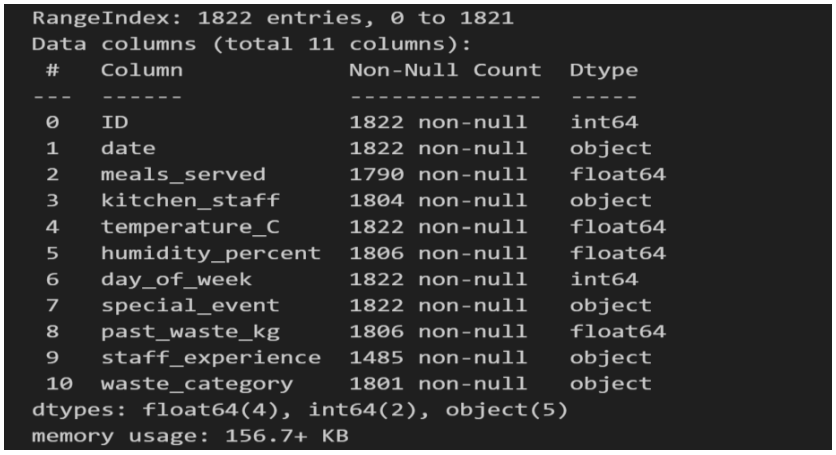
Food Service Data (EDA Report)

1. Introduction

This report provides a comprehensive analysis of data collected from a food service operation, with the primary goal of identifying patterns and relationships that can guide effective strategies for reducing food waste. By examining key operational factors such as staff experience, meal volumes, environmental conditions, and event occurrences, the analysis aims to uncover actionable insights to improve efficiency and sustainability.

The dataset contains 1,822 records with 11 features, including:

- **ID:** Unique identifier for each entry
- **meals_served:** Number of meals served on the recorded day
- **Date:** Date of the observation
- **temperature_C:** Recorded temperature in Celsius
- **day_of_week:** Numeric value representing the day of the week (e.g., Sunday = 0)
- **kitchen_staff:** Number of kitchen staff present
- **humidity_percent:** Recorded humidity percentage
- **waste_category:** Type of food waste (e.g., dairy, meat)
- **staff_experience:** Experience level of kitchen staff (e.g., Beginner, Intermediate)
- **past_waste_kg:** Food waste recorded on previous days
- **food_waste:** Food waste recorded on the current day
- **Special_event:** Binary indicator denoting if a special event took place



```
RangeIndex: 1822 entries, 0 to 1821
Data columns (total 11 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   ID                1822 non-null   int64
 1   date              1822 non-null   object
 2   meals_served      1790 non-null   float64
 3   kitchen_staff     1804 non-null   object
 4   temperature_C     1822 non-null   float64
 5   humidity_percent  1806 non-null   float64
 6   day_of_week       1822 non-null   int64
 7   special_event     1822 non-null   object
 8   past_waste_kg     1806 non-null   float64
 9   staff_experience  1485 non-null   object
 10  waste_category    1801 non-null   object
dtypes: float64(4), int64(2), object(5)
memory usage: 156.7+ KB
```

The analysis explores data quality, trends, statistical relationships, and provides actionable recommendations.

Data Cleaning and Preprocessing

- To ensure the dataset was accurate, consistent, and suitable for analysis, the following preprocessing steps were performed:
- **Irrelevant Column Removal**
- The ID column was removed as it held no analytical significance.
- **Date Parsing and Sorting**
- The Date column was converted to datetime format and the dataset was sorted chronologically to maintain temporal integrity.
- **Handling Missing Values**
    - For **numerical features** (past_waste_kg, humidity_percent, meals_served, and kitchen_staff), missing values were filled using **forward and backward fill methods** within the same date group. This was based on the observation that these values remained constant across a given date.
    - For **categorical variables** such as waste_category and staff_experience, missing values were imputed using the **mode** to preserve the most common category.
- **Data Type Correction and Consistency Adjustments**
    - Textual representations of numbers such as 'ten', 'eleven', and 'one' in the kitchen_staff column were replaced with their corresponding numeric values (10, 11, and 1), and the column was then converted to an integer data type.
    - In the waste_category column, entries like 'barley' and 'wheat' were standardized to 'grain' to maintain categorical consistency.
- **Outlier Detection and Treatment**
- Outliers identified in the meals_served and temperature_C columns were handled by **clipping values** beyond the interquartile range limits, as visualized through boxplots.
- **Categorical Normalization**
- Inconsistencies in categorical labels were corrected. For instance:
    - 'MeAt' was standardized to 'Meat'
    - Variations of 'Intermediate' in staff_experience were unified under a consistent label
- **Duplicate Handling**
- Duplicate rows were identified and removed both **before and after missing value imputation** to ensure data integrity and avoid analytical bias.

