between population and crime rate: -0.87

We aim to remove one variable from the correlated pair, so we remove the variable with less correlation (absolute value) with the food desert rate. We removed the crime rate because it has a lower correlation (absolute value) with the food desert rate than the population factor (0.61 vs. 0.75), as shown in Figure 19. More correlation heat maps and tables are in Appendix C-F.

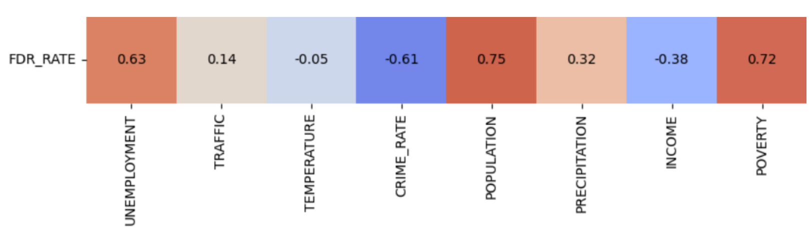


Figure 19: Correlation between Food Desert Rate and 8 factors that are predictive of food deserts

5.2.2 Model Independent Variables

Now, we have seven factors left to go further with multivariable models to predict the prevalence of food deserts. Again, we aim to find out how these underlying factors contribute to the presence of food deserts.

Symbol	Definition	Units
F	Food desert rate	% of people
U_e	Unemployment rate	% of people
T	Traffic	miles
P	Population	people
t	Temperature	Fahrenheit
p	Poverty rate	% of people
P_c	Precipitation	inches
I	Income	dollars

Table 2: Table of Variables for Multivariable Analysis

5.2.3 Regression Models with Model Accuracy Evaluation

We applied the following multivariable modeling methodologies to the normalized data: ridge regression, random forest, and symbolic regression.

The ridge regression model and random forest returned lower Mean Absolute Error (MAE) and Mean Squared Error (MSE) values compared to symbolic regression. Lower MSE and MAE values indicate better accuracy and precision in predicting the food desert rate. Though the random forest model has the least error, it has overfitting risks, because we only have 26 data points and don't have too much room to do hyper-parameter tuning for the machine learning model. The ridge model has more straightforward interpretability and is simple. Therefore, we chose ridge regression as our final multivariable model.

Model Accuracy Metrics	Symbolic Regression	Ridge Regression	Random Forest
MAE	0.01219	0.00621	0.00296
MSE	0.00026	0.00006	0.00002

Table 3: Model Accuracy Comparison Among Three Model Methodologies

5.2.4 Ridge Regression Results

Here are the results from our ridge regression model with seven factors:

- **Ridge Regression Prediction Formula**:

$$\boxed{F = 0.07 + 0.01161 \times U_e + 0.00459 \times T + 0.00298 \times t + 0.01997 \times P + 0.00913 \times P_c - 0.0058 \times i + 0.01553 \times p}$$

Table 4: Ridge Regression Prediction Formula

- Model Performance Evaluation:
  - $R^2 = 0.8010$
- Feature importance list:

Features	Importance
Population	1
Poverty Rate	2
Unemployment Rate	3
Precipitation	4
Income	5
Traffic	6
Temperature	7

Table 5: Feature Importance from Ridge Regression Model

5.2.5 Random Forest Results

Figure 20 illustrates the first decision tree in the random forest model. From its feature importance list (Appendix I), we see that its top 4 important features are: population, poverty rate, income, and unemployment rate. This shows a significant degree of alignment with the ridge regression model's top feature rankings.

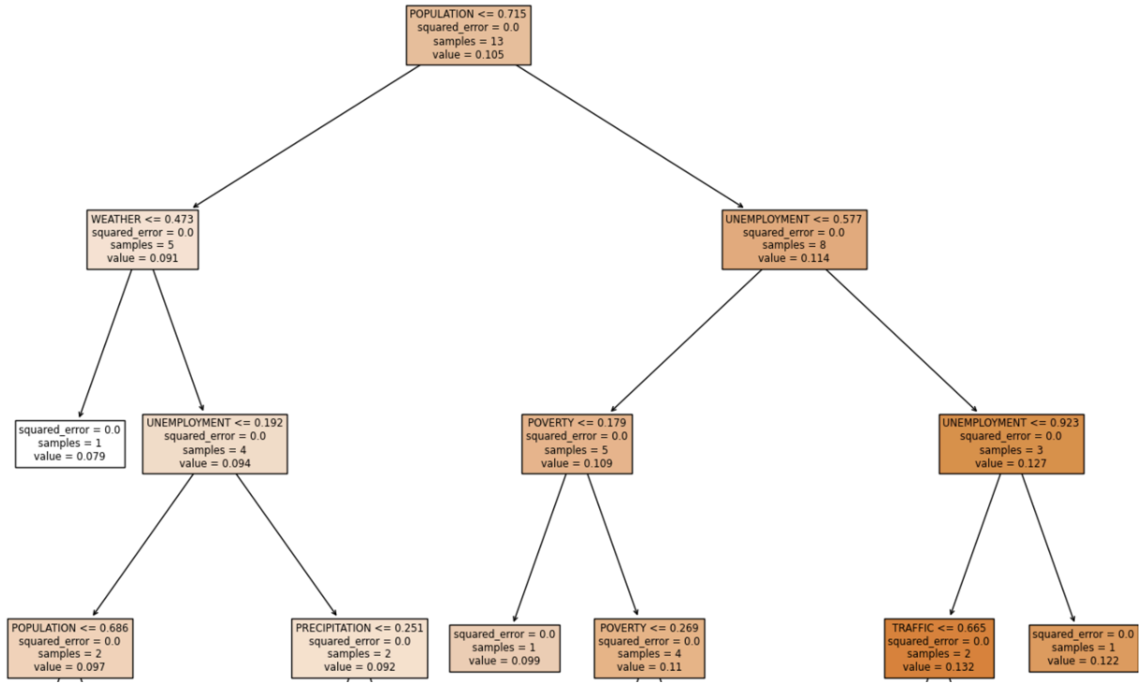


Figure 20: First Decision Tree Exhibit in Random Forest Model

5.2.6 Observation

Ridge regression was chosen as our final multivariable model due to its small MAE/MSE values and better interpretation. We delved into all predictors from our regression analysis:

