World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

Segmenting U.S. households by behavioral patterns to predict food waste: A data-driven approach using public datasets

Adekunbi Bello [1, *], Anita Ogheneochuko Odiete [2] and Friday Anwansedo [3]

[1] *St Mary University, Department of Computer Science, San Antonio, TX. United State.*

[2] *Lancaster University, Department of Information Technology Management, Lancaster, United Kingdom.*

[3] *International Trade Council, SPE, Abuja, Nigeria.*

World Journal of Advanced Research and Reviews, 2025, 26(02), 4328–4347

Publication history: Received on 16 April 2025; revised on 25 May 2025; accepted on 28 May 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.2.2098

Abstract

Household food waste constitutes a significant sustainability and food security challenge in the United States, with substantial environmental and social implications. This study integrates publicly available datasets – including the USDA's Food Acquisition and Purchase Survey (FoodAPS 2012–2013), the 2018 American Community Survey, and ReFED's 2018 regional food waste estimates – with machine learning techniques to model and predict household-level food waste. We applied regression, classification, and clustering approaches to analyze waste behaviors. Among the predictive models tested, a random forest regression provided the most accurate predictions of household food waste, outperforming other methods. Classification models were used to predict Supplemental Nutrition Assistance Program (SNAP) participation and to assign households to waste-behavior clusters identified by unsupervised learning. Demographic factors, particularly household size and poverty ratio were among the strongest predictors of household food waste, while behavioral indicators such as grocery list frequency and food sufficiency played a secondary role. Clustering revealed distinct household profiles with varying waste patterns, differentiating, for example, food-insecure SNAP-dependent families from larger resource-stable households. To ensure alignment with real-world waste quantities, model outputs were calibrated against ReFED's regional waste data. The findings demonstrate the value of integrating diverse public datasets with machine learning to uncover drivers of household food waste. These insights enable more targeted household-level waste reduction interventions and support the development of effective, data-driven policies for food waste mitigation.

Keywords: Food Waste; U.S Households; Behavioural patterns; Data-driven Approach; Sustainable Consumption

1. Introduction

Food waste has emerged as a crucial global sustainability challenge, with an emerging surprising quantity of edible food being discarded, even as millions go hungry. In 2022 alone, an estimated 1.05 billion tons of food was wasted worldwide, while 783 million people suffered from hunger and a third of the global population experienced food insecurity (Change, 2024). Roughly one-third of all food produced is never consumed (Change, 2024)leading to significant environmental and social consequences throughout the food system. Food waste is responsible for about 8–10% of global greenhouse gas emissions (Change, 2024), contributing significantly to climate change. It also represents a massive misuse of resources, draining almost one-third of the world's agricultural land for production that ultimately ends up as waste (Change, 2024). Beyond environmental damage, this lost food could have been a benefit to those in need; the current level of waste is "not only a missed opportunity to feed those in need but also a significant environmental burden" (Change, 2024). The economic costs are astounding as well, food waste exacts an annual toll of around $1 trillion globally (Change, 2024), underscoring the urgency of this issue for sustainability, climate, and food security.

* Corresponding author: Adekunbi Bello

Household food waste is a major driver of this problem. Approximately 60% of global food waste occurs at the household level (Victoria Norton, 2024), making consumers' kitchen a critical intervention point. In the United States, which is one of the largest food-producing and food-consuming countries, 27% of all food (about 63 million tons) is wasted each year (ReFed, 2025). This equates to an estimate of 120 billion meals worth of food being thrown out annually in the U.S.A (ReFed, 2025). Notably, American households are the single biggest source, discarding roughly 43 million tons of food in 2022, nearly half of the nation's total wasted food (Lauria, 2024). The social implications of such waste are stark. Even as trash cans and landfills fill with edible food, tens of millions of Americans struggle with hunger. In 2022, about 44 million people in the U.S. lived in food-insecure households (America, 2023). Food waste occurring alongside widespread food insecurity is both unethical and also highlights potential structural weaknesses in the food system. Reducing household waste would not only improve environmental outcomes but could also help redirect food to those in need of it, thereby strengthening community food security.

A growing body of research informs that household food waste is driven by a complex mix of behavioral and demographic factors, which vary widely from one home to another. Recent analyses show that the average U.S. household wastes about 32% of the food it acquires (Gill, 2020). Even the most frugal households are estimated to throw away approximately 9% of their food (Gill, 2020), while others waste half of all food brought into the home (Gill, 2020). Such differences validates that there is no "typical" food-wasting household, instead, waste behavior varies remarkably across income levels, family sizes, and lifestyles. For example, higher-income families tend to generate significantly more waste than lower-income families (Gill, 2020), possibly because greater disposable income can lead to overbuying or less careful food management (Lauria, 2024). Diet patterns also matter households consuming more fresh and perishable foods (often for health-conscious diets) report higher waste, whereas larger households and food-insecure families tend to waste less per capita (Gill, 2020). These findings indicate that one-size-fits-all solutions are unlikely to succeed, as the drivers of waste differ across household profiles. Instead, experts emphasize the need to investigate different factors specific to a household, that could influence food waste, and tailor interventions accordingly (Gill, 2020). In other words, understanding which types of household waste food, and why, is pivotal for designing effective prevention strategies.

Despite growing efforts to reduce household food waste, current policies often overlook the heterogeneity of household behaviors and various external factors that influence household food waste behaviors, limiting their effectiveness for diverse communities (Leib, 2022). Governments and organizations worldwide have set ambitious targets, for instance, the United Nations Sustainable Development Goal 12.3 and U.S. national initiatives call for halving food loss and waste by 2030 (Change, 2024). To date, much of the effort to curb waste at the consumer level relies on broad campaigns and strategic guidance (e.g. public service announcements, awareness programs, "tips and tricks" for food storage) (Leib, 2022). While well-intentioned, such approaches often treat the public as a monolith and have shown limited effectiveness in changing behavior. Simply providing information or steering people to waste less is not enough to alter complex household habits (Maggie Bain, 2024). For example, a recent community-scale trial found that an informational campaign "was not effective in reducing household food waste", concluding that behavioral change requires more than general tips (Maggie Bain, 2024)Although progress has been made in raising awareness, opportunities remain to tailor interventions more precisely to the behavioral and structural differences among households. Without leveraging behavioral and socioeconomic data to identify which households or communities are most prone to waste and why, it is difficult to tailor solutions or allocate resources effectively. Prior evaluations suggest that broad messaging campaigns often yield limited results, particularly when they are not tailored to individual values or behaviors. Recent findings emphasize that messaging effectiveness varies widely across audiences, and that environmental, financial, or taste-based appeals should be targeted strategically rather than applied to everyone (Christian Bretter, 2023). In essence, current policies seldom differentiate a single-parent urban household with limited support from, say, a large rural family with ample storage space; yet these segments may face very different challenges in minimizing waste. This gap suggests that more data-driven, targeted strategies are needed. Indeed, international climate and sustainability experts have pointed out that tackling food waste requires "dedicated policies informed by data" and new technologies (Change, 2024). By bringing granular, evidence-based insights into who wastes food and under what circumstances, we can design interventions that are far more effective and equitable than one-size-fits-all messaging.

In response to these challenges, the present study aims to advance a household-level, data-informed approach to food waste reduction. We leverage several public datasets – including the USDA's Food Acquisition and Purchase Survey (FoodAPS, 2012–13), the American Community Survey (ACS, 2018), and food waste estimates from ReFED (2018) – to build machine learning models that predict and categorize household food waste behavior. FoodAPS provides detailed data on what foods households obtain and consume, ACS offers socio-demographic context, and ReFED's data serve as an external benchmark for waste quantities. Using these sources, we first apply clustering techniques to segment households into distinct behavioral groups based on their food-related practices and characteristics. Next, we develop regression models to estimate household-level food waste generation, using these behavioral and socioeconomic predictors. Finally, we use classification models to predict cluster membership and SNAP participation from observable household features. This combination of regression, classification, and clustering allows us to not only forecast waste levels but also to uncover patterns: such as clusters with persistently high waste levels (e.g., Underserved & Food Insecure) or low-waste profiles (e.g., Moderate Planners with Basic Reseources)that can inform targeted interventions. By integrating insights from food sustainability and data science, our study frames household food waste as a predictive and preventable outcome rather than an inevitable by-product. In doing so, we seek to demonstrate how advanced analytics on public data can identify significant information for reducing waste at the household scale, thereby contributing to both environmental sustainability and social equity. Ultimately, this research is designed to support smarter, evidence-based policies: by predicting household food waste and classifying waste behaviors, we provide a foundation for interventions tailored to the households that need them most. The following sections provide a review of relevant literature on household food waste and predictive modeling, followed by a description of the datasets used and the data preparation steps undertaken to support our analysis.

2. Literature Review

2.1. Behavioral and Socioeconomic Drivers of Household Food Waste

Households are widely recognized as a leading source of food waste; in developed countries most of the food waste occurs at the consumer stage (Antonia Di Florio, 2016). In the United States, for example, roughly half of all wasted food is generated from residential homes. Understanding why consumers discard food is therefore crucial. A broad array of behavioral habits, social and demographic factors have been linked to household food waste. Extant research has identified over 100 specific drivers of waste across typical in-home food routines; spanning meal planning, food shopping, storage practices, food preparation, and leftover management (National Academies of Sciences, 2020). These behaviors determine how efficiently households use the food they acquire and are influenced by both individual choices and habitual living constraints.

Consumer Behavior Patterns: Certain everyday practices strongly influence household food waste. Households that plan meals in advance and shop with grocery lists tend to waste significantly less food (Lisanne van Geffen, 2019). In contrast, overbuying often driven by impulse purchases or bulk store promotions leads to surplus food that spoils before it can be used (Ludovica Principato a, 2021). Lack of proper planning to cook what can be fully consumed (e.g., preparing more than needed) and failing to repurpose leftovers are also common causes of waste, often resulting in uneaten portions being discarded.

