

TABLE OF CONTENTS:

- 1. Importing libraries**
- 2. Loading and displaying data**
- 3. Data exploration**
- 4. General statistics before removing outliers**
- 5. Cleaning data and removing outliers**
- 6. General statistics after removing outliers**
- 7. Difference between General statistics before and after Data cleaning**
- 8. Categorical analysis**
- 9. Numerical analysis**
- 10. Variation of tips with total bill**
- 11. Variation of tips with Gender**
- 12. Variation of tips by smoker status**
- 13. Variation of day with tips**
- 14. Variation of day with % of tip**
- 15. Variation of tip with size**
- 16. Variation of tips with time of dining**
- 17. Variation of group size with tip %**

1. IMPORTING LIBRARIES

```
IMPORT LIBRARIES

0s  ⏪ import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
     import scipy.stats as stats
```

2. LOADING DATA AND DISPLAY

```
LOAD DATA

[146] import pandas as pd
      #Load the Excel file
      df = pd.read_csv("/content/tipsfunc.csv")
      df.head()
```

3. DATA EXPLORATION

```
Total number of Rows

[163] # Display the rows
      print("TIPS dataset consists of {} rows.".format(rows = len(df)))

→ TIPS dataset consists of 244 rows.
```

4. GENERAL STATISTICS BEFORE REMOVING OUTLIERS

	total_bill	tip	percentage	size
mean	19.785943	2.998279	16.079754	2.569672
median	17.795000	2.900000	15.475000	2.000000
mode	13.420000	2.000000	14.480000	2.000000
stdev	8.902412	1.383638	6.107020	0.951100
min	3.070000	1.000000	3.560000	1.000000
max	50.810000	10.000000	71.030000	6.000000

5. CLEANING DATA AND REMOVING OUTLIERS

```
[149] # Select relevant columns
      df = df[['total_bill', 'tip', 'percentage','size','day','sex','smoker','time']]

      # Remove missing values
      df_clean = df.dropna()

      # Function to remove outliers using IQR
      def remove_outliers_iqr(data, column):
          Q1 = data[column].quantile(0.25)
          Q3 = data[column].quantile(0.75)
          IQR = Q3 - Q1
          lower_bound = Q1 - 1.5 * IQR
          upper_bound = Q3 + 1.5 * IQR
          return data[(data[column] >= lower_bound) & (data[column] <= upper_bound)]

      # Remove outliers for numerical columns
      for col in ['total_bill', 'tip', 'percentage','size']:
          df_clean = remove_outliers_iqr(df_clean, col)

      # Final cleaned dataset
      df_clean.reset_index(drop=True, inplace=True)
      print("TIPS dataset consists of {rows} cleaned rows.".format(rows = len(df_clean)))
```

→ TIPS dataset consists of 216 cleaned rows.

6. GENERAL STATISTICS AFTER REMOVING OUTLIERS

	total_bill	tip	percentage	size
mean	18.341759	2.748843	15.586620	2.425926
median	17.030000	2.580000	15.375000	2.000000
mode	13.420000	2.000000	14.480000	2.000000
stdev	6.999301	1.025754	4.179424	0.749619
min	5.750000	1.000000	5.680000	1.000000
max	40.170000	5.650000	26.630000	4.000000

1. Enhance Experience for Couples and Small Groups

- Most customers dine in pairs or small groups. Focus marketing efforts, promotions, and seating plans around 2–4 person tables.
- Consider offering date night specials, combo meals for two, or split-bill features.

2. Encourage Higher Tips Through Suggested Tipping

- Introduce pre-set tipping options (15%, 18%, 20%) on receipts or digital payments.
- Train staff to engage warmly and explain how tips support team members—this can increase generosity.

3. Target High-Spend Customers with Premium Add-Ons

- While most bills are ~\$17–18, a small segment spends \$40+.
- Introduce upsell opportunities (e.g., premium appetizers, desserts, drink pairings) for these spenders.