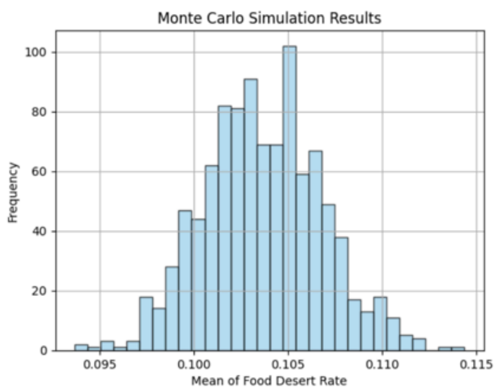
Population has a positive correlation with the occurrence of food deserts, suggesting that areas with larger populations may encounter greater difficulties in accessing fresh and affordable food options. This implies a need for the construction of more grocery stores in such areas. Higher income has a negative impact on the food deserts prevalence, suggesting boosting income levels can help with reducing food deserts. Additionally, higher **poverty** levels are associated with increased instances of food deserts, indicating a potential socioeconomic influence on food accessibility. Higher **traffic** data is associated with increased instances of food deserts, indicating traffic congestion might hinder accessibility to buy food. Moreover, regions with elevated **unemployment** rates tend to exhibit a higher prevalence of food deserts, highlighting a connection between job availability and food security. Lastly, **precipitation and temperature** have a positive model coefficient when predicting food desert prevalence. Areas experiencing higher levels of precipitation tend to have elevated rates of food deserts **[16]**. It's essential to note that while most precipitation falls as rain, it also includes freezing rain, sleet, snow, or hail which can hinder access to grocery stores. We utilize annual precipitation data rather than monthly data for our multivariable regression, as it cannot be interpreted to reflect seasonal influences. Furthermore, our analysis in section 3.8 found positive correlations between temperature, precipitation, and crime rate. Higher temperatures are linked to more precipitation and higher crime rates, potentially impacting access to fresh food.

Overall, these observations underscore the multifaceted nature of factors contributing to food desert prevalence and highlight the importance of addressing socioeconomic, environmental, and infrastructural determinants in efforts to combat food deserts.

5.3 Monte Carlo Simulation to Quantify Risk Mitigation

5.3.1 Monte Carlo Prediction of Food Desert Rate in 2023

We performed a Monte Carlo simulation **[27]** to estimate the mean of food desert rates by randomly sampling from the existing data with replacement. We repeated this process for 10,000 trials and then plotted a histogram below:



Monte Carlo Output	
mean	0.1037
std	0.0031

	Projected Food Desert Rate Range
mean-2*std	0.0975
mean	0.1037
mean+2*std	0.1099

Figure 21: Food Desert Rate Prediction in 2023 via Monte Carlo Simulation

The main takeaway is that for 2023, the mean of the predicted food desert rate in the simulation was 10.4% with a standard deviation of 0.3%. The range of predicted food desert rates is [9.8%, 11.0%]. The simulation Python code is in Appendix H.

5.3.2 Actionable Factors

We identified the factors that contribute to the prevalence of food deserts in Illinois: population, poverty rate, unemployment rate, income, traffic, temperature, and precipitation.

It is challenging to take actionable steps to change precipitation and temperature factors to improve the food desert rate in Illinois as that involves long-term climate patterns. Lowering the population size is also extremely challenging to undergo intentionally.

However, the government can play a role in making interventions to boost personal income, decrease the poverty rate, and create additional job opportunities to reduce unemployment. The government can also play a role in regulating traffic because traffic congestion can hinder access to grocery stores.

5.3.3 Sensitivity Analysis

We conducted several Monte Carlo simulations consisting of 10,000 trials, where we assumed our coefficients followed a normal distribution due to a lack of data. This allowed us to assess how food desert rates vary when adjusting one actionable factor by 10%.

We first take a 10% decrease in the mean unemployment rate in 2023 with 3% as its corresponding standard deviation. We return the left graph in Figure 22 as the simulation result with the mean value marked in the black line. The result of the 10% decrease in the unemployment rate is that the mean food desert rate prediction is now 7.4%. Compared to the previous mean of 10.4%, this was a drop of 28.3% in the food desert rate. This indicates that unemployment has a strong influence on the food desert rate.