Proper food storage plays a key role; storing perishables under the right conditions and paying attention to expiration dates can ensure consumers consume their food before it gets bad and buy produce that would be fresh until they are ready to use it. On the other hand, many consumers discard food once a date label (e.g., "best by") passes out of caution, even when the food is likely still edible reflecting a perceived trade-off between food safety and avoiding waste (Rosetta Newsome, 2014). Surveys show that a large share of people believe discarding food after the printed date reduces the risk of illness or maintains meal quality (Rosetta Newsome, 2014). These safety and convenience motivations often override waste-reduction intentions, especially when households rely more on printed labels than on sensory cues like smell or appearance (Juliane Jörissen, 2015).

In contrast, households that engage in structured habits such as meal planning, preparing shopping lists, freezing leftovers, and creatively using older ingredients consistently report lower waste levels. These behaviors are frequently seen among segments like "Conscious Consumers" and highlight that structured food routines are key to minimizing waste (Theofanis Zacharatos 1, 2024). Information understood from the above research emphasizes that improving planning habits, shopping discipline, and food management skills are effective strategies for reducing household food waste.

Attitudes, Emotions, and Norms: While many consumers report moral discomfort about wasting food, studies suggest that such feelings may not always translate into behavioral change. In one survey, "guilt" emerged as a distinct dimension of food waste attitudes but had limited predictive power regarding actual food management behaviors (Danyi Qi, 2016).

(Violeta Stancu*, 2015)found that moral norms such as guilt or responsibility were not statistically significant predictors of food waste reduction intentions. Instead, their findings emphasized that habitual routines particularly related to shopping, planning, and reuse exerted a stronger influence on behavior than stated intentions or attitudinal variables (Violeta Stancu*, 2015)

Moreover, individuals in higher-income brackets may rationalize discarding food more easily, viewing them as necessary for safety, maintaining good health or meal quality. These beliefs, reinforced by economic comfort, can reduce the perceived urgency to reduce waste (Danyi Qi, 2016). As a result, the psychological "cost" of wasting food is often diminished for wealthier households.

Demographic characteristics add further complexity. Younger adults are frequently linked with higher waste due to convenience-focused habits or limited food planning experience, whereas older adults, while typically more frugal, may waste perishable items if they cannot consume them in time (Lisanne van Geffen, 2019). Single-person households tend to produce more waste per capita due to difficulties managing portion sizes and leftover use, especially when dealing with standard-sized food packaging (Lisanne van Geffen, 2019). Income also plays a nuanced role: high-income households may waste more due to weaker price sensitivity and stronger preferences for food freshness, while low-income households despite (Danyi Qi, 2016).

However, findings on sociodemographic drivers such as age, gender, education, and income remain mixed across studies, indicating that these variables alone do not consistently predict food waste behavior (Ludovica Principato a, 2021). Cultural and national contexts heavily influence outcomes, meaning broad generalizations (e.g., "young people waste more") often fail to capture the variability observed across countries. As (Theofanis Zacharatos 1, 2024) emphasize, it is the intersection of sociodemographic with habits and attitudes such as how affluence interacts with beliefs about food value that ultimately shapes waste behavior.

To address this complexity, segmentation studies have emerged as a promising tool for tailoring strategies to reduce food waste. Rather than assuming a uniform audience, segmentation divides consumers into clusters based on shared behavioral and attitudinal traits. (Ran Li, 2024) used a nationally representative U.S. dataset to identify four segments: "Conscientious Conservers" (low-waste, high-awareness); "Harried Profligates" (busy individuals with good intentions but poor routines); "Unrepentant Drink Wasters" (indifferent toward beverage waste); and "Guilty Carb Wasters" (feel guilty but still discard staple foods like bread or pasta). These profiles enable more focused solutions (e.g., time-saving meal kits for Harried Profligates or educational tools for drink wasters) rather than uniformed messaging.

Outside the U.S., segmentation has also proved effective. In Switzerland, six household types were identified based on food planning and management behaviors; in Turkey, researchers found four clusters aligned with different motivations and habits (Ludovica Principato a, 2021). (Theofanis Zacharatos 1, 2024) analyzed Greek consumers and found notable variation in how food waste clusters align with demographics, indicating that even shared national contexts exhibit internal diversity. Collectively, these findings support the idea that effective food waste interventions must be customized to both behavior and context.

As this literature shows, moving beyond demographic assumptions to cluster-level analysis offers a path toward higher-impact interventions. Segmenting households allows for a deeper understanding of the psychological and structural barriers behind waste, setting the stage for the next phase of research: predictive modeling and data-driven personalization at scale.

2.2. Machine Learning and Predictive Modeling in Food Waste Research

In parallel with advances in behavioral insights, in recent years, growing adoption of machine learning (ML) and predictive analytics have emerged to better understand and forecast household food waste. Traditional research on consumer food waste often relies on surveys and theory-driven statistical models. for example, applying the Theory of Planned Behavior with structural equation modeling to explain self-reported waste intentions (Violeta Stancu*, 2015). Such approaches test hypotheses about specific factors (attitudes, norms, etc.) and generally use linear models or simple regressions. While valuable for confirming theory, they can be limited in handling the complex, non-linear interactions among the dozens of variables that influence waste. Machine learning methods offer a complementary, data-driven approach: they can analyze large, high-dimensional datasets to discover patterns without being constrained to a predetermined model structure. This capability has opened new frontiers in food waste research, from identifying hidden groupings of waste behavior to making granular predictions of waste generation.

One important ML application has been in unsupervised learning to reveal latent structures in consumer behavior. Clustering algorithms like k-means have been used to segment households based on survey responses, as discussed above. These techniques objectively group similar respondents together, an approach that has now been employed in multiple countries to derive data-driven consumer archetypes (e.g. segmenting by attitudes toward leftovers and shopping routines). Similarly, principal component analysis (PCA) and related dimensionality-reduction methods have helped distill complex attitudinal data into core factors. For instance, (Danyi Qi, 2016) applied PCA to a broad battery of food waste attitude questions, finding that three underlying components explained most of the variance: (1) perceived practical benefits of wasting (e.g. throwing away food for safety or taste reasons), (2) guilt about food waste, and (3) felt *self-efficacy* (whether the household believes it could do more to reduce waste) (Danyi Qi, 2016). These kinds of analyses reduce noise and simplify modeling by focusing on a few principal drivers instead of dozens of individual survey items. They also aid interpretation policymakers can target specific attitudinal constructs (such as “waste guilt” or misplaced food safety concerns) that emerge from PCA or factor analysis in multiple studies.

Beyond exploratory analysis, supervised machine learning models are increasingly being developed to predict food waste quantities or classify waste-prone households. Regression-based models have long been used to correlate demographics or self-reported behaviors with waste levels, but newer ML techniques can improve predictive performance by capturing non-linear relationships and complex interactions. For example, ensemble learning methods like random forests have shown promise. In one recent study, a random forest model trained on household data significantly improved the accuracy of forecasting food waste generation, highlighting the power of ML to enhance prediction beyond what traditional linear models can achieve (Yi Yang, 2024). Yang et al. (2024) demonstrate that such models, by learning from patterns in consumption and disposal data, can more efficiently flag when and where waste is likely to occur (Yi Yang, 2024). Compared to a standard regression, the random forest could handle many input features (including household size, purchase frequency, food category consumption, etc.) and automatically discern which factors and interactions mattered most, without overfitting to noise. The result was a more robust prediction of waste outcomes, underscoring ML’s utility in data-rich settings (Yi Yang, 2024). In parallel, researchers have experimented with support vector machines, neural networks, and other algorithms for related tasks such as classifying food items that end up wasted (using image recognition) or predicting the shelf life of perishable products based on sensor inputs (Yi Yang, 2024). Each technique brings trade-offs, but a common theme is the ability to leverage *training data* (historical examples of food being wasted or saved) to *learn* complex decision rules that can then be applied to new situations.

Modern smart kitchen systems leverage IoT sensors and AI to not only monitor food waste but actively predict and prevent it. In a smart kitchen, connected devices like smart bins, sensor-equipped refrigerators, and AI-integrated kitchen tools continuously collect data (e.g. weight of discarded food, temperature/humidity, gas emissions, inventory levels). This real-time data is fed into machine learning models that can detect patterns and anticipate waste before it happens. Crucially, these systems move beyond passive tracking to deliver immediate, data-driven interventions aimed at reducing household food waste.

Smart bins are a key innovation: they automatically measure and classify waste, providing feedback that helps users change their habits. For example, IoT-enabled bins outfitted with weight scales and gas sensors can identify the type and quantity of food being thrown out with high accuracy using ML algorithms (Sahar Ahmadzadeh, 2023). One such system achieved over 92% accuracy in recognizing discarded foods (by detecting spoilage gases like CO_2 and NH_3) and generated daily waste reports for the household (Sahar Ahmadzadeh, 2023). By informing users exactly what and how much they waste each day, these smart bins support personalized interventions. An example is, prompting a family to buy less of an item that frequently ends up in the trash. Networked smart bins can also aid municipal efforts by sending fill-level and composition data to city waste managers, optimizing collection routes and highlighting neighborhoods or foods that could be targets for waste-reduction campaigns (Sahar Ahmadzadeh, 2023).

In tandem, AI-integrated kitchen appliances like smart refrigerators help prevent waste on the front end. A prototype smart fridge system, for instance, tracks food stocks and expiration dates and syncs with a mobile app to guide users in daily food management (proper storage, timely consumption, recipe suggestions, etc.) (Cappelletti F, 2022). Such a system can alert a user that a product is nearing its expiration and recommend meal ideas to use it up, or even adjust grocery lists, thereby bridging the gap between data and action. In a 2022 study, a smart fridge combined with a web app for meal planning and storage tips showed great potential to cut avoidable household waste by improving user awareness and planning (Cappelletti F, 2022). These tools illustrate how continuous sensor data (from RFID tags, weight sensors, cameras, etc.) coupled with predictive analytics can turn a passive kitchen into an active ally against food spoilage and waste.