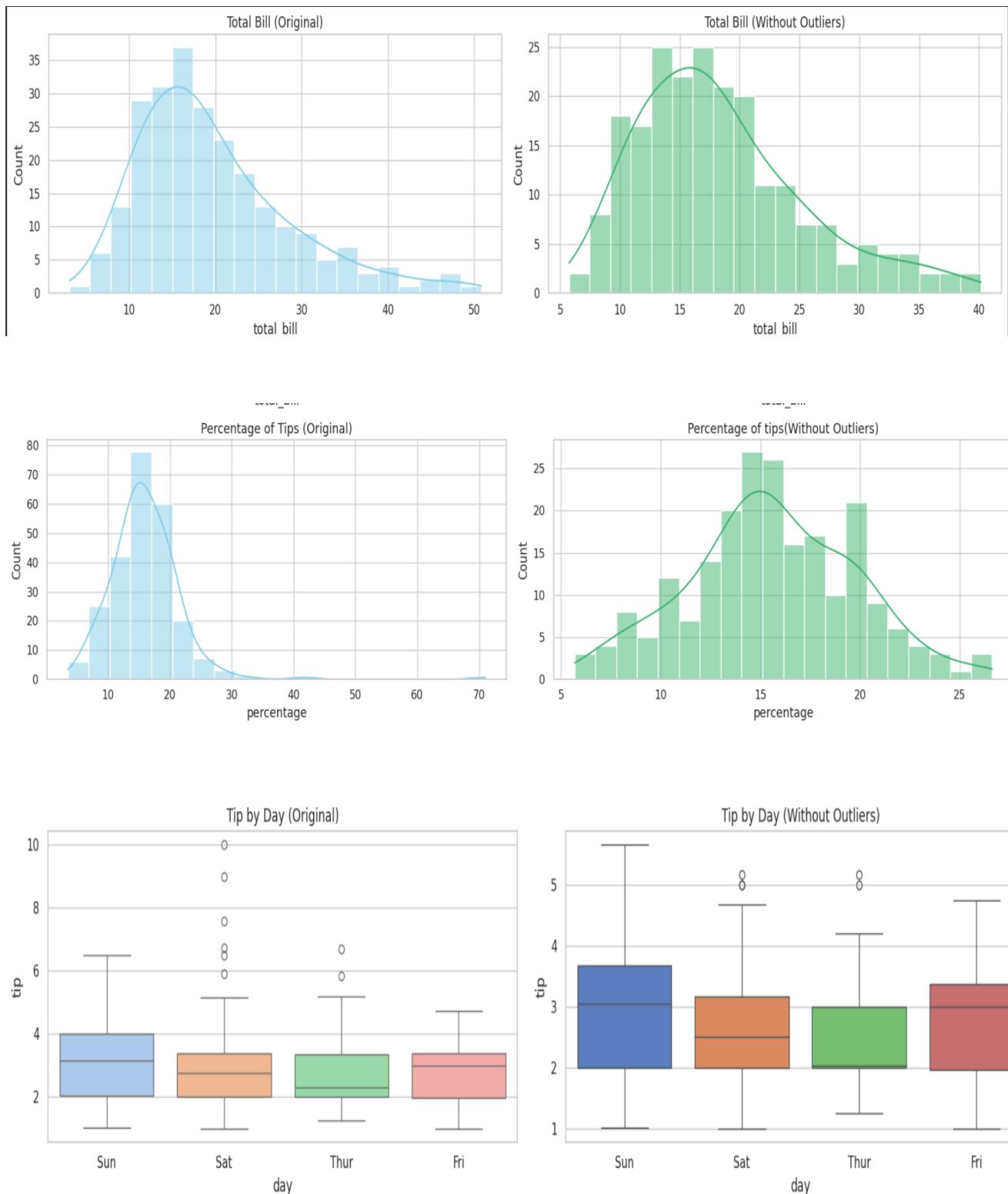
4. Minimize Tip Variability

- A standard deviation of 4.18% in tip percentage shows inconsistency.
- Improve service consistency and provide visual tipping guidance to anchor tipping behavior around higher norms.

