



# Case Study: Predicting Food Assistance in San Diego

DATA SCIENCE ALLIANCE

May 2024

In collaboration with:





# Table of Contents

Executive Summary

Acknowledgements

<b>1</b>	<b>Introduction: Partnering for a Cause</b>	<b>6</b>
<b>2</b>	<b>Project Scope: From Data to Action</b>	<b>7</b>
<b>3</b>	<b>Unlocking Insights through Data Science</b>	<b>11</b>
<b>4</b>	<b>Responsible Data Science in Practice</b>	<b>13</b>
<b>5</b>	<b>Impact and Future Directions</b>	<b>15</b>

---

# Executive Summary

The Data Science Alliance (DSA), in collaboration with Feeding San Diego and the Jacobs & Cushman San Diego Food Bank, has pioneered an initiative to transform the landscape of food assistance in San Diego County. Our collective mission is to address food insecurity using data-driven strategies rooted in DSA's principles of Responsible Data Science: Fairness, Transparency, Privacy, and Veracity.



## The Challenge: Confronting Hunger with Data

Our challenge was clear—help identify areas of need and subsequently optimize food distributions across San Diego County. By blending predictive analytics with practical insights, we aimed to pinpoint areas in need to guide resource allocation, all while adhering to responsible data practices.

## Our Work

Our approach was multifaceted. We developed a sophisticated predictive model and an interactive dashboard to visualize critical data. This powerful combination allows the food banks to quickly assess and respond to the needs of the community, identifying "hunger hotspots" and ensuring that food distribution is efficient.

## Responsible Data Science in Action

Our commitment to responsible data practices has been unwavering. From ensuring fairness and accuracy in our models to maintaining transparency and safeguarding privacy, we are building an ecosystem of positive change through data science community projects.

## Impact and Looking Forward

The impact is tangible. By improving targeting of food assistance, enhancing the understanding of food insecurity dynamics and strengthening partnerships, we have been able to redistribute resources effectively increasing output to those in need. As we look to the future, our focus is on expanding our impact. Upgrading our dashboard, integrating more data sources, refining our models, and potentially expanding to other regions are just some of the ways we're planning to build on our success.



## Join Us in Making a Difference

This project is more than just numbers and data points; it's about real impact on real lives. By supporting our initiative, you are a part of our solution-driven movement in using innovation and data for good. Together, we can continue to make strides in eradicating food insecurity, not only in San Diego County, but potentially in communities worldwide.



# Acknowledgments

We extend our heartfelt gratitude to the Responsible Data Working Group and all the volunteers from the Data Science Alliance who dedicated their time and expertise to this project. Their commitment to upholding the principles of Responsible Data Science—Fairness, Transparency, Privacy, and Veracity—was instrumental in guiding our approaches and methodologies. Special thanks to Maria Lupetini, whose invaluable insights and collaborative spirit were pivotal in optimizing our data-driven strategies. The success of this project reflects all of our collective efforts and deep dedication to making a significant impact in the community through responsible and informed data practices.

# 1 Introduction: Partnering for a Cause



In collaboration with Feeding San Diego and the Jacobs & Cushman San Diego Food Bank, the Data Science Alliance (DSA) embarked on a mission to improve the food assistance landscape in San Diego County. Our goal: to enhance food distribution using data-driven strategies, underpinned by our principles of Responsible Data Science—Fairness, Transparency, Privacy, and Veracity.

Through this partnership, we developed a model to predict food assistance in the region and created a dashboard to provide a visual representation of the food banks' distribution data to monitor “hunger hotspots” and other metrics, allowing the food banks to quickly assess the current status of particular areas to better serve the population in need.

**Confronting Hunger with Data:** By blending predictive analytics with practical insights, we aimed to spotlight areas in need and guide resource allocation and reallocation, all while adhering to responsible data practices.

## 2 Project Scope: From Data to Action

At its heart, this project was about more than just numbers. It involved predictive modeling, identifying areas of need, creating an interactive dashboard, and fostering collaboration. We delved into data—from food distribution patterns to socioeconomic factors—crafting a tool that empowers decision-making.

**79.75M**

Total pounds of food distributed by the food banks (2023)

**\$933.25M**

Total amount of CalFresh assistance (2023)

**561,304**

Population in poverty (below 150% federal poverty level)\*

*\*latest data from 2022, Census Bureau's American Community Survey*

### Development of a Predictive Model:

The foundation of this initiative is the predictive model tailored to forecast food assistance across San Diego County. This model uses historical food bank distribution data as well as economic factors such as unemployment and inflation rate to provide forecasts for various timeframes and geographic areas. This feature is vital for identifying areas of increasing need.