Importantly, machine learning-driven interventions are proving effective not just at home but also in food service and retail settings. AI-powered waste tracking systems employ computer vision and scales to recognize foods being discarded and then provide actionable insights in real time. For example, the Winnow system uses a camera above the trash bin to automatically identify thrown-away foods without disrupting kitchen workflows; this IoT solution has helped some commercial kitchens (e.g. at IKEA cafeterias) cut food waste roughly in half (ReFeD, 2024). Another platform, Leanpath, aggregates kitchen waste data and uses AI (including generative models) to recommend concrete preventive actions – such as adjusting portion sizes, modifying menus, or training staff on wasteful practices, thereby stopping waste before it recurs (ReFeD, 2024). These successes underscore the power of pairing continuous monitoring with intelligent feedback: when kitchen staff or home cooks get timely information (like "X pounds of bread wasted this week" or "leftover cornbread often goes uneaten (Lundeberg, 2023)), they can intervene, bake less bread, serve smaller portions, or repurpose ingredients, to avoid future waste.

The strengths of machine learning in food waste research contrast with more traditional approaches in notable ways. ML excels at prediction and pattern recognition, handling large heterogeneous datasets (e.g. nationwide survey data combined with retail purchase records or sensor readings) and uncovering relationships that a manual analysis might miss. This can yield higher predictive accuracy and novel insights, for example, an algorithm might discover that a combination of subtle factors (like weather, grocery list, and family food sufficiency) predicts a spike in a household's waste, which would be hard to detect via standard statistical tests. ML models can also continuously improve as more data become available, making them well-suited for tracking progress toward waste reduction goals over time. However, these advantages come with challenges. One is interpretability: complex models like random forests or neural networks are often "black boxes," making it difficult to explain their predictions in human terms. This can be a drawback in an academic and policy context, where understanding causation is as important as prediction. There is growing recognition that purely data-driven models should be blended with theory, for instance, incorporating behavioral variables grounded in frameworks like TPB, or using algorithms (such as decision trees) that provide interpretable rules alongside accuracy (László Mucha, 2025). Another challenge is data quality and availability. High-performing ML models require large, reliable datasets; yet measuring household food waste at scale is notoriously difficult. Self-reported data can be biased (e.g. people underestimating their waste due to social desirability), whereas sensor-based approaches are still in pilot stages and not widely deployed. As a result, many ML studies to date have had to make do with limited sample sizes or proxy data, which can constrain model generalizability. Closing this gap will likely involve more investment in data infrastructure – from smart waste measurement systems to integrated databases that link household demographics, purchasing patterns, and waste outputs.

Despite these hurdles, the trajectory of recent research suggests an increasingly important role for machine learning in advancing food waste solutions. Innovative studies have begun to demonstrate the practical benefits of ML-driven insights: from optimizing shopping and menu planning apps with personalized waste forecasts, to improving food donation and redistribution logistics through better demand prediction (Yi Yang, 2024). As the field progresses, we can expect more hybrid approaches that combine the richness of qualitative insights (e.g. why people feel compelled to overshop) with the rigor of quantitative models (e.g. predicting who is likely to overshop and waste).

In summation, the literature on behavioral segmentation and machine learning suggests that integrating large-scale data with predictive analytics can yield actionable insights into food waste behavior, an approach this study now applies to the U.S. households.

2.3. Data Processing and Management

This study integrated three distinct yet complementary datasets FoodAPS (2012–2013) (Agriculture, 2013), ReFED (2018) (ReFED, 2018), and ACS (2018) ((ACS), 2018) to develop a robust foundation for analyzing household food waste in the U.S. ReFED's 2018 estimate of post-consumer food waste was selected because it offers one of the most comprehensive and publicly available national-level assessments prior to the COVID-19 pandemic. This timing avoids pandemic-related behavioral anomalies and aligns well with the demographic structure captured in the 2018 American Community Survey (ACS), which was used for regional and population weighting. Although the FoodAPS data (2012–2013) precedes both sources, key behavioral patterns such as grocery routines, food access, and sufficiency have shown stability over time, making this temporal integration suitable for predictive modeling. This integration allows for a comprehensive and timely analysis that reflects structural patterns still relevant to present-day waste interventions. FoodAPS provides the most comprehensive micro-level data on household food behaviors and socio-demographics, while ReFED and ACS supply aggregate waste and population-level indicators.

All raw survey variables from FoodAPS were carefully cleaned and recoded before analysis. Categorical attributes (such as whether the household currently receives SNAP benefits *snapnowhh*, home ownership status *housingown*, self-reported food sufficiency) were converted into categorical factor variables or binary indicator dummies as appropriate to ensure they could be correctly utilized in the models. Any observations with missing values on the variables of interest were removed using complete-case analysis. This exclusion of incomplete cases yielded a consistent analytic dataset without missing data, preventing potential biases or errors that could arise from imputation or undefined values.

Household Food Insecurity Measure: The USDA 10-item Household Food Security Survey Module was used to construct a composite food insecurity metric for each household. Each yes/no item (e.g. *"we couldn't afford to eat balanced meals* *in the last 30 days – foodsecureq3"*) was first converted to a binary indicator (1 for an affirmative response signaling insecurity, 0 otherwise). These were then summed to produce a continuous *household food insecurity score* ranging from 0 to 10, with higher values indicating greater food insecurity. This score provided a single quantitative measure of a household's food security status (0 meaning fully food secure and 10 meaning very high insecurity) for use as a predictor in the models.

We checked all continuous variables for skewness to see if transformation was needed. Several variables, like household income-to-poverty ratio, number of meals at home, and grocery list frequency were right-skewed (many low values, few very high ones). To correct this and reduce the impact of outliers, we applied log transformations (using log(x+1) to handle zeros). These log-transformed variables were then used in the models to better meet statistical assumptions and improve prediction accuracy.

The analysis dataset was then augmented by merging in relevant external data from the ReFED and ACS sources using region identifiers. Every FoodAPS household record contains a region code, which allowed us to join regional-level statistics. ReFED regional food waste totals (annual residential food waste in tons for each region) were merged onto the household data by matching region. Similarly, key regional socio-economic indicators from the 2018 American Community Survey (ACS) were merged as constant reference features for all households in the same region.

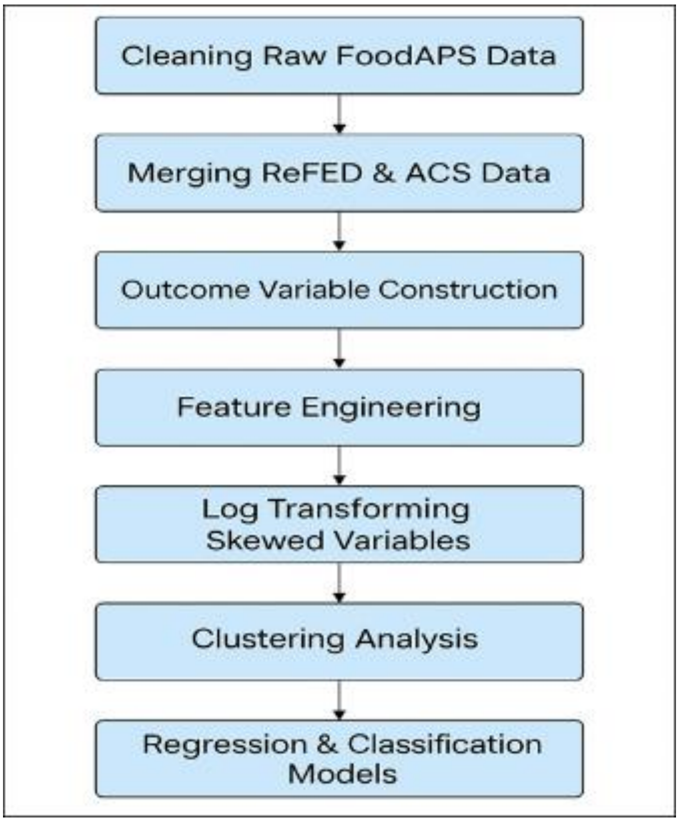

Figure 1 Data flow of the data analysis process

Outcome Variable Construction: The dependent variable—estimated household food waste—was calculated by dividing ReFED's regional household food waste totals (tons/year) by the number of households in each region. Household counts were derived from FoodAPS survey weights (hhwgt), adjusted using ACS 2018 data for accuracy. This yielded a per-household food waste estimate, which was assigned to each FoodAPS household.

Feature Engineering: After cleaning and merging the data, key features were engineered to create cluster profile and model household food waste. These included demographics (e.g., household size, homeownership), food sufficiency, grocery planning (log-transformed), diet quality perceptions, SNAP participation, and store quality. Socioeconomic context was captured using income tier proportions, urban-rural indicators, and log-transformed household weights and income-to-poverty ratios. These variables reflect household behavior, resources, and environment factors known to influence food waste.

Description of variables in the dataset

- hhsize: Household size, used as a basic control for household composition.
- foodsufficient: A categorical indicator of whether the household reports having enough food.
- grocerylistfreq: Frequency of grocery list use, a proxy for planning behavior.
- housingown: Binary variable indicating whether the household owns or rents their home.
- dietstatuspr and dietstatushh: Self-reported and household-rated diet quality levels.
- hhwgt: Household survey weight, used in scaling but excluded from model predictors to prevent bias.
- pctpovguidehh_r: Percentage of the federal poverty guideline for the household used as a normalized income measure.
- Region: Census region
- inchhavg_r : household average monthly income as sum of average inputed income per member
- anyvehicle: Whether anybody in the household owns or leases a vehicle (Y/N)
- caraccess: household has access to car when one is need (Y/N)
- snapnowhh: Binary indicator for current SNAP participation.
- primstorequality: Rating of the household's primary grocery store.
- low_income_pct, mid_income_pct, high_income_pct: Nationally derived ACS-based proportions of households in low-, middle-, and high-income groups, added to provide context.
- nmealshome, nmealstogether: Meal preparation and eating behaviors, indicating how often households eat at home or together.
- nonmetro, rural: Binary indicators based on USDA rurality classifications.
- cluster_label: Result of behavioral clustering (see below); used for descriptive analysis and as a target in classification models.