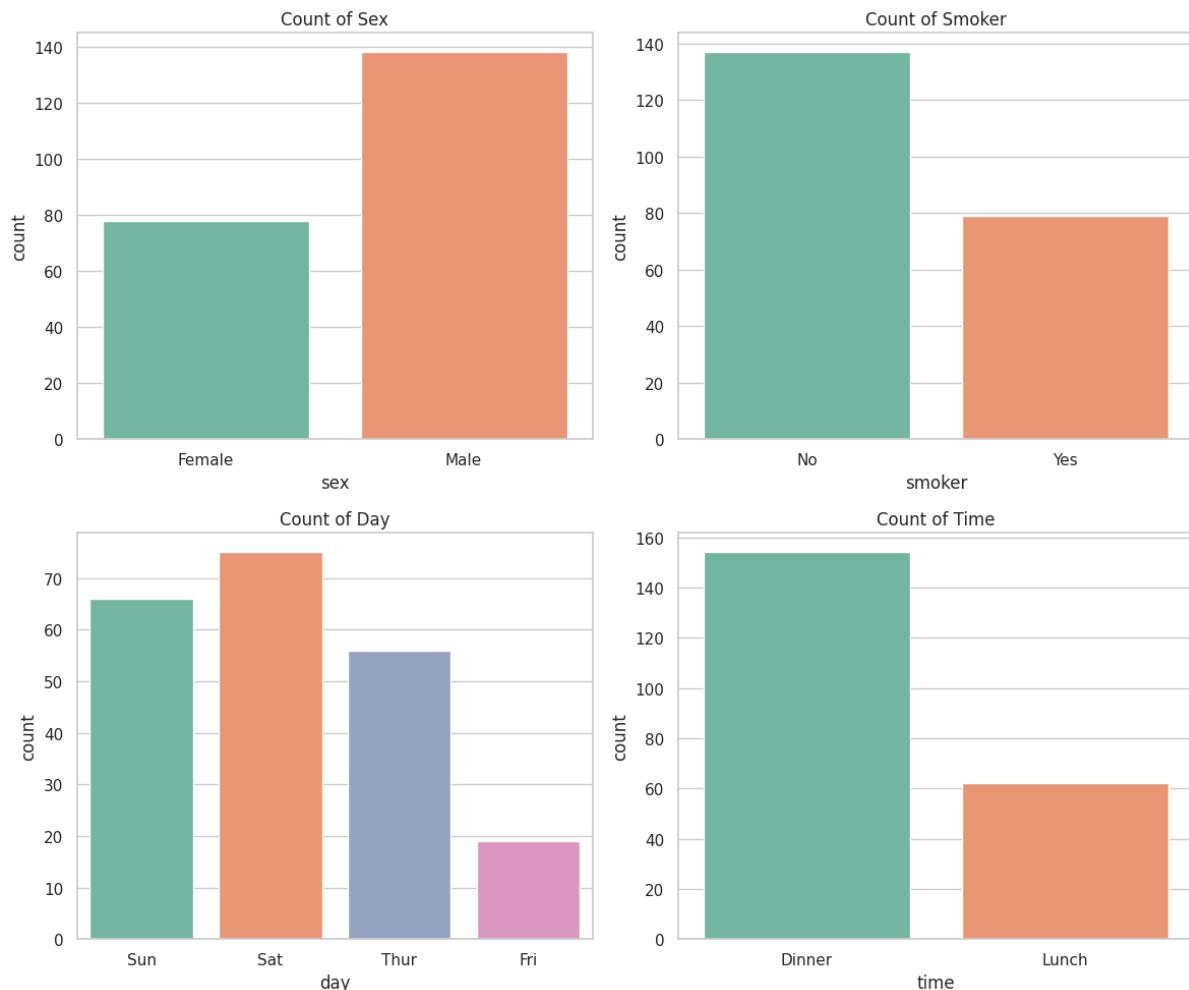
5. Analyze Temporal Trends

- Compare this dataset with time-of-day or day-of-week tipping trends (as you shared earlier). Optimize staffing and offers for peak tipping periods like dinner hours or weekends.

