

Overview of the Project:

A food restaurant dataset was provided having 3-years of data from 2022 to 2024. The main goal of the project is identifying the factors that could enhance operational performance, reduces food wastage, visualize key patterns and suggest actionable insights.

Summary of Dataset and Key Variables:

The data contained 1,822 rows and 11 columns. After data cleaning we have 867 rows with unique date and 10 columns. Key variables are meals_served, kitchen_staff, past_waste_kg and staff_experience.

Data Cleaning:

- **Drop 'ID' column:** as this column does not serve as unique identifier so removing this column.
- **All columns:** Removing White spaces, converting to lower case and replacing blanks/others to NaN
- **meals_served column:** remove “.0” in the end and converted into integer (int)
- **kitchen_staff and special_event:** Alpha-numeric to numeric conversion for example in kitchen_staff convert “ten” and “eleven” to numerical, similarly for special_event convert “One” to numerical.
- **Data Type conversion:** Converted date into consistent datetime format for analysis and meals_served, kitchen_staff, special_event to integer.
- **Duplicates:**
In the context of restaurant business, sales is recorded per day, so by making **date** as unique identifier and using **groupby** function, Out of total 1,822 rows after applying groupby function, rows with unique date are 867, it means 955 rows (52.4% of whole data) was found duplicate and removed.

- **Missing Values**

- As multiple entries of same day were found with slight difference in column values, so by making **date** as unique identifier and applied **groupby** function and selected following for each column to fill the missing value.

Column	function
meals_served	mode
kitchen_staff	mode
temperature_C	max
humidity_percent	max
day_of_week	mode
special_event	mode
past_waste_kg	max
staff_experience	last
waste_category	last

- To find missing values in data, `df.isnull().sum()` was used. It identified 156 missing values in `staff_experience` column. As this is categorical column and we are not sure about expertise of the staff so introducing a new category with name 'unknown' hence missing value were filled with string value 'Unknown' in the column

- **Outliers:** 21 outliers were found in **temperature_C**, as this is an environmental factor and can positively or negatively impact operational efficiency and food wastage, so keeping these outliers.

Summary Statistics

	meals_served	kitchen_staff	temperature_C	humidity_percent	past_waste_kg
mean	277	12	22.3	60.8	27.0
min	1	5	-10	30.1	5
25%	172	8	15.7	46.1	16
50%	285	12	22.4	61.7	26.8
75%	385	15	28.8	75.8	38.1
max	499	19	60	90	49.8

Visualization

To enhance operational efficiency and reduce food wastage, following analysis were done:

1. Univariate analysis

- Histogram - Continuous Numeric Columns
- Box Plot - Continuous Numeric Columns
- Count Plot - Categorical Columns

2. Bivariate analysis

- Scatter Plot - Continuous Numeric Columns (All combinations)
- Box Plot - Categorical Columns vs (One Continuous Numeric Column)
- Correlation – Heatmap
- Spearman Correlation

3. Multivariate Analysis

- Numeric Columns vs One categorical (Text Column)

Refer to appendix for graphical representation of above analysis

Correlation between variables

Column	meals_served	kitchen_staff	temperature_C	humidity_percent	day_of_week	special_event	past_waste_kg
meals_served	1.00	0.02	0.02	-0.03	0.02	0.02	0.02
kitchen_staff	0.02	1.00	-0.01	-0.01	0.03	0.01	-0.08
temperature_C	0.02	-0.01	1.00	-0.01	0.02	-0.01	-0.03
humidity_percent	-0.03	-0.01	-0.01	1.00	0.02	0.01	0.02
day_of_week	0.02	0.03	0.02	0.02	1.00	-0.04	0.00
special_event	0.02	0.01	-0.01	0.01	-0.04	1.00	0.00
past_waste_kg	0.02	-0.08	-0.03	0.02	0.00	0.00	1.00

It is evident from above correlation table that there is no correlation between variables

Key patterns or trends observed

No significant trends, correlation or dependency were found among the variables such as number of meals served with wastage, staff experience with wastage, effect of temperature or humidity on meals served or wastage, and special event effect on meals served or wastage.