Summary Statistics

Before Cleaning

	ID	meals_served	temperature_C	humidity_percent	day_of_week	past_waste_kg
count	1822.000000	1790.000000	1822.000000	1806.000000	1822.00000	1806.000000
mean	910.500000	373.512849	22.189280	60.791257	3.01427	26.997534
std	526.110413	494.791973	8.919939	17.326232	2.00899	12.791891
min	0.000000	100.000000	-10.372207	30.121111	0.00000	5.008394
25%	455.250000	211.250000	15.684259	46.035158	1.00000	15.990760
50%	910.500000	306.000000	22.115040	61.634935	3.00000	26.832569
75%	1365.750000	407.000000	28.807494	75.789317	5.00000	38.116308
max	1821.000000	4730.000000	60.000000	89.982828	6.00000	49.803703

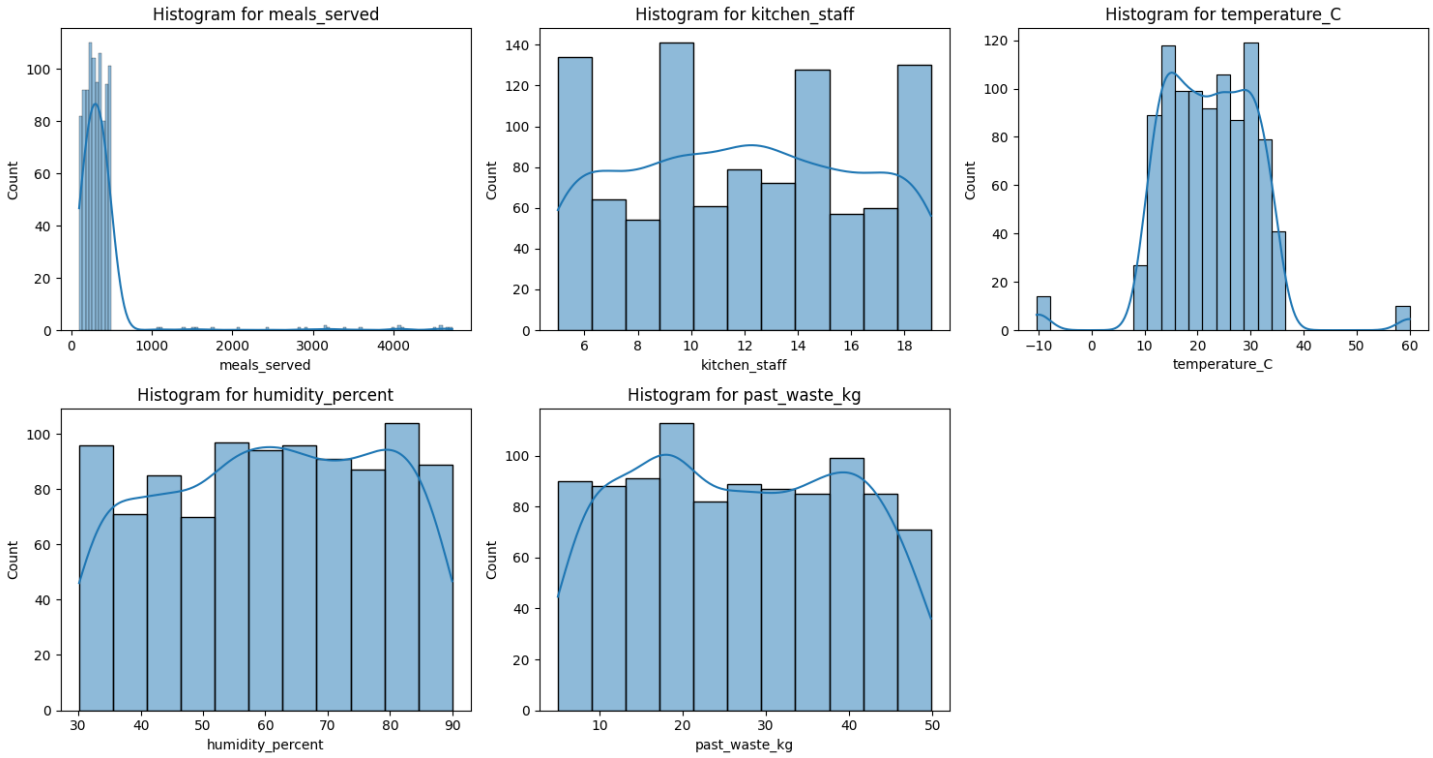
After Cleaning

	meals_served	kitchen_staff	temperature_C	humidity_percent	day_of_week	special_event	past_waste_kg
count	953.000000	953.000000	953.000000	953.000000	953.000000	953.000000	953.000000
mean	372.129066	11.912907	22.103886	60.690869	2.989507	0.088143	26.956453
std	492.424122	4.309674	9.004274	17.299660	2.010449	0.283651	12.687375
min	100.000000	5.000000	-10.372207	30.121111	0.000000	0.000000	5.008394
25%	210.000000	8.000000	15.638504	46.035158	1.000000	0.000000	16.290064
50%	304.000000	12.000000	21.997427	61.548756	3.000000	0.000000	26.541719
75%	409.000000	16.000000	28.732093	75.714409	5.000000	0.000000	37.957480
max	4730.000000	19.000000	60.000000	89.982828	6.000000	1.000000	49.803703

