**The Food Desert Crisis: A Data-Driven Analysis and Modeling**

-Aniruddh Rajagopal, Harshith Maddala, Siddharth Saravanan

**The George Washington University, DC**

In an ideal World, access to affordable and nutritious food would be a fundamental component of public health and community well being. However in the real world, in many regions across the United States, there exists certain regions/communities which have limited access to supermarkets, grocery stores and other sources of healthy food. These areas are known as ‘food desserts’. Residents in these areas often face socio-economic barriers such as high poverty rates, lack of reliable transportation, and long travel distances to reach the nearest full-service grocery outlet. With limited options available, they often tend to rely on convenience stores or fast food outlets that offer fewer healthy options, contributing to health related problems.

But that’s not the worst part the low-income families and elderly individuals had to face in these food desert areas. This inadequacy of healthy food not only increases the severity of health issues, but also plays a subtle role in reinforcing cycles of poverty and poor health outcomes. Over time, they accumulate these adversaries and contribute to systemic inequities, where residents of these neighborhoods face consistent and severe disadvantages in their pursuit of a healthy life.

Understanding the factors which give rise or helps in maintaining these food deserts is crucial for developing effective solutions. Firstly, we need to identify the demographic and structural conditions, that strongly associates with poor food access, stakeholders that policy makers, urban planners and officials can target and craft solutions to resolve

Using these two datasets—the USDA Food Access Research Atlas and the Food Environment Atlas—our group has been tasked with uncovering key indicators that correlate with the inaccessibility of healthy and nutritious food. Through Python-based analysis and visualization, we aim to identify patterns, isolate the most impactful demographic and socioeconomic variables, and ultimately provide actionable policy recommendations at the national level to address and reduce the existence of food deserts.

**1. Objective:**

The primary objective of this project is to analyze and uncover key factors that contribute to the existence of food deserts—low-income areas with limited access to fresh and nutritious food—across the United States. Using the USDA’s Food Research Atlas (tract-level data on supermarket accessibility) and Food Environment Atlas (county-level data on socio-economic and health-related variables), the goal is to identify significant indicators such as poverty rate, vehicle access, SNAP participation, and regional disparities that are strongly correlated with food inaccessibility. Through comprehensive data analysis and visualization, the project aims to reveal patterns that explain why certain communities face greater food access challenges and ultimately provide evidence-based policy recommendations to help mitigate and reduce the prevalence of food deserts on a national scale.

**2. Scope:**

This report focuses on data from the 2019 USDA Food Access Research Atlas, which includes tract-level information from the 2010 U.S. Census. The analysis is conducted entirely in Python using libraries such as pandas, matplotlib, and seaborn. The findings presented here are descriptive and exploratory, providing a foundation for causal and predictive modeling.

**3. Data Sources**

This project uses two publicly available datasets curated by the United States Department of Agriculture (USDA):

1. Food Access Research Atlas (2019): This dataset gives census tract-level data on food access indicators, with focus on helping identifying areas that qualify as food deserts. This helps in detailed spatial and demographic analysis. It also includes numerical information like distance to the nearest supermarket, income levels, population characteristics, and vehicle access.
2. Food Environment Atlas: we do not use this as the primary analysis, but this gives us an understanding of country-level variables that further describe food choices, health and well-being, and community characteristics.

Structure of the Food Access Research Atlas Dataset

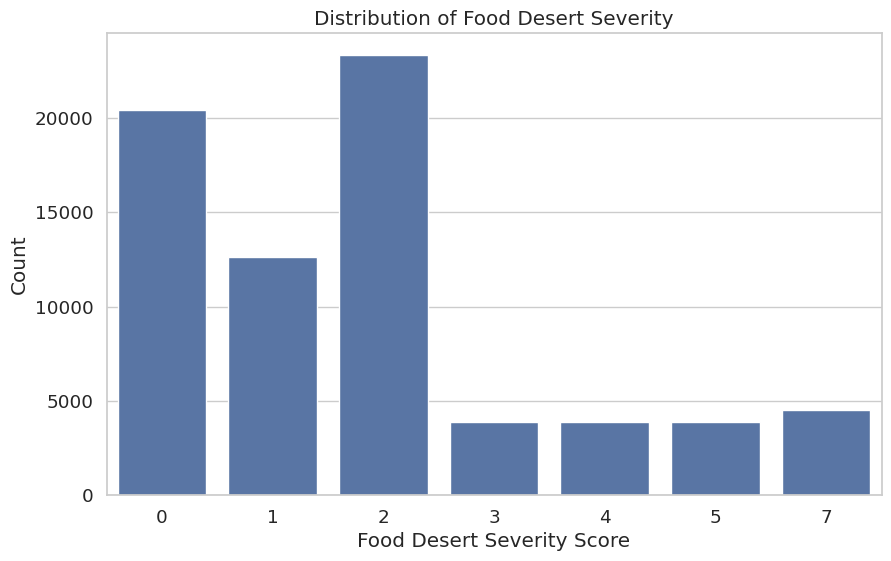
The Food Access Research Atlas dataset consists of over 70,000 records, each representing a single U.S. census tract. Key columns include:

* CensusTract, State, County – Geographic identifiers
* Pop2010, OHU2010 – Population and housing unit counts from the 2010 Census
* PovertyRate – Percentage of individuals living below the federal poverty line
* MedianFamilyIncome – Median income of families in the tract
* TractLOWI, TractSNAP, TractHUNV – Counts of low-income individuals, households receiving SNAP benefits, and households without a vehicle, respectively
* LILATracts\_1And10 – Binary flag identifying tracts that are both low-income and located more than 1 mile (urban) or 10 miles (rural) from the nearest supermarket

**4. Exploratory Data Analysis with Visualizations**

Once we have imported the data, we start working the dataset and find key insights by Visualizing them. Few of the plots which helps us better understand the data and provide helpful insights would be mentioned here.

**4.1. Distribution of Food Desert Severity:**

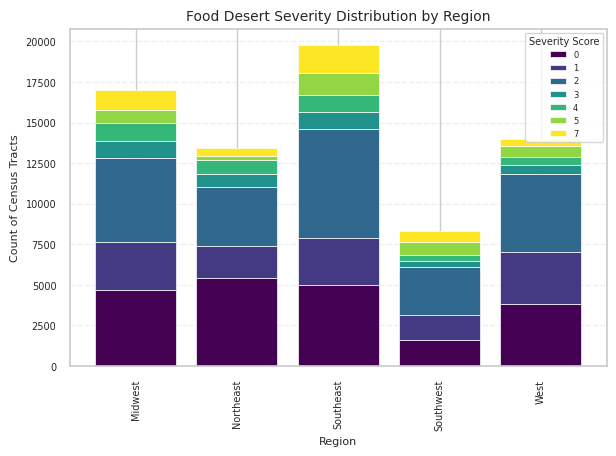


We plotted a Food Desert Severity Score for each census tract which has an aggregate score of USDA-defined low-income and low-access indicators to reflect the cumulative impact of different types of food access challenges.These values have the columns: LILATracts\_1And10, LILATracts\_halfAnd10, LILATracts\_1And20, LILATracts\_Vehicle, LATracts\_half, LATracts1, LATracts10, LATracts20, and LATractsVehicle\_20. These variables generally tells us whether a tract has certain criteria for geographic isolation, low income, or limited vehicle access. Because each indicator is binary (0 or 1), a tract's severity score represents the total number of conditions it meets, with higher values indicating more severe or overlapping food access challenges.

**Key Findings:**Most communities face either no food access issues or only a few – the majority of census tracts have a severity score of 0, 1, or 2. This means they either aren’t food deserts or they meet just one or two conditions for poor food access.

Only a small number of places face many overlapping food challenges – very few tracts have high severity scores (like 5 or more), which means severe food desert conditions are rare but do exist and need focused attention.