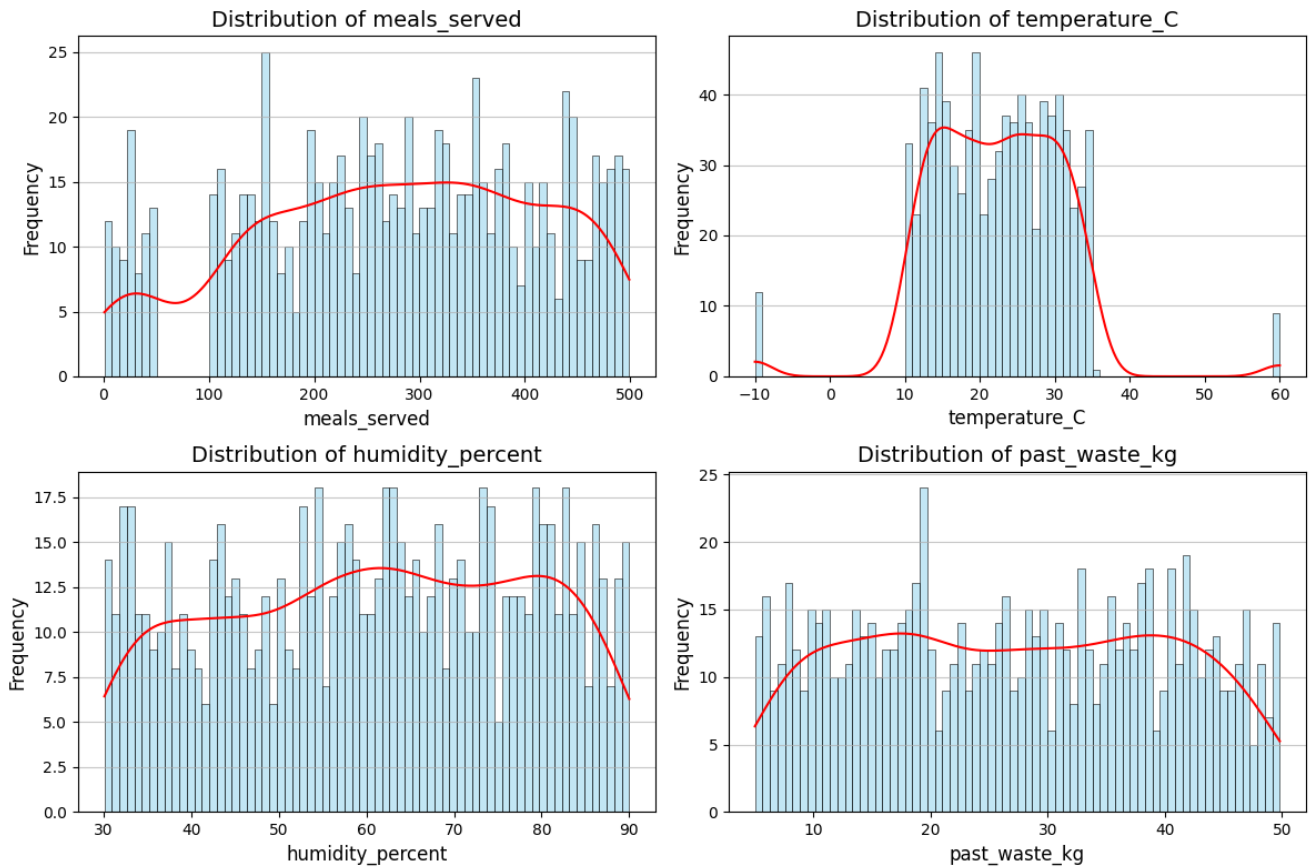
Conclusion

The data provided contained 1,822 rows out of which 955 rows were duplicate, it means more than 50% of the data was removed, and the analysis was done on remaining 867 unique