Visualizations

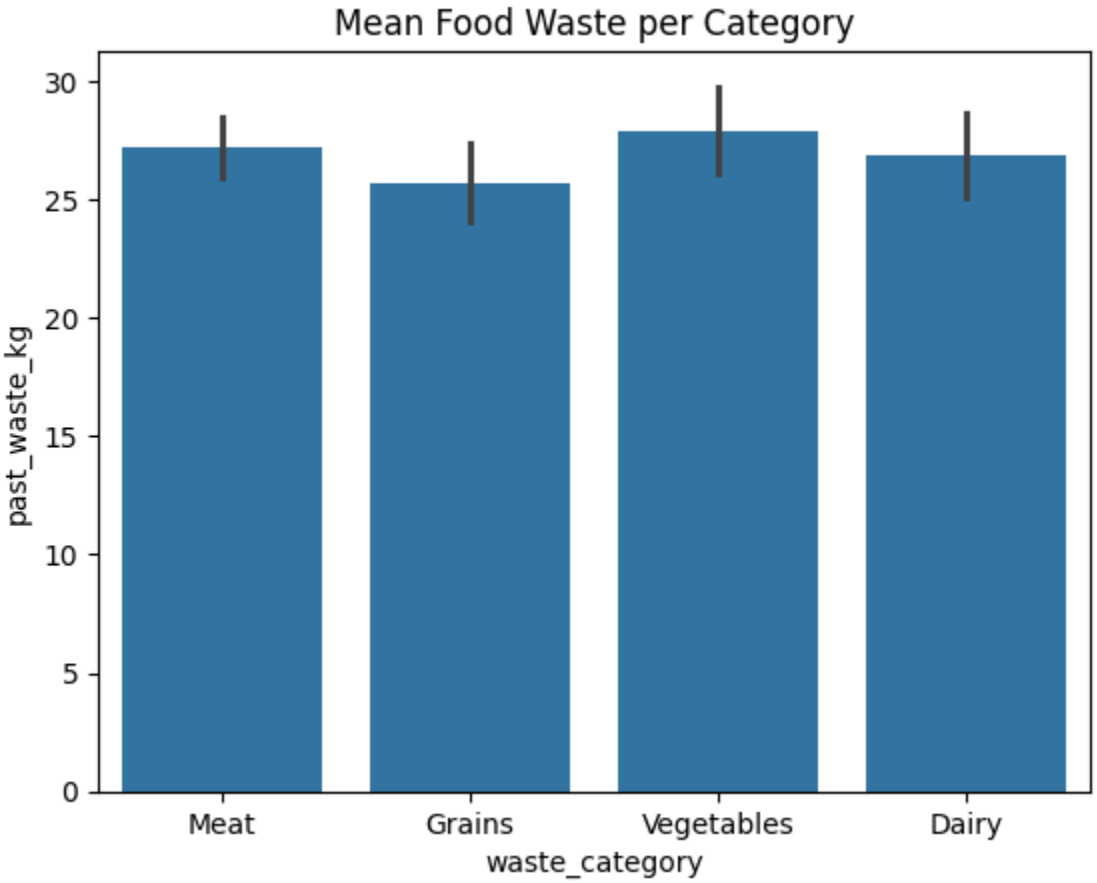
Univariate Analysis:

For the univariate analysis of continuous numerical data, a histogram plot was used. These graphs displayed the distribution and skewness of numerical continuous values in the dataset.



