atomcamp

Project Report

on

Exploratory Data Analysis (EDA) on

Food Service Data

Muhammad Azam Sarwar

atomcamp

Table of Contents

1. Introduction 3
2. Dataset Overview 3
3. Data Cleaning 3
4. Exploratory Data Analysis (EDA) 3
Summary Statistics: 3
Visualizations: 4
5. Correlation Analysis 6
6. Hypothesis Testing 7
Hypothesis 1: Staff size affects food waste 7
Hypothesis 2: Special events increase waste 7
7. Key Insights & Recommendations 7
8. Conclusion 7

1. Introduction

The objective of this project is to analyse a food service dataset to uncover insights into operational efficiency and food waste management. Key variables include meals served, kitchen staff, temperature, humidity, and food waste. The goal is to explore the data, clean it, visualize patterns, test hypotheses, and provide actionable recommendations.

2. Dataset Overview

The dataset contains 1,822 records and 11 features:

- **ID**: Unique identifier
- **date**: Date of observation
- **meals_served**: Meals served per day
- **kitchen_staff**: Staff count per day
- **temperature_C**: Daily temperature (°C)
- **humidity_percent**: Daily humidity (%)
- **day_of_week**: 0 (Sunday) to 6 (Saturday)
- **special_event**: 1 = event day, 0 = normal day
- **past_waste_kg**: Food waste from previous days
- **staff_experience**: e.g., beginner, intermediate
- **waste_category**: Category of waste (e.g., dairy, meat)

3. Data Cleaning

- Converted date to datetime format
- Changed kitchen_staff and special_event to numeric
- Standardized categorical variables to lowercase
- Missing values handled:
    - Dropped rows missing critical values (meals_served, kitchen_staff, past_waste_kg)
    - Imputed humidity_percent with median, special_event with mode
    - Categorical nulls filled with 'unknown'
- No duplicate rows found

4. Exploratory Data Analysis (EDA)

Summary Statistics:

- Numerical features showed reasonable distributions.

- meals_served and past_waste_kg showed variance across the dataset.

	ID	date	meals_served	kitchen_staff	temperature_C	humidity_percent	day_of_week	special_event	past_waste_kg
count	1822.000000	1822	1822.000000	1822.000000	1822.000000	1822.000000	1822.00000	1822.000000	1822.000000
mean	910.500000	2023-04-22 09:30:37.541163520	372.327113	11.905598	22.189280	60.791257	3.01427	0.085620	26.997534
min	0.000000	2022-01-01 00:00:00	100.000000	5.000000	-10.372207	30.121111	0.00000	0.000000	5.008394
25%	455.250000	2022-07-25 06:00:00	212.250000	8.000000	15.684259	46.137537	1.00000	0.000000	16.148956
50%	910.500000	2023-04-23 00:00:00	306.000000	12.000000	22.115040	61.514385	3.00000	0.000000	26.997534
75%	1365.750000	2024-01-07 18:00:00	405.750000	15.000000	28.807494	75.755784	5.00000	0.000000	37.978663
max	1821.000000	2024-09-26 00:00:00	4730.000000	19.000000	60.000000	89.982828	6.00000	1.000000	49.803703
std	526.110413	NaN	490.505492	4.270492	8.919939	17.249947	2.00899	0.279879	12.735569