7. DIFFERENCE BETWEEN GENERAL STATISTICS BEFORE AND AFTER DATA CLEANING

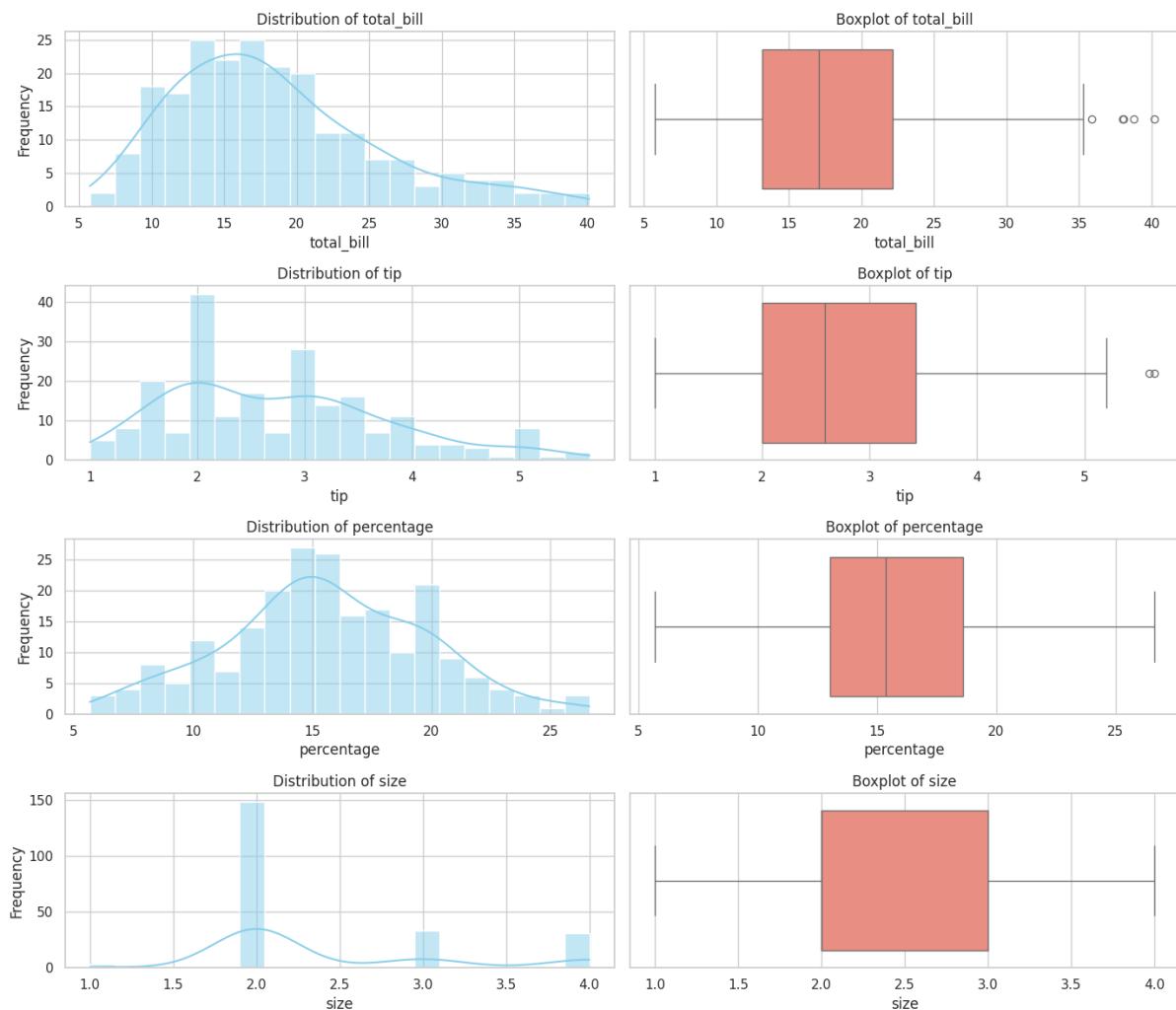


8. CATEGORICAL ANALYSIS



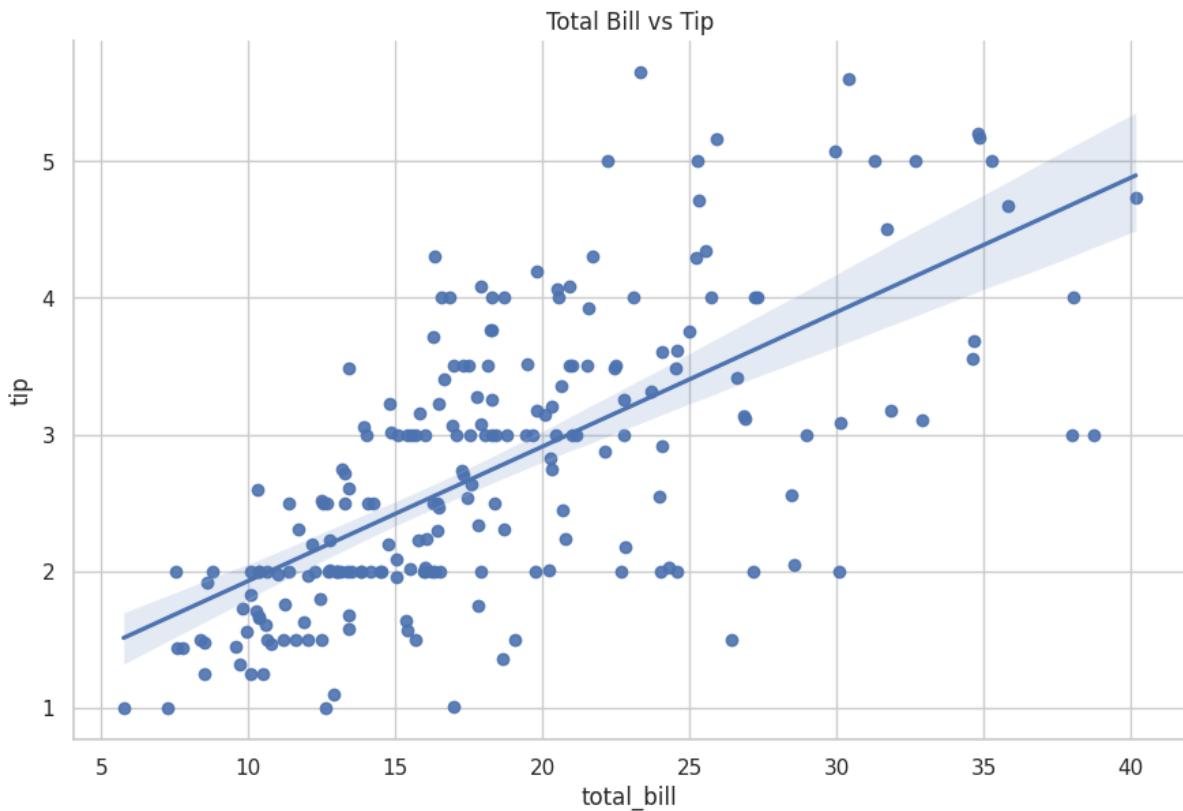
- Gender distribution isn't balanced. It may influence tipping behaviour or be linked to dining times (e.g., males may dine more at dinner).
- Smoking customers are a minority. This could influence tipping patterns or party size (e.g., smokers may dine in smaller groups).
- Highest restaurant traffic is during the weekend (Sat & Sun). Business is lowest on Fridays, which may need promotions or review.
- Customers prefer dinner over lunch at the restaurant.

9. NUMERICAL ANALYSIS



- Most customers spend moderately, but a small number spend significantly more, potentially due to larger group sizes or special occasions.
- Most people tip around \$2–\$3 regardless of the bill size, with some being exceptionally generous.
- Most customers tip between 12%–20%, which is typical in many tipping cultures.
- The restaurant primarily serves couples or small groups, and large group bills (or tips) are rare but may contribute to the outliers in other variables.