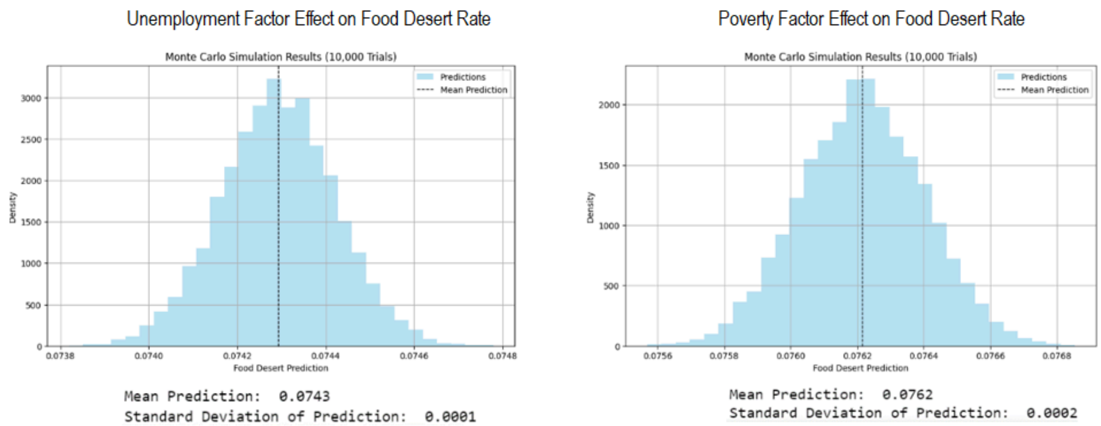


Figure 22: Monte Carlo Simulation - Predicting Food Desert Rate with a 10% Decrease in Unemployment Rate (Left Graph) & Poverty Rate (Right Graph)

The graph on the right in Figure 22 illustrates that a 10% decrease in the poverty rate results in a food desert rate prediction mean of 7.6%. Compared to the previous mean of 10.4%, this was a drop of 26.5% in the food desert rate. This reduction is attributed to a decrease in the poverty rate by 10%, while other factors remained constant. This suggests that addressing the poverty rate can also have a notable impact on reducing food deserts.

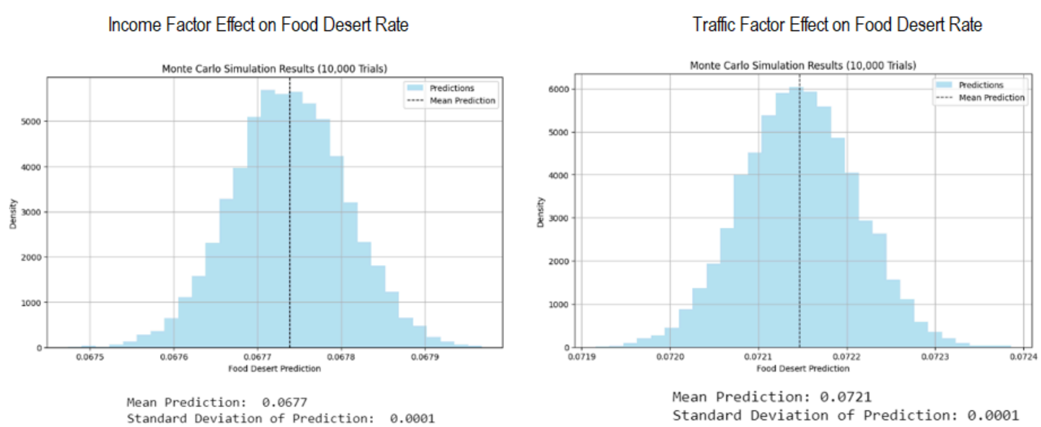


Figure 23: Monte Carlo Simulation: Predicting Food Desert Rate with a 10% Decrease in Income (Left Graph) & Traffic (Right Graph)

The result of the 10% increase in income is that the mean food desert rate prediction is now 6.8%, as shown in the left graph in Figure 23. Compared to the previous mean of 10.4%, this was a drop of 34.7% in the food desert rate. This reduction is attributed to an increase in income by 10%, while other factors remained constant. This highlights the importance of improving income levels to mitigate food deserts significantly.

The right graph in Figure 23 shows when traffic decreases by 10%, the food desert rate decreases to 7.2%, with an impact of 30.5%. This shows that factors related to transportation and accessibility also play a role in food desert prevalence.

Original Mean of Food Desert Rate: 10.4%			
Factors	Changes	Food Desert Rate Changed to	Food Desert Rate Impact
Unemployment	decrease 10%	7.43%	decrease 28.3%
Poverty	decrease 10%	7.62%	decrease 26.5%
Income	increase 10%	6.77%	decrease 34.7%
Traffic	decrease 10%	7.21%	decrease 30.5%

Table 6: Sensitivity Analysis Results Over All Actionable Factors

Overall, addressing the unemployment rate, poverty rate, income levels, and transportation accessibility can have substantial impacts on reducing food deserts [7]. These factors are interconnected and should be considered comprehensively in strategies to address food deserts and enhance access to nutritious food in affected areas.

5.3.4 Expected Costs of Food Deserts

We found healthcare costs of food deserts in Illinois. Although we researched cost studies for the actionable mitigation factors, we only found the costs of the SNAP program which is related to the poverty factor, giving us specific financial dollar amounts for a mitigation cost-benefit analysis.

5.3.4.1 Healthcare Costs of Food Deserts in Illinois

Illinois annual health costs due to food deserts are roughly \$1.7 billion. This is derived from US costs proportioned down to Illinois people living in a food desert. About \$71 billion is expended each year to manage chronic diseases in the US that could be effectively addressed or mitigated through a nutritious diet [25]. The number of food desert households in the US was 44.2 million in 2022 [5]. The average American family consisted of 3.13 persons in 2022 [47]. Thus the US has 138.3 million persons (*138.3M = 44.2M * 3.13*) in food deserts. 3.3 million Illinois people are living in a food desert in 2021[1], assuming the number is the same for 2022, which means Illinois annual healthcare costs due to food deserts are roughly \$1.7 billion (*\$1.7B = 3.3M * \$71B / 138.3M)*. If the mitigation efforts cost less than \$1.7 billion per year then a positive cost-benefit should result. Table 7 shows the chronic disease management cost reduction by mitigating one actionable factor at a time:

