Food desert Analysis

# Question 1:

**\*\*What is the geographic distribution of food deserts in the dataset by state and county?\*\***

**- How many unique states and counties are represented in the dataset? - What is the total number of census tracts included?**

* There are a total of 72,864 entries in the dataset.
* **POP2010** column: The average population in 2010 for the census tracts is around 4,237, with a minimum of 0 and a maximum of 37,341.
* **OHU2010** column: The average housing units in 2010 is approximately 1,602.
* There are 1,837 unique counties present in the dataset.
* County "0" (might be a placeholder or missing label) has 7 non-urban areas.
* **Number of Unique States**: There are 51 unique states present in the dataset. This count includes all 50 US states and the District of Columbia.
* Some of the census tracts in Alaska have a population (**POP2010**) as low as 87.

# Question 2

* **Urban** column: About 75.7% of the entries are labeled as urban, given its mean value.

**Percentage of Urban Areas with LILATracts\_1And10 value of 1**:

* About 14.33% of urban areas are classified as low income and low access tracts based on the **LILATracts\_1And10** column.
* Out of the 75% urban tracts, 14% are food deserts

**Percentage of Non-Urban Areas with LILATracts\_1And10 value of 1**:

* Approximately 7.57% of non-urban areas are classified as low income and low access tracts based on the **LILATracts\_1And10** column.
* Out of the 25% Rural tracts, 7% are food deserts

# Question 3 :

**Group quarters with food deserts -** 22%

**group quarters without food deserts – 13%**

**percentage of group quarters in overall tracts**

* **0.7 %**
* **The tracts with group quarters are more in urban and so the percentage of food deserts in high in urban compared to rural. 95% of the food desert tracts with groupquarters are in urban.**
* Percentage of rural people in group quarters – 0.4%
* Percentage of urban people in group quarters – 0.8%
* **# poverty rate is high in group quarters tracts - 80 % higher than non-group quarters tract**
* **Ratio of low income tracts with group quarters – 53%**
* **Ratio of low income tracts with non- group quarters – 42%**

# Question 4 :- Scatter plot- between poverty rate in all states and count of food desert.

**\*\*What is the relationship between poverty rates (PovertyRate) and the presence of food deserts in different regions (State/County)?\*\***

**- Are there states/counties with higher poverty rates that also have a higher prevalence of food deserts?**

**- Is there a correlation between poverty rates and food desert status?**

* Low access is not correlated with low income  
   \* For tracts with non-low access the low income tracts are evenly distributed
* # Low access people are rich, rich people wants to stay outside.
  + 66% of the non-low income tracts are low access tracts
* Share of the rich and urban tracts is 72%
* Share of the rich and rural tracts is 28%
* Poverty rate is high in food deserted areas
  + In food deserts , poverty rate is 75% higher than non-food deserts
* ## Poverty rate is a little high in non-low access areas
  + 26% higher
* **Are there states/counties with higher poverty rates that also have a higher prevalence of food deserts?** There is no correlation between high poverty rate states/counties and food desert status –
  + **State -** 0.21
  + **County -** 0.01
* There is very less correlation between poverty rate and food deserts in all states.
  + **State -** 0.41
  + **County -** 0.08

#### Hypothesis testing for states :

##### Null hypothesis :

There is no relationship between poverty rate and food deserts counts of the states

##### alternative hypothesis :

There is some linear correlation, either positive or negative between poverty rate and food deserts counts of the states

### Correlation between High poverty rate states with count of their food desert

We fail to reject the null hypothesis, The evidence is not strong enough to conclude that there is a statistically significant linear correlation between the two variables in the population.

A computer code with black text

Description automatically generated

### Correlation between poverty rate of all states with count of their food desert.

We reject the null hypothesis, in favour of the alternative, suggesting that there is a statistically significant linear correlation between the two variables in the population.

A computer code with black text

Description automatically generated

#### Hypothesis testing for counties :

##### Null hypothesis :

There is no relationship between poverty rate and food deserts counts of the counties

##### alternative hypothesis :

There is some linear correlation, either positive or negative between poverty rate and food deserts counts of the counties

### High poverty County correlation with count of their food desert

A computer screen shot of a number

Description automatically generated

We fail to reject the null hypothesis, The evidence is not strong enough to conclude that there is a statistically significant linear correlation between the two variables in the population.