Behavioral Clustering: In addition to the continuous outcome modeling, we performed an unsupervised clustering analysis to segment households into distinct behavioral profiles. Using K-means clustering on standardized household behavior and perception variables (including food sufficiency status, primary store quality, grocery planning frequency), we identified groups of households with similar characteristics. Based on examination of cluster centers and summary statistics, a five-cluster solution was selected as the most interpretable. Each cluster was assigned a descriptive name reflecting its defining attributes. The final clusters were labeled *Moderate Planners with Basic Resources*, *Underserved & Food Insecure*, *Well-Equipped & Food Secure*, *Least Supported with Poor Food Routines*, and *Routine Shoppers with Mild Constraints*. These labels summarize the general profile of each group (for example, the "Underserved & Food Insecure" cluster had high food insecurity scores and limited resources, whereas the "Well-Equipped & Food Secure" cluster had ample resources and good access to food). Cluster membership was subsequently used as an alternative target variable in classification models to predict which profile a household belonged to, offering a complementary analysis of the data. Notably, the cluster labels were not used as predictors in the food waste regression; instead, they served to highlight latent group differences and to develop a classification modeling exercise distinct from the main waste estimation task.

Modeling Preparation: With a clean, enriched dataset in place, we proceeded to model development. Both regression and classification models were built using the preprocessed data, ensuring that all skewed variables were in their log-transformed form and all categorical variables were encoded as described. For the continuous outcome (household food waste), we fitted a standard linear regression model to establish a baseline and then explored more flexible machine learning methods, including a Random Forest regressor and a gradient boosting regressor, to capture non-linear relationships and interactions among predictors. For the multi-class classification task of predicting household cluster membership, we analogously trained machine learning classifiers (Random Forest and gradient boosting classification models) using the same set of features (excluding the waste outcome). In all cases, the data were randomly split into a training set (70% of households) and a testing set (30%) prior to modeling. This 70/30 split was used to train the models on one subset while reserving a separate hold-out subset for evaluating out-of-sample performance. This approach helps prevent overfitting and provides an unbiased assessment of how well each model generalizes to new data. Overall, the extensive preprocessing and data management steps outlined above were essential to ensure model accuracy and consistency. By carefully cleaning the data, engineering meaningful features, and integrating external information, we aimed to maximize the reliability of the regression estimates and classification results in our household food waste modeling project

3. Results

3.1. Behavioral Clustering

To better understand household food waste behavior, we employed K-means clustering on a subset of behavioral and access-related variables in the FoodAPS dataset. These included indicators such as grocery list usage, food sufficiency, and store quality. After testing multiple cluster solutions, a 5-cluster model was selected based on interpretability and alignment with theoretical distinctions in food access and management behavior. This segmentation contributes directly to our objective classifying household types based on socio economic and behavioral traits and supports the other objective of identifying behavioral patterns that help explain variation in food waste predictions.

The clusters were named based on their dominant characteristics:

- Moderate Planners with Basic Resources: households who consistently use grocery lists and have steady but modest access to reliable stores and are averagely food sufficient.
- Underserved & Food Insecure: households with low food sufficiency, lower access to quality stores compared to cluster 1 and inconsistent planning habits. “Underserved” signals (poor store access, poverty), while “Food Insecure” indicates struggle to get enough food.
- Well-Equipped & Food Secure: households had the highest access to quality stores, reported to have sufficient food, and moderate-to-structured food routines.
- Least Supported with Poor Food Routines: households with weak planning habits, limited access to quality food, and often low food sufficiency. “Least Supported” highlights the lack of both material and behavioral support
- Routine Shoppers with Mild Constraints: households that engage in consistent shopping habits and basic planning but face moderate limitations, such as suboptimal store access. “Mild Constraints” signals manageable barriers that do not severely disrupt food access.

Table 1 Food Consumption Behavioral Result

Cluster Label	**Avg Store Quality**	**Avg Grocery Freq(log)**	**Avg Food Sufficiency**	**Description**
Moderate Planners with Basic Resources	0.206	1.42	0.567	Moderate food store quality and planning routines; moderate food sufficiency.
Underserved & Food Insecure	0.170	1.34	0.374	Lower store quality and grocery planning; weakest food sufficiency.
Well-Equipped & Food Secure	0.246	1.43	0.719	Highest store quality and food sufficiency; strongest planning behavior.
Least Supported with Poor Food Routines	0.162	1.27	0.409	Lowest grocery planning and food quality ratings; food routines are weaker.
Routine Shoppers with Mild Constraints	0.194	1.28	0.436	Somewhat structured grocery habits but modest store quality and food access.

3.1.1. Avg Store Quality

Scale: 0 to 1

Higher values indicate better perceived food store quality, based on factors like product availability, affordability, and proximity.

Example interpretation:

~0.25 = highest store quality (e.g., Well-Equipped cluster)

~0.16 = lowest store quality (e.g., Least Supported cluster)

3.1.2. Avg Grocery Frequency (log)

Log-transformed frequency of household grocery trips per week.

Higher values represent more frequent shopping, which may indicate stronger planning routines or better access.

Approximate scale:

1.27–1.43 = about 3 to 4 grocery trips per week (after back-transformation from log)

Avg Food Sufficiency

Scale: 0 to 1

Measures how often households report having enough of the kinds of food they want to eat.

Higher values indicate greater food sufficiency and security.

Example interpretation:

~0.72 = high sufficiency (Well-Equipped)

~0.37–0.41 = low sufficiency (Underserved & Least Supported)

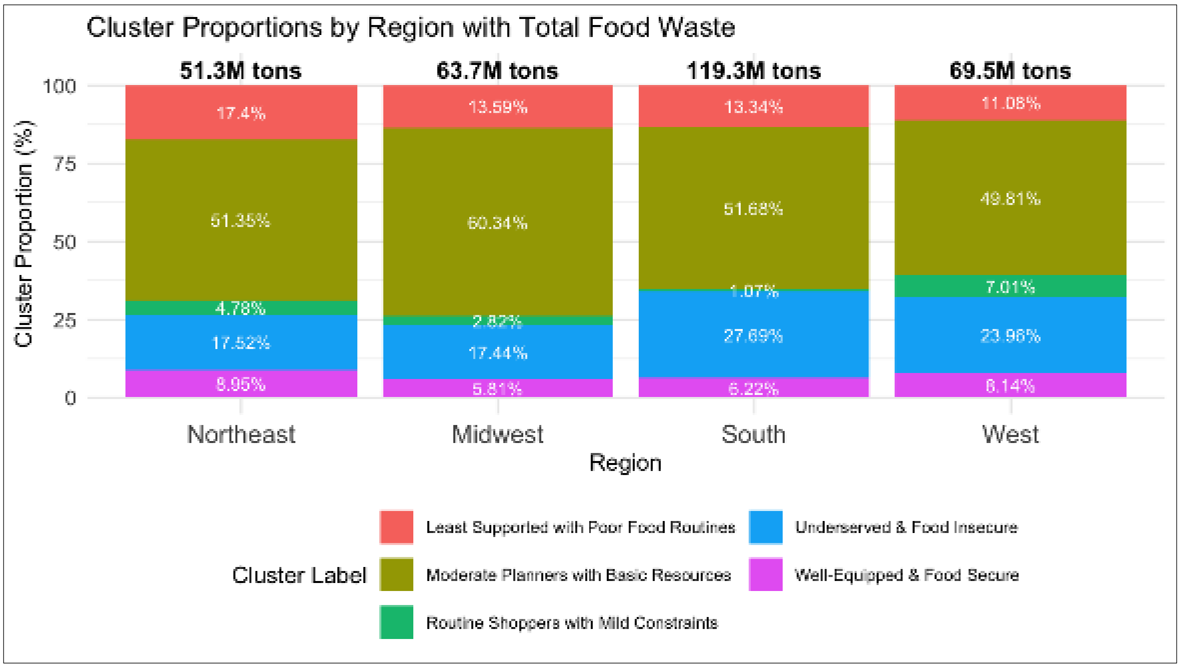

Figure 2 Stacked bar chart of cluster proportions by region with total household food waste

We applied unsupervised clustering to FoodAPS household data (the first nationally representative U.S. food acquisition survey, including 4,826 households with SNAP and nonSNAP participants (Miller, 2025) using variables such as self-reported food sufficiency, perceived store quality, and shopping frequency. Five distinct clusters emerged, each characterized by different combinations of these factors. For example, one cluster comprised households with high food security and good store access, while the *Underserved & Food Insecure households* showed high SNAP usage and limited food sufficiency, but mobility limitations were not explicitly observed in this segment. These clusters were validated by comparing summary statistics (e.g. average waste) and by examining their geographic distribution. We found clear differences in waste outcomes: some clusters, despite being smaller, generated disproportionately more total waste. The regional breakdown (Figure above) shows how cluster membership varies across census regions and highlights each cluster's contribution to total household food waste. This clustering lays the groundwork for understanding patterned differences in household waste behaviors.

3.2. Predicting Household Food Waste

We then built predictive models for each household's total food waste. Using linear regression, random forest, and gradient boosting, we aimed to improve on a simple baseline (MAPE ≈19.7%). Gradient boosting delivered the strongest performance with a Root Mean Squared Error (RMSE) of 0.52 and Mean Absolute Percentage Error (MAPE) of 18.60%, outperforming the benchmark RMSE of 0.54 and benchmark MAPE of 19.71% obtained by predicting the mean waste per household for all cases. Importantly, these errors indicate reasonably accurate predictions of waste amounts at the household level.

