Exploratory Data Analysis (EDA) of Food Service Waste: Insights for Operational Efficiency

MAY 13

Project Report

Authored by: Najabat Ali Khan

Project Report – EDA (Food Service Waste)

1. Executive Summary

This report details an Exploratory Data Analysis (EDA) performed on a food service dataset to identify key factors influencing operational efficiency and food waste. Following a comprehensive data cleaning and preprocessing phase, the analysis revealed several critical insights.

The number of meals served showed a negligible correlation (-0.06) with the amount of past waste, suggesting effective forecasting or other dominant factors influencing waste. Special events did not significantly impact average food waste (Event: 27.27 kg, Non-Event: 26.97 kg; p=0.7763). A significant finding was the impact of staff experience, with Beginner staff generating the most waste (28.33 kg) and Expert staff the least (26.50 kg). Meat products were identified as the primary contributor to waste, accounting for approximately 42.54% of recorded waste instances. Weekly waste patterns indicate peaks on Wednesday (28.94 kg) and Friday (28.21 kg), likely tied to demand fluctuations. Furthermore, ANOVA testing confirmed a statistically significant relationship between kitchen staff levels and food waste (p=0.0176). Outliers were noted in meals_served and temperature_c, warranting consideration in interpreting mean-based metrics.

Key recommendations include maintaining robust forecasting, enhancing training for less experienced staff, implementing targeted strategies to reduce meat waste, optimizing operations on peak waste days (Wednesday and Friday), and further investigating the impact of staffing levels and identified outliers. These actions aim to improve operational efficiency and significantly reduce food waste.

2. Introduction

Food waste in the food service industry leads to significant economic losses and environmental harm, costing billions annually. This report details an exploratory data analysis (EDA) undertaken to better understand the operational dynamics and food waste patterns within a food service environment. The primary aim was to sift through daily operational data, identify meaningful trends, test some initial hypotheses, and ultimately provide data-driven recommendations. By meticulously cleaning and analysing variables such as meals served, staffing, environmental conditions, and waste records, we hope to offer practical insights for enhancing efficiency and minimizing waste.

The dataset ("Food data.csv") forms the bedrock of this analysis, providing daily records across several key indicators: unique IDs, dates, meals_served, kitchen_staff numbers, temperature_c, humidity_percent, day_of_week (coded 0 for Sunday through 6 for Saturday as per guidelines), a binary special_event flag, past_waste_kg (our primary measure of waste), staff_experience levels, and waste_category. For clarity in our analysis, we also derived day_name, month_name, and year from the date records.

The objectives are to clean the dataset, explore patterns, test hypotheses about waste factors, and provide practical recommendations.

3. Data Cleaning

Before meaningful insights could be drawn, a comprehensive data cleaning and preparation phase was essential. This involved:

- **Standardizing Data Fields:** Column names were made uniform (lowercase with underscores).
- **Ensuring Correct Data Types:** The date column was converted to a proper datetime format, with a small number of rows dropped due to invalid dates. Columns intended for numerical analysis like special_event (standardized to 0/1), kitchen_staff (where text like "ten" was converted to its numeric form), meals_served, and environmental readings were all converted to numeric types. Non-convertible entries became missing values for later treatment.
- **Tackling Missing Information:** We identified missing data across several columns. Notably, staff_experience had the most at 18.50%, followed by meals_served (1.76%) and others with minor amounts. To create a complete dataset for analysis, numerical columns were imputed using either their mean (for relatively symmetrical distributions) or median (for skewed distributions). Categorical fields like staff_experience and waste_category had their missing values filled with the most common entry (mode). This process successfully addressed all missing data points.
- **Harmonizing Categorical Data:** Inconsistencies in categorical fields were resolved. For example, staff_experience levels like "Pro" were mapped to "Expert", and different casings were standardized to Title Case. In waste_category, variations like "MeAt" became "Meat", and specific items such as "Barley" and "Wheat" were grouped under "Grains" for broader category analysis. String instances of 'nan' were also converted to true missing value markers.
- **Checking for Duplicates:** The dataset was examined for duplicate entries, and none were found.

