Predicting Food Bank Demand: A Socioeconomic Analysis and Forecasting Model Investigation

By Kaitlyn Rakestraw
Advisor: Inma Borrella, PhD

Topic Areas: Demand Planning, Forecasting

Summary: This research evaluated the demand for food bank services from the Mid-Ohio Food Collective (MOFC), highlighting how demand varied based on the socioeconomic vulnerability of counties. A framework was created to cluster counties as "more vulnerable" or "less vulnerable", and forecasting models were tested on demand data from each cluster. It was discovered that a naive forecasting model effectively predicted demand for the more vulnerable regions, but an exponential smoothing model with level and trend components significantly improved forecast accuracy for the less vulnerable regions. The insights from this study are applicable to all United States counties, providing a valuable tool for food banks across the country to enhance their demand forecast accuracy.

Prior to attending MIT, Kaitlyn Rakestraw worked in inventory planning within the retail industry for 3 years. She received her Bachelor of Science in Supply Chain and Information Systems from Pennsylvania State University.

KEY INSIGHTS

1. In less vulnerable counties, switching to an exponential smoothing model with level and trend components will lead to significant improvements in MAPE.

2. For more vulnerable counties, a naïve forecasting model can be used with reliable accuracy.

3. Utilizing United States county-level socioeconomic data to cluster the service area of a food bank can lead to valuable insights about demand patterns and forecasting.

Introduction

In 2021, over 53 million people in the United States received assistance from charitable food organizations. Food banks, large warehouses that support smaller community-serving agencies, face uncertainty in estimating demand over time. This research aimed to understand how socioeconomic status affects demand for food bank services, identify effective forecasting models for this demand, and analyze the implications of improved demand forecasting for food bank operations. Using data from the Mid-Ohio Food Collective (MOFC) and relevant socioeconomic information, the study tested various forecasting models and proposed recommendations for each cluster to improve forecasting accuracy. The hypothesis was that by segmenting service areas based on socioeconomic factors and applying a demand forecasting model, food banks could more accurately predict community needs and improve operations.

Data and Methodology

This research analyzed demand for the MOFC and proposed a forecasting method based on the performance of various models. Data was gathered from MOFC and public sources to classify counties by socioeconomic vulnerability, and a clustering analysis was conducted using a k-means approach. The study tested nine forecasting models on the MOFC demand data and evaluated them using three error metrics – RMSE, MAE, and MAPE. A vulnerability assessment was conducted for every county in the United States, and the findings were applied to the MOFC's service area. This approach was used to test the hypothesis that food bank demand differs based on the vulnerability of the area, and to determine if the forecasting models performed differently based on these vulnerability levels.

County-Level Vulnerability Assessment

A k-means clustering analysis was used to evaluate vulnerability across US counties based on socioeconomic factors: median household income, poverty percentage, food insecurity rate, and two education-based factors. The optimal number of clusters was found to be two, with a moderate silhouette score of .38, indicating distinctiveness between clusters.

The clusters were classified as Less Vulnerable (Cluster 0) and More Vulnerable (Cluster 1). The less vulnerable cluster had a lower poverty percentage, food insecurity rate, and a lower percentage of adults with less than a high school diploma, along with a higher median household income and a higher percentage of adults with a bachelor's degree or higher. The more vulnerable cluster had the opposite characteristics.

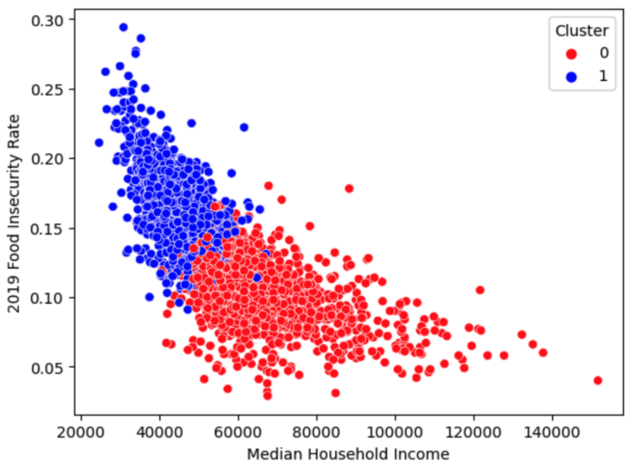

Figure 1: Visualization of Clustering Analysis

The same clustering was applied to the service area of the MOFC. Out of the 20 counties served by MOFC, nine were classified as less vulnerable and eleven as more vulnerable, based on the same socioeconomic metrics. This information highlighted disparities among counties and between the clusters, enabling an understanding of variations in these metrics across the service area.

Demand Characterization for MOFC Services

It was hypothesized that the vulnerability differences identified in the clustering analysis would impact the MOFC's demand. The demand data from all 20 counties serviced by MOFC was analyzed, revealing that demand peaked in March 2023, with most high-demand months falling in 2022 or 2023. There also appeared to be a seasonality in the data, with lowest demand consistently occurring in winter and peaking towards the end of the year.

The average monthly services from January 2019 to March 2023 were 93,127, with a standard deviation of 16,393, implying a low to moderate level of variability. However, demand consistently exceeded one standard deviation from the mean from August 2022 to March 2023.

The analysis also highlighted significant variations in demand at the county level. The county-level data was used to examine the trend of MOFC services over time, with a linear trendline moderately fitting the data with an R-squared value of 0.41. When data from January 2021 to March 2023 was isolated to analyze data after the initial effects of the COVID-19 pandemic, the trendline fit the data significantly better with an R-squared value of 0.87.