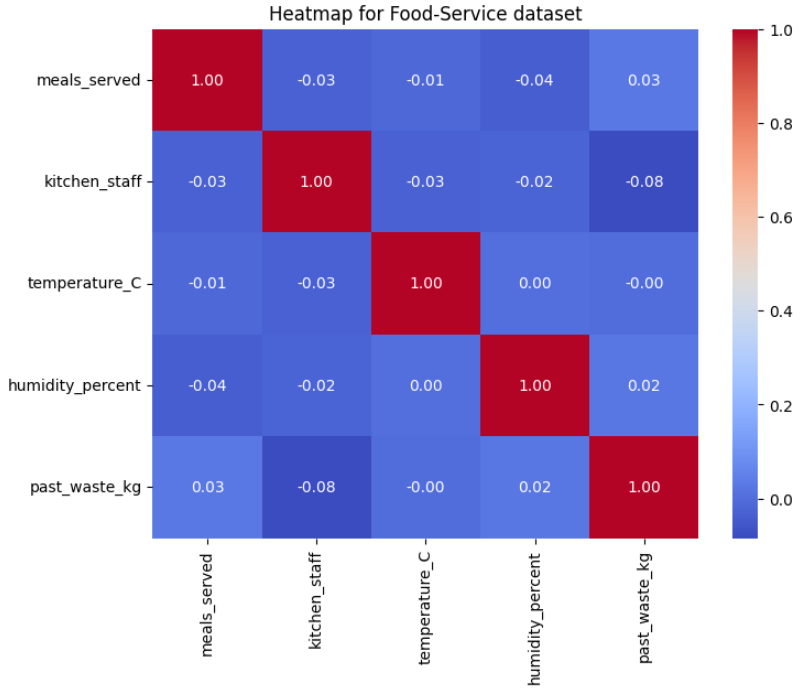
Five histogram plots were created to examine the distribution of the following features: kitchen staff, meals served, temperature, humidity percentage, and food waste (in kg). Among these, only the "meals served" feature displayed a right-skewed distribution, indicating a concentration of lower values with a long tail towards the higher end. In contrast, the distributions of kitchen staff, temperature, humidity percentage, and food waste appeared to be approximately normal, showing a more symmetric spread around their central values.

The bar chart displays food waste in kilograms across different staff experience levels—Beginner, Intermediate, Pro, and Expert—with an additional hue distinguishing between event and non-event days. On event days, Pro-level staff recorded the highest food waste, reaching approximately 40 kg, whereas Beginners contributed the least, with around 26 kg of waste. In contrast, on non-event days, the pattern shifts; Beginners reported the highest food waste at approximately 29 kg, while Pro-level staff generated the lowest, with about 25 kg. Intermediate and Expert levels show values falling between these extremes for both event and non-event days.



The bar graph illustrates the amount of food wasted, measured in kilograms, across different categories including grains, dairy, meat, and vegetables. Among these, vegetables account for the highest level of waste, reaching approximately 28 **kg**, suggesting they are more perishable or over-purchased. Meat follows closely behind at around 27 **kg**, indicating a significant portion of staple foods being discarded. Dairy products and grains show relatively lower levels of waste, at approximately **26 kg** and **25 kg** respectively, which may reflect their higher cost or shorter shelf life prompting more cautious consumption. The graph highlights the need for better storage, planning, and consumption habits, particularly in the case of vegetables and meat, to reduce overall food waste.

Correlation Analysis:



The heatmap indicates that there were no strong correlations among the variables. Environmental factors such as temperature and humidity did not influence the amount of food wasted or the meals served.

Insights

- There was **no significant correlation** between environmental factors (temperature, humidity) and food waste.
- The number of meals served also **did not correlate strongly** with food waste, implying **inefficiencies are not driven by volume alone**.
- Among food categories, **grains and vegetables** exhibited the **highest levels of waste**, indicating potential over-preparation or storage issues.
- **Dairy and meat** showed relatively lower waste quantities, possibly due to stricter handling or smaller batch sizes.
- On **event days**, *Pro-level staff* were responsible for the **highest food waste (~40 kg)**, while *Beginner-level staff* wasted the least (~26 kg).
- On **non-event days**, the pattern reversed: *Beginner staff* wasted the most (~29 kg), and *Pro staff* wasted the least (~25 kg).
- This inconsistency suggests a **context-dependent efficiency** among staff types.

Recommendations

- Implement tighter inventory and portion control for **grains and vegetables**, which showed the highest levels of waste.
- Allocate **Pro and Expert staff on non-event days** where they have shown higher efficiency.
- Assign **Beginners or Intermediate staff under supervision on event days**, as Pro staff tended to waste more under pressure.
- Conduct **post-event reviews** with kitchen staff to assess what worked and what led to excess waste, fostering continuous improvement.

Conclusion

This report analyzed key variables influencing food waste in a food service operation to identify actionable insights for reducing inefficiencies. Through exploration of staff experience, event types, and operational metrics, it was found that staff allocation plays a significant role in waste patterns—particularly with Pro staff contributing more waste on event days and Beginners on regular days. Environmental conditions such as temperature and humidity showed no significant impact on food waste, highlighting that internal operations are the primary drivers of inefficiency. Overall, the findings emphasize the need for more strategic staff deployment, refined preparation protocols, and stronger monitoring practices to enhance operational efficiency and reduce waste.

For further analysis, it is recommended to conduct a time-series analysis to identify seasonal or weekly trends in food waste, investigate the impact of specific event types on waste generation to tailor management strategies more effectively, and perform a detailed study of staff performance over time to uncover training needs and recognize best practices. These additional analyses will help deepen understanding and support more targeted waste reduction initiatives.