rows. The data is not sufficient to address issues related to operational efficiency and food wastage.

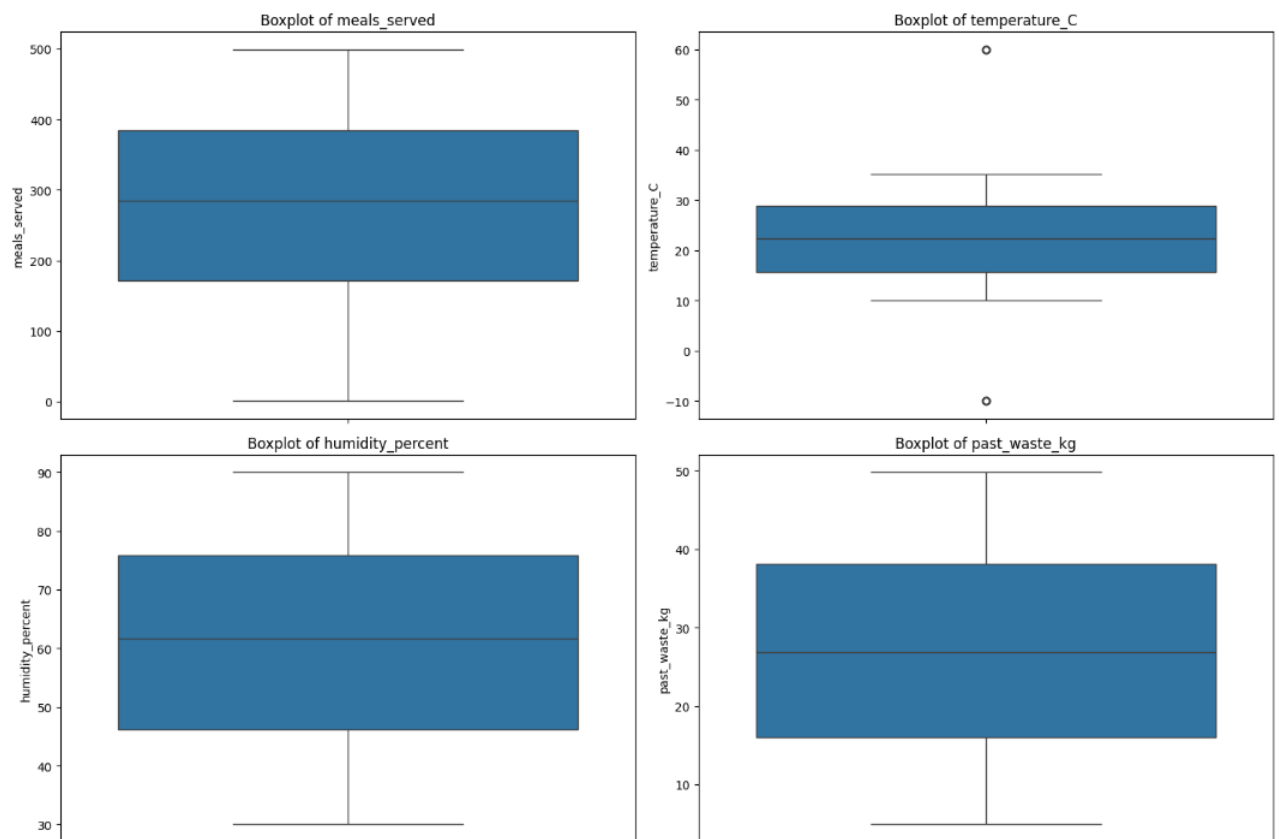
The analysis could have been better if more quality data would have been provided.