**4.2. Food Desert Severity Distribution by Region:**

****

To understand how food deserts were spread out across different parts of the United States, we categorized states into five geographic regions: Northeast, Southeast, Midwest, Southwest, and West. Each census tract was assigned a region based on its state, and the distribution of Food Desert Severity Scores was visualized in a stacked bar chart. This visualization displays the number of census tracts within each region, broken down by severity level (ranging from 0 to 7 we obtained earlier).

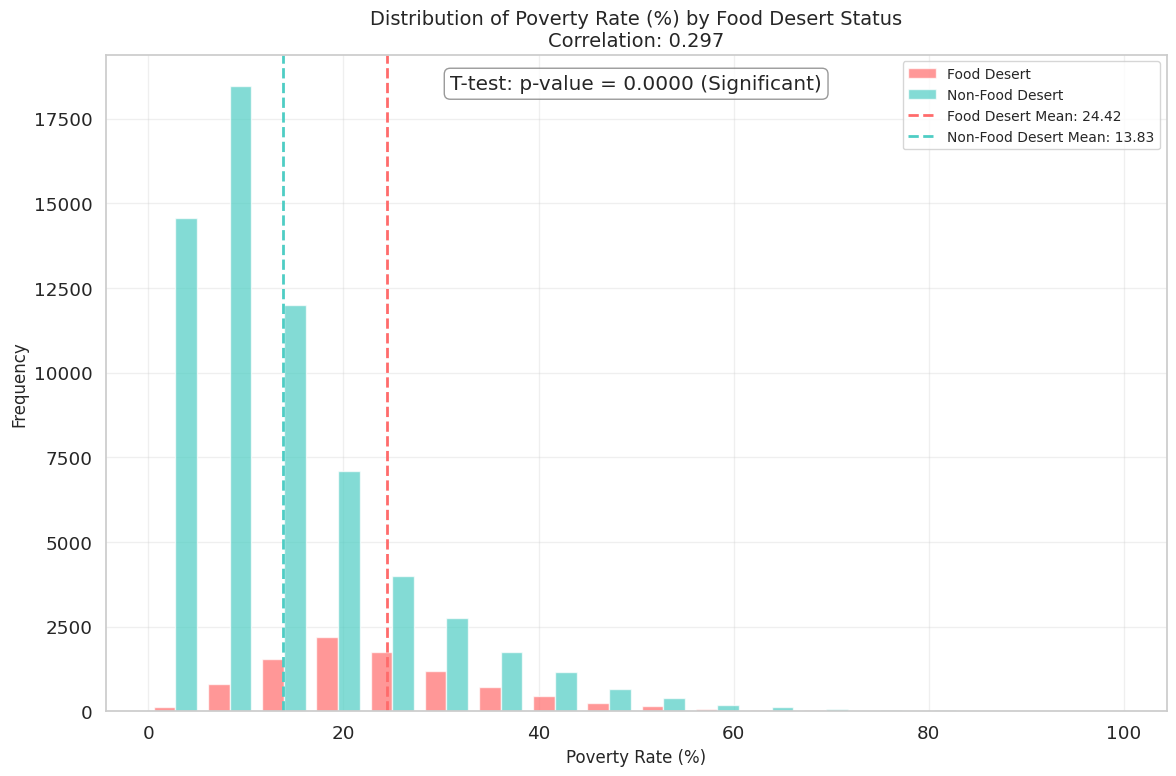
**Key Findings:**

The Southeast region has the highest number of census tracts with food desert indicators, including many with higher severity scores—suggesting both widespread and more intense food access challenges in that area.

While the Southwest has fewer census tracts overall, a noticeable portion of them have moderate to high severity scores, indicating that food deserts there may be more concentrated and impactful in specific locations.

**4.3. Food desert status by each Demographic factor:**

**4.3.1. Poverty Rate**

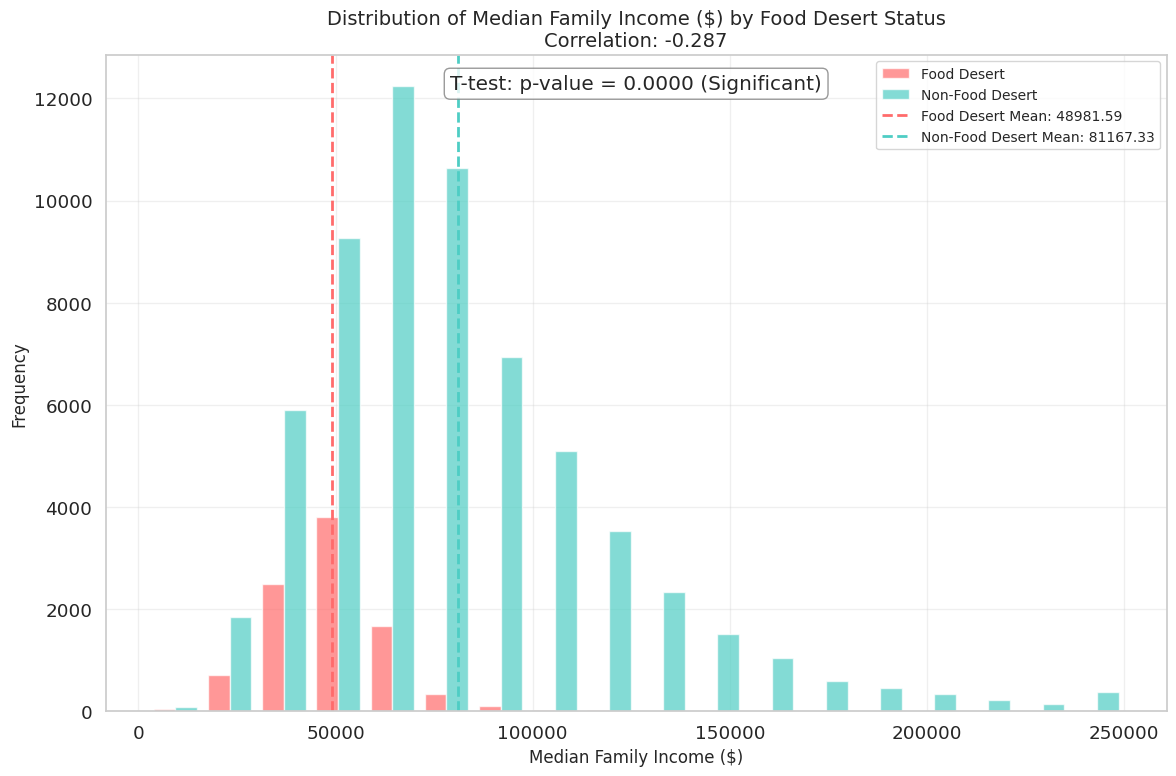
****

This graph compares the distribution of poverty rates between food desert and non-food desert census tracts, and includes the results of a t-test and correlation analysis.

**Key Insights:**

1. Census tracts classified as food deserts have a significantly higher average poverty rate (24.42%) compared to non-food deserts (13.83%), supported by a highly significant t-test H₀ (Null Hypothesis): There is no difference in the mean poverty rate between food desert and non-food desert census tracts. Since p-value is 0.0000, we reject the Null Hypothesis.
2. Although there is some overlap, food deserts are more commonly found in areas with higher poverty levels, as shown by the positive correlation of 0.297.

