Abrar, Sehr

Mentor Me Collective

Advanced Data Analytics

November 20, 2025

NYC Food Poverty Analysis

Food insecurity has always struck me as one of the most urgent social issues in urban environments. I've seen firsthand how some neighborhoods have easy access to nutritious food while others face persistent barriers. My motivation for this project stemmed from a desire to understand these patterns more deeply and explore how data can inform solutions. In researching this topic, I found that food insecurity in the U.S. has been increasing in recent years, with millions of households struggling to afford adequate nutrition. By combining publicly available datasets, I aimed to identify where these disparities are most pronounced and what factors might predict food insecurity at the neighborhood level.

Understanding the history of food assistance programs in the United States provided valuable context for this project. SNAP, formerly known as the Food Stamp Program, has evolved over decades in response to economic shifts, urbanization, and public health concerns. I explored legislative milestones, such as expansions during economic recessions and adjustments to eligibility criteria, to better interpret trends in program participation. This historical perspective helped me appreciate how policy decisions interact with demographic and economic factors to influence food insecurity patterns at the neighborhood level.

Living in New York City, I noticed stark differences between neighborhoods in terms of access to fresh and affordable food. Some areas have an abundance of grocery stores and farmers' markets, while others rely heavily on convenience stores and fast food. I wanted to explore how these disparities correlate with socioeconomic and demographic variables, and whether public data could reveal hidden patterns. This project allowed me to connect what I saw in daily life with measurable data trends, making the problem tangible and actionable.

To carry out this analysis, I relied on several datasets. The primary sources included NYC Open Data for SNAP (Supplemental Nutrition Assistance Program) participation, census tract-level demographic data from the U.S. Census Bureau, and geospatial boundaries of neighborhoods. I also referenced secondary sources such as USDA reports and research articles to understand underlying trends and validate the data. Gathering these datasets involved ensuring compatibility in terms of geographic resolution and temporal alignment, which required careful preprocessing and cross-referencing to maintain accuracy.

The data acquisition process itself posed several challenges. Different datasets used varying identifiers for neighborhoods, some census tracts were missing data, and certain demographic variables were inconsistently reported across years. I handled these issues by standardizing identifiers, imputing missing values where reasonable, and excluding a few outliers that could distort analysis. I used Python extensively for this work, with libraries such as pandas and GeoPandas for data manipulation, and matplotlib and seaborn for visual exploration. These tools allowed me to transform raw data into structured formats suitable for analysis, such as merging SNAP participation rates with median household income and education levels by tract.

Once the data was cleaned, I began exploratory data analysis (EDA) to uncover trends and patterns. I visualized the geographic distribution of food insecurity using choropleth maps, which revealed clusters in areas like the Bronx and parts of Brooklyn. I also plotted correlations between income, educational attainment, and SNAP participation, noticing that tracts with lower median income and lower high school graduation rates tended to have higher food assistance needs. Temporal patterns were also interesting; some neighborhoods saw increased participation in recent years, possibly reflecting broader economic pressures. These visualizations not only helped me understand the landscape of food insecurity but also guided decisions for feature selection in modeling.

Mapping the data geographically revealed not only high-risk neighborhoods but also patterns linked to urban planning. Areas with limited public transportation access often coincided with higher SNAP participation, suggesting mobility barriers contribute to food insecurity. I also noticed that food deserts often overlapped with historically underserved neighborhoods, raising questions about systemic inequities. Incorporating this spatial lens highlighted the importance of considering both physical infrastructure and socioeconomic factors when analyzing urban food access.

After EDA, I moved into predictive modeling to estimate food insecurity levels based on demographic and socioeconomic variables. I experimented with linear regression as a baseline and then applied tree-based models like Random Forests for more complex relationships. Feature selection was informed by both domain knowledge and statistical measures; variables like median household income, unemployment rate, and percentage of households with children emerged as the most predictive. I split the data into training and test sets, applied cross-validation, and tuned hyperparameters to improve model performance. Metrics such as R^2 and mean absolute error helped me evaluate the models, and I found that tree-based methods captured non-linear relationships more effectively than linear approaches.

Working with demographic and socioeconomic data raised questions about privacy and ethical responsibility. Although all datasets were publicly available, I made sure to anonymize any tract-level identifiers in my analysis and avoid conclusions that could stigmatize communities. I also reflected on how predictive modeling can influence real-world decisions: while identifying high-risk areas is valuable, it is critical that data-driven recommendations are paired with human judgment to avoid unintended consequences.

During the modeling process, I encountered several challenges. The data was sparse in certain neighborhoods, which limited the model's predictive power for those areas. Additionally, some important factors influencing food insecurity—such as informal community support networks or temporary food drives—were not captured in public datasets. I documented these limitations to ensure transparency and considered how future research could incorporate richer data sources. Despite these constraints, the models were able to highlight high-risk neighborhoods and suggest patterns that could inform policy decisions.

Analyzing feature importance from the models revealed some clear insights. Income consistently appeared as a dominant factor, followed by educational attainment and household composition. These results reinforced my expectations but also highlighted subtler patterns, such as the influence of population density and age distribution. Interpreting these findings made me think critically about how socioeconomic variables intersect to produce disparities in access to food and how interventions could be targeted more effectively.

To validate the data-driven insights, I looked for qualitative information from community organizations and local food banks. Reports from nonprofits in the Bronx and Queens corroborated many of the patterns I observed, such as neighborhoods with high SNAP participation also having frequent community food drives. Engaging with these perspectives, even indirectly through public reports, helped me contextualize the numbers and reinforced the practical relevance of my analysis.

Beyond technical outcomes, I also reflected on the broader implications of this project. My findings suggest that policymakers and nonprofits could prioritize neighborhoods with high predicted food insecurity for new programs or funding. Predictive tools could help allocate resources efficiently and measure the impact of interventions over time. Additionally, this project could serve as a model for other cities, demonstrating how publicly available data can support data-driven decision-making in social policy.

Completing this project taught me a great deal about both technical and real-world aspects of data analysis. I deepened my skills in data cleaning, visualization, and modeling, while also gaining a richer understanding of urban social issues. Working with geospatial data and demographic datasets allowed me to connect abstract numbers to tangible communities, which reinforced my motivation to apply data science for social good. I also learned to navigate challenges in messy real-world data, from missing values to inconsistent identifiers, which will be invaluable in future projects.

This project pushed me to combine technical problem-solving with social awareness. I developed patience and creativity in handling messy datasets and learned to approach challenges from multiple perspectives: statistical, geographic, and ethical. Each stage reinforced the importance of curiosity, persistence, and reflection in meaningful data science work. I now feel better prepared to tackle complex, real-world problems where solutions require more than coding skills, they require thoughtful integration of data, context, and human impact.

In conclusion, this project gave me the opportunity to explore food insecurity in New York City through a data-driven lens. By combining public datasets, systematic analysis, and predictive modeling, I was able to identify patterns, highlight high-risk neighborhoods, and suggest avenues for targeted interventions. The experience reinforced my belief in the power of data to uncover insights, guide policy, and ultimately make a positive social impact. Moving forward, I hope to continue applying these skills to projects that address pressing societal challenges and to expand this work with richer data sources and more advanced modeling techniques.