Appendix

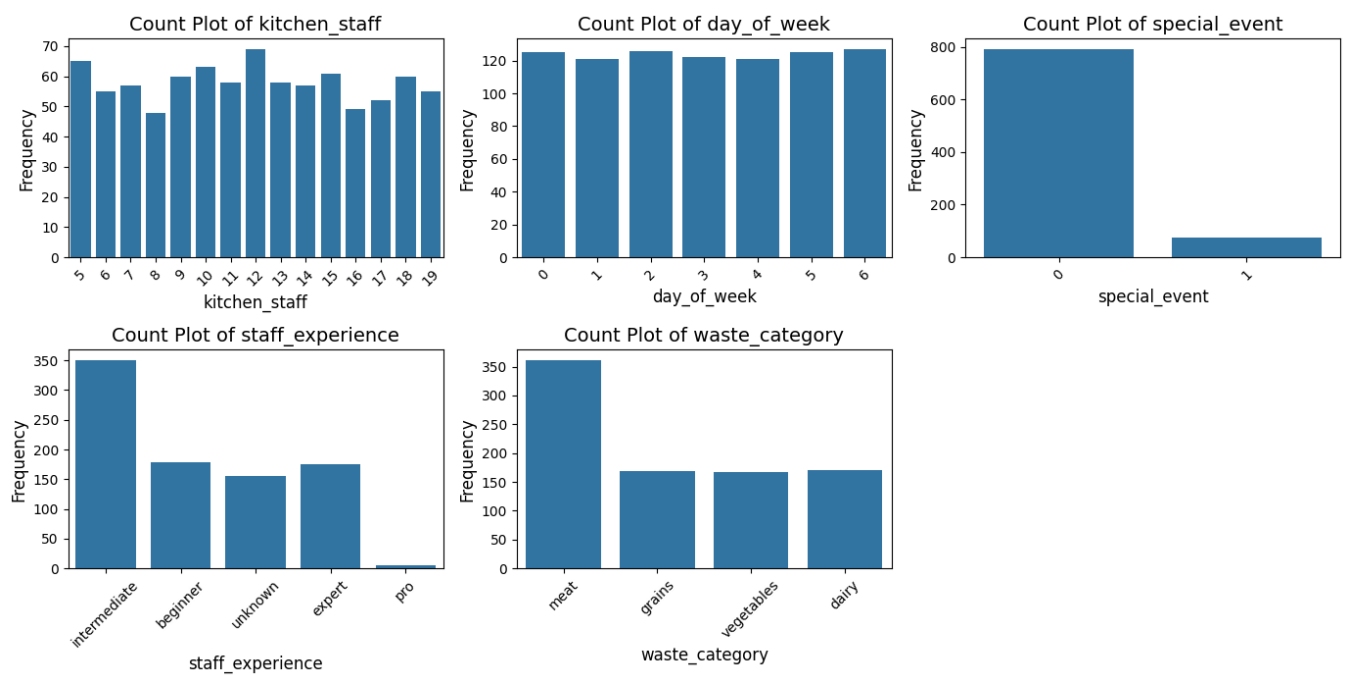
Univariate Analysis: Histogram - Continuous Numeric Columns