10. VARIATION OF TIPS WITH TOTAL BILL

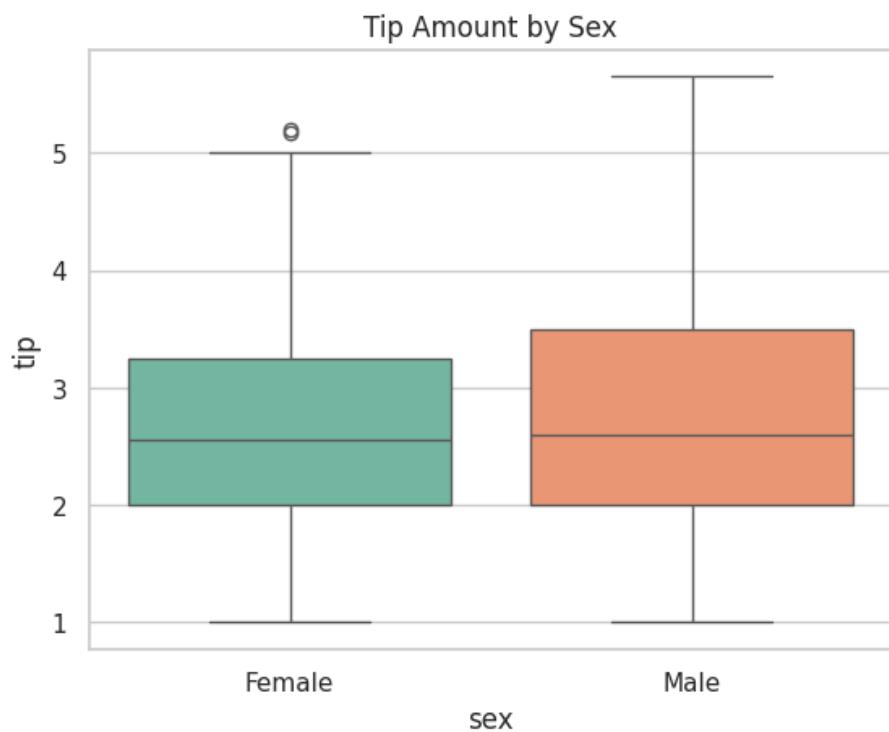


Key Insights:

- There is a positive linear relationship between total bill and tip.
- As the total bill increases, the tip amount also tends to increase.
- This trend suggests that customers usually tip more when they spend more.
- The scatter is fairly wide around the regression line, suggesting a moderate (not strong) correlation.
- There's variability in tips even for similar total bill values, meaning tipping isn't strictly proportional.
- Some points are far from the regression line:
 - Some low tips for high bills (under-tipping).
 - Some high tips for moderate bills (generous tipping).
- These may be important to flag as outliers for deeper analysis.

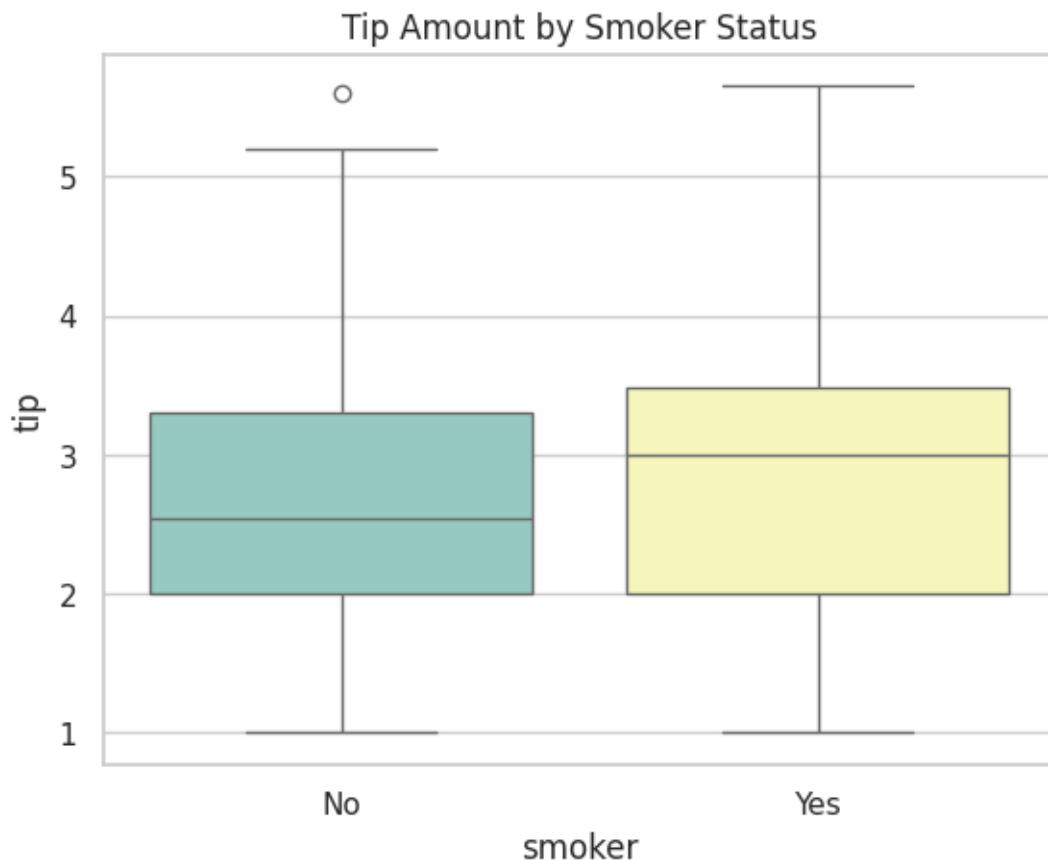
- At the higher end of the total_bill axis (~\$30-\$40), there's slightly more scatter and fewer data points.
- This could mean:
 - Fewer high-value transactions.
 - Tips are less predictable for high bills.
- The shaded area around the regression line shows a wider confidence interval at the ends, indicating less certainty in predictions for very low or very high total bills.
- Tipping behavior follows total bill but with variation — could be influenced by:
 - Time of day (lunch vs. dinner)
 - Service quality
 - Group size
 - Gender or smoking section
- A deeper model (e.g., multivariate regression) might improve predictability by including more features.