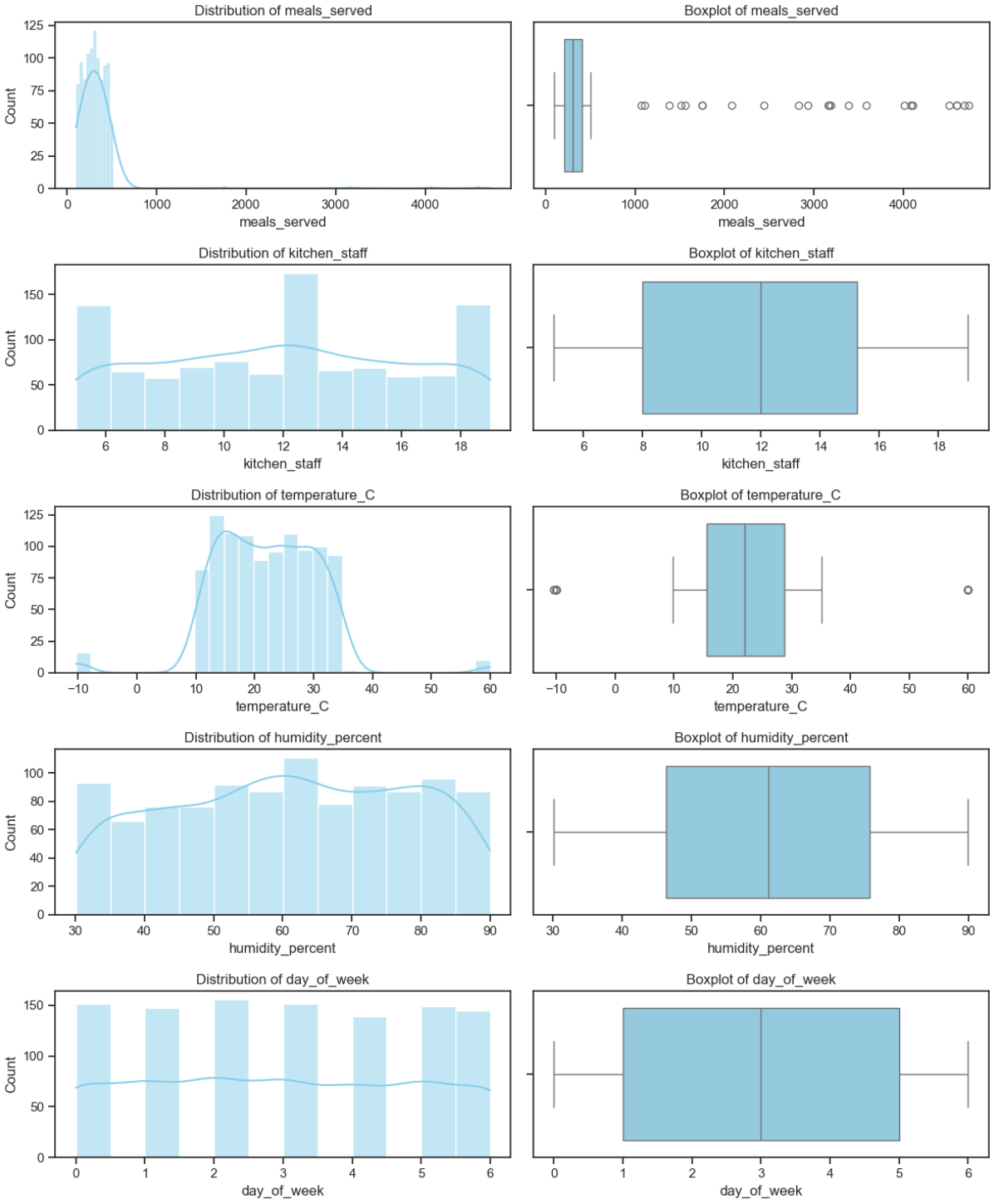
	ID	date	meals_served	kitchen_staff	temperature_C	humidity_percent	day_of_week	special_event	past_waste_kg	staff_experience	waste_category
129	129	2022-11-16	2930.0	19	17.929893	33.464463	2	1	20.743447	Intermediate	Meat
176	176	2023-09-15	3160.0	14	29.529261	38.219944	4	1	24.266983	Intermediate	Meat
1040	1040	2022-11-16	2930.0	19	17.929893	33.464463	2	1	20.743447	Intermediate	Meat
1087	1087	2023-09-15	3160.0	14	29.529261	38.219944	4	1	24.266983	Intermediate	Meat

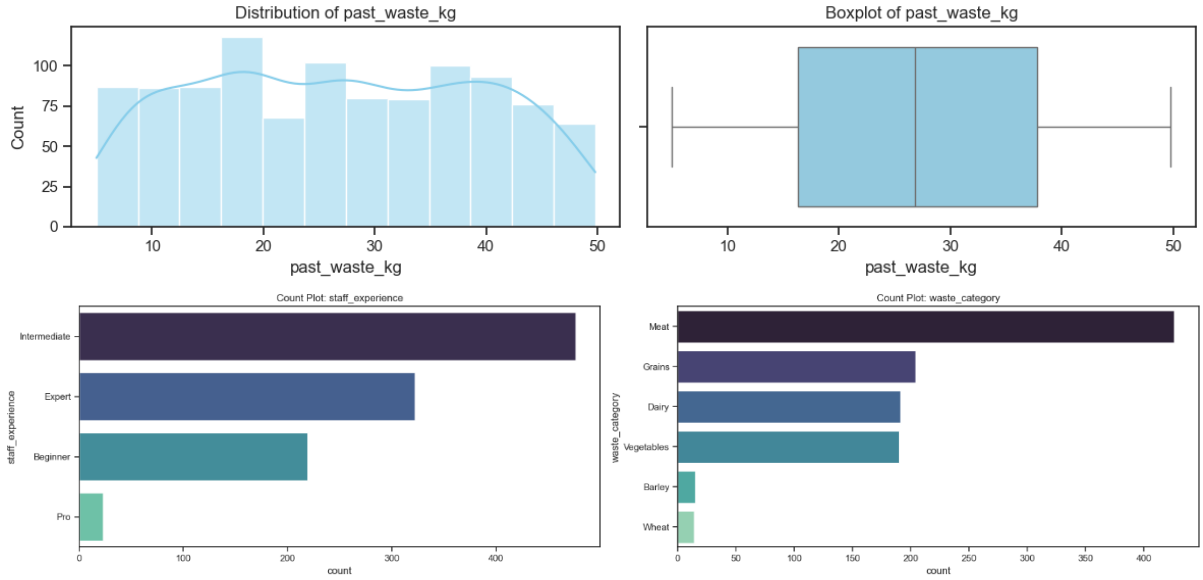
	staff_experience	waste_category
count	1040	1040
unique	4	6
top	Intermediate	Meat
freq	476	426

There were approximately 800 duplicated rows which were removed!!!