**4.3.2. Median Family Income**

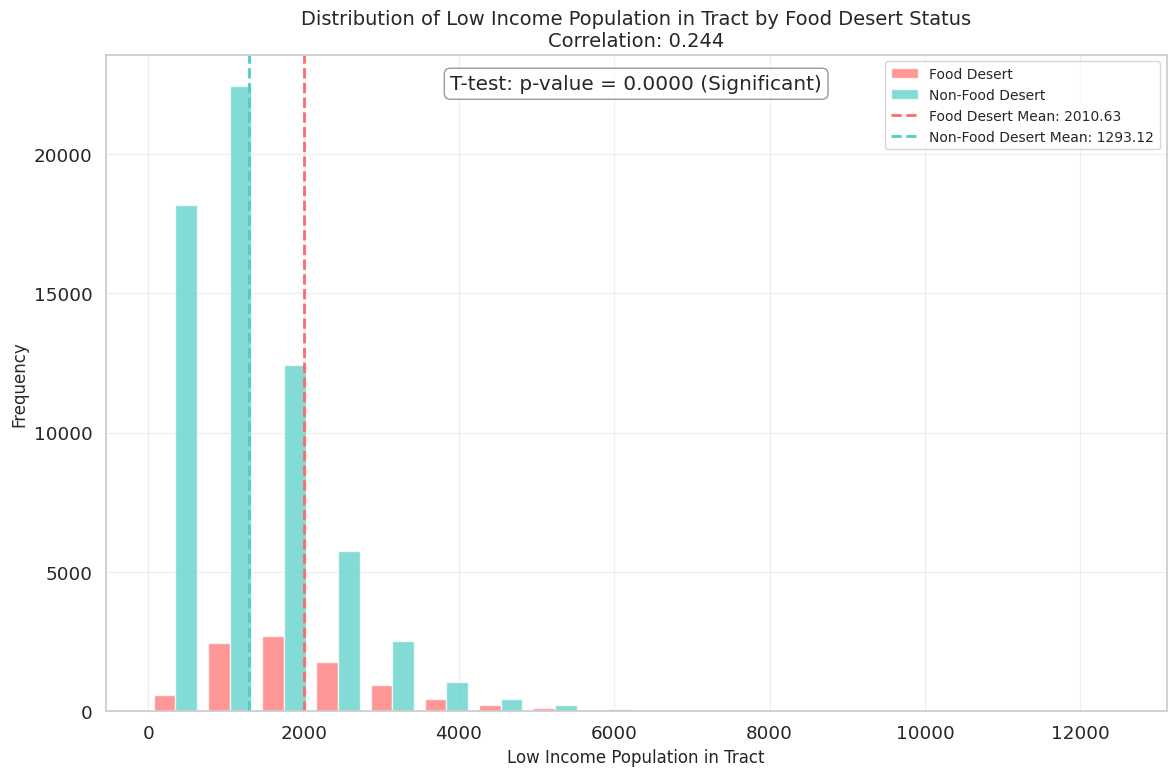
****

This graph shows the distribution of median family income across food desert and non-food desert census tracts, with a comparison of their means and results from a t-test and correlation analysis.

**Key Insights:**

1. Census tracts classified as food deserts have significantly lower median family incomes ($48,982) compared to non-food desert tracts ($81,167), indicating a strong economic divide.
2. The negative correlation (−0.287) and a highly significant p-value (0.0000) confirm that lower median income is closely associated with the likelihood of being a food desert. Null Hypothesis would be; There is no difference in the mean median family income between food desert and non-food desert tracts. Since the p-value is essentially zero, we reject the null hypothesis.

**4.3.3. Low Income Population**

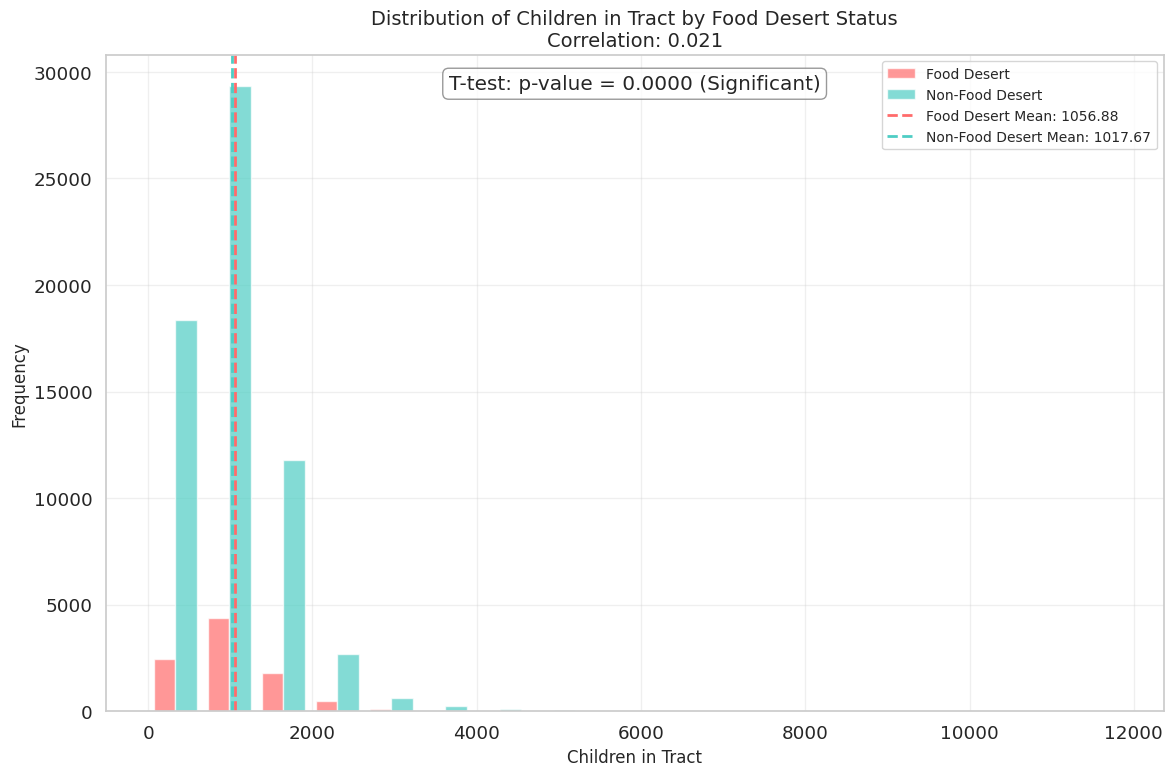
****

This graph compares the distribution of low-income population counts between food desert and non-food desert census tracts, alongside the results of a t-test and correlation analysis.

**Key Insights:**

1. On average, food desert tracts have a significantly higher low-income population (mean = 2,010.63) compared to non-food desert tracts (mean = 1,293.12). Null Hypothesis would be there is no difference in the mean low-income population between food desert and non-food desert tracts. Since the p-value is 0.0000, we reject the null hypothesis, confirming that the low-income population is significantly associated with food desert status.
2. The positive correlation (0.244) suggests that tracts with more low-income residents are more likely to be classified as food deserts.