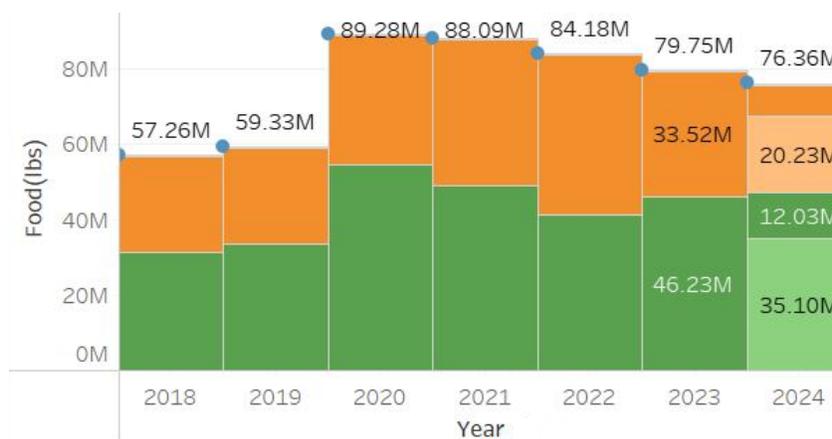


Figure 1: Historical and forecasted annual food distribution of both food banks in San Diego County. The blue dot represents both food banks combined, while the green bar represents Jacobs & Cushman San Diego Food Bank and the orange bar represents Feeding San Diego. The dark color marks the historical data and the light color displays the forecast.

# Definition of Hunger Hotspots:

Hunger hotspots within the county are defined through the establishment of specific criteria and indicators. These hot spots are clearly delineated and regularly updated on the dashboard, ensuring focused monitoring and response to these critical areas.



# Creation of a Dynamic Dashboard:

The interactive, user-friendly dashboard we've created for this project features three tabs:

## 1. Food Bank Distribution

The first tab showcases the combined efforts of Feeding San Diego and the Jacobs & Cushman San Diego Food Bank. It displays the historical volume of food distributed and the forecasted need across different regions in different timeframes, as well as the average number of meals distributed to a person living in poverty in a zip code. Tracking distribution changes over time and planning future resource allocation come from this tab.

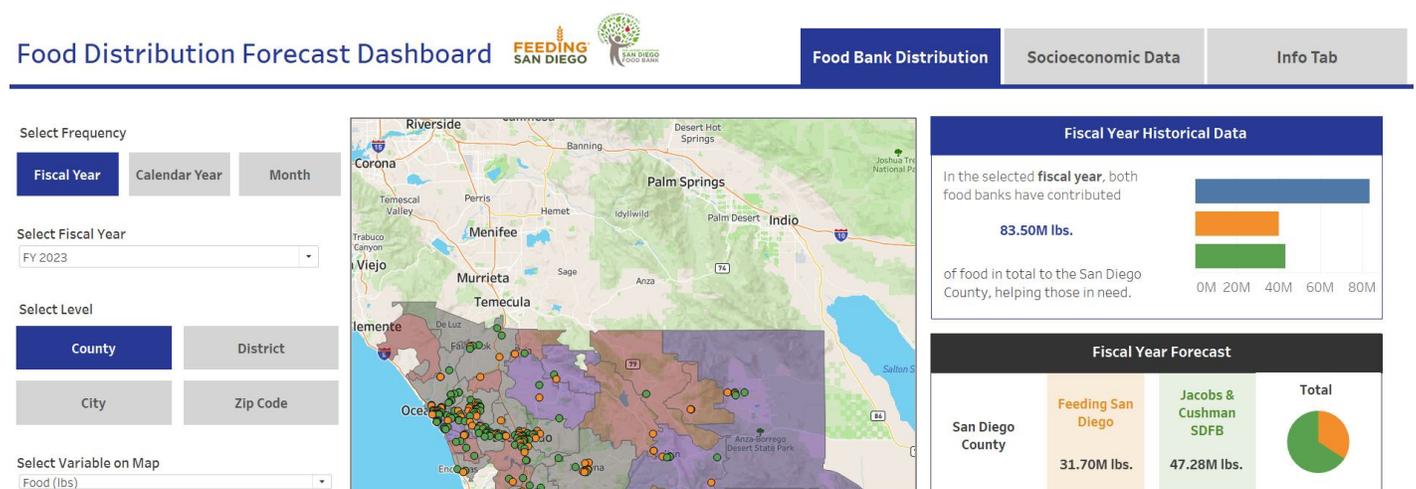


Figure 2: A partial representation of the Food Bank Distribution tab of the dashboard.