Univariate Analysis : Box Plot - Continuous Numeric Columns

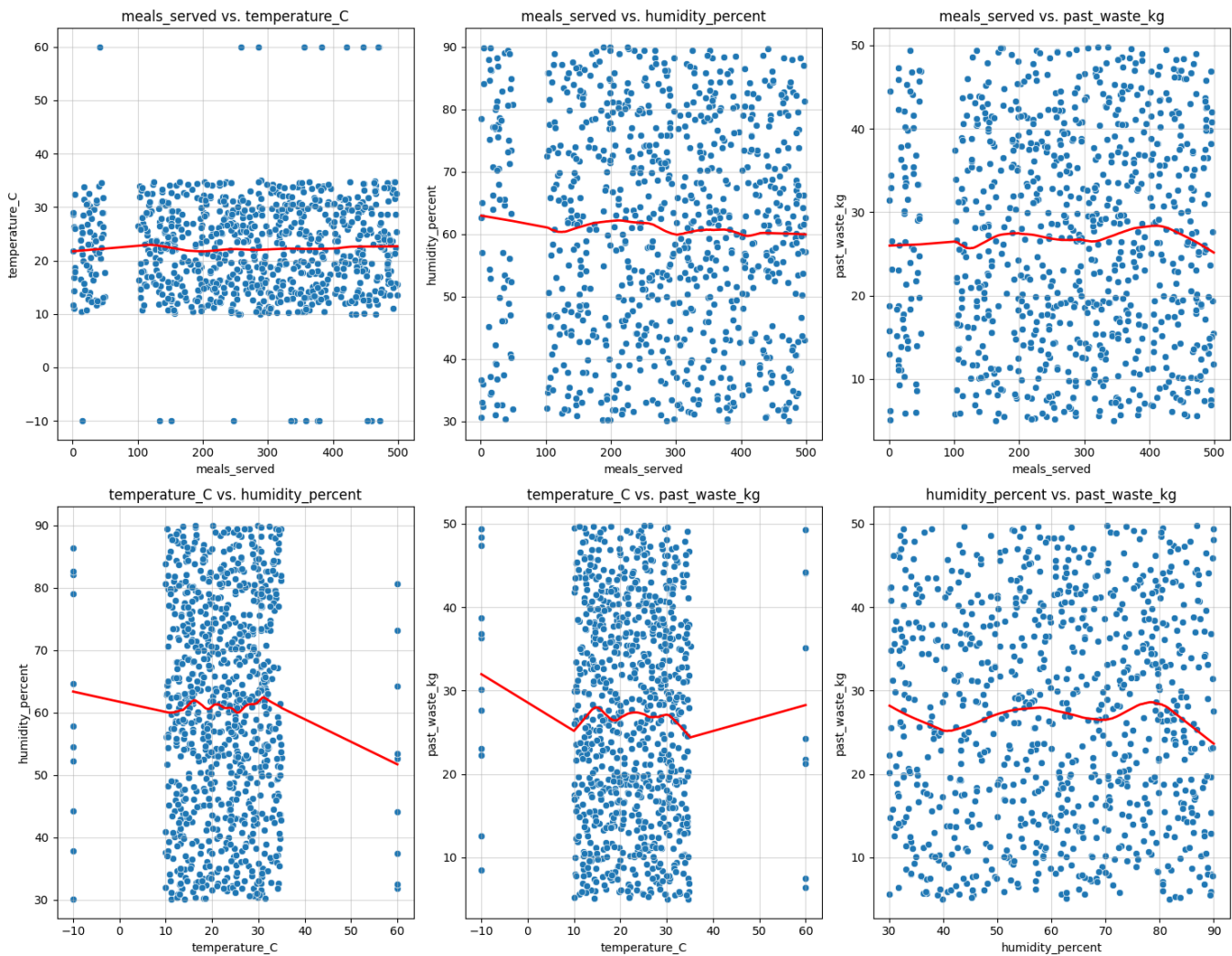


Univariate Analysis: Count Plot - Categorical Columns

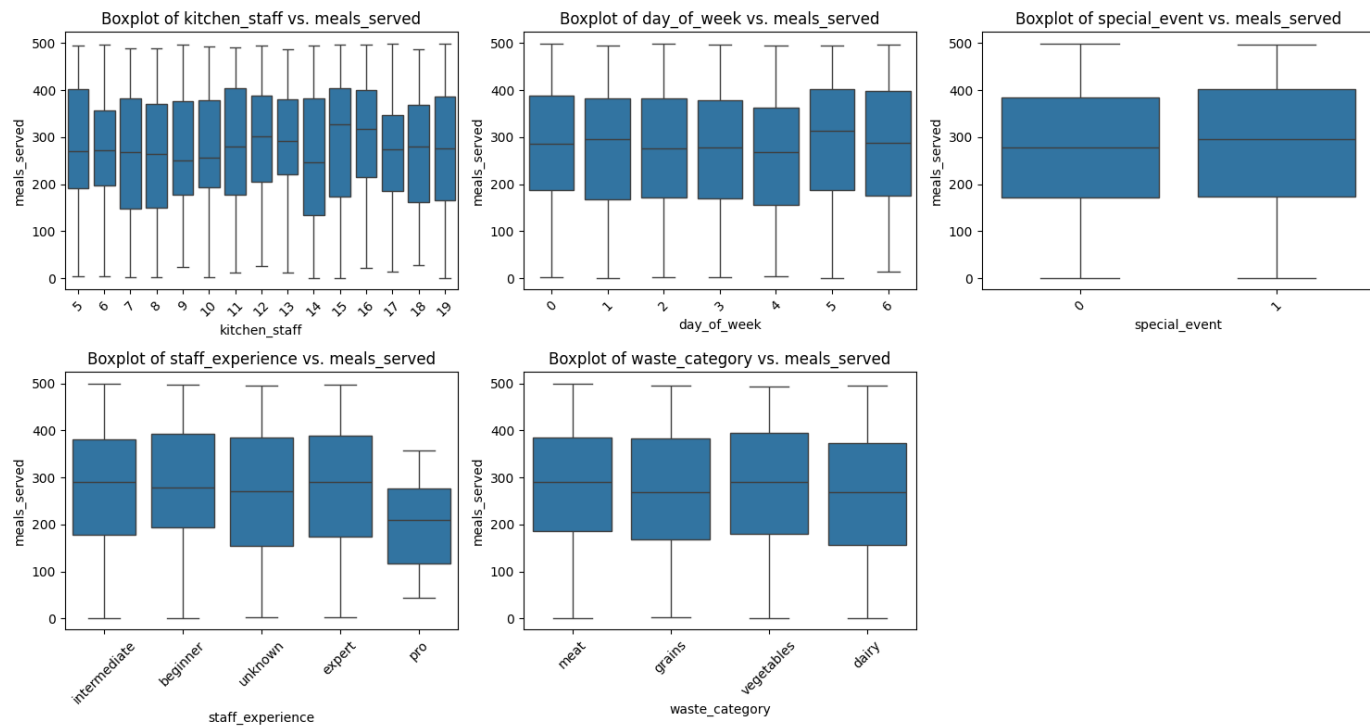


Bivariate Analysis: Scatter Plot- Continious Numeric Columns (All combinations)

