EDA
PROJECT REPORT
FOOD DATA ANALYSIS

Shehla Shireen

DS & AI - 11

Objective

The goal is to analyze operational data from a food service setting to identify patterns in food waste and efficiency. The dataset includes environmental factors, staff details, and service volume, and the analysis focuses on:

- Understanding key drivers of food waste.
- Assessing operational trends (e.g., staffing, environmental conditions).
- Identifying areas for optimization in food preparation and resource allocation.

---

1. Data Cleaning and Preprocessing

Actions Taken:

- **Removed Unnecessary Columns**: Dropped ID as it did not add analytical value.
- **Type Corrections**:
  - Converted date to datetime format.
  - Cast meals_served, kitchen_staff, special_event to appropriate integer types.
- **Standardized Categorical Values**:
  - Uniform letter casing in staff_experience and waste_category.
- **Removed Duplicates**:
  - 777 exact duplicate rows removed.
  - Duplicates within groups of date and waste_category also filtered.
- **Missing Values**:
  - No column had >50% missing data.
  - For continuous columns, mean/mode-based imputations applied.
- **Outlier Treatment**:
  - Identified and removed outliers in meals_served and temperature_C using domain knowledge and IQR method.
  - Justification: No strong correlation with special events for meals_served → outliers deemed invalid. Temperature outliers were impractical and did not synchronize with rest of the dataset's temperature range.

---

2. Data Overview & Summary

Column	Type	Notes
meals_served	Continuous	Avg: ~298; Range: 100–489; Left-skewed; strong operational indicator
kitchen_staff	Categorical	Range: 5–19; important for workload/waste mapping
temperature_C	Continuous	Cleaned outliers; normal distribution with some tails

Column	Type	Notes
humidity_percent	Continuous	Normal distribution; filled missing with mean
past_waste_kg	Continuous	Avg: ~26.7 kg; left-skewed; core variable for optimization
special_event	Binary	Highly imbalanced (mostly 0s); weak correlation with meals served
staff_experience	Categorical	4 categories: Beginner, Intermediate, Expert, Pro
waste_category	Categorical	Various types (e.g., dairy, meat)

---

3. Key Visual Insights

Distribution Plots (Histograms)

- **Meals Served**: Slightly left-skewed, large variation.
- **Temperature**: Normal distribution; a few temperature outliers removed.
- **Humidity**: Close to normal distribution.
- **Past Waste (kg)**: Left-skewed; moderate variation.

Boxplots

- Confirmed presence of:
  - Moderate spread in past_waste_kg.
  - Slight right-skew in humidity_percent.

---

4. Statistical & Correlation Insights

Correlation Matrix (Pearson)

- temperature_C vs. past_waste_kg: **-0.013** → negligible correlation.
- meals_served shows **no strong correlation** with past_waste_kg.
- humidity_percent has **weak to no correlation** with waste.

---

5. Staff Experience Impact

Summary Stats of **past_waste_kg** by **staff_experience**:

Experience	Count	Mean Waste (kg)	Std Dev
Beginner	174	**28.96**	12.77
Intermediate	515	26.69	12.96

Experience	Count	Mean Waste (kg)	Std Dev
Expert	163	25.84	12.25
Pro	19	**25.31**	10.44

- **Insight**: Waste **decreases with higher experience**.
- **Actionable Point**: Staff training or experience balancing may improve waste efficiency.

---

6. Special Events Insight

- Special events **do not significantly influence**:
  - Number of meals served
  - Amount of food waste
- Events are rare (imbalanced), making it statistically difficult to derive trends.

---

7. Detailed Insights

Meals Served vs. Food Waste

Plot Type: Scatter Plot

- Insight:
  There is a weak correlation between meals served and past waste, with values between -0.02 and 0.07. This indicates that the volume of meals served does not directly influence food waste. Both high and low waste can occur regardless of the number of meals served. This suggests that other factors, such as kitchen operations or waste management practices, play a more significant role than meal volume.

---

2. Temperature vs. Food Waste

Plot Type: Scatter Plot

- Insight:
  The correlation between temperature (°C) and past waste is extremely low (-0.01), suggesting that temperature does not significantly impact food waste. Despite temperature fluctuations, waste levels remain relatively unaffected.

However, temperature extremes (e.g., below 0°C or above 40°C) could still pose a risk for food safety and spoilage, warranting further investigation in specific operational contexts.

---

3. Staff Experience vs. Food Waste

Plot Type: Box Plot

- Insight:
  Beginner staff members tend to have a higher average food waste (28.96 kg), with a wider range of waste values. Intermediate and expert staff have lower average waste (26.69 kg and 25.84 kg), indicating that more experienced staff tend to produce less waste. However, the differences are small, and the waste levels across all experience groups have a substantial spread, suggesting that experience has a minimal impact on waste reduction.