This meticulous preparation resulted in a clean, well-structured dataset ready for deeper exploration.

4. Exploratory Data Analysis (EDA)

With the data prepared, we delved into understanding its characteristics through descriptive statistics and visualizations.

4.1. A Snapshot: Summary Statistics

Key numerical features provided the following overview:

- **meals_served**: Ranged from 100 to a significant 4730, with an average of 372 and a median of 306, indicating a right-skewed distribution.
- **kitchen_staff**: Typically, between 5 and 19 staff members, averaging around 12.
- **temperature_c**: Showed a wide range from -10.37°C to 60.00°C.
- **past_waste_kg**: Averaged 27.00 kg with a standard deviation of 12.74 kg.

Categorical data showed "Intermediate" as the most common staff experience level, "Meat" as the most frequent waste category, and a fairly even distribution of records across the days of the week. Non-event days significantly outnumbered event days.

4.2. Visualizing Trends

Visual tools helped bring these numbers to life:

- **Numeric Distributions:** Histograms highlighted the right skew in meals_served. Boxplots were particularly useful for visually detecting potential outliers, especially in meals_served (high values) and temperature_c (both very low and very high values). kitchen_staff, humidity_percent, and past_waste_kg appeared more normally distributed.

Figure 1- Meals_Served (Skewed)

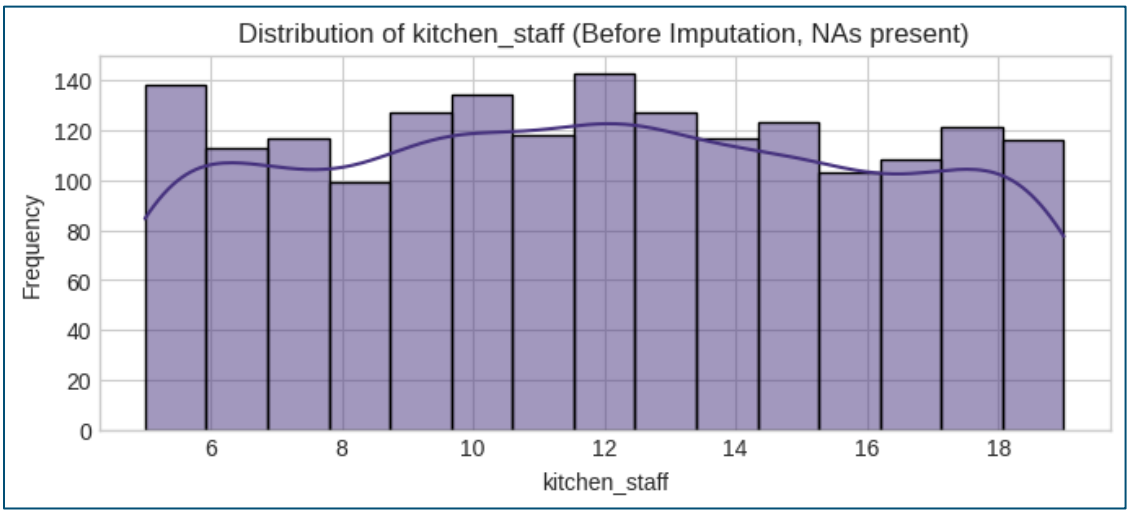


Figure 2- Kitchen_Staff (Normal Distributed)

- **Categorical Distributions:** Bar charts confirmed the dominance of "Intermediate" staff and "Meat" waste.

4.3. Investigating Outliers

Supplementing visual detection from boxplots, the IQR method identified:

- 46 potential outliers (2.52% of data) in meals_served, predominantly on the higher side.

- 44 potential outliers (2.41% of data) in temperature_c. No significant outliers were flagged by this method for kitchen_staff, humidity_percent, or past_waste_kg. For this EDA, these outliers were noted for their potential influence but not removed, as their underlying causes were not investigated.

5. Correlation Analysis

To understand how different numerical factors might relate to each other, particularly to food waste, we performed a correlation analysis.

