

The dataset has 244 entries and includes the following 7 columns:

1. **total\_bill**: The total bill amount (numeric, continuous).
2. **tip**: The tip amount (numeric, continuous).
3. **sex**: Gender of the person paying (categorical, **Male** or **Female**).
4. **smoker**: Whether the customer is a smoker (categorical, **Yes** or **No**).
5. **day**: Day of the week (categorical, e.g., **Sun**, **Sat**, **Thur**, **Fri**).
6. **time**: Time of day (categorical, **Lunch** or **Dinner**).
7. **size**: Size of the group (numeric, discrete).

The dataset has no missing values. I will now conduct a detailed analysis by exploring distributions, correlations, and patterns across these variables. Let's start with a summary of key statistics for numerical and categorical columns.

## Summary Statistics:

### Numerical Variables:

1. **Total Bill**:
  - Mean: \$19.79
  - Standard Deviation: \$8.90
  - Range: \$3.07 (min) to \$50.81 (max)
  - Interquartile Range: \$13.35 (25th percentile) to \$24.13 (75th percentile)
2. **Tip**:
  - Mean: \$3.00
  - Standard Deviation: \$1.38
  - Range: \$1.00 (min) to \$10.00 (max)
  - Interquartile Range: \$2.00 to \$3.56
3. **Group Size**:
  - Mean: 2.57
  - Standard Deviation: 0.95
  - Range: 1 to 6

### Categorical Variables:

1. **Sex**:
  - Male: Most frequent (157 entries)
  - Female: 87 entries
2. **Smoker**:
  - Non-smokers: Most frequent (151 entries)
  - Smokers: 93 entries
3. **Day**:
  - Most frequent: Saturday (87 entries)
  - Distribution across days: Sunday, Saturday, Thursday, and Friday.
4. **Time**:
  - Most frequent: Dinner (176 entries)
  - Lunch: 68 entries

## Correlation Analysis:

1. **Total Bill and Tip:** Strong positive correlation (0.68). As the total bill increases, the tip tends to increase proportionally.
2. **Total Bill and Group Size:** Moderate positive correlation (0.60). Larger groups tend to have higher total bills.
3. **Tip and Group Size:** Moderate positive correlation (0.49). Larger groups also tend to leave larger tips.