## 2. Socioeconomic Data

The second tab offers insights into the correlation between socioeconomic factors and food assistance needs, aiding in the development of targeted programs and policy decisions.

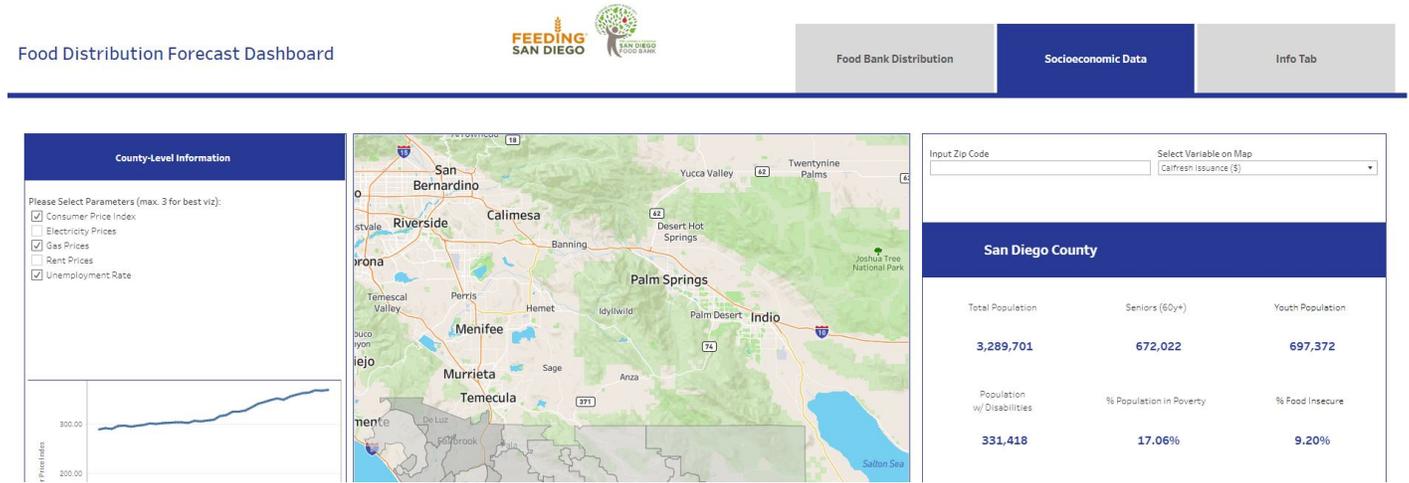


Figure 3: A partial representation of the Socioeconomic Data tab of the dashboard.

## 3. Info Tab

The third tab provides essential transparency about data sources, methodologies, and assumptions, serving as a critical resource for data validation and educational purposes.