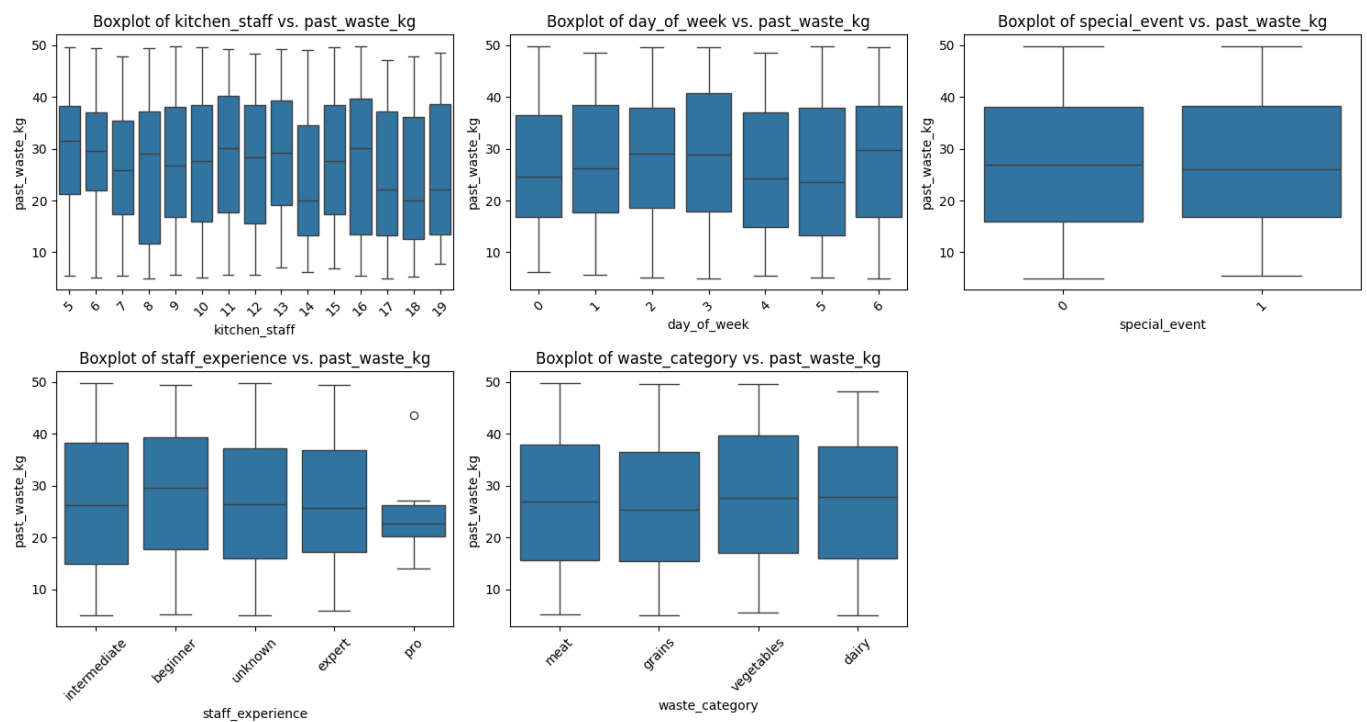
Scatter Plots with Smoothing



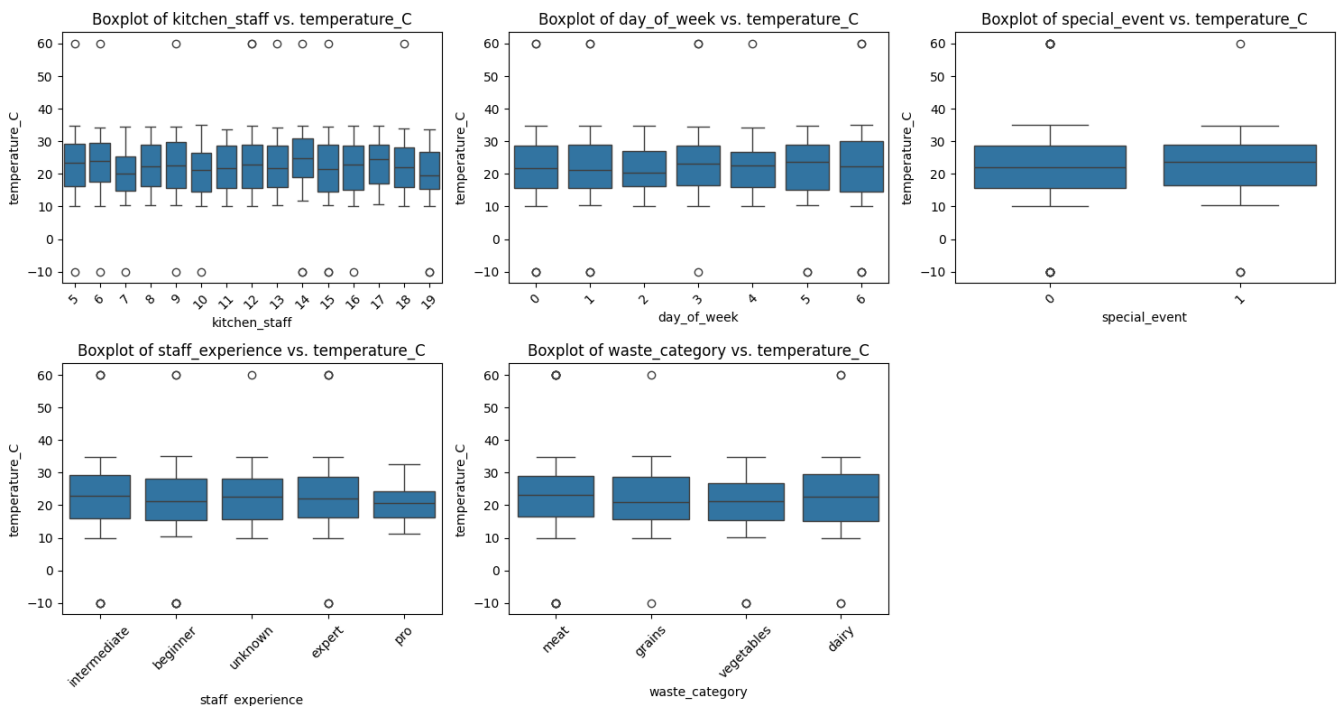
Bivariate Analysis: Box Plot - Categorical Columns vs (meals_served)