### poverty County correlation with count of their food desert

# A computer screen shot of a number Description automatically generated

We fail to reject the null hypothesis, The evidence is not strong enough to conclude that there is a statistically significant linear correlation between the two variables in the population.

# So in brief :

There is no correlation between counties and poverty rate, there is also no correlation between high poverty states and their food desert counts but there is a linear correlation between all states and their food desert counts.

# Question 5 :

**5. \*\*Can we measure the impact of vehicle availability (LILATracts\_Vehicle) on food desert designation? \*\***

**- What percentage of food deserts have limited access to vehicles?**

* By using LILATRACT\_Vehicle we found that food desert people has less vehicle access and by using "lahunv1share" we found that percentage of people who doesn't have a vehicle is high for food deserts.
* As expected the percentage of people who doesn't have a vehicle is high for food deserts both in urban and rural areas
  + Food deserts and urban

A graph with a line graph

Description automatically generated

Food desert and rural

A graph with a line

Description automatically generated

* + Non - Food deserts and urban
* A graph with a line graph

  Description automatically generated

Non- food desert and rural

A graph with a line graph

Description automatically generated

### The percentage of people who doesn't have a car for 10 miles in rural and 1 mile in urban for each state

#### is there a significant difference in the prevalence of food deserts in areas with and without vehicle access issues?

Yes from the below analysis we find that food deserts are higher in areas without vehicles

* The percentage of people who doesn't have a car in a food desert is higher than a non-food desert for urban population
* Comparing the lahunv1share for urban "0.054564" and lahunv10share for rural "0.040797", the percentage of population who doesn't have vehicles is slightly more in urban food deserts

A graph of food desert and food desert

Description automatically generated with medium confidence

### #### Observations from the box plot – Do 2 sample tests

Central Tendency: The median line inside the box for the "food desert" category appears to be higher than the "non-food desert" category. This suggests that, on average, a higher percentage of the population in food deserts lack vehicles compared to those in non-food deserts.

Spread & Variability: The interquartile range (box height) for both categories is roughly similar, indicating a similar spread or variability in the data for both food and non-food deserts.

Outliers: There are some outliers in both categories, but it's particularly noticeable in the "non-food desert" category. This indicates that while most non-food desert tracts have a relatively lower percentage of people without vehicles, there are a few tracts where this isn't the case.

Skewness: The median line appears to be roughly in the middle of the box for food desert suggesting that the distribution of the share of the population without vehicles is symmetric for food desert and non-symmetric for non-food deserts.

Overall Comparison: The entire box (representing the middle 50% of the data) for the "food desert" category is higher on the y-axis compared to the "non-food desert" category. This indicates that a larger share of the population in food deserts typically lacks vehicles compared to those in non-food deserts. This could be concerning, as people in food deserts without vehicles may have even more difficulty accessing fresh and healthy food.

## ### Vehicle access in rural and urban tracts

A graph of a number of people

Description automatically generated with medium confidence

## Observation from the box plot: DO anova

**Non-Food Desert in Urban** (3rd category):

* **Central Tendency**: The median suggests that a relatively low percentage of the population in non-food desert urban areas lack vehicle access, even lower than the non-food desert rural areas.

**Urban Areas**:

The non-food desert urban areas have the smallest median percentage of people without vehicle access. This is likely because urban areas often have alternative modes of transportation, such as public transit, walking, or biking, making vehicle ownership less necessary. Moreover, urban areas often have amenities and services within walking distance, further reducing the need for a vehicle.

* **Impact of Food Deserts**: In both rural and urban contexts, food deserts show a higher median percentage of people without vehicle access compared to non-food deserts. This is a significant concern because it suggests that people in food deserts, who already might have limited access to healthy food options, might also face transportation challenges.
* **Rural vs. Urban**: While non-food desert urban areas have the lowest median percentage of people without vehicle access, food desert urban areas have the highest. This highlights the impact of food desert status on transportation barriers, irrespective of the urban or rural context.

# Question 6

**6. \*\*How do income-related variables (MedianFamilyIncome, LowIncomeTracts) relate to food deserts? \*\***

**- What is the median family income in food desert areas compared to non-food desert areas? - What percentage of food deserts are classified as low-income tracts?**

**Median Income in food deserts**

Look how the graph is skewed.