11. VARIATION OF TIPS WITH GENDER



- No major difference in tipping behaviour by sex in terms of central tendency (median).
- Males tend to have a slightly higher maximum tip value and more variability in tip amounts.
- Females show a few high outliers, which might need further exploration (e.g., if affected by party size or total bill).

12. VARIATION OF TIPS BY SMOKER STATUS

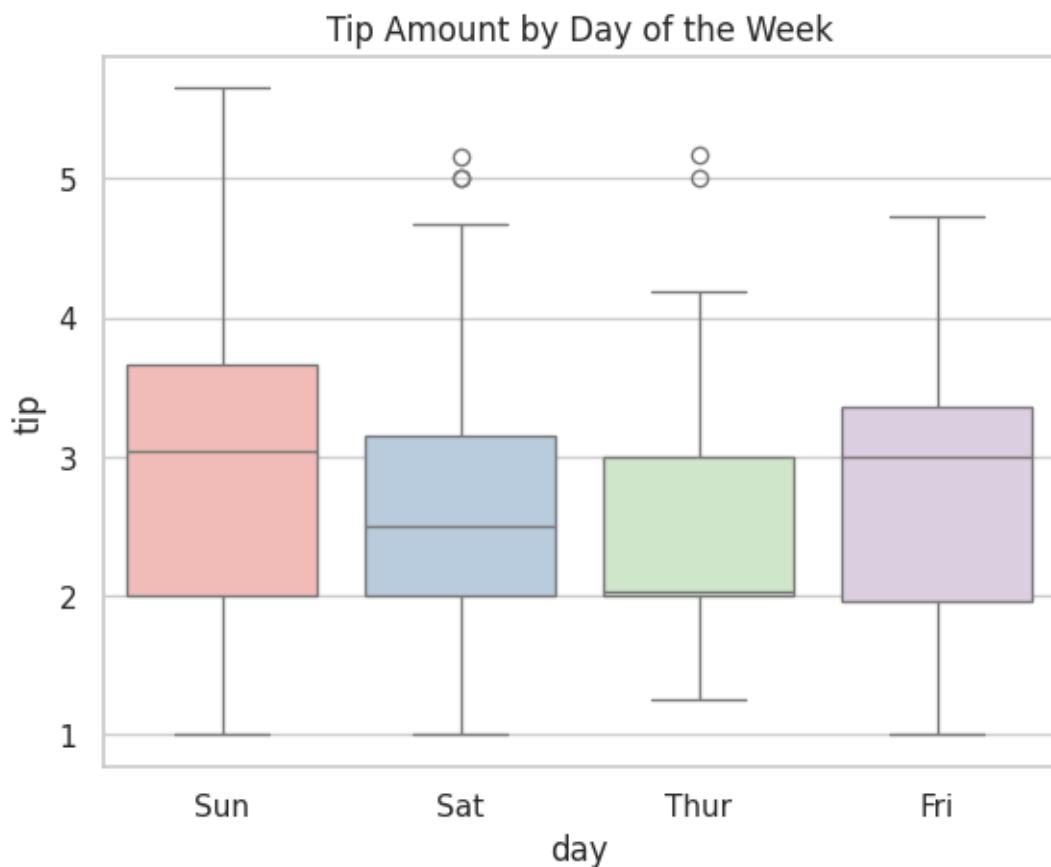


1. Median Comparison:
 - Smokers tip slightly more on average.
 - The difference is small but noticeable.
 2. Distribution Spread:
 - Smokers have a wider IQR, suggesting greater variation in tipping behaviour.
 - Both groups have similar minimum tips (~\$1), indicating consistent lower bounds.
 3. Outliers:
 - Non-smokers have one visible outlier above \$5.3.
 - Smokers do not show obvious outliers, but they do tip higher in general.
 4. Maximum Tips:
 - Smokers have slightly higher maximum tips compared to non-smokers.
- Behavioural Insight: Smokers might be slightly more generous or influenced by social/friend group dynamics (e.g., group dining, bar settings).

- Business Application: For restaurants or service settings, understanding these differences could inform:
 - Targeted promotions
 - Staff allocation during smoker-friendly hours or zones
 - Designing loyalty programs

While both groups have similar tipping ranges, smokers tip slightly more on average and show greater variability.

13. VARIATION OF TIPS WITH DAY



Key Insights:

1. Sunday Stands Out:
 - Highest median and widest range of tip values.
 - Suggests higher generosity or group/family dining behaviour.

2. Friday Shows Strong Tips Too:

- Median tip matches Sunday (~\$3), indicating generous tipping behaviour.
- Less variability than Sunday, suggesting more consistent tipping amounts.

3. Saturday & Thursday Show More Outliers:

- More variability, with outliers above \$4.5.
- These days may include both low and high tipping events—perhaps due to varying group sizes or occasions.

4. Comparative Variability:

- Sunday and Friday appear more favourable for higher and stable tip earnings.
- Saturday has notable highs but also shows more fluctuation.

