Exploratory Data Analysis (EDA) on Food Service Data

1. Introduction

Objective of the Project

The goal of this Exploratory Data Analysis (EDA) project is to analyse a food service dataset to identify patterns and insights that can help improve operational efficiency and reduce food waste. The analysis explores relationships between variables such as meals served, kitchen staff, environmental conditions, and food waste.

Dataset Overview

The dataset contains the following columns:

- ID: Unique identifier for each record
- date: Date of the observation
- meals served: Number of meals served on a given day
- kitchen staff: Number of kitchen staff on duty
- temperature: Temperature on the day (in Celsius)
- humidity percent: Humidity percentage
- day_of_week: Day of the week (0 = Sunday, 6 = Saturday)
- special event: Indicator of whether a special event occurred (1 = Yes, 0 = No)
- past_waste_kg: Food waste recorded (in kilograms)
- staff experience: Experience level of the kitchen staff (e.g., Beginner, Intermediate)
- waste category: Type of food waste (e.g., meat, dairy)

2. Data Cleaning

Missing Values

- meals served and past_waste_kg had missing values which were imputed using their respective median values.
- staff experience and waste category had missing entries which were filled using the mode.

Duplicate Rows

- Duplicate entries were checked using df.duplicated () and removed.

Categorical Data

- The staff experience column was encoded using a mapping: Beginner = 0, Intermediate = 1, Expert = 2.

Data Types

- The date column was converted to datetime format using pd.to_datetime() to facilitate time-based analysis.

3. Exploratory Data Analysis (EDA)

Summary Statistics

- Descriptive statistics showed the central tendency and dispersion of numerical features such as meals served, temperature, and past_waste_kg.

Visualizations

- **Histograms** showed distribution of meals served, temperature, humidity percent, and past_waste_kg.
- **Boxplots** revealed outliers and spread across numeric features.
- **Bar plots** visualized counts of staff experience levels and waste category.

Key Patterns

- Days with special events showed higher variance in waste.
- Temperature and humidity varied across days but did not show extremely skewed patterns.

4. Correlation Analysis

Heatmap

- A correlation heatmap revealed relationships among numeric features.

Key Correlations

- Moderate positive correlation between meals served and past_waste_kg.
- kitchen staff showed some association with meals served.
- Environmental variables had minimal correlation with waste.

5. Hypothesis Testing

Kitchen Staff vs Food Waste

- **Hypotheses**:
    - H0: Kitchen staff count has no effect on food waste.
    - H1: Kitchen staff count significantly affects food waste.
- **Test**: One-way ANOVA was performed across staff count groups (low, medium, high).
- **Result**: p-value indicated whether differences in mean waste were statistically significant.

Special Events vs Food Waste

- **Hypotheses**:
    - H0: No difference in waste between event and non-event days.
    - H1: Food waste is higher on event days.
- **Test**: Independent t-test compared average waste on event vs. non-event days.
- **Result**: Statistically significant difference found if p-value < 0.05.

6. Key Insights and Recommendations

Operational Insights

- Staff numbers influence food waste, suggesting over- or under-staffing can impact efficiency.
- Special events significantly increase waste, highlighting a need for better planning.

Recommendations

- Optimize staff allocation based on predicted meal counts.
- Implement better planning for special events to reduce excess preparation.
- Use temperature/humidity data to adjust storage and preparation if strong patterns emerge.

7. Conclusion

Summary

This analysis identified meaningful patterns linking staff levels and special events to food waste. Several numeric and categorical variables were found to influence operational outcomes.

Limitations and Future Work

- Dataset lacks detailed time information (e.g., shift-level data).
- No cost variables available to assess economic impact.
- Future work could include predictive modelling for food waste forecasting.