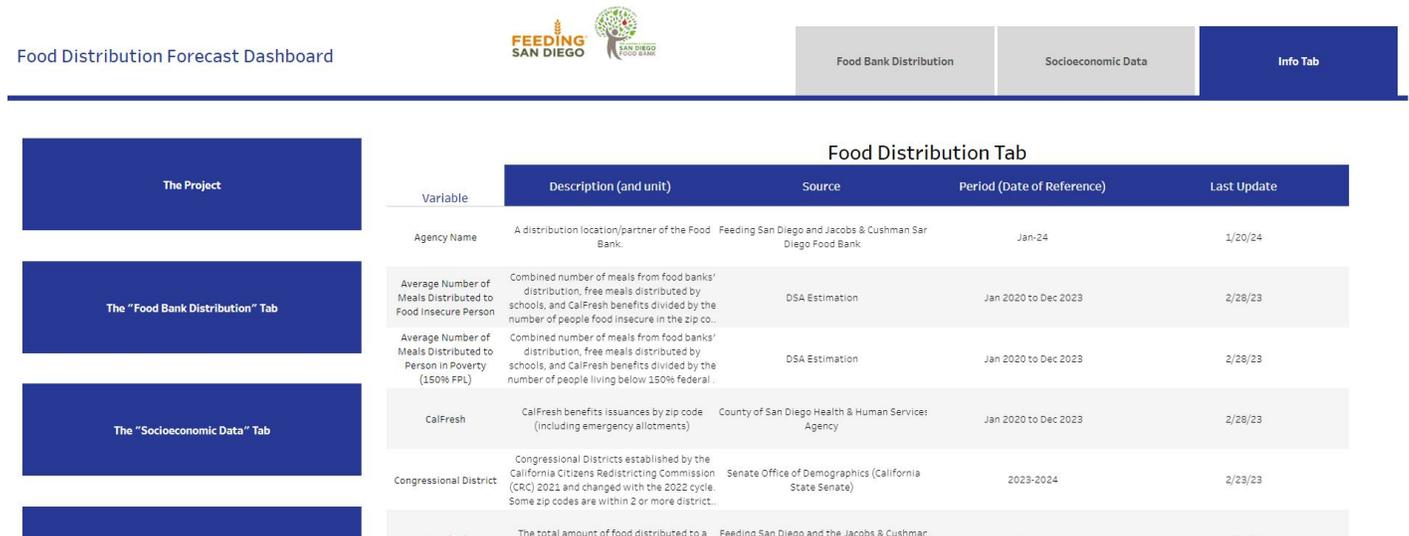


Figure 4: A partial representation of the Info tab of the dashboard.

Together, these tabs facilitate strategic planning, impact assessment, and an in-depth understanding of the dynamics of food insecurity.

## Technical Details:

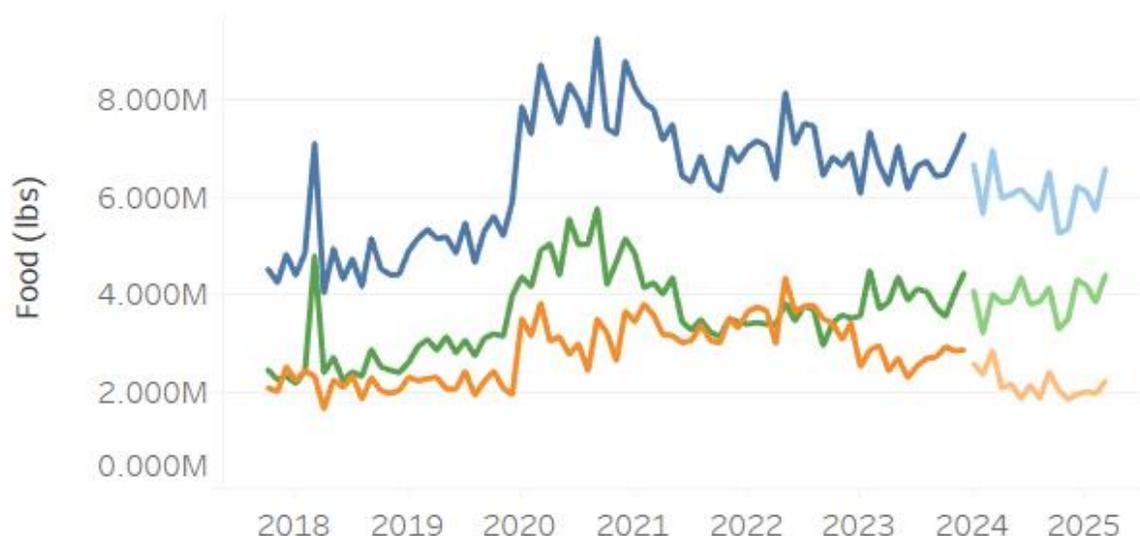
All data collection, cleaning, and processing tasks are executed using Python. The models are developed using both Python and R, leveraging the strengths of each programming language for optimal results. All files and data are securely stored in a private GitHub repository and in DSA's private Google Drive account, ensuring organized collaboration and data integrity. The dashboard was developed in Tableau and hosted on DSA's website, protected by password access, ensuring data security and controlled accessibility.

## Stakeholder Engagement:

This project thrives on collaboration with Feeding San Diego and the Jacobs & Cushman San Diego Food Bank. We are committed to providing comprehensive support to these organizations, enabling them to effectively use the dashboard and interpret the model outputs.

## Monthly Dashboard Update, Continuous Monitoring and Improvement:

The dashboard is updated and the model is retrained every month with the most recent data available. To acknowledge the dynamic nature of data and community needs, we are dedicated to the ongoing monitoring and refinement of the models and dashboard, ensuring they remain relevant and effective in addressing the evolving challenges of food insecurity.