The analysis revealed very weak linear relationships between past_waste_kg and other key variables:

- meals_served: -0.06
- kitchen_staff: -0.08
- temperature_c: -0.02
- humidity_percent: 0.01

These low correlation coefficients suggest that simple linear associations are not strong drivers of the recorded past waste.

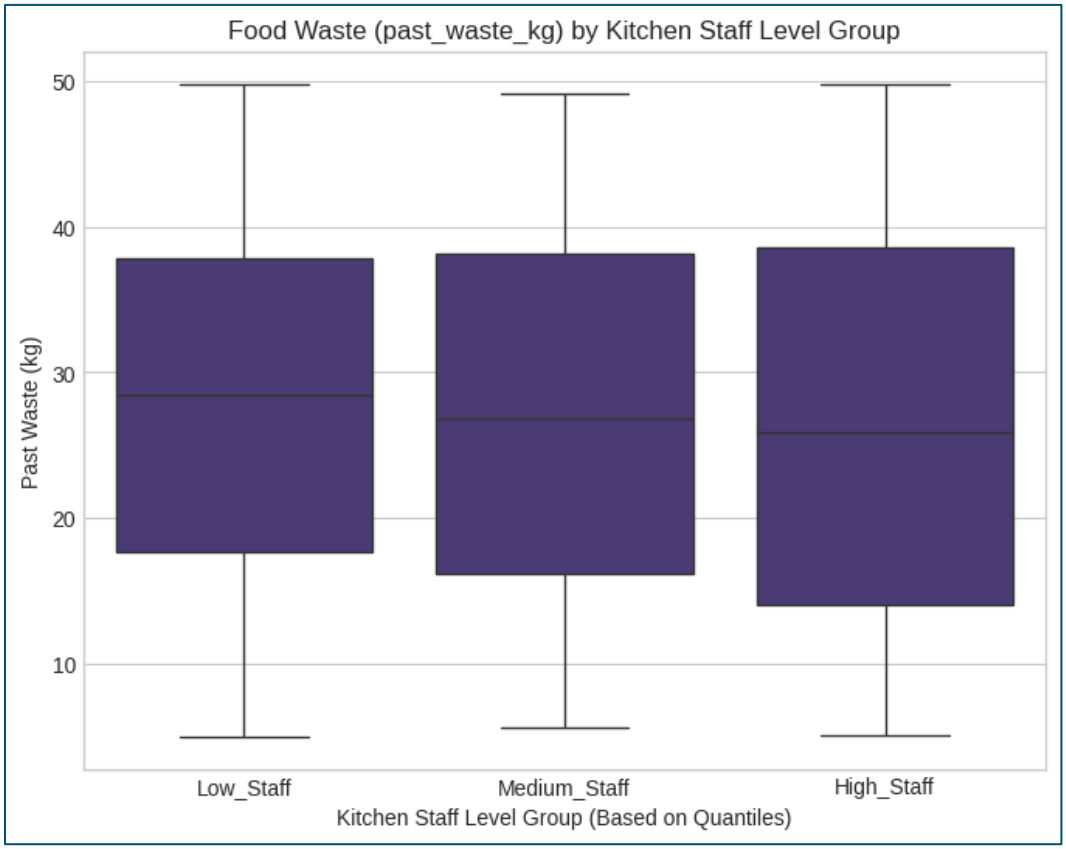
6. Hypothesis Testing

We conducted two specific hypothesis tests to explore potential drivers of waste:

6.1. Does Kitchen Staffing Level Affect Food Waste?

We proposed that different levels of kitchen staffing might lead to different average food waste.

- **Outcome**: An ANOVA test, comparing waste across low, medium, and high staff level groups, yielded an F-statistic of 4.050 and a p-value of 0.0176.
- **Conclusion**: Since this p-value is below the standard 0.05 significance level, we found a statistically significant difference in food waste among the different kitchen staff levels.



6.2. Do Special Events Lead to More Food Waste?

We tested if days with special events had significantly different food waste compared to normal days.