**4.3.4. Distribution of Children**

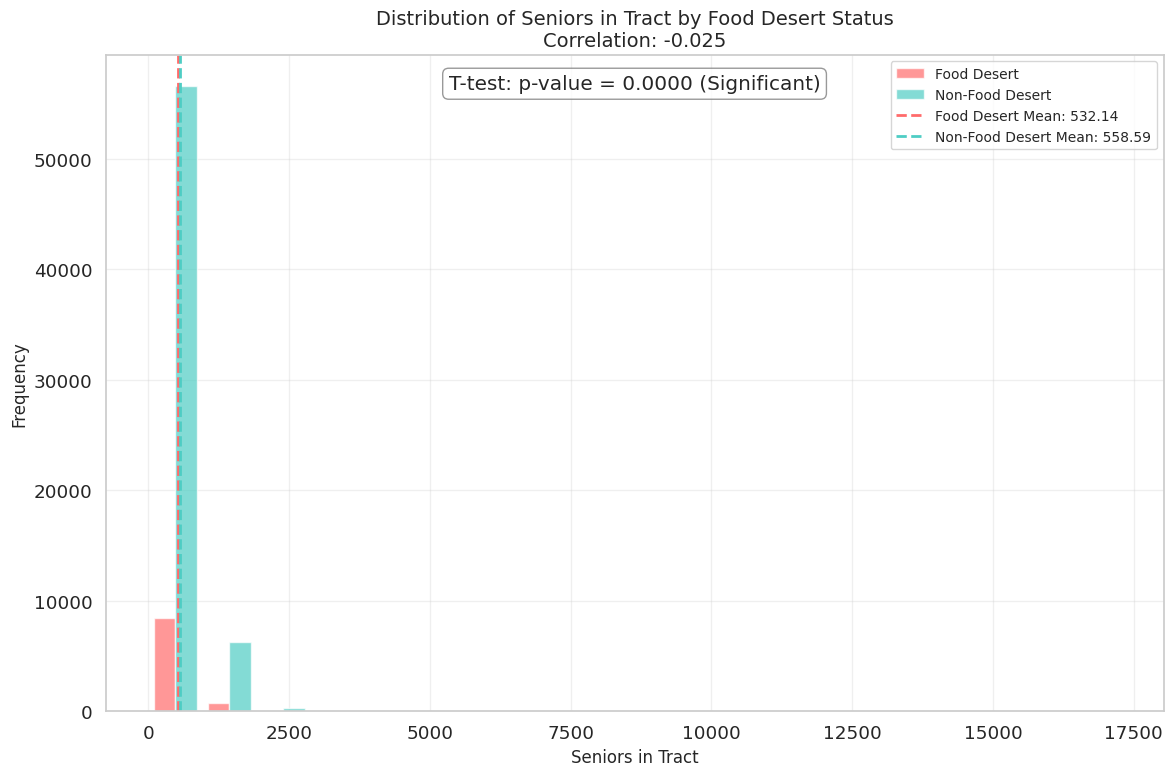


This graph shows the distribution of children (under 18) across census tracts, comparing those classified as food deserts and non-food deserts. It includes a t-test result and correlation analysis to assess statistical significance.

**Key Insights:**

1. The average number of children per tract is very similar between food desert (1,056.88) and non-food desert tracts (1,017.67), with only a small difference in means. Null Hypothesis would be: There is no difference in the average number of children between food desert and non-food desert tracts The null hypothesis is rejected due to the low p-value
2. Despite a statistically significant p-value (0.0000), the correlation between child population and food desert status is extremely weak (0.021), suggesting little to no practical relationship.

**4.3.5. Distribution of Seniors**

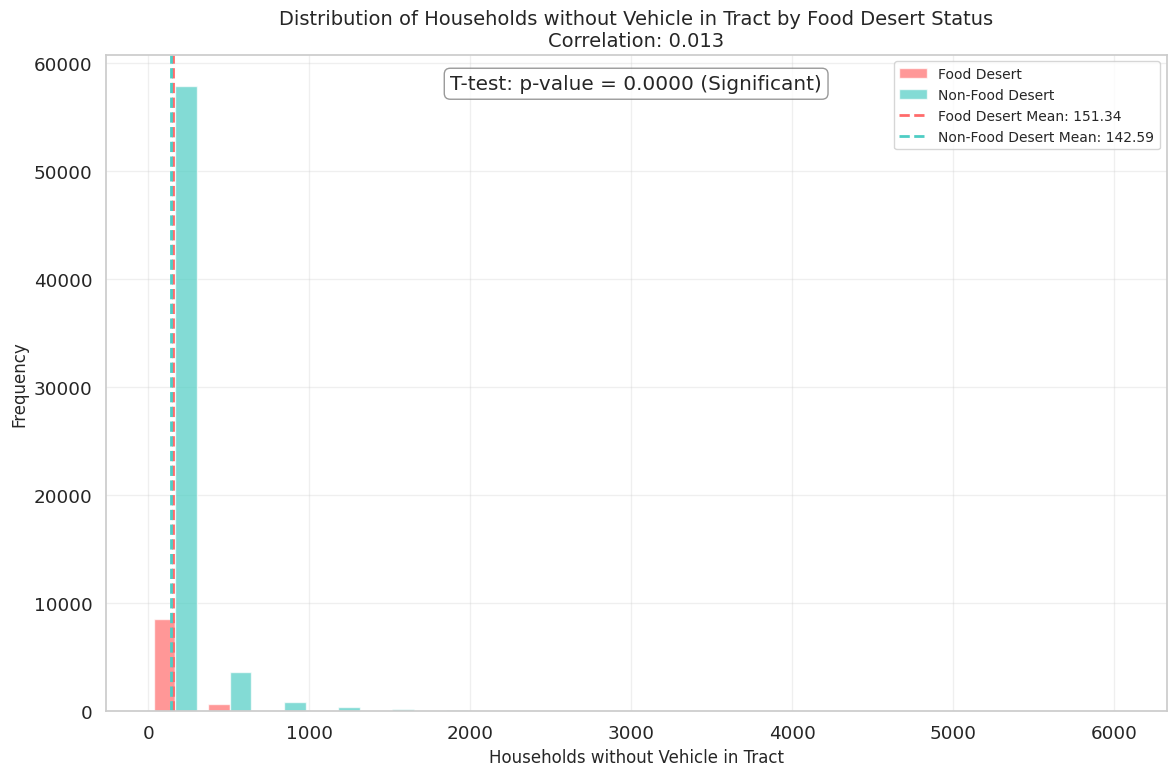


This graph illustrates the distribution of the senior population (typically age 65+) in food desert and non-food desert census tracts, along with a t-test and correlation analysis.

**Key Insights:**

1. The average number of seniors is slightly lower in food desert tracts (532.14) compared to non-food desert tracts (558.59), though the difference is small. Null Hypothesis would be: There is no difference in the mean number of seniors between food desert and non-food desert tracts .
2. Despite a statistically significant p-value (0.0000), the correlation is very weak (−0.025), indicating no meaningful relationship between the number of seniors in a tract and its likelihood of being a food desert.