*Figure 5: Historical and forecasted monthly food distribution of both food banks in San Diego County. The blue line represents both food banks combined, while the green line represents Jacobs & Cushman San Diego Food Bank and the orange line represents Feeding San Diego. The dark color marks the historical data and the light color displays the forecast.*

# 3 Unlocking Insights through Data Science



In order to help both food banks tackle food insecurity, we harnessed the power of the Prophet algorithm for predictive forecasting. In addition, a key focus of our strategy was the identification of hunger hotspots, aiding our partners to respond effectively and precisely where needed most.

## Forecasting Model:

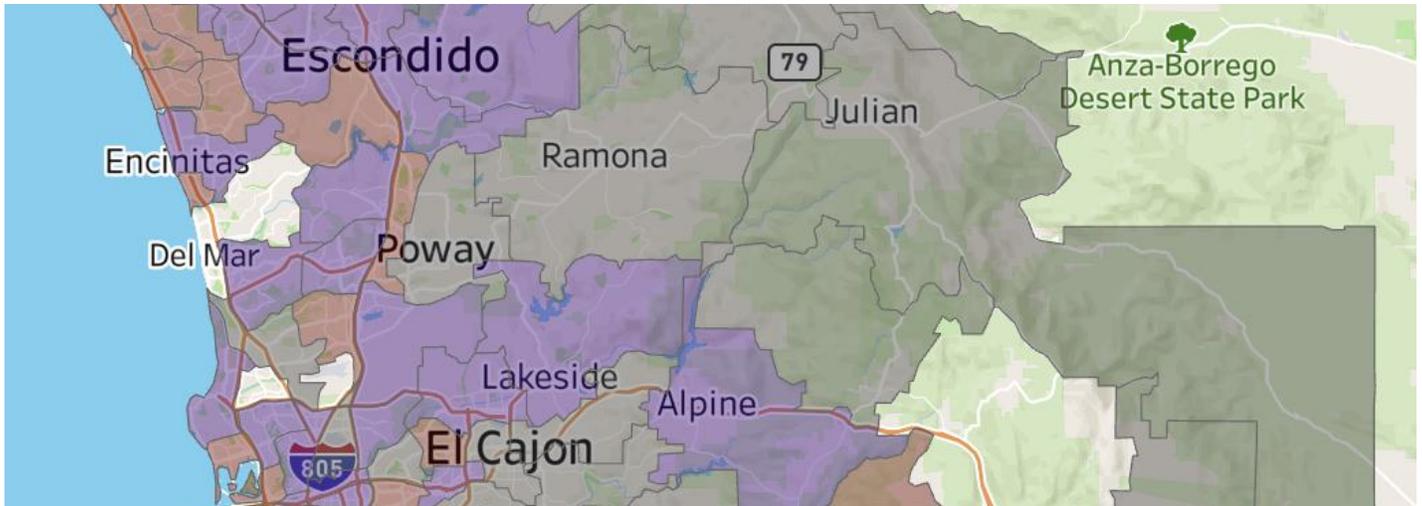
We utilize the Prophet algorithm, a brainchild of Facebook's Core Data Science team, known for its proficiency in handling intricate data trends, including nonlinear patterns and seasonal fluctuations. Our approach used time series data beginning in January 2018 to generate detailed forecasts at various geographic granularities, from zip codes to the county level. We enriched our model with external variables like unemployment rates, inflation, and notable events including the Food Banks' Together Tour and the Farmers to Families Program during the Covid pandemic.



Figure 6: A table from the dashboard displaying the food distribution forecast for both food banks.

## Hunger Hotspots Definition:

Hunger hotspots were defined using the average number of meals distributed to a person in poverty. To quantify this number, we aggregated various measures of food assistance (from the food banks, CalFresh, and school programs) and translated the result into units of meals. This conversion took into account the unique economic factors of San Diego County, such as the higher cost of living and inflation rates. This thorough and careful process enabled a comprehensive understanding of the county's food assistance needs and the ability to pinpoint areas in urgent need of intervention.



*Figure 7: The “hotspot” areas colored ■ in the map show the zip code locations where a person experiencing poverty is receiving, on average, less meals than people in other locations.*

The tools and methods employed underscore our commitment to using data science for strategic planning and informed decision-making in addressing food insecurity. The integration of sophisticated modeling techniques, along with a keen understanding of local economic conditions, ensures that our efforts are both impactful and targeted towards the most critical areas of need in San Diego County.

