ATOMCAMP

PROJECT 02: EDA on Food Service Data

1. Introduction

1.1 Brief Overview of the Objective of the Project

The objective of this project is to conduct an exploratory data analysis (EDA) on a food service dataset. The primary goals are to understand the characteristics of key variables, identify potential factors influencing food waste, and provide actionable recommendations for operational improvements, particularly in food waste reduction, staffing, and event management.

1.2 Summary of the Dataset and Key Variables

The dataset contains information on daily operations in a food service establishment. Key variables include:

- date: Date of the observation.
- meals_served: Number of meals served.
- kitchen_staff: Number of kitchen staff on duty.
- temperature_C: Temperature in Celsius.
- humidity_percent: Humidity percentage.
- past_waste_kg: Amount of food waste in kilograms.
- staff_experience: Kitchen staff experience level (e.g., Beginner, Intermediate, Expert, Pro).
- waste_category: Category of food waste (e.g., Wheat, Vegetables, Dairy).
- special_event: Indicator for whether a special event occurred (1 = Yes, 0 = No).

2. Data Cleaning

2.1 Description of Steps Taken to Handle Missing Values, Duplicates, and Data Types

Due to the nature of the original data source, significant initial cleaning was required. The following steps were taken:

- Checking for Unique Values to Handle Missing Data
- Fixing Inconsistent Categorical & Numerical Values
- Assigning Correct Data Types
- Data was checked for Duplicate rows to avoid skewing the analysis.
- Converted the data to correct data types like date, kitchen_staff, and special_event
- Categorical columns were inspected for inconsistency, like spellings, and different formats for one category.

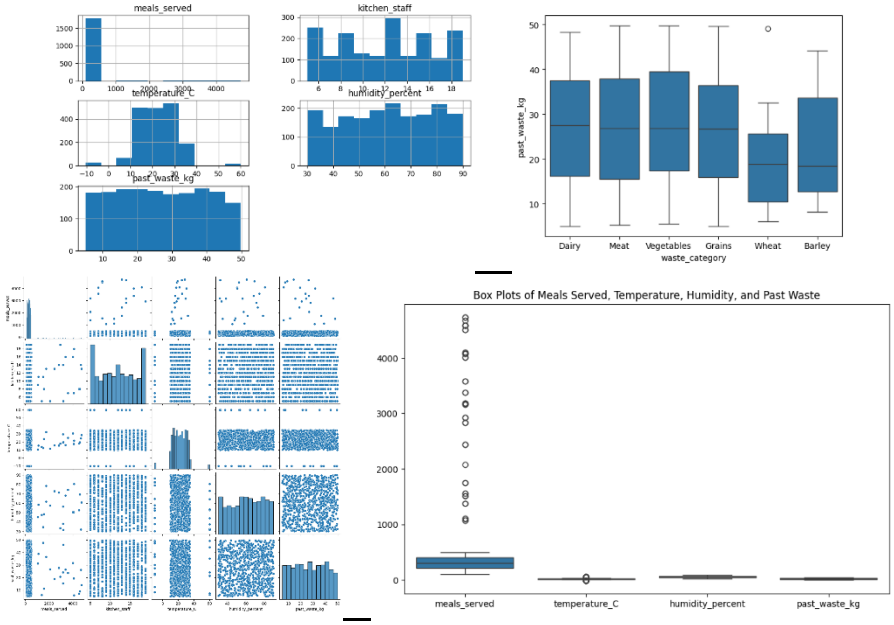
3. Exploratory Data Analysis (EDA)

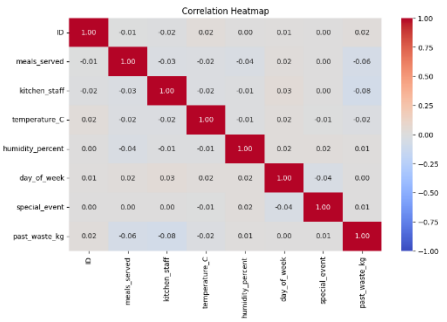
3.1 Summary Statistics

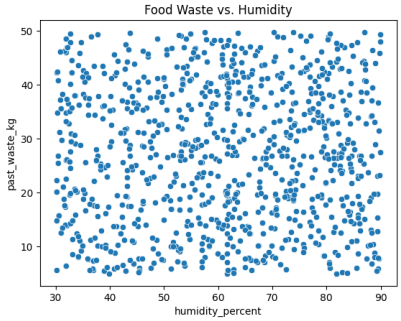
- ID: This column is the identifier.
- Meals_served: The mean (372.33) is greater than the median (306.00), indicating a right-skewed distribution. This confirms that most days have a moderate number of meals served, with occasional days having a much higher volume (up to 4730). The standard deviation (490.51) is also high, further suggesting significant variability.
- Kitchen_staff: The mean and median are very close, suggesting a relatively symmetric distribution. The values range from 5 to 19, indicating the typical range of kitchen staff numbers. The standard deviation (4.27) shows moderate variability.
- Temperature_c: The mean and median are very close, indicating a roughly symmetric distribution. The temperature ranges from -10.37°C to 60.00°C, suggesting data from a location with significant temperature variation or potential outliers. The standard deviation (8.92) shows moderate variability.