• Business Strategy:

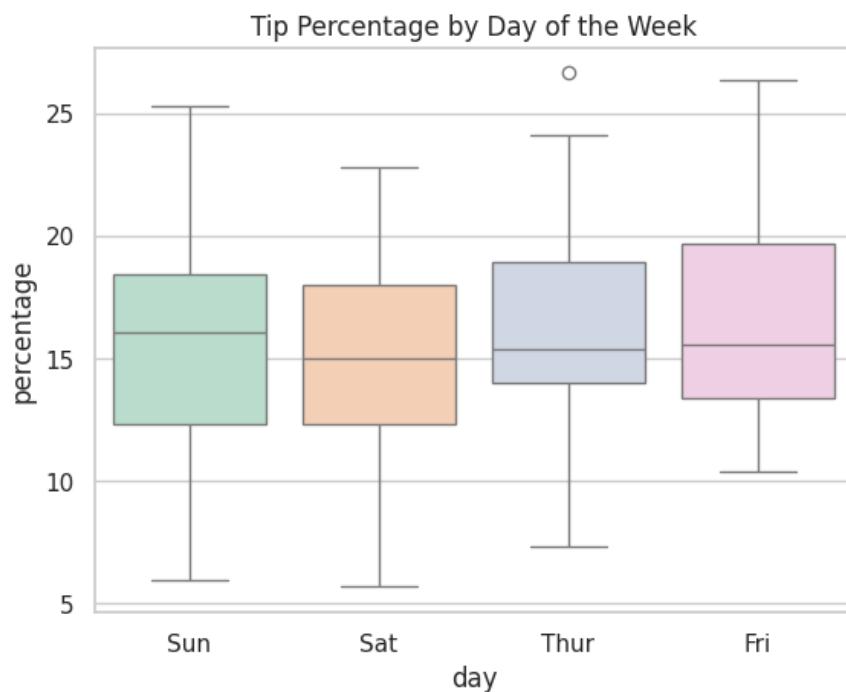
- Sundays and Fridays may be high-revenue tipping days; servers should be strategically scheduled.
- Promotions or staff incentives might be more effective on Thursdays and Saturdays to boost lower or inconsistent tipping patterns.

• Operational Insight:

- Staff allocation and customer service efforts can be optimized by recognizing these trends.
- For training or tip pooling, insights can help create fair distribution mode

- Sunday and Friday are prime days for tipping, both in terms of median and range.
- Saturday and Thursday show slightly lower medians and more outliers, indicating more variable tipping behaviour.
- A deeper statistical test (e.g. ANOVA) could confirm if these differences are statistically significant.

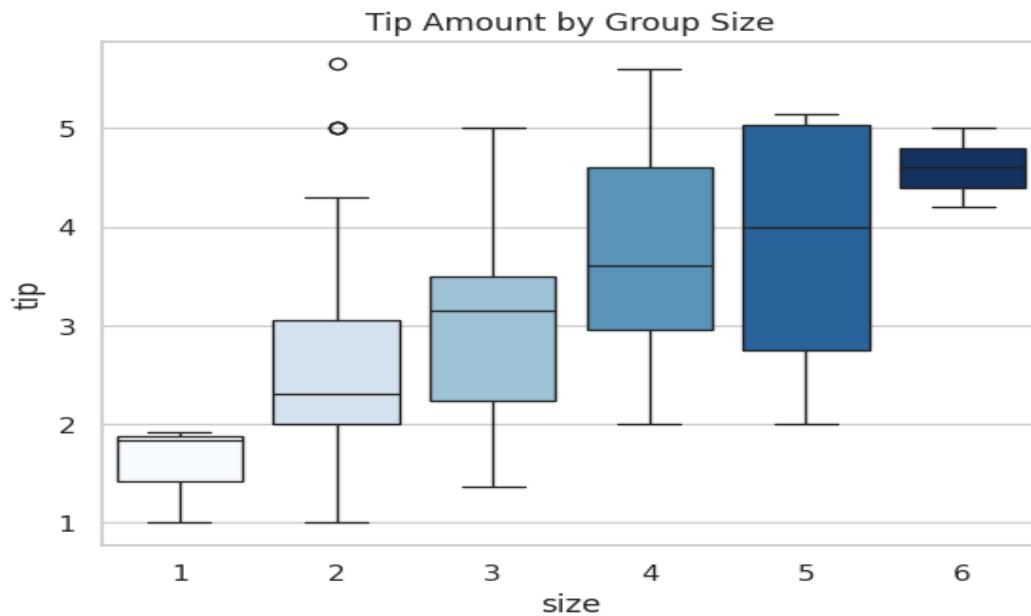
14. VARIATION % OF TIP WITH DAY



Key Takeaways:

1. Thursday has the most consistent tipping (smallest IQR), despite a notable outlier.
 2. Friday has the highest overall tip percentages, indicating potential higher customer generosity.
 3. Sunday and Saturday show more variability, suggesting inconsistent tipping behaviors.
 4. No clear day dominates in tipping percentage, but Friday shows a slightly higher upper range and median is comparable.
- Staffing Strategy: Consider allocating experienced servers to Fridays, where tipping is higher.
 - Promotion Timing: Leverage Thursdays for promotions as customer behavior is stable.
 - Training Focus: Improve service quality on weekends (Sat/Sun) to reduce tipping variability.

15. VARIATION OF TIP WITH SIZE



Key Insights for Business Development:

1. Target Larger Groups for Higher Revenue

- Tip amount increases consistently with group size.
- This suggests larger groups also spend more overall.
- Strategy: Offer promotions or reservation incentives for groups of 4+.