# 4 Responsible Data Science in Practice

Our project's backbone is its commitment to Responsible Data Science. From understanding food bank operations to ensuring data security and privacy, our principles were woven into every aspect of our work. The following outlines examples of how each principle of Responsible Data Science—Fairness, Transparency, Privacy, and Veracity—was integrated into our project:

## Fairness



- **Collaboration with experts and stakeholders:** Working alongside data science luminaries and regularly meeting with executives from both food banks helped us refine our models to prevent biases and promote fairness in predicting food distribution and identifying hunger hot spots.
- **Understanding operations:** By volunteering at the food banks, we gained firsthand insight into their operations, giving us confidence that our analyses fairly represent the realities of food distribution and need.

## Transparency



- **Datasheet and documentation:** The datasheet and thorough documentation provide transparency to all stakeholders.
- **Info tab development:** Our commitment to transparency is the direct motivation for the info tab on our dashboard. The tab illuminates our process by providing clear information on all data sources, methodologies, and assumptions behind our model.
- **Regular Updates:** We keep stakeholders updated through regular reports and well-documented methodologies to maintain transparency throughout the entire project lifecycle.

## Privacy



- **Securing collected data:** We have taken rigorous measures to secure the data collected. Data is stored in secured environments and access is strictly controlled to protect the privacy of individuals and organizations involved.
- **Ethical data handling:** All data-handling procedures—from collection to processing—have been conducted with the utmost respect for privacy concerns. If any stakeholders or participants withdraw participation in the future, we are able to remove their data from consideration quickly.

## Veracity



- **Collaboration with stakeholders:** Regular check-ins with both collaborating food banks ensured the accuracy and relevance of our data. This continuous engagement allowed us to better understand the use of the dashboard and deliver improvements, guaranteeing the veracity of our work.
- **Collaboration with experts:** Experienced data science professionals helped us with building and testing our models, certifying that our work was accurate.
- **Datasheet:** We maintained a comprehensive datasheet to track the data used in the project, ensuring the veracity and credibility of our data sources.
- **Documentation:** All the steps of the project development process were documented, guaranteeing the reproducibility of the project. The documentation includes the data collected, the methodologies of the models, and the technology used to build the dashboard.

By embedding these principles into the core of our project, we have positioned this project to not only achieve our objective of combating food insecurity but to also set a standard for responsible data science practices in community-oriented projects.

# 5 Impact and Future Directions

Our project has made significant strides in addressing food insecurity in San Diego County. By aggregating the data from both food banks and applying data science techniques to predict food distribution and identify hunger hotspots, we have enabled the local food banks to optimize their operations and more efficiently reach those in greatest need. The insights gained from our models have informed strategic decision-making, leading to more targeted and effective food distribution efforts. Furthermore, our commitment to Responsible Data Science has fostered trust and collaboration among stakeholders, enhancing the overall impact of the project.



## Key impacts include:

- Improved targeting of food assistance to areas identified as hunger hotspots
- Identification of distribution overlaps, leading to more efficient service provision and reduced redundancy in food assistance
- Enhanced understanding of the dynamics of food insecurity within the county
- Strengthened partnership between both food banks in San Diego County



## Looking ahead:

there are several avenues for expanding and enhancing the project:

- **Dashboard upgrade:** Create an executive summary tab displaying key metrics and relevant analyses.
- **Integration of additional data sources:** Incorporate new data sources such as CALPADS to improve our understanding of the need for food assistance at a more granular level.
- **Forecasting hunger hotspots:** Develop a model to forecast the demand for food assistance at the zip code level to help food banks proactively address needs.
- **Neighborhood clusters:** Incorporate new clusters of food bank activity by combining neighborhoods with similar levels of poverty.
- **Enhancing model accuracy:** Continuous refinement of our models to increase accuracy and more precise predictions and interventions.
- **Expansion to other regions:** Building on the success in San Diego County, there is potential to apply our models and methodologies to other regions, adapting to local contexts and data availability.

In summary, this project has not only provided immediate benefits in terms of enabling improved food distribution strategies but has also laid a foundation for future initiatives aimed at eradicating food insecurity. By continuing to leverage data science responsibly and innovatively, we can broaden our impact and contribute to lasting change in communities within and beyond San Diego County.