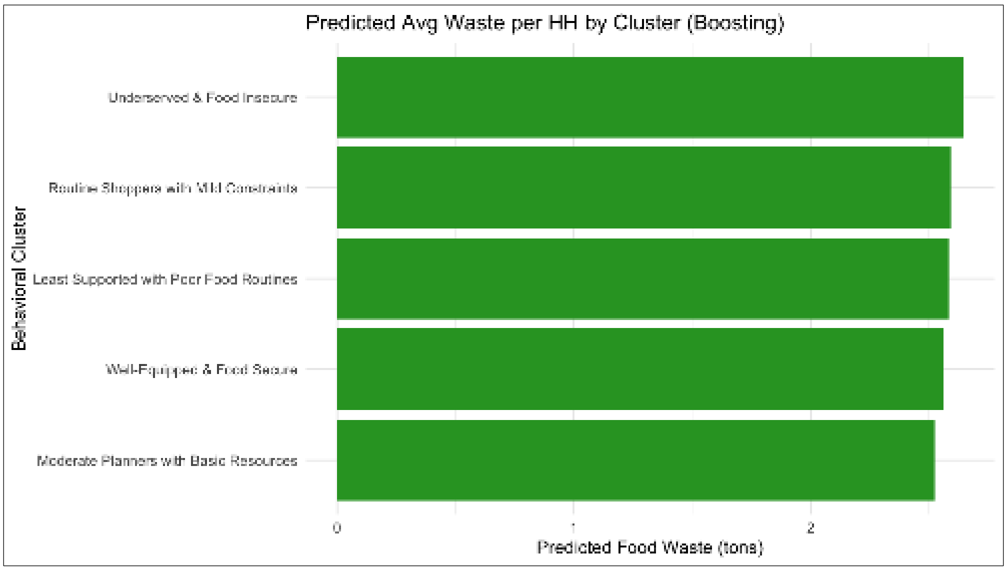

Figure 3 Predicted average food waste per household by behavioral cluster (from the boosting model)

Boosting model predictions revealed clear differences in estimated food waste across behavioral clusters. Households in the 'Underserved & Food Insecure' cluster had the highest predicted average waste at approximately 2.64 tons per household annually, while the 'Moderate Planners with Basic Resources' cluster had the lowest, at around 2.53 tons. This pattern aligns with the behavioral characteristics identified in the clustering stage, suggesting that resource limitations and planning behaviors are associated with food waste outcomes.

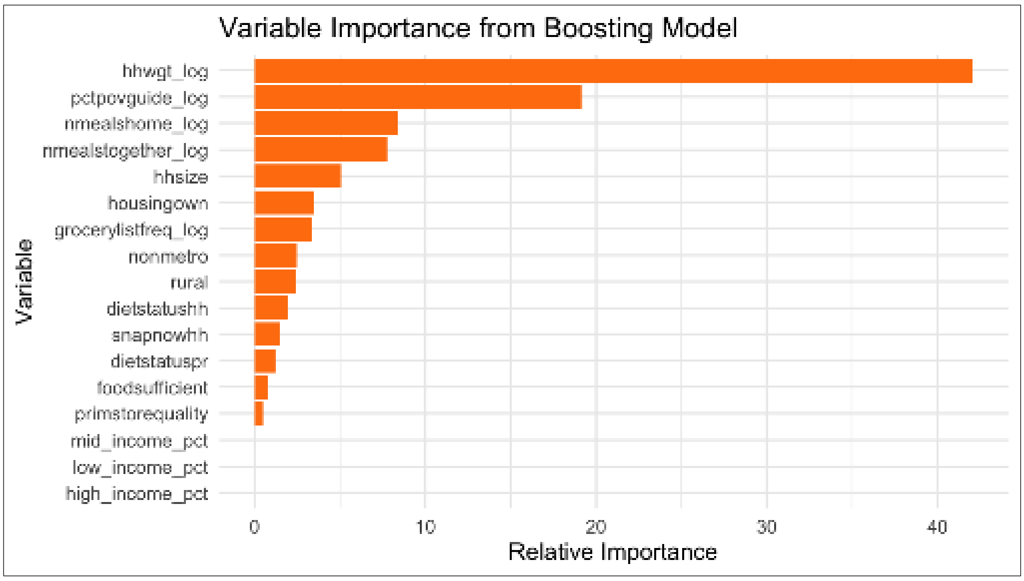

Figure 4 Variable importance from the gradient boosting model predicting household waste

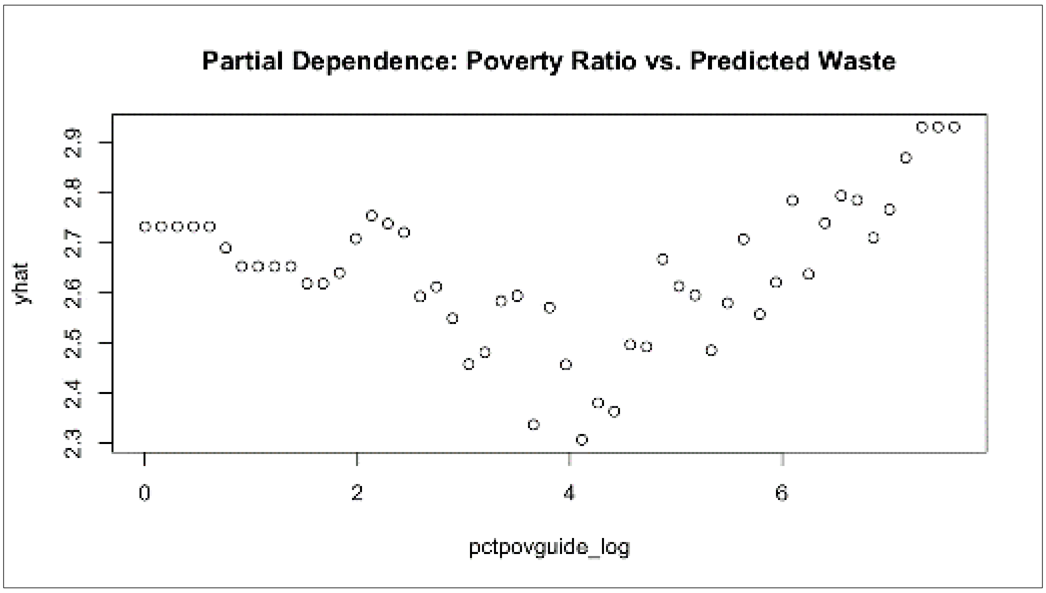

Figure 5 Partial dependence of predicted food waste on household poverty ratio (log transformed)

As shown in Figure 5, the model captured a U-shaped partial dependence between poverty ratio and food waste. This suggests that both high-income and very low-income households are associated with increased waste, albeit likely driven by different mechanisms. In the random forest model, similar variables topped the importance list, confirming the robustness of these predictors across the different algorithms. Both models emphasize socioeconomic and behavioral features. Larger households were associated with greater total food waste. Additionally, food waste was higher among households at the lowest and highest ends of the income distribution, suggesting that both economic hardship and abundance may lead to inefficiencies in food use. This model-driven analysis supports the idea that demographic and planning behaviors are key to forecasting household waste.

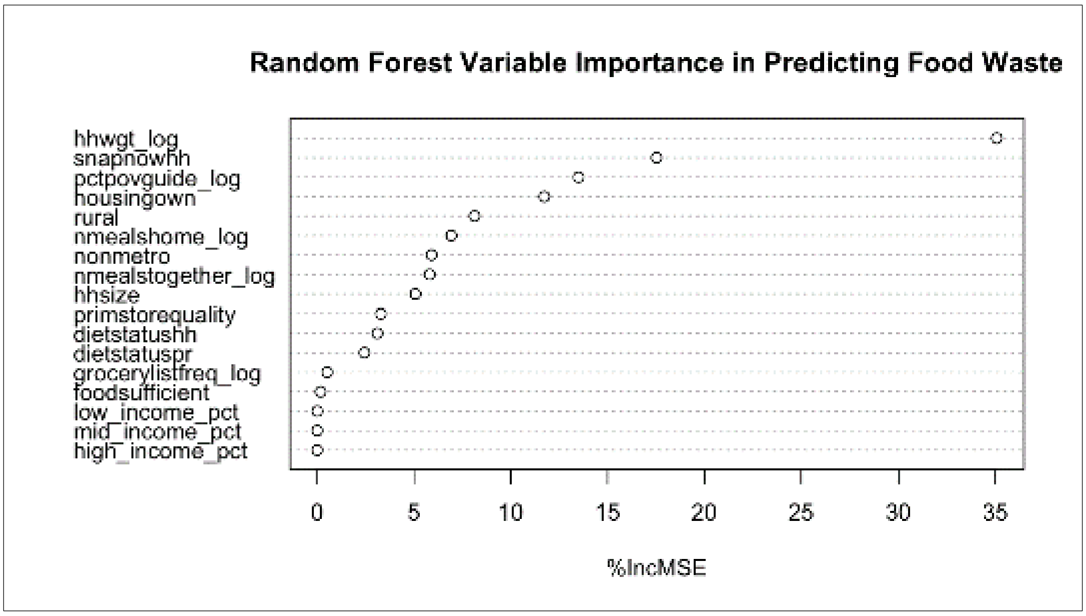

Figure 6 Variable importance from the random forest model predicting household waste

3.3. Classification Tasks

3.3.1. SNAP Participation

Although logistic regression achieved a slightly higher AUC (0.838) than the random forest model (AUC = 0.827) in predicting SNAP participation, we relied on the variable importance results from the random forest classifier to interpret key predictors. This is because random forests naturally capture nonlinear relationships and interactions among variables, making their importance scores more robust in high-dimensional, behaviorally complex datasets like FoodAPS. In contrast, logistic regression coefficients assume linear effects and may be less informative when variables interact or exhibit threshold behaviors. Thus, for interpretability and policy relevance, we used the random forest's ranked importance to highlight which household features, such as poverty ratio, household size, and household ownership, most strongly influence SNAP classification.