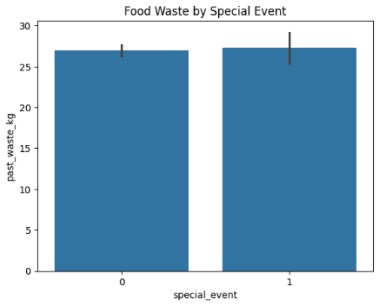
- Humidity_percent: The mean and median are close, indicating a roughly symmetric distribution. The humidity ranges from 30.12% to 89.98%, showing a wide range of humidity conditions. The standard deviation (17.25) suggests moderate variability.
- Day_of_week: This likely represents days of the week (e.g., 0 = Sunday, 6 = Saturday). The mean and median are close, suggesting a relatively uniform distribution across the week.
- Special_event: This is a binary variable (0 or 1). The mean of 0.09 indicates that special events are relatively infrequent (9% of the days).
- Past_waste_kg: The mean and median are very close, indicating a roughly symmetric distribution. The waste ranges from 5.01 kg to 49.80 kg, with a standard deviation of 12.74 kg.

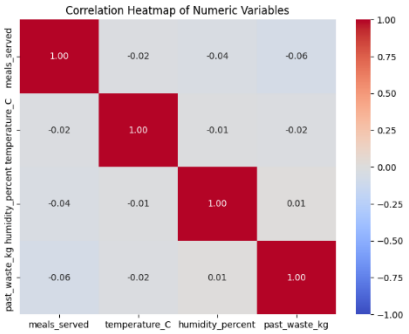
3.2 Visualizations

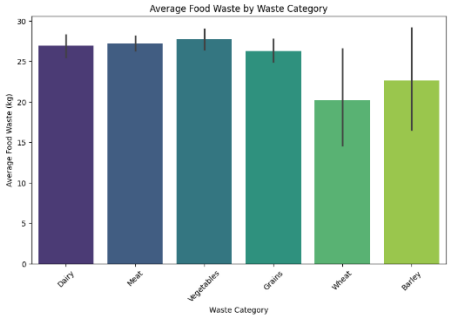












3.3 Key Patterns or Trends Observed

- Food waste varies significantly across different food categories.
- There is no strong linear correlation between food waste and operational/environmental factors like the number of meals served, kitchen staff, temperature, or humidity.

- Special events may contribute to a slight increase in food waste.
- Staff experience, as broadly categorized, does not appear to have a strong direct impact on the amount of food waste.

4. Correlation Analysis

4.1 Heatmap or Table Showing Relationships Between Numeric Variables

The correlation heatmaps visualized the relationships between the numerical variables: meals_served, kitchen_staff, temperature_C, humidity_percent, and past_waste_kg.

4.2 Interpretation of Key Correlations

The analysis revealed weak linear correlations between most numerical variables. Specifically, past_waste_kg showed little to no linear correlation with meals_served, kitchen_staff, temperature_C, and humidity_percent. This suggests that food waste is likely influenced by a combination of factors, including non-numerical variables like waste_category, or by non-linear relationships.

5. Key Insights and Recommendations

5.1 Operational or Strategic Insights

- Food waste is strongly influenced by the type of food.
- Simple operational factors (staffing levels) and environmental conditions (temperature, humidity) do not appear to be primary drivers of food waste.
- Special events may contribute to increased waste.

5.2 Recommendations for Food Waste Reduction, Staffing, or Event Management

- Food Waste Reduction:

        - Implement targeted waste reduction strategies for high-waste categories like Vegetables and Dairy.
        - Analyze the causes of waste in these categories (e.g., spoilage, portion control, preparation methods).
        - Optimize ordering, inventory management, portion sizes, food handling, and storage practices.
    - Staffing:
        - While staff experience level (as categorized) does not show a strong correlation with waste, continue to emphasize staff training on waste reduction best practices.
        - Focus training on proper food handling, portion control, and waste tracking.
    - Event Management:
        - Develop specific protocols for food planning and preparation during special events.
        - Improve demand forecasting for events.
        - Consider offering limited menus during events to reduce complexity and waste.

6. Key Insights and Recommendations

6.1 Brief Summary of Findings

The EDA revealed that food waste in the analyzed food service establishment is most strongly related to the type of food. Operational and environmental factors showed weak correlations with waste. Special events may contribute to a small increase in waste. Special events may contribute to increased waste.

6.2 Limitations and Suggestions for Further Analysis

- Causality: Correlation does not imply causation. Further research is needed to establish causal relationships between variables.
- Granularity: More granular data (e.g., detailed staff performance metrics) could provide deeper insights.
- Time Series Analysis: A time series analysis could reveal trends and seasonality in food waste.
- Performing formal hypothesis testing.
- Building predictive models for food waste.