When the data was grouped by cluster, the coefficient of variation for Cluster 0, the less vulnerable counties, was 0.19 and for more vulnerable Cluster 1, it was 0.14. The trend analysis for each cluster showed similar patterns to the overall data – the trendline fits improved significantly when the data was isolated to only January 2021 through March 2023.

Time Series Forecasting Results

Time series models were used to analyze the demand data for the MOFC service area from January 2019 to March 2023. The data was split, with 80% used for training and 20% for testing. The exponential smoothing with level and trend components and SARIMA models performed well for the full service area based on MAPE, MAE, and RMSE values.

These models were also applied to data for Clusters 0 and 1. For Cluster 0 (less vulnerable counties), the results were similar to those for the full MOFC Service Area, with SARIMA and the exponential smoothing with level and trend models performing best.

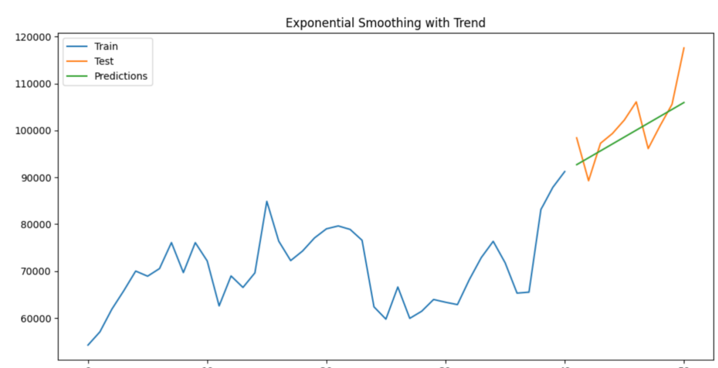

Figure 2: Cluster 0 Forecasting Results

For Cluster 1 (more vulnerable counties), the SARIMA model performed the best. The exponential smoothing with level and trend model also performed well for this cluster, but the parameter weights were optimized to an alpha of 0.995 and beta of 0.0001, effectively making this a simple exponential smoothing model.

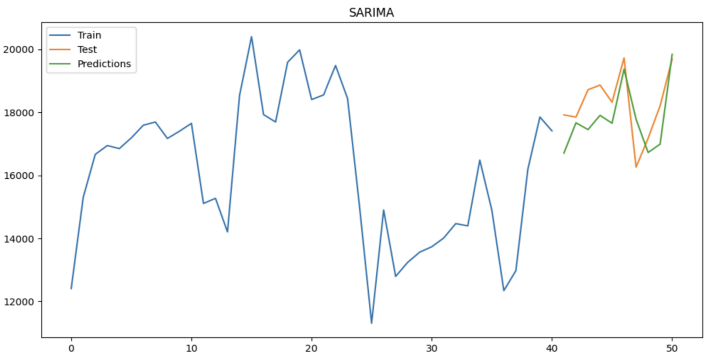

Figure 3: Cluster 1 Forecasting Results

The two most promising forecasting models for this data were the exponential smoothing with level and trend components and the SARIMA model. Interestingly, the naive forecast performed significantly better for Cluster 1 than for Cluster 0, with about a 4-percentage point difference in MAPE.

Results

The vulnerability assessment and demand analysis for the MOFC revealed seasonal fluctuations and trends in the data after 2020. The best forecasting models were the exponential smoothing with level and trend components for Cluster 0 (Less Vulnerable) and SARIMA for Cluster 1 (More Vulnerable). The naive forecasting model performed better for Cluster 1 than Cluster 0.

The MOFC service area's best forecasting models were the exponential smoothing with level and trend components and SARIMA models. The exponential smoothing with level and trend model yielded a MAPE of 4.2% for the MOFC service area, compared to the naïve model which produced a MAPE of 9.0%.

Cluster 0 had similar results to the full service area, with the exponential smoothing with level and trend performing the best. The SARIMA model performed best for Cluster 1. The baseline naive forecasting model performs better for Cluster 1 with only a 6.1% MAPE versus a 9.9% MAPE for Cluster 0. This takeaway suggests that the adoption of a new forecasting technique will have a larger impact on the accuracy of forecasts in Cluster 0.

Given these findings, it is recommended that MOFC uses the naive model for Cluster 1 due to its simplicity and performance. For Cluster 0, a more sophisticated model, the exponential smoothing with level and trend components, is recommended due to the improvement in forecast accuracy and ease of implementation.

Improved demand forecasting can significantly impact MOFC's operations, helping with financial planning, volunteer scheduling, food sourcing, partner selection, staff allocation, budgeting decisions, and communications. This will result in better resource allocation, increased support from donors, and improved service delivery, benefiting the communities served by food banks.

Conclusion

Food insecurity is a significant issue in the United States, with food banks like the Mid-Ohio Food Collective (MOFC) providing crucial support to their communities. However, demand for these services is variable, necessitating accurate forecasting. The segmentation produced by a k-means clustering analysis was used to assess the performance of various time series forecasting models.

The results suggest that applying demand forecasting models based on socioeconomic vulnerability will enhance forecast accuracy. For more vulnerable areas, a naïve forecasting model was effective, but for less vulnerable areas and aggregated forecasts, the exponential smoothing with level and trend model was recommended due to its higher accuracy. Improved forecasts would aid MOFC in decision-making and resource allocation, improving their ability to effectively serve their community. This research supports the hypothesis that socioeconomic segmentation and demand forecasting can improve the operations of food banks like the MOFC.