**4.3.6. Distribution of Households without vehicle in Tract**

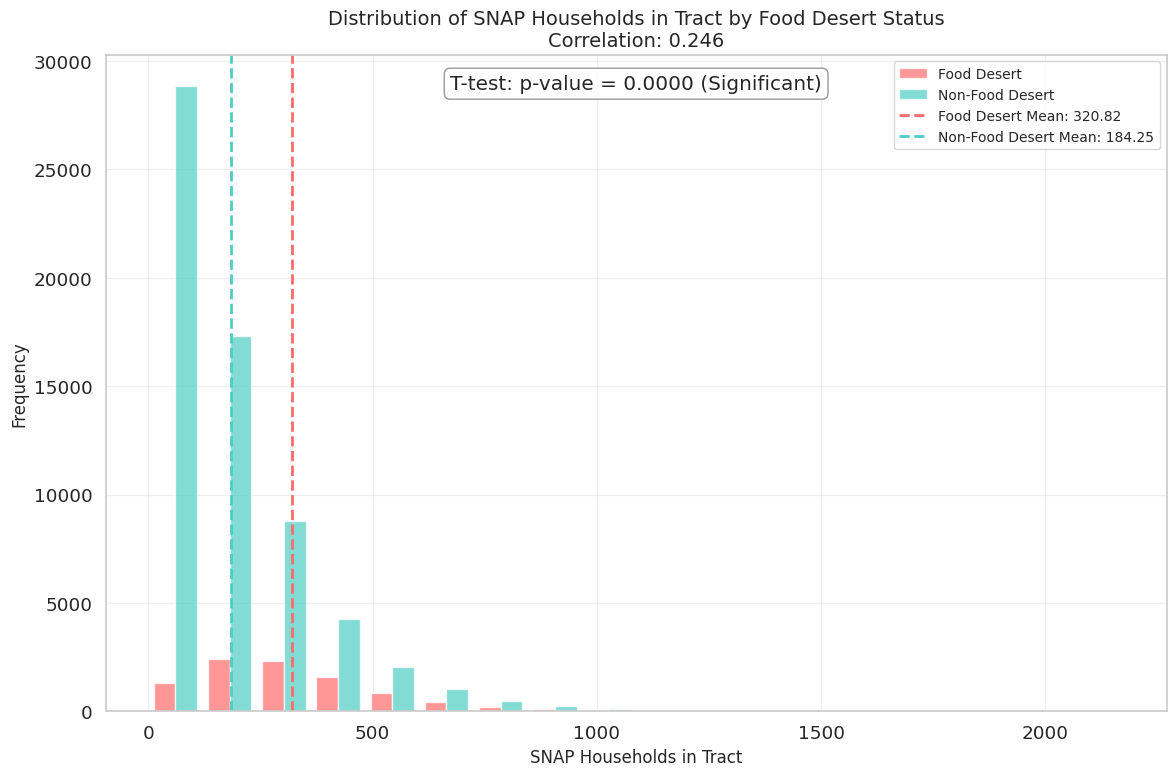
****

This graph displays the distribution of households without a vehicle in census tracts, comparing food desert and non-food desert areas. It includes statistical insights from a t-test and a correlation analysis.

**Key Insights:**

1. Food desert tracts have a slightly higher average number of vehicle-less households (151.34) compared to non-food desert tracts (142.59), but the difference is minimal. Null Hypothesis would be: There is no difference in the mean number of households without a vehicle between food desert and non-food desert tracts.
2. The correlation is extremely weak (0.013), suggesting that although vehicle access is a known barrier, the raw count of vehicle-less households alone does not strongly differentiate food desert areas.

**4.3.7. Distribution of SNAP Household in Tract**

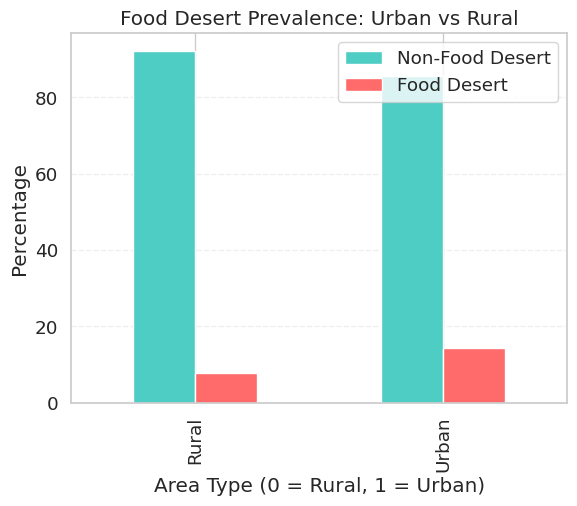


This graph compares the distribution of SNAP (Supplemental Nutrition Assistance Program) households between food desert and non-food desert census tracts, supported by a t-test and correlation analysis.

**Key Insights:**

1. Food desert tracts have a notably higher average number of SNAP households (320.82) than non-food desert tracts (184.25), indicating that food-insecure areas often rely more heavily on federal food assistance. Null Hypothesis would be: There is no difference in the mean number of SNAP households between food desert and non-food desert tracts.
2. With a correlation of 0.246 and a highly significant p-value (0.0000), the number of SNAP households is moderately and meaningfully associated with food desert status.

**4.4. Rural VS Urban Food Desert Comparison**

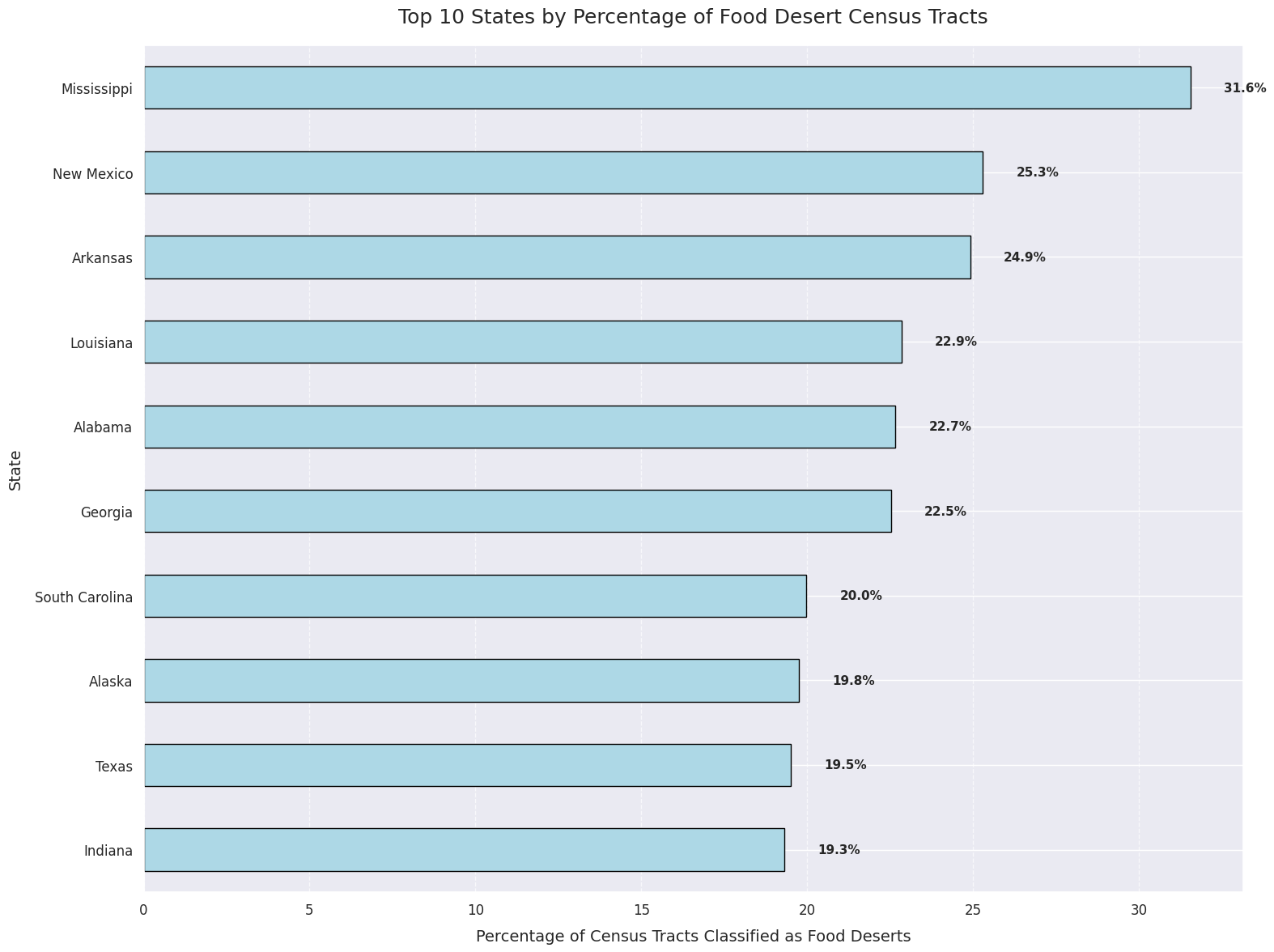
****

This bar chart shows the percentage of census tracts classified as food deserts versus non-food deserts, separated by area type: rural (0) and urban (1).

**Key Insights:**

1. In both urban and rural areas, the majority of tracts are not food deserts; however, the share of food desert tracts is higher in urban areas (around 15%) compared to rural areas (less than 10%).
2. Although food deserts exist in both settings, urban areas tend to have a slightly greater prevalence, likely due to higher population density and income disparities concentrated in certain neighborhoods.