Chronic Disease Management Costs in Illinois Food Deserts is \$1.7 Billion per Year				
Factors	Changes	Food Desert Rate Changed to	Food Desert Rate Impact	Chronic Disease Management Costs Reduction
Unemployment	decrease 10%	7.43%	decrease 28.3%	\$0.479 B / Year
Poverty	decrease 10%	7.62%	decrease 26.5%	\$0.449 B / Year
Income	increase 10%	6.77%	decrease 34.7%	\$0.588 B / Year
Traffic	decrease 10%	7.21%	decrease 30.5%	\$0.517 B / Year

Table 7: Cost Reduction Estimation from Actionable Factors

5.3.4.2 Mitigation Costs of SNAP Program in Illinois

SNAP annual costs for Illinois are estimated to be \$0.417 billion achieving a benefit of \$0.449 billion in reduced chronic disease management costs, a \$32 million annual benefit. The cost-benefit calculation is based on the following:

1. The poverty factor measurement unit is the percent of the population in poverty. When we change that by 10% we change the average poverty rate from 12.6% to 11.3%, a 1.3% poverty rate reduction
2. The estimated Illinois population in 2023 is 12,549,689 according to the US Census Bureau [13]

3. The poverty factor can be mitigated in part with the SNAP program. When we change the poverty rate factor by 10%, we find that 163,146 people will benefit (*163,146 = 1.3% * 12,549,689*)
4. Illinois' SNAP program costs $2,556 per person per year**,** based on 1.993 million people in SNAP for an annual cost of $5.092 billion **[48]** in 2021 (*$2,556 = $5,092 / 1.993*)
5. The cost of SNAP to mitigate the poverty rate factor by 10% is $417 million per year (*$417M = $2,556 * 163,146*)
6. Table 7 shows that 10% poverty mitigation could result in $449 million in reduced chronic disease management costs
7. This benefit of 10% poverty rate mitigation from SNAP is $32M per year (*$32M=$449M-$417M*)

5.4 Discussion

We identified the factors that contribute to the prevalence of food deserts in Illinois: population, poverty rate, unemployment rate, income, traffic, temperature, and precipitation. The actionable factors are poverty rate, unemployment rate, income, and traffic. Based on these factors, we will establish four main recommendations to decrease the number of food deserts and mitigate the effects of the lack of access to nutritious food. Our study showed over the last 26 years the trend in food desert rate in Illinois is slightly increasing or flat. However, the good news is that within the most recent 13 years, the trend is decreasing. Our geospatial data analysis also reveals a predicted decrease in the number of food deserts within Cook County Illinois during 2010, 2015, and 2019, along with the areas most in need of support. Finally, the current estimated annual healthcare cost is $1.7 billion in Illinois from poor nutrition, which we hope to reduce with our recommendations.

5.4.1 Study Strengths

The evident strength of our study lies in our comprehensive approach:

1. Thorough Exploratory Data Analysis (EDA) establishes a solid foundation of clean and relevant data to use in our subsequent geospatial analysis and multivariable regression models. We collected data from nine data sources and conducted extensive EDAs, including trend and correlation analysis.
2. The strength of multiple forecast methods resolves the divergent forecast results. The forecast comparing Illinois State' 26-year to Cook County's 13-year history differ, however, both forecasts indicate the food desert trend is not increasing. Our geospatial analysis revealed a consistent decrease in food deserts across Cook County, aligning with the results of Cook County's 13-year ARIMA forecast. The multiple forecasts lead us to one picture of a decreasing trend.
3. Geospatial analysis adds a unique view **[24]** by identifying persistent food desert hotspots in Cook County for the years 2010, 2015, and 2019.
4. Monte Carlo simulations add strength to the study. Monte Carlo quantifies the effect on food desert rates from a 10% change in actionable mitigation factors, one factor at a time, which contributes to the subsequent full cost-benefit analysis. Monte Carlo accommodates complex and non-linear relationships and provides upper and lower confidence limits.

5. We integrated multifaceted modeling methods to gain a deep understanding of food desert presence. The multiple models let us see which factors were common across the models giving us more confidence in finding the significant actionable factors to mitigate food desert rates.

5.4.2 Study Weaknesses

Some limitations of our study stem from the constrained availability of up-to-date historical data. For instance, while Illinois data offers a 26-year historical record, Cook County only provides 13-year historical data. Furthermore, the latest food insecurity information extends only to 2022 for Illinois and 2021 for Cook County. Ideally, access to historical 30-year data up to 2023 would greatly enhance the accuracy of predictions for the current year.

Additionally, we only have 2010, 2015, and 2019 geospatial data, which hinders our ability to analyze the COVID-19 impact in the Cook County area food deserts geographically. And this scarcity of recent data limits our confidence in the downward trend. Because of a lack of data, the accuracy of our models may be limited, especially when attempting to make estimates far into the future. Specifically, in our Geospatial model, we lacked data after 2020, which most likely saw an increase in food deserts due to the COVID-19 pandemic **[46]**. Also, food insecurity data in general is limited after 2020, meaning that our data is unlikely to be completely accurate to reality. Therefore, the downward trend of food deserts may be too optimistic, since the extraordinary circumstances of the pandemic have caused a drastic resurgence in food deserts in recent years **[36].**

The literature identifies that race has a relationship to food desert rates **[7]**, and food deserts will increase the incidence of diabetes **[29]** and obesity **[16]**. We did consider race, diabetes, and obesity in our EDA phase. There is insufficient data to complete any analyses for food desert rates.

Although we researched cost studies for all the actionable mitigation factors, we only found the costs of the SNAP program to mitigate the poverty factor, allowing a full cost-benefit analysis. Our analysis also estimates the reduction in chronic disease management costs due to food deserts.

