**Project Report: Exploratory Data Analysis (EDA) on Food Service Data**

# 1. Introduction

This project aims to explore a food service dataset to gain insights into operational efficiency and food waste management. The dataset includes variables like meals served, kitchen staff, environmental conditions (temperature and humidity), and food waste. The goal is to clean the data, perform EDA, test hypothesis, and generate actionable insights.

# 2. Dataset Overview

The dataset contains the following columns:  
**- ID**: Unique identifier  
**- date**: Date of record  
**- meals\_served**: Number of meals served  
**- kitchen\_staff**: Number of staff working  
**- temperature\_C**: Temperature in Celsius  
**- humidity\_percent**: Humidity percentage  
**- day\_of\_week**: 0 = Sunday to 6 = Saturday  
**- special\_event**: 0 = No event, 1 = Event  
**- past\_waste\_kg**: Previous food waste in kilograms  
**- staff\_experience**: Experience level ("Beginner", "Intermediate", etc.)  
**- waste\_category**: Category of food waste ("meat", "dairy", etc.)

# 3. Data Cleaning

**3.1 Handling Missing Values:**  
- Checked for null values using df.isnull().sum().  
- Imputed missing categorical values with the mode.  
- Removed rows with excessive missing numerical data.

**3.2 Duplicate Rows:**  
- Dropped exact duplicates using df.drop\_duplicates().

**3.3 Categorical Data:**  
- Verified categories for staff\_experience and waste\_category.  
- Used one-hot encoding or label encoding for modeling if needed.

**3.4 Data Types:**  
- Converted date to datetime format.  
- Verified numerical columns are in proper float/integer format.

# 4. Exploratory Data Analysis (EDA)

**4.1 Summary Statistics:**  
- meals\_served: Mean ≈ 490, Range: 100–1000  
- past\_waste\_kg: Mean ≈ 25, Range: 0–80  
- temperature\_C: Mean ≈ 23°C  
- humidity\_percent: Avg ≈ 60%

**4.2 Visualizations:**  
- Histograms: Meals served, temperature, humidity showed near-normal distribution.  
- Boxplots: Revealed some outliers in food waste and temperature.  
- Bar Plots:  
 • staff\_experience: Most frequent = Intermediate  
 • waste\_category: Most common = Vegetables

**4.3 Correlation Analysis:**  
- Strong correlation: meals\_served and past\_waste\_kg (r ≈ 0.6)  
- Weak correlation: temperature\_C and past\_waste\_kg (r ≈ 0.2)

# 5. Hypothesis Testing

**5.1 Impact of Kitchen Staff on Food Waste:**  
- H₀: No relationship between kitchen staff and food waste  
- H₁: Number of staff significantly affects food waste  
- Method: ANOVA  
- Result: p-value < 0.05 → Reject H₀  
✅ More staff correlates with increased food waste (possibly due to overproduction).

**5.2 Special Events and Food Waste:**  
- H₀: No difference in waste between event and non-event days  
- H₁: Waste is higher on event days  
- Method: Independent t-test  
- Result: p-value < 0.05 → Reject H₀  
✅ Special events lead to significantly more food waste.

# 6. Key Insights and Recommendations

- Staffing: Optimize staff allocation based on historical waste trends to reduce overproduction.  
- Event Management: Plan portion sizes more accurately during special events.  
- Environment: Adjust storage/prep during hot, humid days.  
- Training: Upskill beginner staff to reduce waste.

# 7. Conclusion

The analysis revealed key operational inefficiencies such as overstaffing and unoptimized event management contributing to food waste. Data-driven recommendations can help reduce costs and environmental impact.  
  
**Limitations:**  
- The dataset lacks detailed customer or menu data.  
- Does not account for spoilage vs. leftover distinction.  
  
**Future Work:**  
- Include food type/recipe-level data for more granular analysis.  
- Track real-time consumption and leftovers for predictive waste modeling.

# 8. Appendix

- Additional graphs (boxplots, heatmaps, etc.)  
- Code snippets for imputation, visualizations, and statistical tests