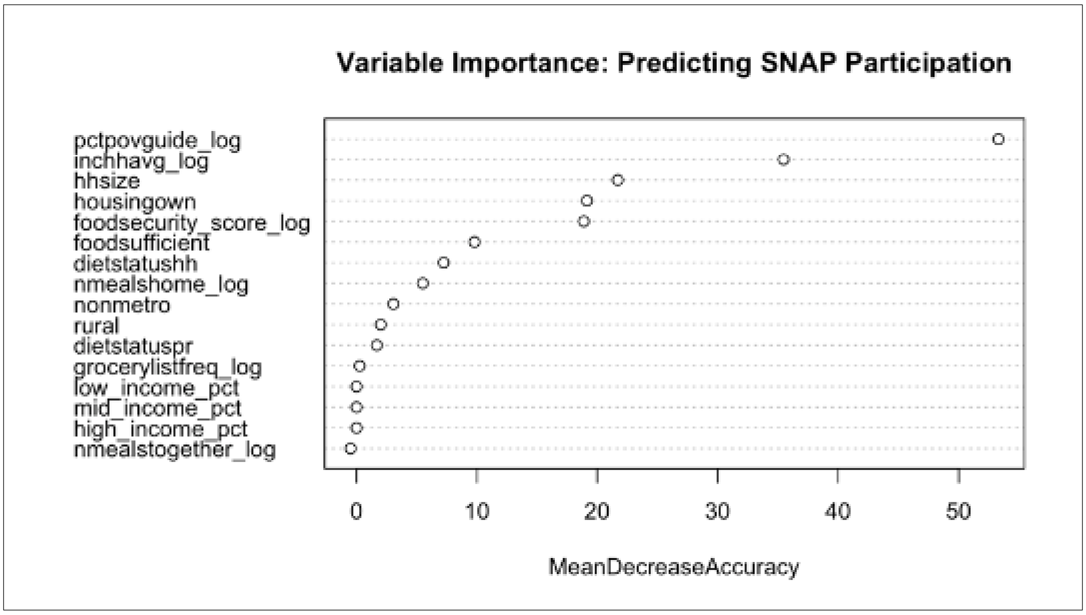

Figure 7 Variable importance from the random forest model predicting SNAP participation

Predicted SNAP probabilities also differed systematically across our behavioral clusters. Clusters characterized by higher need (e.g. lower food sufficiency and income) showed substantially higher predicted SNAP probabilities (Figure below). This clustering-based breakdown confirms that the unsupervised segments correlate with program participation, consistent with their socioeconomic profiles.

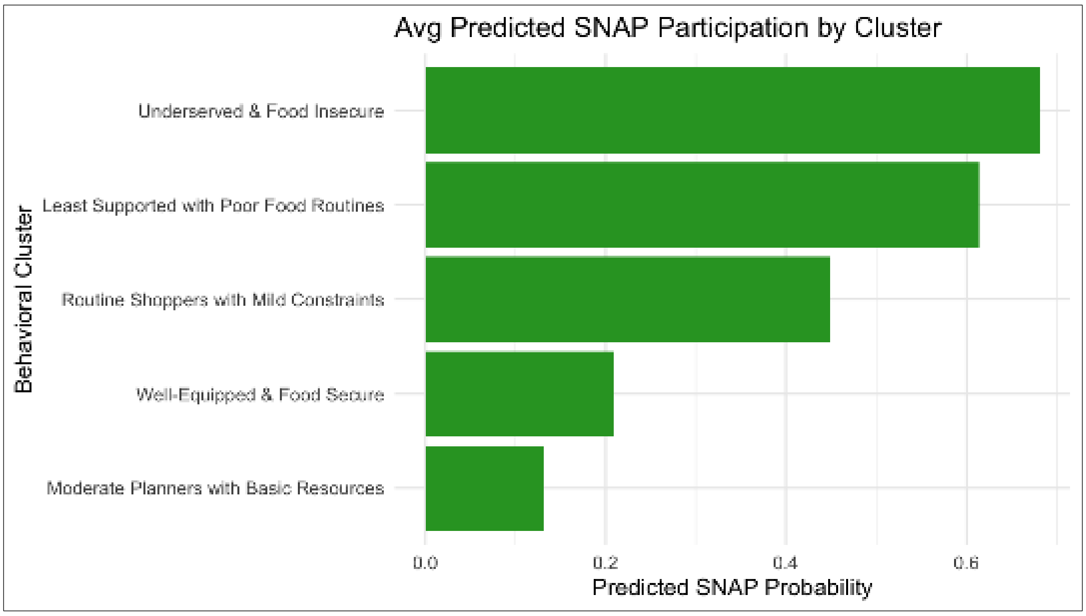

Figure 8 Predicted probability of SNAP participation by behavioral cluster (bar chart)

3.3.2. Behavioral Cluster Membership

Finally, we treated cluster membership as a multi-class outcome to see how well households could be classified into the five behavioral clusters from their features. A random forest model achieved about 72.7% accuracy in predicting cluster labels: well above the 20% baseline for random guessing. The confusion heatmap (Figure below) shows that most households were correctly classified (major diagonal), though there was some confusion between similar clusters. This suggests the clusters are reasonably distinct but have overlapping attributes.

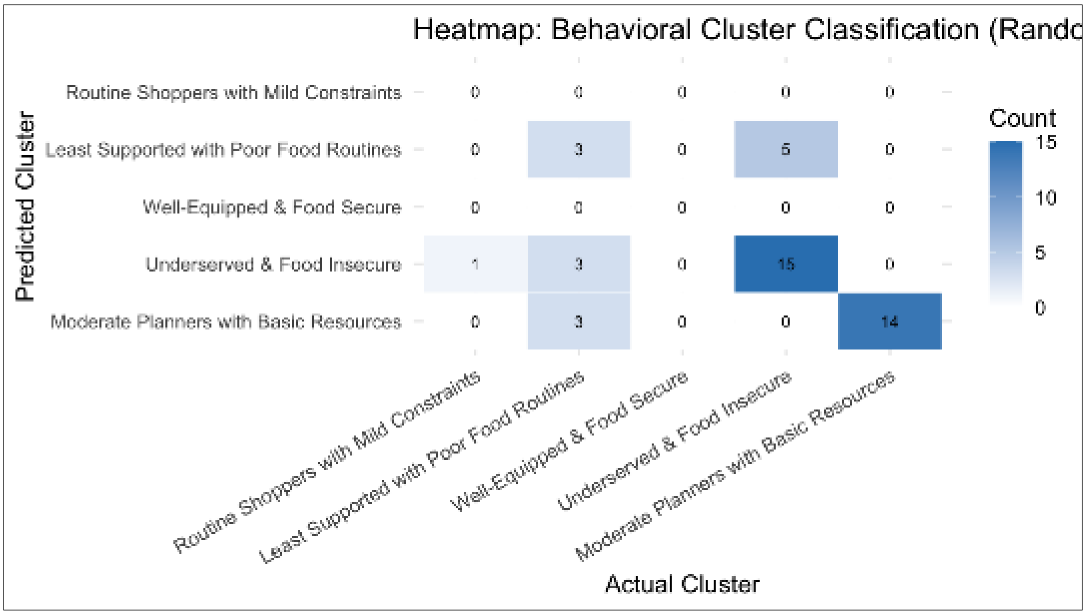

Figure 9 Heatmap of behavioral cluster classification (true vs. predicted cluster counts) from the random forest model

The feature importance for this multi-class classification again highlighted the same key drivers: socioeconomic and food security factors dominated (e.g. household size, snap participants) along with some diet status and routine meal dynamic indicators in the household. These results indicate that the clustering captures real, predictable patterns: knowing a household's demographics and consumption patterns allows fairly accurate assignment to the correct behavior-based segment.

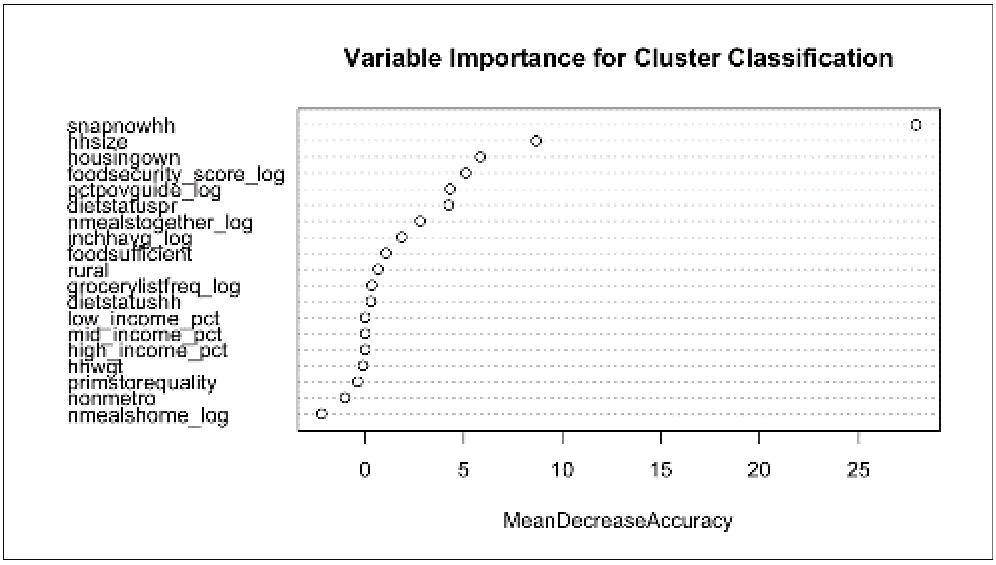

Figure 10 Variable importance for predicting behavioral cluster membership (random forest model). This highlights which features most distinguish the five clusters

Overall, the clustering and predictive analyses together paint a coherent picture: distinct household segments with different needs and behaviors can be identified, and these segments help explain and predict food waste patterns. The visualized results above link clusters to regions and waste levels and show that models using demographic and behavioral variables can effectively predict both waste quantity and related outcomes (like SNAP participation), supporting the goal of leveraging public data to understand household food waste.

