

New York State Farmer's Market

Code & Visualizations

A 2016 Siena College report defined food deserts as “often situated in low-income [communities] and are often the result of unfortunate suburban or rural planning.” Communities in food deserts have limited access to grocery stores that provide fresh food on a consistent basis. In New York State, 32 out of 62 counties are identified as having food deserts¹. There is a need that farmer's markets can fill to provide healthy fresh food to low-income communities across the state of New York. Are they filling that need? Are they serving the regions and counties of NYS that are the most in need?

Datasets

Three datasets were merged for analysis. The Farmer's Market in New York State dataset was collected by the NYS Department of Agriculture and Markets provides a listing of farmer's markets operating year-round. It lists 579 farmer's markets in 61 out of the 62 counties. The data is supplied by part-time market managers who are often volunteers and often have inconsistencies that have to be fixed when discovered.

The second dataset is the 2020 Small Area Income and Poverty Estimates (SAIPE) from the United States Census Bureau. Collected by combining survey data, population estimates and administrative records for each state and county and is used to provide estimates of poverty and income to federal programs to assist in the allocating government funds. It uses tax return data to estimate the median household income and the poverty universe, people in and not in poverty. It is not the same as the population estimates and is often larger than population estimates even though it excludes groups like foster children, and folks who live in college dormitories or military housing. Poverty status is determined by adding the incomes of all of the family members of a household factoring in size to determine if the total is over the poverty threshold which is represented in the dataset as Number in Poverty.² The third dataset is the 2019 Labor Market Regions from the New York State Department of Labor. It lists the labor market regions and corresponding counties of the state.

Others' Findings

There were several findings related to the US Farmer's Market dataset from the USDA Agricultural Marketing Service but little on the NYS Farmers Market specifically³. The US Farmer's Market dataset is more comprehensive as it includes economic data (ex: population and median household income). The studies I found were varied and mostly focused on the question of the accessibility of farmer's markets for low-economic Americans. Some found farmer's markets to be more frequent in more densely populated and higher-income areas.

Coding Process

The Python libraries Pandas and Matplotlib were imported into Jupiter Notebook and used to import and read the datasets' CSV files. A review of each dataset illustrated the shapes, column

¹ <https://www.siena.edu/files/resources/ace-cpi-library-food-access-and-security.pdf>

² <https://www.census.gov/topics/income-poverty/poverty/guidance/poverty-measures.html>

³ <https://www.kaggle.com/datasets/madeleineferguson/farmers-markets-in-the-united-states/code>

names, and data types. They all contain a column with county names used to join the datasets together. They were all individually cleaned and minor changes were made to each as well to the merged dataset. Outliers were detected but not removed from the datasets.

A review of the datasets showed no duplicate data in any of the datasets and null values in farmer's market and merged datasets. In the column, 'Market Link' from the farmer's market dataset 146 null values were dropped. The other null values were filled with the column mode and the value zero. Additional columns and header row summary data that were not relevant for analysis were removed from each dataset. Columns were renamed in the county and merged datasets to make it more clear for displaying the data. Several column types were changed to integers and floats for easier calculations and charting. The county columns were reformatted in the farmer's market and regions dataset to include additional text to match with the format in the county dataset to allow for all three datasets to be merged on the county column. Two new columns were created for the merged dataset that calculated the percent of the total population in poverty for NYS and the percent of the total farmer's markets for NYS for each row.

There were several outliers in the final dataset for the columns, "farmer's Market Count", "Median Household Income" and "Number in Poverty". The outliers were not removed from the dataset because they present relevant data that would be lost without them. For example, removing the outliers for "Number in Poverty" and "farmer's Market Count" would remove 3 of 5 and 4 of 5 counties in the New York City region, respectively. The five counties in the NYC region account for 55% of the total NYS population in poverty. The outlier data is critical and even minor removals could impact the final results.

Visualizations

The data visualization and analysis consisted of two major sections: the percentage of the population in poverty and median household income. The visualization for the population in poverty was primarily pie charts and a bar chart while the median household income visualizations consisted of a scatter plot and boxplots. Histograms were selected to illustrate the dataset outliers as they would be easier to spot them.

Key Takeaways

The analysis successfully showed farmer's markets are serving communities with higher median household income and those with a lesser percentage of the population are in poverty.

The population in poverty section showed the counties in downstate NYS accounting for 70.5% of the state's total population in poverty yet it only accounted for 34% of the farmer's markets in the state. The counties and regions with the highest number of people in poverty did not similarly have the highest number of farmer's markets. For example, the NYC region accounts for 55% of the state's total population in poverty but only 19% of farmer's markets. Additionally, the counties with the lowest median household incomes consistently had fewer farmer's markets on average by about 10 than the counties with the highest median household income.

I would like to do further analysis of the county data to see if there are any possible connections between population size, food deserts, and farmer's market count.