**4.5. Top 10 states with Food Desert areas**

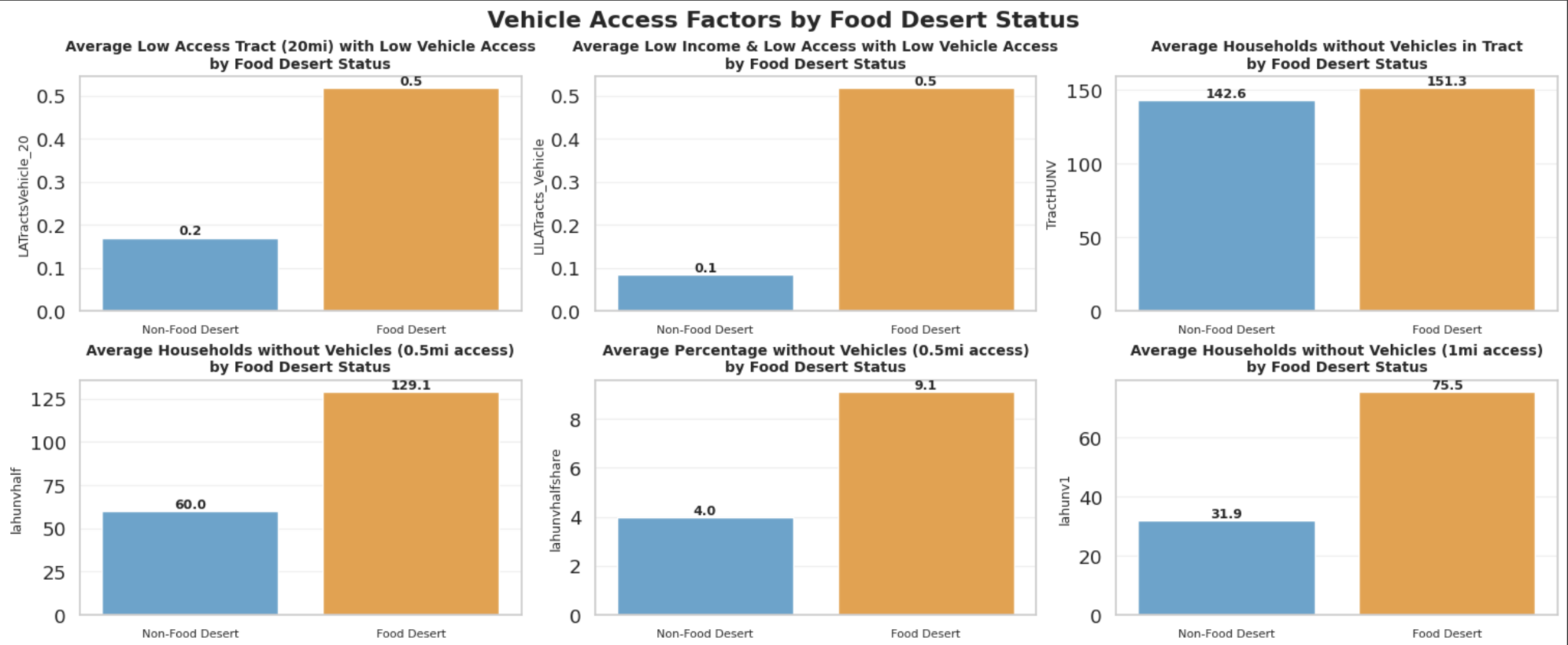
****

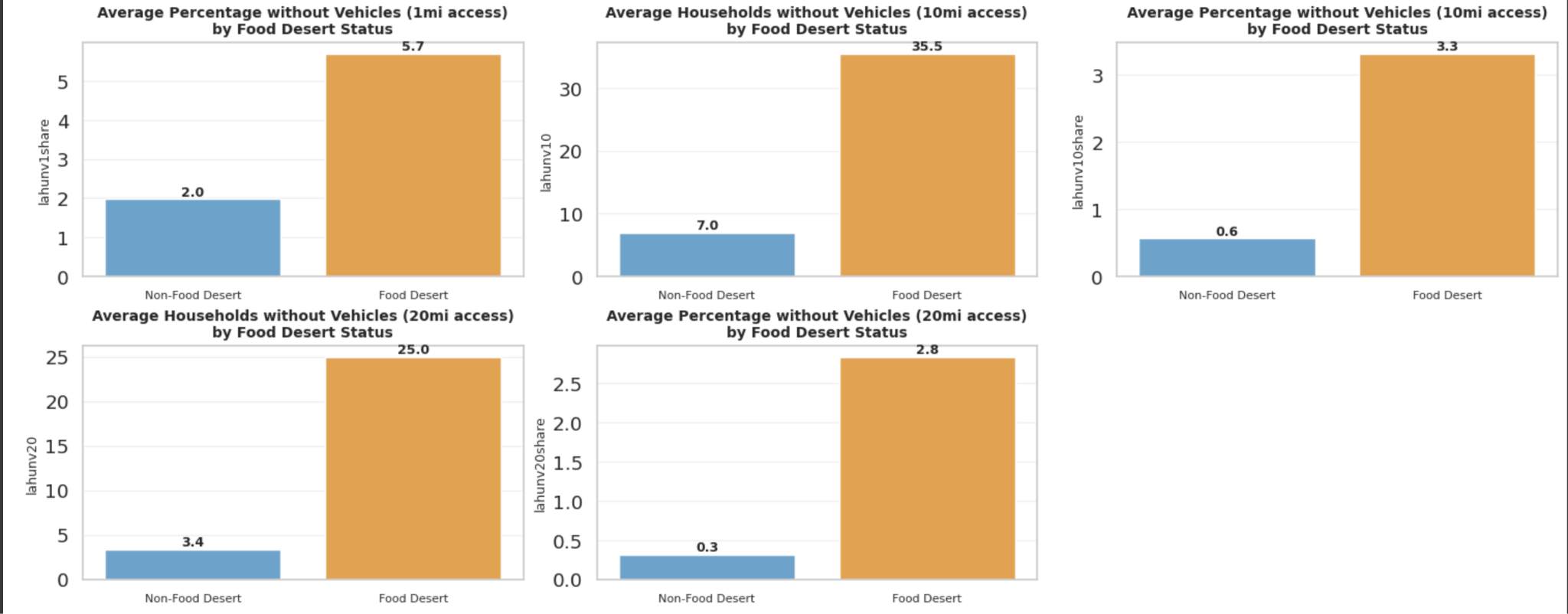
This horizontal bar chart highlights the top 10 U.S. states with the highest percentage of census tracts classified as food deserts.

**Key Insights:**

Mississippi leads with 31.6%, followed by New Mexico (25.3%), Arkansas (24.9%), and Louisiana (22.9%). The data clearly shows that the food desert burden is concentrated in the Southern region of the United States. These high percentages point to systemic challenges in food access, transportation, and poverty that are more prevalent in these states. Targeted policy interventions and infrastructure investment are urgently needed in these areas to improve access to fresh and nutritious food.