4. Discussion

4.1. Model Insights

The predictive modeling in this study provided important insights into which household characteristics are most useful in estimating food waste and classifying household behaviors. Among the three regression models tested, linear regression, random forest, and gradient boosting. The Gradient boosting model produced the most accurate predictions, based on evaluation metrics like RMSE and MAPE. This reflects findings in sustainability research where advanced modeling techniques are used to handle behavioral and environmental complexity. For example, Yu and (Fan, 2023)modeled household-level food waste responses to SNAP using structural estimation, highlighting the importance of accounting for heterogeneous behaviors in food policy design. (Fan, 2023). This model highlighted behavioral and demographic features such as household size, poverty ratio, and frequency of meals at home as key drivers of food waste outcomes. These findings support past research showing that food waste tends to rise with family size and income, and that food planning and cooking habits play a crucial role in how food is used or discarded (Violeta Stancu*, 2015)

Beyond predicting food waste, we developed classification models to examine how household characteristics relate to SNAP participation. Among the two models tested logistic regression and random forest logistic regression achieved slightly better performance, with an AUC of 0.838 and an accuracy of 76.9%. However, we relied on the variable importance output from the random forest model to interpret the most influential predictors. This decision reflects the strength of tree-based models in capturing nonlinear relationships and interaction effects, which are often present in high-dimensional behavioral data like FoodAPS. The top predictors of SNAP participation included poverty ratio, household size, and housing status (own vs. rent), highlighting that socioeconomic and household structure indicators can reliably signal assistance needs. These results reinforce the paper's hypothesis: behavioral and demographic patterns can be leveraged to proactively identify food-insecure households, even in the absence of direct income questions an insight valuable for targeting outreach and program eligibility in practice.

The behavioral cluster classification model also used a random forest approach and achieved moderate accuracy (72.7%) in assigning households to one of five defined clusters. While the model classified some groups such as "Moderate Planners with Basic Resources" and "Underserved & Food Insecure" with high sensitivity and specificity, others like "Routine Shoppers with Mild Constraints" were more difficult to distinguish. This reinforces the need for segmentation methods that accommodate overlapping or blended household profiles.

4.2. Interpretation of Findings

The gradient boosting regression model highlighted several key predictors of household food waste. Notably, a household's socioeconomic status and composition emerged as dominant factors. Household income relative to the poverty line (the FoodAPS variable pctpovguidehh_r) was among the most influential features. Partial dependence results from the boosting model show a nonlinear relationship, where both the lowest and highest poverty ratio households exhibit higher predicted waste. This partially supports prior research suggesting that wealthier households may waste more (Izzy Klugman, 2024), but also highlights complexities, with lower-income groups potentially facing other waste-related inefficiencies. Household size (hhsize) was another strong predictor: larger families generated more total waste, although waste per capita tends to be lower in bigger households (Petra Nováková, 2021), whereas single-person households often have higher per-person waste due to difficulties in purchasing and preparing food in small quantities (Petra Nováková, 2021). This aligns with studies finding that families with children particularly young children, waste more food than those without (Pietro Tonini, 2023), likely because managing kids' diets leads to uneaten leftovers (Barcelona, 2024). Behavioral factors related to food management also showed high importance. For example, meals consumed at home (nmealshome) and meals eaten together as a family (nmealstogether) had substantial predictive power, indicating that more frequent home cooking and family dining can increase opportunities for leftovers and spoilage if not managed well. At the same time, planning-oriented behaviors like using grocery lists (grocerylistfreq) were associated with lower waste; a finding in line with evidence that households who regularly plan meals and shop with a list report significantly less wasted food (Pietro Tonini, 2023). The model also captured urban–rural differences: the rural indicator had notable importance, with rural households predicted to waste less on average than urban ones, consistent with studies that urban residents tend to discard more edible food than their rural counterparts (Aakanksha

Bhatia, 2024) (possibly due to more limited food access or stronger preservation practices in rural areas). Interestingly, home ownership (housingown) and indicators of food security or nutrition attitudes (e.g. foodsufficient, dietstatus) were moderate predictors as well. This suggests an interplay between food waste and household stability or health consciousness. In fact, households striving for healthier diets often purchase more perishable produce, which can inadvertently raise waste if fruits and vegetables spoil. Thus, the model's top predictors reinforce known drivers of waste: higher income and larger households tend to waste more overall, while proactive food management behaviors (planning, inventory awareness) can mitigate waste.

Beyond individual variables, the behavioral segmentation of households provides further insight into waste patterns. The predicted average food waste per household varied widely across the identified clusters. Notably, some of the highest waste levels were found in clusters characterized by large family size or a combination of behavioral and socioeconomic vulnerability. For example, the model projected the "Underserved & Food Insecure" cluster to have among the highest average predicted food waste. Thus, estimating that families within the "Underserved & Food Insecure" for example have more individuals in their household. While this may seem counterintuitive, it aligns with research suggesting that increased food access through SNAP or food pantries does not always reduce waste, especially when not accompanied by effective food planning, storage, or preparation skills. Households receiving large volumes of food (e.g., during monthly SNAP disbursements) may face spoilage if storage capacity or meal planning is inadequate. Figure 3: predicted Avg Waste per Household by Cluster

In contrast, the Well-Equipped & Food Secure cluster had the lowest predicted food waste. These households showed signs of stability, resource access, and moderate planning behavior, which likely contributed to reduced spoilage. Still, in other groups, low predicted waste may reflect gaps in data rather than actual lower waste. Prior research shows that tracking challenges especially among lower-income or vulnerable populations can lead to underreporting of food routines or purchases, underestimating actual waste. (Juliane Jörissen, 2015) (Roni A. Neff, 2015)

The Least Supported with Poor Food Routines cluster also exhibited high predicted food waste. This aligns with existing literature suggesting that households facing economic and logistical constraints especially those with children may waste food due to over-preparation, inconsistent planning, or difficulty matching meals to preferences (Violeta Stancu*, 2015; Ran Li, 2024). These time-constrained households often purchase in bulk or make infrequent trips to low-quality stores, increasing the likelihood of spoilage and waste.

Conversely, The Moderate Planners with Basic Resources cluster was the dominant segment across all U.S. regions, accounting for approximately 50% to 60% of households, with the highest share in the Midwest. Despite their scale, these households showed lower predicted waste per household, likely due to their regular food planning and consistent store access. This is consistent with research showing that rural or modest households can exhibit more conservative food use habits, especially when budgets are tight (Roni A. Neff, 2015)

4.3. Policy Implications

These findings offer important implications for policy targeting and program design. A central insight is that a small number of household segments drive a disproportionate share of food waste. Targeting these high-waste clusters—especially those with structural and behavioral constraints—can deliver meaningful reductions.

The Least Supported with Poor Food Routines cluster contributes substantially to household-level waste. These time-constrained households face limited food access and report minimal planning behavior, increasing the likelihood of over-purchasing and spoilage. They mirror the 'Harried Profligates' segment identified by (Ran Li, 2024); households that waste ~45% more than average but respond positively to convenience-enhancing solutions (Ran Li, 2024)Prior work by (Violeta Stancu*, 2015) similarly finds that impulsive and unstructured shopping habits are key predictors of higher household food waste. Programs offering meal-planning apps, pre-portioned kits, or peer support could improve routines without adding burden.

Equally important is the *Underserved & Food Insecure* cluster, which had among the highest predicted waste levels despite receiving food assistance. This reflects a critical policy gap: food distribution (e.g., via SNAP or pantries) does not always reduce hunger if recipients cannot properly store or plan meals. Solutions could include expanding SNAP-Ed programs to cover meal planning and food storage skills, offering low-cost storage tools through food banks (e.g., insulated bags, containers), testing biweekly SNAP disbursements to reduce end-of-month shortages and spoilage, a strategy that helps smooth consumption and avoid end-of-month spoilage (Fan, 2023)

A third insight concerns Routine Shoppers with Mild Constraints; a moderate-income group with moderate predicted waste. These households may not qualify for SNAP but still experience food insecurity and make planning tradeoffs due to work schedules or stigma. They often fall through the cracks of both food waste and food security programs. Policy tools like educational outreach through workplaces or clinics, food budgeting classes, or public messaging campaigns can reach this group in ways traditional programs do not.

Limitations

- While this study offers important insights into household food waste, it is essential to acknowledge key limitations and outline areas for future research.

Age of the data: One major limitation is the age of the data. This analysis is based on the USDA FoodAPS dataset from 2012–2013. Since then, household behaviors have likely changed due to factors such as inflation, increased public awareness of food waste, and the COVID-19 pandemic, which affected how people shop, cook, and store food (USDA ERS, 2023). As a result, the patterns we identified may not fully reflect current realities.

To better reflect today's food landscape, the USDA is currently developing a second wave of the National Household Food Acquisition and Purchase Survey, known as FoodAPS-2 (Maguire, 2025)This updated survey is designed to address limitations in the original 2012–2013 data by incorporating changes in consumer behavior, such as the rise of online grocery shopping, meal kits, food delivery, and the impacts of COVID-19 and inflation on food acquisition.

Lack of direct household waste measurement: We used ReFED's regional estimates to calibrate our model but lacked direct observations of food waste per household. This means our predictions are based on indirect behavioral signals (e.g., meals cooked or food security scores) rather than recorded waste quantities. New tools are under development to address this. At Oregon State University, researchers are building smart compost bins equipped with sensors and imaging technology to automate food waste tracking (Food Waste, 2023) .These tools allow passive monitoring and could improve data accuracy over traditional self-reporting.

Self-report bias in survey variables: Behavioral predictors like meal planning or food sufficiency come from self-reported surveys, which may suffer from social desirability bias. Participants might overstate positive behaviors or underreport waste-prone actions (Roni A. Neff, 2015).Omitted variables and proxy features: Our model relied on observable predictors, but did not capture nuanced drivers like cultural attitudes, waste norms, or household technology use. This limitation is common in food waste studies that lack detailed consumption data (Luca Secondi, 2015).

Cluster segmentation refinement: While our five behavioral clusters provided a useful starting point, they could be refined. The segmentation was based on available behavioral and demographic features, but future work might incorporate attitudinal or psychographic variables. Reviews have shown that cluster types like "conscientious avoiders" or "carefree wasters" are common across cultures, but cluster definitions vary depending on input data (Victoria Norton, 2024). Mixed methods clustering or longitudinal segmentation could track how households evolve across life stages or economic conditions.

Policy and modeling feedback loop: The U.S. has committed to halving food waste by 2030 (EPA/USDA goal). To achieve this, we need more than awareness campaigns we need predictive tools that guide targeted interventions. For example, if meal-planning interventions lower waste for one segment, models can be retrained to reflect the shift. This iterative, data-driven refinement process aligns with best practices in sustainability and public health research.