6. Recommendations

Based on our analysis of food deserts, we established four main recommendations to decrease the number of food deserts and mitigate the effects of the lack of affordable and nutritious food.

6.1 Increase Access to Supermarkets in Low-Income Areas

A straightforward method to decrease food deserts is to ensure that communities have access to a variety of healthy food options by increasing access to supermarkets. Ensure grocery stores are close to the population by encouraging more grocery stores in the food desert areas. Since our model shows that food deserts often appear in low-income areas, we recommend that Chicago incentivizes the development of grocery stores and supermarkets with grants, loans, tax credits, and other economic benefits. The city could also enforce zoning regulations that require supermarkets to be built in underserved areas. For example, a similar bill was passed by Gov. J.B. Pritzker, known as the Illinois Grocery Initiative, to subsidize grocery stores to move into existing food deserts. This initiative costing $20 million is an excellent starting point for a thorough solution to food deserts in Illinois **[31]**.

However, the development of grocery stores is ineffective without a viable method to reach them. Therefore, we suggest that Chicago realigns bus routes and provides other methods of transportation or delivery to ensure supermarkets are easily accessible with affordable public transportation **[30]**.

Additionally, high crime rates were found to have a strong correlation with food deserts, likely because they deterred residents from traveling outside freely. We propose that the city invest in public safety efforts, such as lighting and increased police presence to increase accessibility to supermarkets.

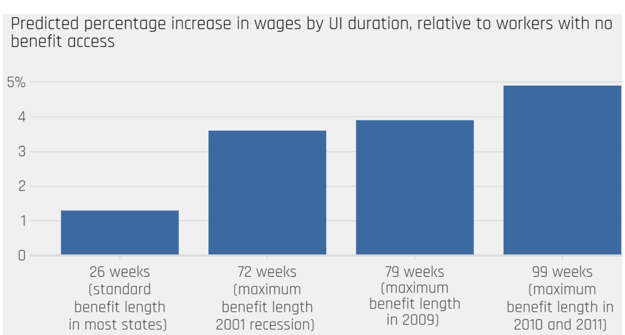
6.2 SNAP, Healthy Food Incentives, Education

Even with access to supermarkets, people may still gravitate towards unhealthy options due to costs or preferences. This is evident in the positive correlation between poverty rate and poor diet, because unhealthy options like fast food are often cheaper. Accordingly, strategies must be implemented to promote healthier diets through pricing policy **[32]**.

We suggest that the city management of Chicago works with the federal government's various food assistance programs, such as the Supplemental Nutrition Assistance Program (SNAP) which already supports a variety of supermarkets and local farmers. SNAP has demonstrated great success concerning aiding the vulnerable in society. By offering direct assistance to those who have less money for fresh foods, SNAP has been able to target one of the root causes of food deserts. Although SNAP has seen budget concerns in recent years due to lawmakers finding the push for healthier alternatives to be non-essential, it offers a baseline for what an effective health-conscious food program can do **[26]**.

Chicago can also encourage publicity and endorsement of stores that meet certain health criteria. Restaurants can be required to provide customers with calorie information on menus and be offered coupon incentives for healthier options. Additionally, since eating habits are best established at a young age, Chicago can invest in educational programs in schools teaching the importance of having a mindful, balanced diet and how to properly read nutrition labels and avoid deceptive advertising.

6.3 Unemployment Benefits to Those Struggling with the Cost of Living



Increasing unemployment benefits can create a ripple effect that positively influences the choice of housing near a supermarket, access to fresh food, and ultimately, an individual's ability to make healthier food choices. Furthermore, increasing the period that one receives unemployment benefits has also been shown to increase one's re-employment income and increase the effective use of one's skills and training **[50].** Having an increased salary after re-employment then contributes to increasing average personal income, which we identified as having a significant effect on food deserts.

Increasing the length of unemployment benefits received further serves to decrease the economic disparity between marginalized groups (e.g., women and people of color) with members of more privileged demographic groups **[50].** They found that a 53-week increase in unemployment benefits improved the job match for workers of color more than White workers and that marginalized workers have larger returns in terms of salary levels. As the prevalence of food deserts in Illinois has also been attributed to racial disparities, this may also aid in decreasing the food desert rate.

6.4 Increase Access to Farmers' Markets

The benefit of farmers' markets is that they are often a more practical solution to food deserts, being less expensive, more mobile, and taking up less space than a supermarket. For those predicted to be living in food deserts by our geospatial model, such as South Shore, West Englewood, and Austin, farmers' markets also serve as an educational resource where workers and volunteers can guide their community members through how to store and prepare fresh produce.

Farmers' markets have already shown success in decreasing the presence of food deserts, with the Ron Finley Project **[33]** and the Fresh Moves Mobile Market **[34]** serving as examples of very successful initiatives, reaching over 15k consumers with fresh, accessible produce. Hence, we recommend that Chicago modifies their land use policies to include new spaces for farmers' markets, as well as subsidizing the development of new markets with increased funding.

Acknowledgments

We would like to express our sincere gratitude for the invaluable support we received from our mentor Paul Heffernan and coach Paul Kim in reviewing our paper. We also extend our thanks to Nicholas Tschaika and Yan Zhao for their guidance in answering our questions and providing encouragement. Finally, we would like to thank The Actuarial Foundation and the Institute of Competition Sciences for organizing this Modeling The Future Competition and providing essential resources that have greatly contributed to the development of our model.