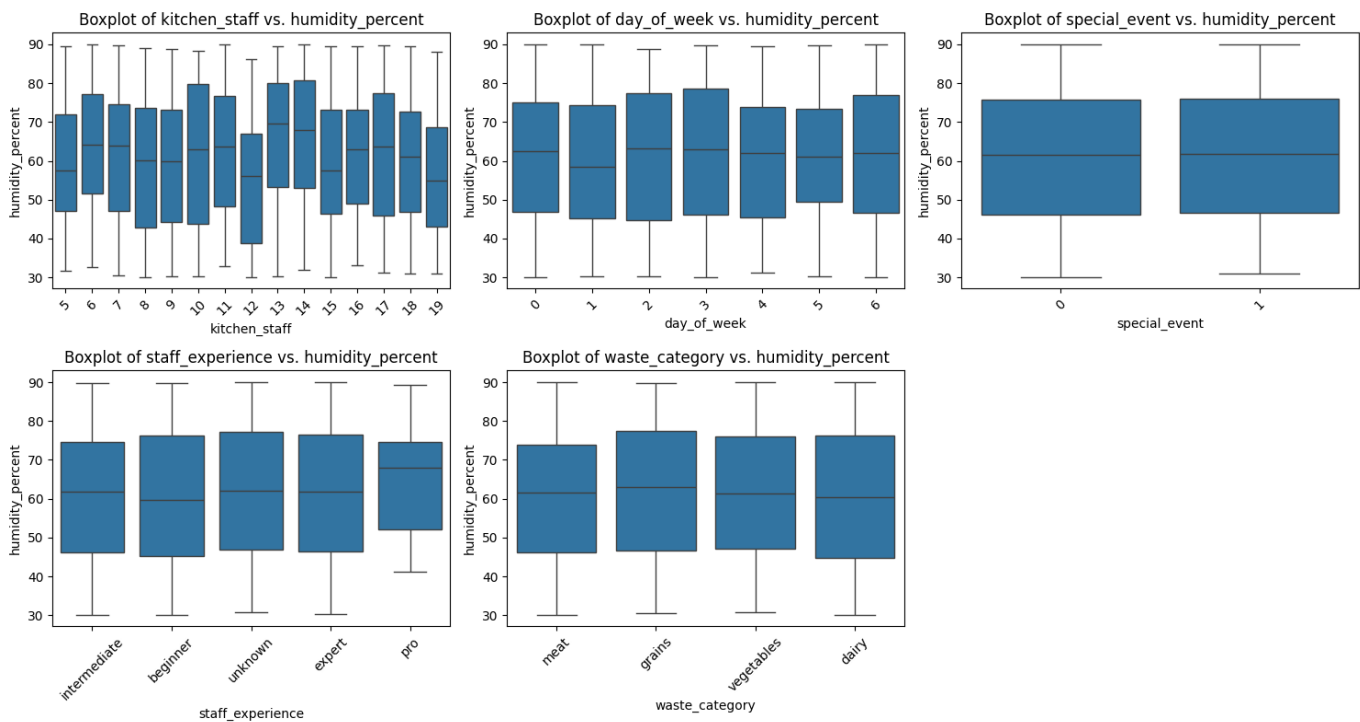
Bivariate Analysis: Box Plot - Categorical Columns vs (past_waste_kg)



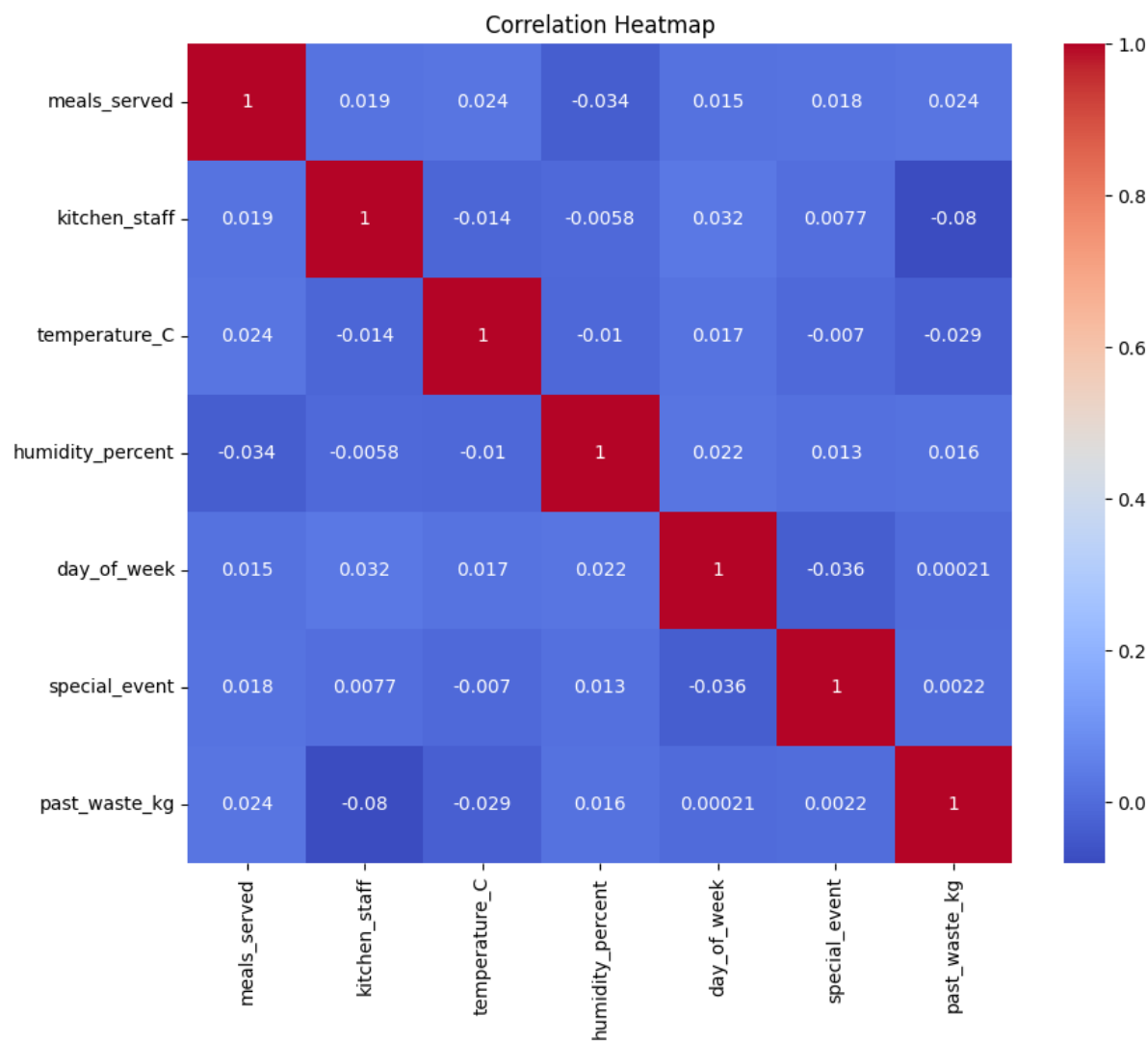
Bivariate Analysis: Box Plot - Categorical Columns vs (temperature_C)



