
📥 1. Setup and Data Loading

- Imported libraries: pandas, numpy, matplotlib, seaborn, warnings.
- Loaded the pre-cleaned CSV file: Fully_Cleaned_Halls_Data.csv into DataFrame df.

🔍 2. Initial Data Checks

- Checked for duplicate rows using:
- df.duplicated().sum()
- Displayed first 10 rows of the dataset to inspect structure using df.head(10).

📊 3. Column Classification

- Segregated column types:
  - num_df: Numerical columns (int, float)
  - cat_df: Categorical columns (object)
- Printed both lists to identify types of features for further analysis.

📐 4. Skewness Analysis

- Computed skewness for all numeric columns to understand distribution characteristics:
- df.skew(numeric_only=True)

⚠️ 5. Outlier Removal Using IQR Method

- Identified numerical columns **excluding** 'Menu_Price', 'Ratings', and 'Rooms'.
- Applied IQR method (Interquartile Range) for robust outlier removal:
  - Calculated Q1 (25th percentile) and Q3 (75th percentile)
  - Removed rows where values were outside [Q1 - 1.5×IQR, Q3 + 1.5×IQR] range **for selected features**
- Created a new cleaned DataFrame df2.

🔁 6. Post-Cleaning Checks

- Reset index of df2 for clean sequencing.
- Verified that the cleaned DataFrame had **no duplicate rows**.