References:

[1] Gov. Pritzker signs bill combating food deserts across Illinois. URL
https://chicago.suntimes.com/politics/2023/8/18/23837236/food-deserts-illinois-pritzker-signs-bill

[2] FOOD DESERTS: CAUSES, CONSEQUENCES AND SOLUTIONS URL
https://www.learningforjustice.org/classroom-resources/lessons/food-deserts-causes-consequences-and-solutions

[3] Food insecurity magnifies inequity in Chicago. URL
https://www.chicagobusiness.com/equity/how-food-access-highlights-economic-inequity-chicago

[4] Food Deserts and Inequality. URL https://www.socialpolicylab.org/post/grow-your-blog-community

[5]The report finds that millions of American families struggle to put food on the table. URL
https://sph.lsuhsc.edu/press/millions-of-american-families-struggle-to-get-food-on-the-table-report-finds/#:~:text=The%20report%20found%20that%2044.2,million%20people%20the%20year%20prior

[6] The US Is Facing Its Worst Hunger Crisis in Generations. URL
https://youtu.be/p2MoB27rAbY?si=84X0yHa6Uky01xZF

[7] Characteristics and Influential Factors of Food Deserts. URL
https://www.ers.usda.gov/webdocs/publications/45014/30940_err140.pdf

[8] Fast Facts About Food Deserts in Chicago. URL
https://food-deserts.com/food-deserts-in-chicago/#Fast_Facts_About_Food_Deserts_in_Chicago

[9] Redlining and Food Justice in America. URL
https://moveforhunger.org/blog/redlining-and-food-justice-america#:~:text=This%20process%20results%20in%20neighborhood,of%20redlining%20are%20food%20deserts.

[10] Map the Meal Gap Data. URL https://www.feedingamerica.org/research/map-the-meal-gap/by-county

[11] Personal Income by County, Metro, and Other Areas. URL
(https://www.bea.gov/data/income-saving/personal-income-county-metro-and-other-areas)

[12] SAIPE Datasets. URL https://www.census.gov/programs-surveys/saipe/data/datasets.html

[13] Population. URL https://www.census.gov/topics/population.html

[14] Crimes - 2001 to Present. URL
https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/about_data

[15] National Centers for Environmental Information. URL
https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/county/mapping

[16] Climate Implications – Food Distribution and Access. URL
https://eri.iu.edu/erit/implications/food-distribution-and-access.html

[17] Per Capita Income URL
https://www.census.gov/quickfacts/fact/note/US/INC910222#:~:text=Per%20capita%20income%20is%20the,total%20population%20in%20that%20group

[18] Vehicle Miles Traveled by County. URL
https://idot.illinois.gov/transportation-system/network-overview/highway-system/reports/illinois-travel-statistics.html

[19] Data Downloads - Unemployment Insurance. URL https://oui.doleta.gov/unemploy/DataDownloads.asp

[20] U.S. Census Bureau. TIGER/Line Shapefiles. URL
https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2010.html#list-tab-790442341
[21] Geospatial Data Analytics: What It Is, Benefits, and Top Use Cases. URL
https://www.safegraph.com/guides/geospatial-data-analytics
[22] Illinois Metra Areas' Unemployment Rates Among Worst in U.S. URL
https://www.illinoispolicy.org/illinois-metro-areas-unemployment-rates-among-worst-in-u-s/
[23] How Seasons and Weather Affect Crime Trends. URL
https://esfandilawfirm.com/do-seasons-weather-affect-crime-trends/
[24] Geospatial Analytics: What Is It & How Can It Give You a Competitive Edge.URL
https://capeanalytics.com/blog/geospatial-analytics-what-is-it-how-can-it-give-you-a-competitive-edge/#Geospatial_Analytics_What_It_Is_Why_It_Matters
[25] Life In America's Food Deserts, How Poverty And Suburbanization Have Impacted Public Health. URL
https://onlinedegrees.kent.edu/college-of-public-health/community/causes-of-food-deserts
[26] The New Face of Hunger. URL https://www.nationalgeographic.com/foodfeatures/hunger/
[27] What is Monte Carlo Simulation? URL https://www.ibm.com/topics/monte-carlo-simulation
[28] Food Security. URL
https://www.usda.gov/topics/food-and-nutrition/food-security#:~:text=Given%20population%20growth%20and%20rising,will%20need%20to%20almost%20double
[29] The Intersection between Food Insecurity and Diabetes: A Review. URL
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4218969/
[30] Public Transit Access to Full-Service Grocery Stores Will Help Address Country's Obesity Crisis. URL
https://www.tfah.org/story/public-transit-access-full-service-grocery/
[31] New Illinois Law Invests $20 million for new grocery stores, eliminating food deserts. URL
https://www.illinois.gov/news/press-release.26902.html
[32] USDA Economic Research Service. URL
https://www.ers.usda.gov/data-products/food-access-research-atlas/documentation/
[33] Ron Finley Project. URL https://ronfinley.com/
[34] Food on the Move. URL https://www.urbangrowerscollective.org/fresh-moves-mobile-market
[35] Personal vehicle use and food security among US adults who are primary shoppers for households with children URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10258743/
[36] Effect of COVID-19 on food security, hunger, and food crisis URL
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9335023/
[37] What are food deserts, and how do they impact health? URL
https://www.medicalnewstoday.com/articles/what-are-food-deserts
[38] Supplemental Nutrition Assistance Program (SNAP) URL
https://www.fns.usda.gov/snap/supplemental-nutrition-assistance-program
[39] Prevalence of Municipal-Level Policies Dedicated to Transportation That Consider Food Access. URL

https://www.cdc.gov/pcd/issues/2021/21_0193.htm

[40] Pearson correlation coefficient. URL https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

[41] Wilstrup C. and Kasak J. Symbolic regression outperforms other models for small data sets. URL https://arxiv.org/pdf/2103.15147.pdf

[42] Hyndman R. and Athanasopoulos G. Forecasting: Principles and Practice. URL https://arxiv.org/pdf/2103.15147.pdf