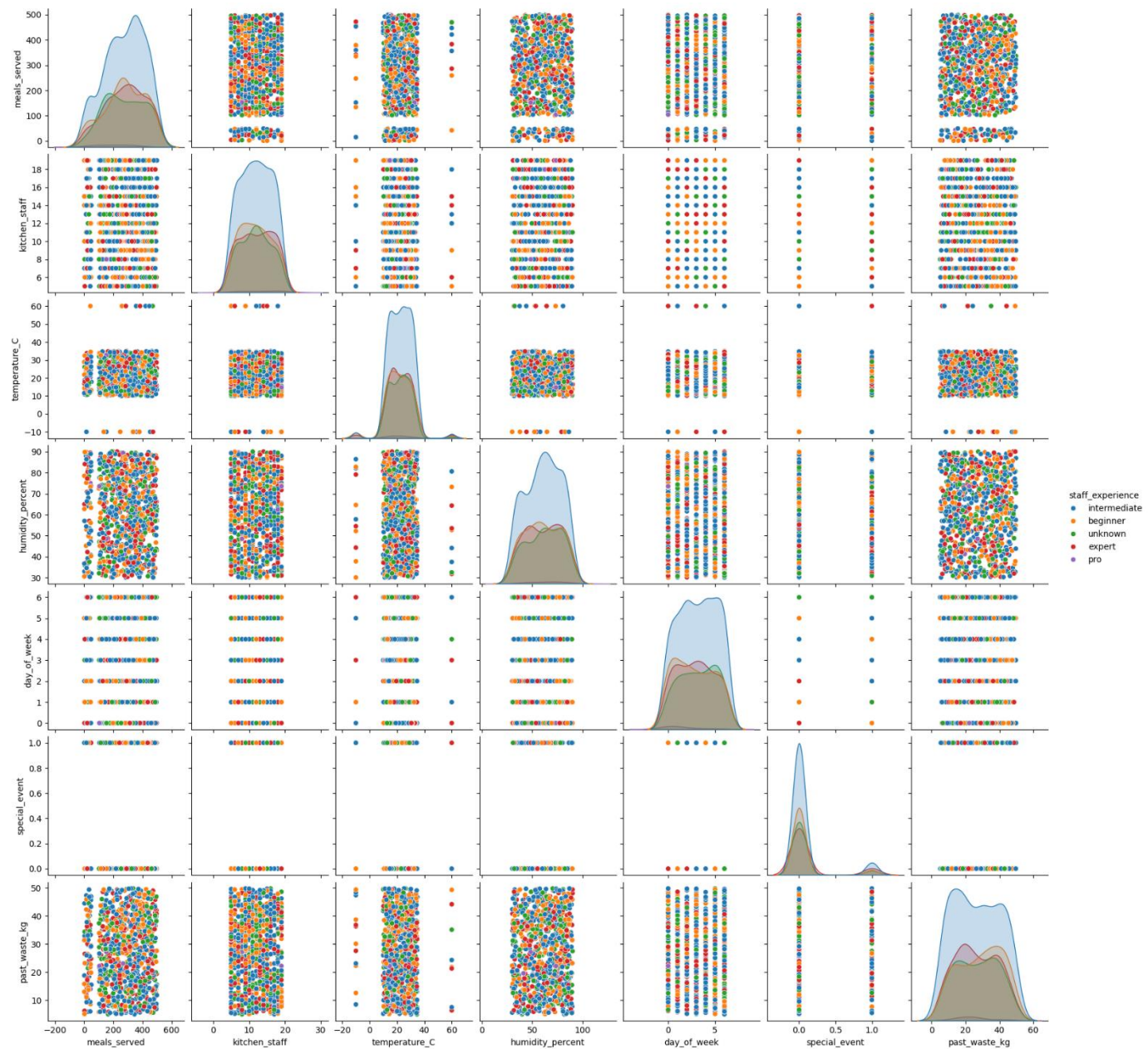
Bivariate Analysis: Box Plot - Categorical Columns vs (humidity_percent)



Bivariate Analysis: Correlation Heatmap



Multivariate Analysis : Numeric Columns vs staff_experience



Multivariate Analysis : Numeric Columns vs waste_category