**4.6. Vehicle Access Factors by Food Desert Status**

****

****

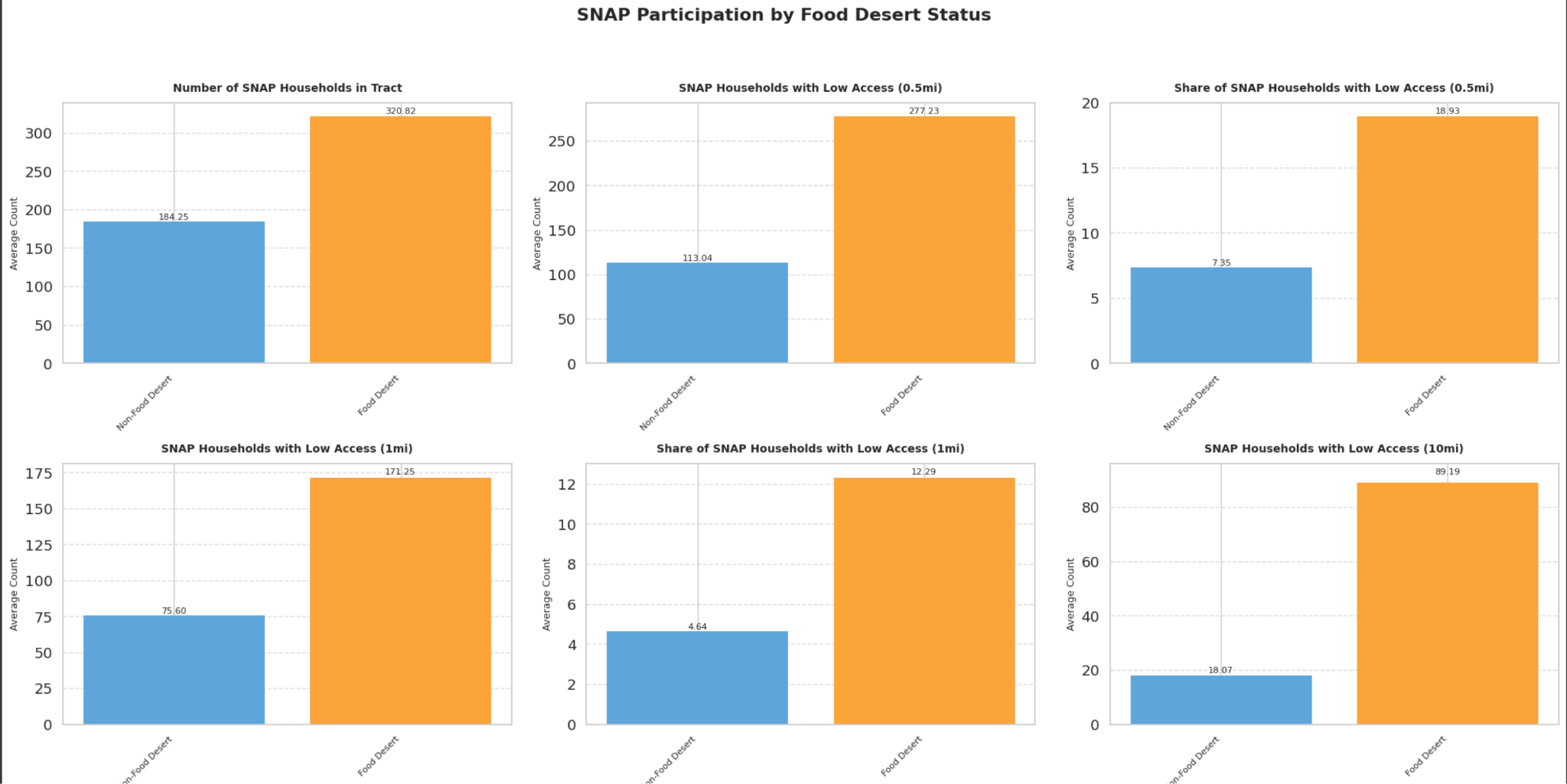
These visualizations compare vehicle access-related indicators between food desert and non-food desert census tracts across multiple distance thresholds to grocery stores (0.5, 1, 10, and 20 miles). Across all measures, food desert tracts consistently exhibit more severe transportation challenges.

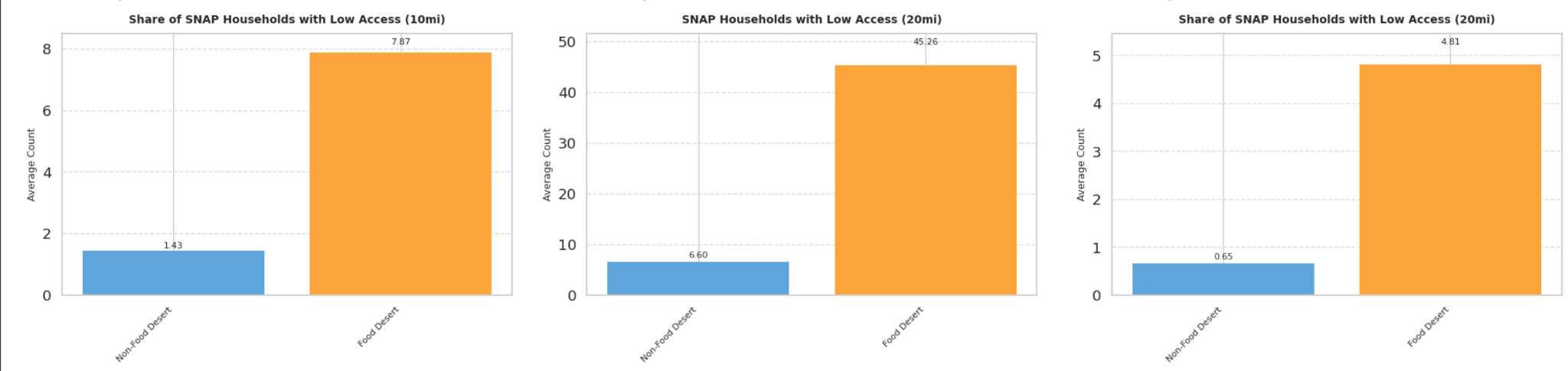
**Key Insights:**

The average number of households without vehicles is significantly higher in food deserts—for example, nearly 129 households within 0.5 miles lack a vehicle in food desert areas, compared to just 60 in non-food deserts. A similar pattern emerges at greater distances: 75.5 vs. 31.9 at 1 mile, and 25 vs. 3.4 at 20 miles. Additionally, the proportion of the population without vehicle access is more than twice as high in food deserts across all ranges, highlighting a systemic disparity in mobility and grocery access.

Further reinforcing this trend, binary indicators such as LATractsVehicle\_20 and LILATracts\_Vehicle show that food desert tracts are far more likely to meet combined criteria for low income, low vehicle access, and being located far from supermarkets. These compounding barriers are not isolated occurrences but are widespread across food desert communities. The clear and consistent gap in both household counts and population percentages underscores the critical role that transportation plays in defining food access inequality. Overall, this analysis illustrates that food deserts are not just shaped by geographic distance but also by the mobility limitations of the communities living within them.

**4.7. Distribution of SNAP household**





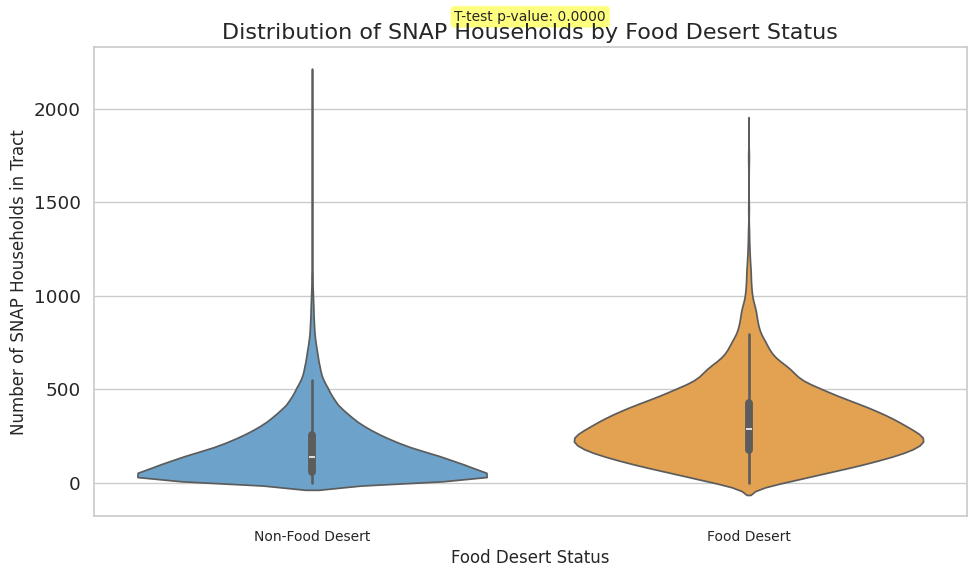
These sets of charts compare SNAP (Supplemental Nutrition Assistance Program) participation between food desert and non-food desert census tracts, focusing on both absolute counts and proportional shares across various distance thresholds from grocery stores (0.5, 1, 10, and 20 miles).