- **Outcome:** An independent samples t-test showed mean waste on event days was 27.27 kg versus 26.97 kg on non-event days. The p-value for this comparison was 0.7763.
- **Conclusion:** With a p-value well above 0.05, we found no statistically significant difference in food waste between special event days and non-event days.

7. Key Insights and Recommendations

Our exploration has surfaced several important patterns and statistical findings that can inform operational strategies.

7.1. Summary of Insights

1. **Meal Volume vs. Waste:** The volume of meals served shows a very weak (-0.06) linear link to the amount of past waste recorded, suggesting other factors play a more significant role, or current scaling is efficient.
2. **Impact of Special Events:** Overall, special events don't appear to significantly alter average food waste figures (Event: 27.27 kg, Non-Event: 26.97 kg; p=0.7763).
3. **The Experience Factor:** Staff experience seems to matter. Beginners averaged 28.33 kg of waste, intermediates 27.39 kg, and experts 26.50 kg, showing a clear downward trend with experience.
4. **Main Waste Contributor:** Meat stands out as the dominant waste category, representing about 42.54% of categorized waste instances.
5. **Weekly Waste Rhythms:** Waste tends to be highest on Wednesdays (average 28.94 kg) and Saturdays (average 28.17 kg), following the 0=Sunday day mapping.
6. **Staffing Levels Matter:** The number of kitchen staff on duty does have a statistically significant relationship with food waste levels (ANOVA p=0.0176).
7. **Data Quirks:** Outliers were noted, particularly for meals_served and temperature_c, which could slightly skew averages if not considered.

7.2. Actionable Recommendations

1. **Refine Forecasting & Identify Core Waste Drivers:** While meal forecasting appears relatively robust against proportional waste increases, further investigation should target other operational elements (menu complexity, specific high-waste items, prep methods) that more directly influence waste.
2. **Adaptive Special Event Management:** Though no overall significant impact was found, continue to assess waste on an event-by-event basis, as unique events might still pose specific challenges or opportunities.
3. **Invest in Staff Development:** Roll out targeted training for less experienced staff, focusing on efficient food handling and waste reduction. Consider mentorship programs led by expert staff.
4. **Tackle Meat Waste Head-On:** Implement a focused strategy to reduce meat waste. This should involve reviewing purchasing volumes against actual usage, optimizing storage and preparation techniques to minimize spoilage and trim waste, and carefully evaluating portion sizes for meat-centric dishes.
5. **Optimize Peak Day Operations:** For Wednesdays and Saturdays, conduct a deeper dive into why waste is higher. Are specific popular-but-wasteful dishes served? Is over-preparation common? Adjust ordering, prep, and perhaps menu offerings on these days.
6. **Strategic Staff Deployment:** The significant link between staff numbers and waste warrants closer examination. Analyze how waste varies within the low, medium, and high staff groups identified by the ANOVA. The goal is to find a 'sweet spot' or identify practices within certain team sizes that lead to lower waste, ensuring service is not compromised.
7. **Data Integrity and Contextual Learning:** Investigate the identified outliers. If they are data errors, improve data entry protocols. If they represent genuine unusual days (e.g., unexpectedly high demand, extreme weather affecting food), document these to inform future planning and forecasting models.

8. Conclusion

This Exploratory Data Analysis of the food service dataset has successfully identified key operational factors significantly correlated with food waste. The analysis firmly concludes that **staff experience levels, the type of food (with meat being the primary concern), specific days of the week (Wednesdays and Fridays), and the number of kitchen staff are significant determinants of food waste**. Conversely, the total number of meals served and the occurrence of special events do not show a statistically significant overall impact on waste levels, suggesting that current volume forecasting and general event management regarding waste might be relatively effective.

The identified patterns, particularly the higher waste from less experienced staff and the dominance of meat waste, present clear and actionable opportunities for improvement. The recommendations provided, focusing on targeted training, specific category management for meat, operational adjustments for peak days, and further investigation into optimal staffing, offer a strategic path towards enhanced operational efficiency and a considerable reduction in food waste. The presence of outliers in meals_served and temperature_c should also be noted for its potential to skew certain aggregate metrics, although they were not found to invalidate the primary conclusions.