2. Optimize Staffing for Peak Tipping Opportunities

- Higher group sizes = higher tips, which may increase server motivation.
- Ensure experienced staff handle large tables to maximize customer satisfaction and sales.

3. Customize Offers by Group Size

- Solo diners: Quick-meal promotions or bundled offers.
- Groups: Combo platters, shared appetizers, and volume discounts.

4. Encourage Group Reservations

- Implement group booking systems and offer discounts or free items for groups of 5+.
- Promote via social media, especially during peak hours and weekends.

5. Staff Incentive Programs

- Align server bonuses or incentives with handling larger parties effectively.
- This supports better service delivery and potentially increases return visits.

Larger group sizes directly correlate with higher tipping behaviour. By tailoring marketing, staffing, and service offerings to attract and accommodate these larger groups, the business can increase both tip earnings for staff and overall revenue.

16. VARIATION OF TIPS WITH TIME OF DINING



1. Dinner Offers Higher Revenue Potential

- Higher median and upper tip amounts at dinner suggest larger bills and more generous tipping behavior.
- Business Action: Focus marketing efforts, premium offerings, and up-selling strategies during dinner service.

2. Lunch Crowds Are Predictable but Less Profitable

- While lunch tipping is steady, it trends lower.
- Business Action: Offer quick-turnover deals or set lunch menus to maintain profitability and efficiency.

3. Tailored Promotions by Time of Day

- Dinner: Emphasize multi-course meals, family combos, or special occasion packages.

- Lunch: Market to office workers or quick diners with speedy service promotions or meal deals.

4. Staffing Strategy

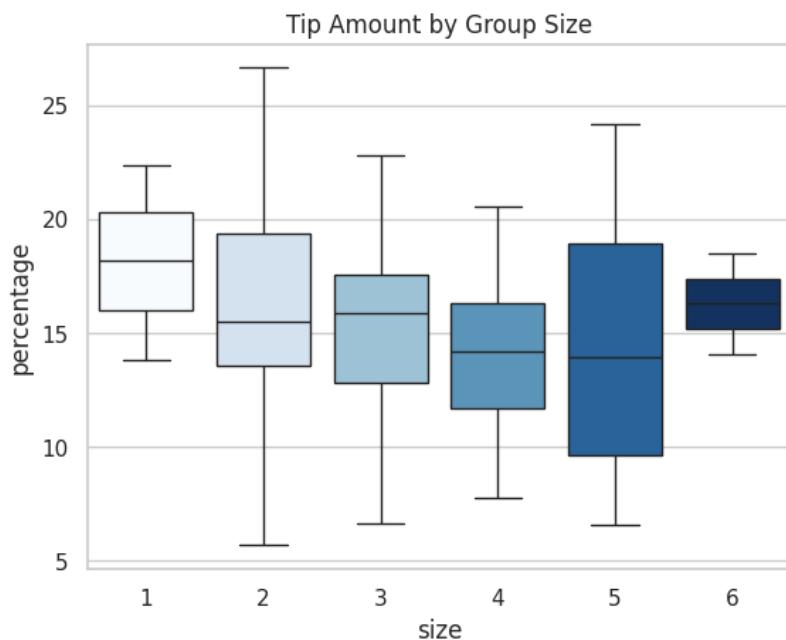
- Dinner requires more experienced servers who can handle upselling and larger tables.
- Lunch can be handled by smaller, faster teams for operational efficiency.

5. Monitor Outlier Behavior

- Occasional high tips during lunch (outliers) indicate opportunities for exceptional service to impact even short meals.
- Action: Train staff to maintain service quality across both shifts to capture these opportunities.

Dinner brings in higher and more varied tip amounts, implying better revenue opportunities. Lunch, while steady, offers less tip potential. Aligning marketing, staffing, and menu strategies with these patterns can help maximize profitability across both time slots.

17. VARIATION TIP % WITH GROUP SIZE



1. Tip Percentage Declines with Group Size (Until 6)

- Smaller parties (especially size 1) are more generous per dollar spent, suggesting high value in solo or couple diners.

- Action: Reward solo diners or couples with loyalty offers or personalized service to encourage retention.

2. Larger Groups Tip Less Consistently

- Tip percentages drop and vary more in mid-sized groups (3–5).
- Action:
 - Train servers to subtly remind about customary tipping.
 - Consider adding auto-gratuity for tables of 5+ to stabilize earnings.

3. Size 6 Shows Stabilization — Possibly Due to Policy

- Tipping % recovers and becomes more consistent.
- May suggest automatic gratuity kicks in at this point.
- Recommendation: If not already implemented, formalize automatic gratuity for groups of 6 or more.

4. Menu and Service Strategy by Group Size

- Solo/Couples: Upsell premium items, desserts, or drinks.
- Groups: Promote shared platters, group discounts, and split-bill ease to drive satisfaction and improve tip likelihood.

Summary:

As group size increases, tip percentages tend to decline and become more variable, particularly for sizes 3–5. Businesses can benefit by:

- Encouraging solo/couple diners through loyalty perks,
- Training staff to manage group dynamics effectively,
- Implementing automatic gratuity policies for larger parties, and
- Tailoring service approaches by group size.