**Key Insights:**

The data clearly shows that food desert areas have significantly higher average numbers of SNAP households (320.82 vs. 184.25). More importantly, food desert tracts have nearly double or more SNAP households with low access at every distance category. For example, at the 0.5-mile range, food desert tracts average 277.23 households with low access, compared to only 113.04 in non-food deserts. This trend continues at 1 mile (171.25 vs. 75.60), 10 miles (89.19 vs. 18.07), and 20 miles (45.26 vs. 6.60), underscoring a severe access gap for SNAP-dependent communities.

In addition to the raw counts, the share of SNAP households facing access issues is also much higher in food deserts. At 0.5 miles, 18.93% of SNAP households in food deserts experience low access, compared to just 7.35% in non-food deserts. By the 20-mile mark, this gap remains stark (4.81% vs. 0.65%). These differences emphasize that not only do more people in food deserts rely on SNAP, but they are also disproportionately affected by geographic food access barriers. Together, these insights highlight the urgent need for targeted interventions—such as improved transportation options or mobile grocery programs—to address nutritional inequities in SNAP-reliant communities.

**4.8. Distribution of SNAP Household by Food Desert Status**

****

This violin plot shows the number of SNAP (food assistance) households in areas that are food deserts versus those that are not. The orange section (food deserts) is clearly wider and higher, meaning these areas have more SNAP households on average. The shape also shows that the distribution of SNAP households is consistently higher in food deserts than in non-food desert tracts.

**Key Insights:**

The t-test result confirms this difference is statistically significant (p-value = 0.0000). This means that food deserts tend to have many more people relying on SNAP benefits, and it’s not just by chance. It highlights that food-insecure areas are not only low on access but also house more residents who depend on food assistance.

**5. Modeling**

### **5.1 Classification Model**

#### **Objective**

The classification model aims to predict whether a given census tract qualifies as a food desert based on a set of socio-economic and geographical features. This binary classification task helps in identifying high-risk areas that may require intervention.

#### **Model Selection**

We employed a Random Forest Classifier due to its robustness, ability to handle high-dimensional data, and built-in feature importance ranking. This model was chosen because:

It effectively manages a large number of independent features without overfitting.

It provides interpretability through feature importance scores.

It performs well with missing data and categorical variables.

#### **Feature Engineering**

The target variable, FoodDesert, was derived using the median value of TractSNAP, where tracts with TractSNAP above the median were labeled as 1 (Food Desert) and others as 0 (Not a Food Desert).

* Categorical variables were handled using one-hot encoding.
* Features were standardized to improve model performance.

#### **Model Training and Evaluation**

* Data Split: 80% training, 20% testing.
* Classifier: Random Forest with default hyperparameters.
* Evaluation Metrics:  
  + Accuracy: 95%
  + Confusion Matrix:  
    - True Positives (TP): 6780
    - True Negatives (TN): 7013
    - False Positives (FP): 361
    - False Negatives (FN): 353
  + Precision & Recall: Both at 95%, indicating strong predictive power with minimal false classifications.

#### **Insights from Classification Model**

* The model accurately identifies food deserts, enabling policymakers to focus on vulnerable regions.
* The most important features influencing classification include socioeconomic indicators such as median income, access to grocery stores, and percentage of households receiving SNAP benefits.
* The low false negative rate suggests the model is highly sensitive in detecting food deserts.

### **5.2 Regression Model**

#### **Objective**

The regression model predicts the value, which represents the number of households in a census tract participating in the Supplemental Nutrition Assistance Program (SNAP). This helps in understanding which factors drive higher food insecurity levels.

#### **Model Selection**

We used a Random Forest Regressor, selected for:

* Its ability to handle high-dimensional data efficiently.
* Its robustness to outliers and missing data.
* Its capability to capture non-linear relationships in the dataset.

#### Feature Selection and Engineering

* The dataset initially contained 146 independent features.
* Feature importance analysis reduced the model to 20 key predictors, improving performance and reducing overfitting.
* The dataset was standardized before training the regression model.

#### Model Training and Evaluation

* Data Split: 80% training, 20% testing.
* Regressor: Random Forest with default hyperparameters.
* Evaluation Metrics:  
  + Mean Squared Error (MSE): 1894.77
  + Root Mean Squared Error (RMSE): 43.53
  + R² Score: 0.9458 (94.58% of the variance in TractSNAP is explained by the model)

#### Insights from Regression Model

* The model effectively predicts SNAP participation, with high accuracy (R² = 0.9458).
* Key predictors include household income, unemployment rate, vehicle availability, and distance to grocery stores.
* The model can be used for forecasting food insecurity trends and guiding resource allocation efforts.

**6. Proposed Solutions**

* Analysis of Food desert severity reveals that the Southeast and Midwest have the highest concentration of food desert severity, highlighting the urgent need for targeted intervention and resource allocation in these areas. Prioritize focused intervention in the top 10 states with the highest concentration of food deserts (e.g., Mississippi, New Mexico, Arkansas, Louisiana, etc.) Addressing regional disparities is essential to making meaningful progress toward nationwide food equity.
* Establish subsidized, community-run grocery stores or cooperatives in high-poverty areas, supported by government grants, local farm partnerships, and SNAP integration, to ensure affordable and consistent access to healthy food while empowering local residents.
* To address transportation barriers, initiatives that enhance mobility—such as community shuttle services, subsidized ride programs, or strategic placement of mobile grocery units—can help ensure that households without vehicles, especially those living farther from food sources, still have reliable access to healthy and affordable groceries.
* Restructure SNAP distribution infrastructure to prioritize expansion in areas farther from grocery outlets, ensuring that outreach and benefit support extends beyond the typical 0.5-mile radius to reach SNAP-eligible households living 10 or even 20 miles away from food sources. This could involve collaboration with mobile markets, delivery services, and satellite SNAP centers to serve remote food-insecure populations more effectively.
* By using the classification and regression models built in this study, stakeholders can proactively identify census tracts at risk of becoming food deserts or experiencing heightened food insecurity. These models enable data-driven decision-making, allowing targeted policy responses, efficient allocation of SNAP resources, and strategic placement of grocery services in areas predicted to show increased need.

**7. Conclusion**

In conclusion, this report aims to highlight the complex and interlinked factors that contribute to the existence and preservation of food deserts across the United States. Through a detailed examination of socio-economic, geographic, and access-related variables—including poverty levels, vehicle ownership, SNAP dependency, and regional disparities—it becomes evident that food deserts are shaped by systemic inequities that go beyond simple proximity to grocery stores. The findings underscore the urgent need for interventions involving government investment, targeted policies, community-driven programs, and individual action. Addressing these issues requires not only improving food access infrastructure but also empowering and encouraging low-income communities through education, transportation support, and sustainable local food systems. By leveraging data-driven insights and promoting collaborative solutions at national, state, and local levels, we can take meaningful steps toward reducing the prevalence of food deserts and building a healthier, more equitable food environment for all Americans.

**8. References**

1. [U.S. DEPARTMENT OF AGRICULTURE](https://www.ers.usda.gov)
2. [Communities With Limited Food Access in the United States](https://www.aecf.org/blog/communities-with-limited-food-access-in-the-united-states)
3. [How to Code the Student’s t-Test from Scratch in Python](https://machinelearningmastery.com/how-to-code-the-students-t-test-from-scratch-in-python/?utm_source=chatgpt.com)