[43] Time Series Analysis using Facebook Prophet. URL https://www.geeksforgeeks.org/time-series-analysis-using-facebook-prophet/

[44] Breiman, L.: Random forests. Machine Learning 45 (2001). doi:10.1023/A:1010933404324

[45] Ridge Regression. URL https://en.wikipedia.org/wiki/Ridge_regression

[46] Haskell S. How the Covid-19 Pandemic Affects Food Deserts. URL https://www.canr.msu.edu/news/how-the-covid-19-pandemic-affects-food-deserts

[47] Average number of people per family in the United States from 1960 to 2022. URL https://www.statista.com/statistics/183657/average-size-of-a-family-in-the-us/

[48] HOW SNAP BECAME THE MOST EXPENSIVE FEDERAL FOOD ASSISTANCE PROGRAM. URL https://allen.house.gov/news/documentsingle.aspx?DocumentID=5173

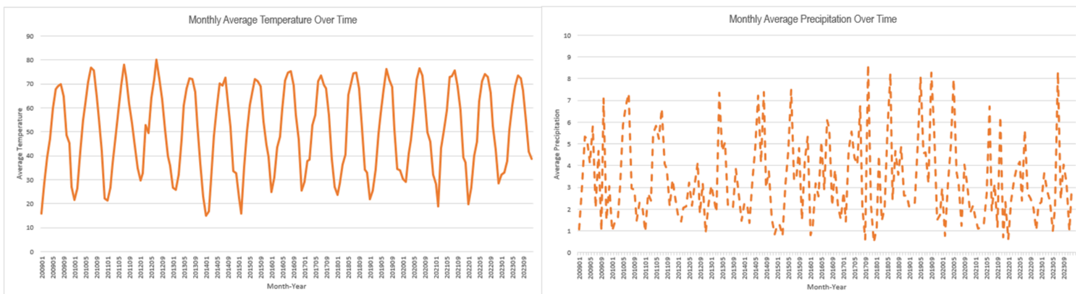
[49] The Relationship Between Food Deserts, Farmers' Markets, and Food Assistance Programs in Hawai'i Census Tracts. URL https://pubmed.ncbi.nlm.nih.gov/32043088/

[50] The long-run implications of extending unemployment benefits in the United States for workers, firms, and the economy. URL https://equitablegrowth.org/the-long-run-implications-of-extending-unemployment-benefits-in-the-united-states-for-workers-firms-and-the-economy/

Appendix:

A. SPY ETF History to Verify Unemployment Spikes. URL finance.yahoo.com/quote/SPY/history

B. Cook County IL Average Monthly Temperature (Left) and Precipitation (Right) Over Time



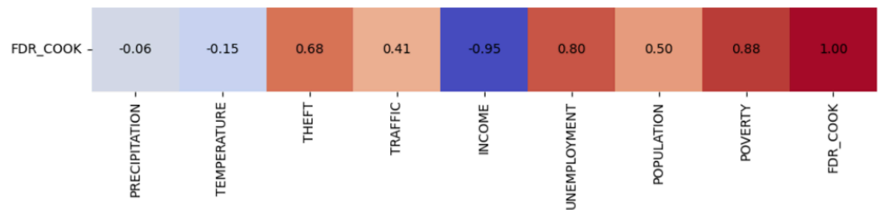
C. Pairwise Correlation Table Using 26-year Illinois Data

	POVERTY	UNEMPLOYMENT	TEMP	CRIME	POPULATION	PRECIPITATION	INCOME	TRAFFIC	FDR_RATE
POVERTY	1.000000	0.602516	-0.093239	-0.284799	0.514324	0.013962	-0.521680	0.101280	0.687838
UNEMPLOYMENT	0.602516	1.000000	-0.069339	-0.291962	0.535001	-0.026881	-0.199045	0.301549	0.599850
TEMP	-0.093239	-0.069339	1.000000	0.125779	-0.252000	-0.399196	0.107895	-0.205605	0.027092
CRIME	-0.284799	-0.291962	0.125779	1.000000	-0.811022	-0.113891	-0.133108	-0.603964	-0.404115
POPULATION	0.514324	0.535001	-0.252000	-0.811022	1.000000	0.189051	-0.220128	0.735710	0.493365
PRECIPITATION	0.013962	-0.026881	-0.399196	-0.113891	0.189051	1.000000	-0.113074	0.163948	0.190022
INCOME	-0.521680	-0.199045	0.107895	-0.133108	-0.220128	-0.113074	1.000000	-0.086594	-0.359808
TRAFFIC	0.101280	0.301549	-0.205605	-0.603964	0.735710	0.163948	-0.086594	1.000000	0.131325
FDR_RATE	0.687838	0.599850	0.027092	-0.404115	0.493365	0.190022	-0.359808	0.131325	1.000000