The urban population is skewing the data with high median family income, compared to the rural population

Median family income for all food deserts

A graph of a number of people

Description automatically generated

Median family income for rural deserts

A graph of a normal distribution

Description automatically generated

Median family income for urban deserts

A graph of a number of people

Description automatically generated

**Median Income in non-food deserts**

* In non-food desert areas the rural population is more richer than the urban population.

Non-food desert and urban

A graph of a normal distribution

Description automatically generated

Non-food desert and rural

A graph of a normal distribution

Description automatically generated

#### Medianfamily income Between urban and rural

#### Median family income in all urban tracts

A graph of a normal distribution

Description automatically generated

#### Median family income in all rural tracts

A graph of a normal distribution

Description automatically generated

**Median family income between urban, rural and food desert, non deserts**

**Observations:**

**Median Family Income in Food Deserts (Rural)**:

The distribution appears to be left-skewed, not right-skewed. The tail of the histogram extends more to the left, indicating a concentration of data on the right. This means there's a higher proportion of families in rural food deserts with higher incomes, and fewer families with lower incomes.

**2. Median Family Income in Food Deserts (Urban)**: This histogram appears to be right-skewed, suggesting a concentration of families with lower incomes in urban food deserts. The peak is towards the lower income values, with a tail extending to the higher income values.

**3. Median Family Income in Non-Food Deserts (Rural)**: This histogram seems more symmetrical, indicating a more even distribution of incomes. The peak is around the middle values, suggesting that families in rural areas that are not food deserts have a balanced mix of incomes.

**4. Median Family Income in Non-Food Deserts (Urban)**: This histogram appears roughly bell-shaped, indicating a normal distribution of incomes. The peak is around the middle values, suggesting that families in urban areas that are not food deserts also have a balanced mix of incomes.

**More inferences:**

* **Rural Food Deserts**: Contrary to common perceptions, rural food deserts appear to have a concentration of families with higher incomes. This suggests that the presence of food deserts in rural areas might not solely be due to economic reasons but could be influenced by other factors like distribution logistics, infrastructure, or availability of food retailers.
* **Urban Food Deserts**: As expected, urban food deserts have a concentration of families with lower incomes, indicating that economic factors play a significant role in the presence of urban food deserts.
* **Non-Food Deserts**: Whether in urban or rural settings, areas that are not classified as food deserts appear to have a more balanced mix of family incomes, suggesting that economic disparities might be less pronounced in these areas.

#### Medianfamily income Between food desert and non-food desert

* Non-food desert areas is more richer than food-desert tracts

Medianfamily income in Non food desert tracts

A graph of a normal distribution

Description automatically generated

Medianfamily income in food desert tracts

A graph of a number of people

Description automatically generated

1. **Median Family Income in Food Deserts**: This histogram appears to be skewed to the right, indicating that a significant portion of families in food deserts have lower incomes. There seems to be a peak around the lower income values, which tapers off as income increases.
2. **Median Family Income in Non-Food Deserts**: This histogram looks more symmetrical and possibly bell-shaped, suggesting a more normal distribution. The peak is closer to the middle income values, and the tails decrease symmetrically on both sides.
3. From this visual inspection, it seems that food deserts tend to have a higher proportion of families with lower incomes, while non-food deserts have a more even distribution of incomes, with a peak around the middle values.

### **What percentage of low-income tracts are classified as food deserts ?**[**¶**](http://127.0.0.1:8888/notebooks/EDA.ipynb#What-percentage-of-low-income-tracts-are-classified-as-food-deserts-?) **-**30%

# Box plot comparison

A graph showing a number of food and desserts

Description automatically generated with medium confidence

95th level Confidence interval of mean for Food deserts

### 43566.32 44127.11

Mean estimate

43846.72

95th level Confidence interval of mean for non- Food deserts

### 70588.70 71129.46

### Mean estimate

70859.08

A graph showing a number of food groups

Description automatically generated with medium confidence

Observations:

As we saw from the histograms, the median family incomes is about the same between food deserts in rual and urban places. And non-food deserts in urban have higher median income than others.

## ANOVA ON median income

Null hypothesis : There is no difference in the median income between food desert and non-food desert

Alternate hypothesis : There is difference in the median income between food desert and non-food desert

A screenshot of a computer

Description automatically generated

At level of significance alpha set at = 0.05, as The p-value 2e-16 is less than the 0.05, we reject the null hypothesis "There is no difference in the median income between food desert and non-food desert"

The anova results tell us that there is statistically significant difference in median income among food desert and non -deserts