Diet-health tradeoff caution: Finally, our model's partial dependence plot for "dietstatushh" suggests a positive association between self-reported diet quality and predicted household food waste. This trend may reflect an increased risk of spoilage when households pursue healthier diets, especially by purchasing more fresh, perishable foods. While we did not directly model nutritional quality or diet content, the inclusion of "dietstatushh" as a predictor allowed us to observe this potential tradeoff. This aligns with prior studies suggesting that health-conscious behavior, without adequate storage or planning, may inadvertently raise waste levels (Roni A. Neff, 2015). Future work should investigate how to align food waste reduction with nutrition improvement especially in low-income settings where affordability, perishability, and food security intersect. We underscore the importance of continuing to tie this work back to national policy and global sustainability goals. The United States has committed to halving food loss and waste by 2030, in line with UN Sustainable Development Goal 12.3, which aims to "halve per capita global food waste at the retail and consumer levels" and reduce food losses along production and supply chains. Achieving this target will require more than public awareness it will demand data-driven, targeted interventions informed by predictive models like those developed in this study.

For instance, if a community program targeting households classified as "Routine Shoppers with Mild Constraints" improves meal planning practices and in-store education, can we track a measurable reduction in predicted waste for that segment? Such iterative evaluation using model predictions to guide interventions, and then refining models based on observed outcomes can create a feedback loop that accelerates national progress on waste reduction.

Additionally, this research highlights the intersection between food waste and nutrition. Our findings suggest that while healthier diets (e.g., those high in fresh produce) are beneficial, they may lead to higher waste if households lack the tools to manage perishables effectively. Future research should explore how to achieve both nutrition security and food waste reduction, especially for low-income households. This calls for multidisciplinary collaboration among nutritionists, behavioral scientists, economists, and machine learning experts to design interventions that are both health-promoting and waste-conscious.

5. Conclusion

This study shows that machine learning can help us better understand and predict household food waste, offering a clearer view into which behaviors and households drive the most waste. By combining behavioral patterns with demographic and access-related data, we can move beyond broad assumptions and begin designing smarter, more targeted solutions.

At the same time, the work highlights important limitations in current data and modeling assumptions. Predictive insights are only as strong as the data they rely on, and food waste, by nature, is difficult to observe directly. Moving forward, advances in data sources such as real-time monitoring, household-level tracking tools, and digital receipts could enable even deeper, more precise models.

The path ahead is not just technical, it's practical and urgent. As policymakers and community leaders work to reduce food waste, predictive models like this one can serve as a compass. They help us prioritize resources, tailor interventions, and ultimately ensure that the food we produce nourishes people rather than filling landfills. Each step toward better data and smarter modeling brings us closer to a food system that is more efficient, equitable, and sustainable.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] (ACS), A. C. (2018). Retrieved from United states Census : https://www.census.gov/programs-surveys/acs/

[2] Aakanksha Bhatia, S. S. (2024). Identifying determinants of household food waste behavior in urban India. Cleaner Waste Systems.

[3] Agriculture, U. D. (2013). Retrieved from ers.usda.gov: https://www.ers.usda.gov/data-products/foodaps-national-household-food-acquisition-and-purchase-survey/

[4] America, F. (2023, October 25). Feeding America. Retrieved from Feeding America: https://www.feedingamerica.org/about-us/press-room/USDA-food-security-report-2022#:~:text=The%20number%20of%20people%20living,children%20from%20the%20previous%20year

[5] Antonia Di Florio, R. M. (2016). Food waste in the phase of domestic consumption: the causes and preventive actions. Rivista di Economia Agraria, 1.

[6] Barcelona, A. U. (2024, July 15). Phys.org. Retrieved from https://phys.org/news/2024-07-households-children-regularly-food-spanish.html#:~:text=70,dietary%20needs%20and%20managing

[7] Cappelletti F, P. A. (2022). Smart strategies for household food waste management. IRIS Institutional Research Information System.

[8] Change, U. C. (2024, September 30). Retrieved from UN Climate Change: https://unfccc.int/news/food-loss-and-waste-account-for-8-10-of-annual-global-greenhouse-gas-emissions-cost-usd-1-trillion

[9] Christian Bretter, K. L. (2023). Food waste interventions: Experimental evidence of the effectiveness of environmental messages. Journal of Cleaner Production , 5-6.

[10] Danyi Qi, B. E. (2016). Household Food Waste: Multivariate Regression and Principal Components Analyses of Awareness and Attitudes among U.S. Consumers. PLOS One.

[11] Fan, Y. Y. (2023). Evaluating Food Assistance Programs in the Presence of Food Waste: The Case of SNAP. Social Science Research Network .

[12] Food Waste, H. i. (2023, May 3). Retrieved from Foodfight.news: https://foodfight.news/the-future-of-composting-smart-bins-to-monitor-and-minimize-waste

[13] Gill, C. (2020, Jan 23). Penn State. Retrieved from Penn State: https://www.psu.edu/news/research/story/study-suggests-us-households-waste-nearly-third-food-they-acquire

[14] Izzy Klugman, M. S. (2024). Retrieved from Wastedfood.american.edu: https://edspace.american.edu/foodfueled/issues/volume-i/the- correlation-between-household-income-and-food-wastage-a-case-study-of- food-wastage-quantity-and-consumer-attitudes/

[15] Juliane Jörissen, C. P.-R. (2015). Food Waste Generation at Household Level: Results of a Survey among Employees of Two European Research Centers in Italy and Germany. Sustainability.

[16] László Mucha, T. O. (2025). Assumptions and perceptions of food wasting behavior and intention to reduce food waste in the case of Generation Y and Generation X. Scientific Reports.

[17] Lauria, M. (2024, March 14). One5c . Retrieved from One5c: https://one5c.com/food-waste-america-136939685/#:~:text=Together%2C%20non,7

[18] Leib, E. M. (2022, April). Retrieved from Chlpi.org: https://chlpi.org/wp-content/uploads/2022/04/2023-Farm-Bill-Food-Waste.pdf?

[19] Lisanne van Geffen, E. v. (2019). Household Food Waste—How to Avoid It? An Integrative Review. Springer Nature Link, 34.

[20] Luca Secondi, L. P. (2015). Household food waste behaviour in EU-27 countries: A multilevel analysis. Food Policy.

[21] Ludovica Principato a, *. G. (2021). The household wasteful behaviour framework: A systematic review of consumer food waste. Industrial Marketing Management .

[22] Lundeberg, S. (2023, September 13). Oregon State College of Engineering. Retrieved from Oregon State College of Engineering: https://engineering.oregonstate.edu/all-stories/smart-bin-tracks-household-food-waste#:~:text=Additionally%2C%20he%20said%2C%20the%20device,"

[23] Maggie Bain, D. S. (2024). The limitations of an informational campaign to reduce household food waste at the community scale. Cleaner Waste Systems.

[24] Maguire, K. B. (2025, April 30). Retrieved from ers.usda.gov: https://www.ers.usda.gov/privacy/ers-annual-report-to-omb-on-the-implementation-of-cipsea

[25] Miller, L. (2025, March 27). ers.usda.gov. Retrieved from https://www.ers.usda.gov/data-products/foodaps-national-household-food-acquisition-and-purchase-survey

[26] National Academies of Sciences, E. a. (2020). Retrieved from National Academies Press: https://nap.nationalacademies.org/catalog/25876/a-national-strategy-to-reduce-food-waste-at-the-consumer-level

[27] Petra Nováková, T. H. (2021). An Analysis of Food Waste in Czech Households—A Contribution to the International Reporting Effort. MDPI.

[28] Pietro Tonini, P. M. (2023). Predicting food waste in households with children: socio-economic and food-related behavior factors. PubMed Central.

[29] Ran Li, B. E. (2024). Segmenting U.S. consumers by food waste attitudes and behaviors: Opportunities for targeting reduction interventions . Sustainable Production and Consumption.

[30] ReFED. (2018). Retrieved from refed.org: https://refed.org/data-explorer/food-waste-by-sector/

[31] ReFeD. (2024, October 22). Retrieved from https://refed.org/articles/three-ways-ai-is-driving-reductions-in-food-loss-and-waste/#:~:text=Waste%20Tracking%20and%20Analytics%3A%20This,waste%20tracking%20and%20analytics%20provider

[32] ReFed. (2025, Feb). Retrieved from ReFED: https://refed.org/food-waste/the-problem/#:~:text=In%202023%2C%20the%20U,tons—goes%20to%20these%20waste%20destinations

[33] Roni A. Neff, M. L. (2015). Wasted Food: U.S. Consumers' Reported Awareness, Attitudes, and Behaviors. Plos One .

[34] Rosetta Newsome, C. G. (2014). Applications and Perceptions of Date Labeling of Food. Comprehensive Review in Food Science and Food Safety.

[35] Sahar Ahmadzadeh, T. A. (2023). A Comprehensive Review on Food Waste Reduction Based on IoT and Big Data Technologies. MDPI - Sustainability.

[36] Theofanis Zacharatos 1, 2. a. (2024). From Plate to Bin: Consumer Segments and Food Waste in Greece. MDPI.

[37] UN Climate Change. (2024, September 30). Retrieved from UN Climate Change: https://unfccc.int/news/food-loss-and-waste-account-for-8-10-of-annual-global-greenhouse-gas-emissions-cost-usd-1-trillion#:~:text=In%202022%2C%20according%20to%20UNEP%2C,also%20a%20significant%20environmental%20burden

[38] Victoria Norton, S. L. (2024). Exploring Food Waste from a Segmentation and Intervention Perspective—What Design Cues Matter? A Narrative Review. MDPI, 1.

[39] Violeta Stancu*, P. H. (2015). Determinants of consumer food waste behaviour: Two routes to food waste. Appetite, 9.

[40] Yi Yang, C. Z. (2024). Leveraging Machine Learning for Food Waste Reduction: An Analysis of Predictive Models. nternational Conference on Signal Processing and Machine Learning.