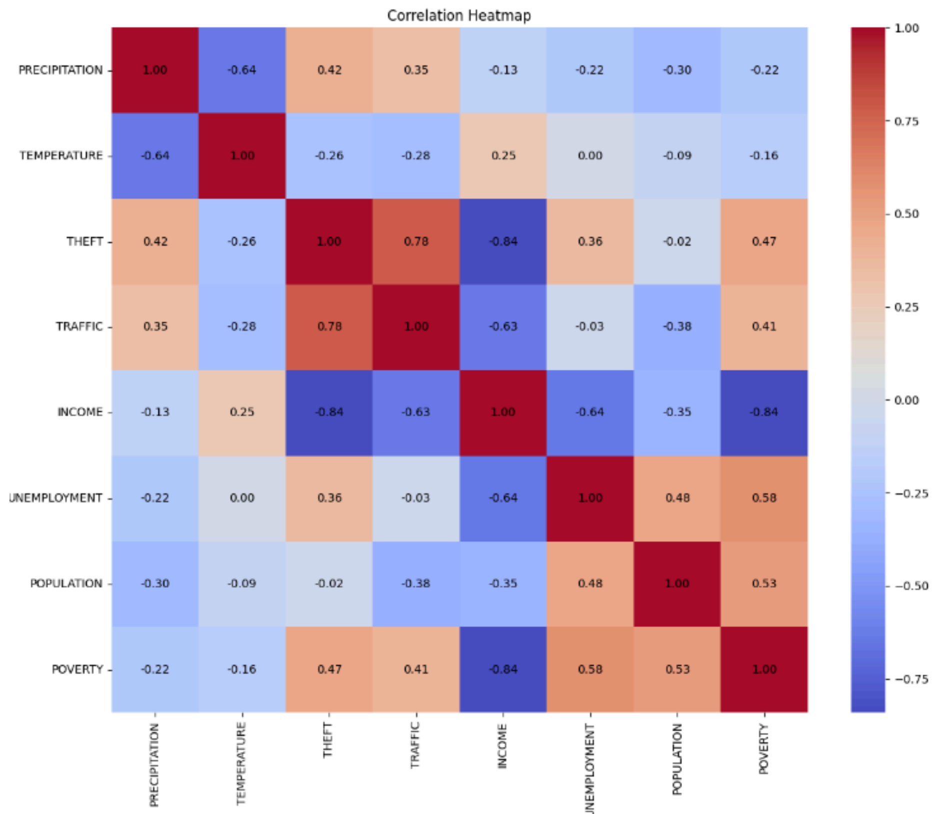
D. Pairwise Correlation Table Using 13-year Cook County IL Data

	WPRECIPITATION	WAVGTEMP_	THEFT	TRAFFIC	INCOME	UNEMPLOYMENT	TOT_POP	POVERTY
WPRECIPITATION	1.000000	-0.638206	0.420657	0.345161	-0.125361	-0.217208	-0.302421	-0.221627
WAVGTEMP_	-0.638206	1.000000	-0.257316	-0.279703	0.254664	0.002734	-0.089025	-0.159435
THEFT	0.420657	-0.257316	1.000000	0.781970	-0.840939	0.364849	-0.023393	0.474050
TRAFFIC	0.345161	-0.279703	0.781970	1.000000	-0.633200	-0.029774	-0.381234	0.409898
INCOME	-0.125361	0.254664	-0.840939	-0.633200	1.000000	-0.641511	-0.349688	-0.838958
UNEMPLOYMENT	-0.217208	0.002734	0.364849	-0.029774	-0.641511	1.000000	0.476521	0.581526
TOT_POP	-0.302421	-0.089025	-0.023393	-0.381234	-0.349688	0.476521	1.000000	0.528481
POVERTY	-0.221627	-0.159435	0.474050	0.409898	-0.838958	0.581526	0.528481	1.000000

E. Pairwise Correlation Heatmap with Target Variable Using 13-year Cook County IL Data



F. Pairwise Correlation Heatmap for 9 Independent Variables Using 13-year Cook County IL Data



G. Python Code for ARIMA Forecasting

```python
# function adapted from https://scikit-learn.org/stable/modules/generate
def dist_in_meters(point_1, point_2):
    point_1 = [math.radians(l) for l in [point_1.y, point_1.x]]
    point_2 = [math.radians(l) for l in [point_2.y, point_2.x]]
    dist_array_m = haversine_distances([point_1, point_2])*6371000
    return dist_array_m[0][1]
```

```python
import itertools
import numpy as np
import pandas as pd
from statsmodels.tsa.arima.model import ARIMA

# Create a DataFrame with the provided data
data = {
    'ds': list(range(2009, 2022)),
    'y': [16.1, 15.6, 16.6, 15.3, 14.6, 13.9, 12.6, 12.6, 12.0, 10.1, 9
}
df = pd.DataFrame(data)

# Convert 'ds' to datetime
df['ds'] = pd.to_datetime(df['ds'], format='%Y')

# Set 'ds' as the index
df = df.set_index('ds')

# Function to evaluate ARIMA model for given parameters
def evaluate_arima(order, data):
    model = ARIMA(data, order=order)
    try:
        results = model.fit()
        return results.aic
    except:
        return np.nan  # Return NaN for invalid parameter combinations
```

H. Python Code for Monte Carlo Simulation

```python
import numpy as np
import matplotlib.pyplot as plt

# Food Desert Rates
data=X.FDR

# Run 10,000 trials
num_trials = 10000
num_samples = len(data)

# to store Monte Carlo simulation results
simulation_results=np.zeros(num_simulations)

# Monte Carlo Simulation
for i in range(num_trials):
    # Generate random sample with replacement from the given data
    random_sample=np.random.choice(data,size=num_samples, replace=True)
    simulation_results[i]=np.mean(random_sample)

# Plot histogram of simulated results
plt.hist(simulated_results, bins=30, color='skyblue', edgecolor='black', alpha=0.7)
plt.title('Monte Carlo Simulation Results (10,000 Trials)')
plt.xlabel('Mean of Food Desert Rate')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```

I. Random Forest Model Feature Importance Results:

```python
# Print out the sorted feature importances
print(sorted_feature_importances)
```

```
POPULATION       0.496216
POVERTY          0.222687
INCOME           0.124283
UNEMPLOYMENT     0.093810
PRECIPITATION    0.024491
WEATHER          0.023589
TRAFFIC          0.014924
dtype: float64
```

J. List of Figures in the Main Body of the Paper

K. List of Tables in the Main Body of the Paper