---

4. Waste Category Analysis

Plot Type: Box Plot or Bar Plot

- Insight:
  Meat is the most wasted category (360 kg), significantly outpacing grains, dairy, and vegetables (160 kg each). This indicates that the focus should be on managing meat waste more effectively. Barley and wheat have minimal waste (around 15 kg each), showing less concern for these categories. However, the variability within the waste amounts suggests that waste management practices could be improved, particularly in meat-related categories.

---

5. Meals Served vs. Kitchen Staff Size

Plot Type: Line Plot or Scatter Plot

- Insight:
  Meals served generally increase with the number of kitchen staff,

particularly between 6 and 16 staff members. Within this range, meal service is most efficient, with fewer variances. In contrast, staff sizes of 5 or 9 tend to result in fewer meals served with higher variability, indicating potential inefficiencies at these staff levels. Therefore, the optimal staff size for maximizing meal capacity appears to be around 6-16 members.

---

6. Special Event vs. Food Waste

Plot Type: Box Plot

- Insight:
  Special events cause a slight increase in meals served but also lead to a small uptick in food waste. Despite this, the overall pattern and variability are similar to non-special event days. Special events appear to have a minimal impact on waste management strategies, but further exploration of their specific operational context may reveal more significant effects.

---

7. Humidity vs. Food Waste

Plot Type: Scatter Plot or Correlation Matrix

- Insight:
  Humidity percent has very low correlations with all variables, including food waste. While humidity could affect food spoilage and storage conditions, its effect is not statistically significant within the context of this dataset. This suggests that humidity management may not be a primary factor in reducing food waste under normal operational conditions.

---

8. Correlation Insights:

Plot Type: Heatmap or Correlation Matrix

- Insight:
  The correlation analysis revealed weak or no significant relationships between key variables, such as meals served, special events, humidity, and

past waste. This suggests that most operations and outcomes, like meal service or waste levels, are not influenced by day of the week or special events. For example, no meaningful correlation exists between meal service and waste, which implies that more meals do not inherently cause more waste. Similarly, past waste shows only weak correlations with staff levels, temperature, or humidity, indicating that the waste levels are not easily predictable by these factors alone.

---

9. Waste Distribution by Staff Experience

Plot Type: Box Plot

- Insight:
  The distribution of waste by staff experience shows that beginners tend to produce more waste on average (28.96 kg), but the variation in waste is quite broad. Expert staff show lower average waste (25.84 kg) and exhibit more consistency in waste levels. This suggests that while experience might reduce variability in waste, its impact on the total quantity of waste is minimal. More training or refined waste reduction strategies might be more impactful than relying solely on staff experience.

---

10. Waste Category Variability

Plot Type: Box Plot or Histogram

- Insight:
  The variation in food waste is most pronounced in meat, dairy, and vegetables, whereas barley and wheat exhibit relatively little variation due to fewer data points. The wide range of waste within categories like meat (with some extreme waste values approaching 50 kg) indicates that there are inefficiencies that need to be addressed, particularly in meat-related waste. Targeted strategies for reducing waste in these categories could lead to significant improvements.

8. Final Recommendations

Operational Optimization

- Focus training and staffing strategy to **prioritize experienced staff** during high service days.
- Use predictive scheduling based on **meals served trends** to optimize staffing.

Waste Management

- Track waste categories across time to target high-waste groups (e.g., meat, dairy).
- Establish baseline temperature/humidity conditions to ensure environment stability, though current correlation is weak.

Data Enhancements

- Include **real-time waste logs** or **meal type data** in future datasets.
- Consider **time-of-day granularity** for better operational insights.

Conclusion

The EDA revealed key patterns in food waste generation and operational behavior. Despite weak environmental correlations, **staff experience and service volume** emerged as critical dimensions to address food waste and improve efficiency. This analysis provides a strong base for implementing data-driven process improvements in kitchen operations.

To minimize food waste and optimize meal service efficiency, it is recommended to focus on improving waste management strategies, particularly for meat, which is the most wasted category. While staff experience and temperature have minimal direct impact on waste levels, operational practices such as staff training, efficient food handling, and waste protocols should be prioritized. Maintaining an optimal kitchen staff size of 6 to 16 members is crucial for maximizing meal service capacity and minimizing waste. Special events, though infrequent, show a slight increase in both meals served and waste, so it's important to plan waste reduction efforts accordingly. Given the operational consistency observed, a more targeted approach to waste management, focusing on specific food categories and improving staff efficiency, will likely yield the most substantial reductions in food waste. Ultimately, leveraging data to monitor and adjust these factors in real-time can lead to more sustainable and cost-effective kitchen operations.