Visualizations:

- **Histograms** for all numeric features
- **Boxplots** identified outliers in past_waste_kg
- **Bar plots** for staff_experience and waste_category revealed most common categories


Distribution of meals_served
Count
meals_served
Boxplot of meals_served
meals_served
Distribution of kitchen_staff
Count
kitchen_staff
Boxplot of kitchen_staff
kitchen_staff
Distribution of temperature_C
Count
temperature_C
Boxplot of temperature_C
temperature_C
Distribution of humidity_percent
Count
humidity_percent
Boxplot of humidity_percent
humidity_percent
Distribution of day_of_week
Count
day_of_week
Boxplot of day_of_week
day_of_week

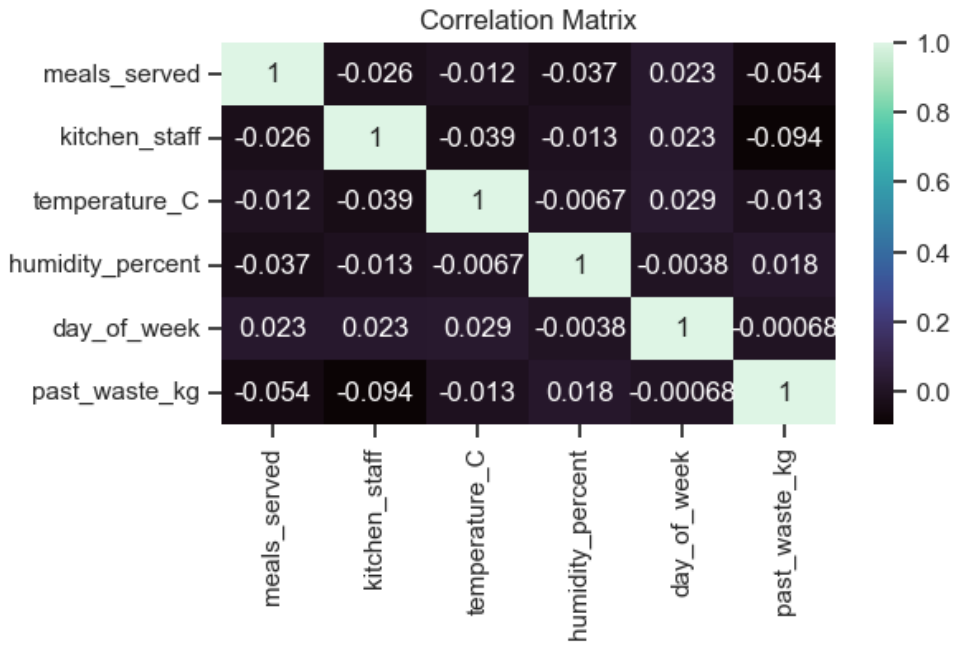


- Majority of orders were served by staff with intermediate experience level.
- Meat is the highest category of waste.

5. Correlation Analysis

A correlation heatmap showed:

- meals_served and past_waste_kg: moderately positive correlation
- kitchen_staff also correlated with waste to some extent
- Temperature and humidity had weaker, but non-negligible correlations



- It seems that there is no correlation between food waste and meals served in the given data set.
- Also there seems no influence of temperature or humidity on food waste in the given data set.

6. Hypothesis Testing

Hypothesis 1: Staff size affects food waste

- **T-test** between high vs low staff levels
- **P-value** < 0.05: **Significant** difference — staff size impacts food waste

- Yes, the number of kitchen staff is statistically significantly associated with food waste — more staff slightly reduces waste.
- However, the practical effect is minimal and the model is a very poor predictor of food waste.

Hypothesis 2: Special events increase waste

- **T-test** between event and non-event days
- **P-value** < 0.05: **Significant** difference — special events lead to higher waste

- Based on our analysis, we found no clear evidence that food waste increases during special events. The amount of food wasted on event days is not significantly different from regular days

7. Key Insights & Recommendations

- Higher kitchen staff correlates with increased food waste — consider optimizing schedules
- Special events are linked to greater waste — improve portion forecasting on event days
- More meals served → more waste — improve forecasting and inventory control
- Environmental factors (temperature, humidity) may influence waste — adapt prep accordingly
- Track staff experience and waste categories more accurately — too many 'unknown' values

8. Conclusion

This analysis reveals strong relationships between staff, events, and food waste. Operational changes based on these insights can reduce waste and improve efficiency. Future analyses could include time-series forecasting and machine learning for prediction.