An Analysis of Tipping Behavior Using the “Tips” Dataset from Seaborn

By

Ousman Gaye

Abstract

This study explores tipping behavior across various factors, including gender, smoking status, day of the week, mealtime, and party size. Using data analysis techniques such as T-tests and ANOVA, we examined the differences in tipping amounts. Our findings indicate that males tend to tip slightly more generously than females, as evidenced by a higher median tip amount and the presence of higher-value outliers. Analysis by smoking status revealed no significant difference in tipping behaviors between smokers and non-smokers. Additionally, tips were observed to vary by day of the week, with a trend suggesting slightly higher tipping on weekends. When examining tipping behavior by mealtime, we found that dinner tips were generally higher than lunch tips, potentially due to larger total bills during dinner. Further, party size was found to significantly influence tipping behavior, with larger parties generally providing higher tips. However, simple linear regression analysis indicated a weak correlation between total bill amounts and tip percentages, suggesting other factors may also play a significant role in determining tip amounts. The results provide insights into the factors influencing tipping behavior, offering valuable information for businesses and servers to optimize their service strategies.

Introduction

Tipping behavior is a crucial aspect of the service industry, reflecting customer satisfaction, service quality, and cultural norms. The amount a customer chooses to tip can vary significantly based on a variety of factors, including the total bill amount, the gender of the customer, whether the customer is a smoker, the day of the week, the time of day, and the size of the dining party (Lynn, et al. 1993). These factors can provide insights into customer preferences

and behaviors, helping businesses tailor their services to enhance customer satisfaction and potentially increase revenue (Azar, 2005).

Understanding the factors that influence tipping is essential for restaurant managers and staff. By analyzing these factors, businesses can identify key drivers of tipping behavior and implement strategies to improve service quality and customer experience (Lynn, 2001). For instance, knowing that tips tend to be higher during dinner service might encourage a restaurant to allocate more experienced staff during these times. Similarly, understanding how tipping behavior varies between smokers and non-smokers or between different days of the week can help in designing targeted marketing and service improvement initiatives.

This study aims to analyze tipping behavior using the `tips` dataset available in Seaborn, a popular Python data visualization library. The dataset includes several variables that are likely to influence tipping behavior: total bill, tip, sex, smoker status, day of the week, time of day, and party size. By conducting a thorough analysis of this dataset, we can gain valuable insights into how these factors interact and affect the amount of tip given. The findings from this analysis can inform strategies for improving customer service and optimizing restaurant operations.

Problem Statement

Despite the significant role of tipping in the restaurant industry, there remains a lack of comprehensive understanding of the various factors that influence tipping behavior. Restaurant managers and staff often rely on intuition and anecdotal evidence to gauge how different variables, such as the total bill amount, customer demographics, dining time, and party size, impact the tips they receive. This lack of data-driven insights can lead to suboptimal service

strategies and missed opportunities to enhance customer satisfaction and increase revenue (Azar, 2005).

The purpose of this study is to address this gap by systematically analyzing tipping behavior using the `tips` dataset from Seaborn. The dataset provides a robust framework to examine how multiple factors—including total bill, customer gender, smoker status, day of the week, time of day, and party size—affect tipping amounts. By leveraging statistical tests and regression analysis, this study aims to uncover the underlying patterns and relationships that drive tipping behavior. The findings will offer actionable insights for restaurant managers, enabling them to implement targeted improvements in service delivery and customer experience.

Objectives

The primary objectives of this study are:

1. To explore the relationship between the total bill and the tip amount.
2. To examine how tipping behavior varies by customer gender.
3. To investigate if there is a difference in tipping patterns between smokers and non-smokers.
4. To analyze tipping trends across different days of the week.
5. To assess the impact of mealtime (lunch vs. dinner) on tipping behavior.
6. To understand how party size influences the amount of tip given.

Research Questions

Based on the objectives outlined, the following research questions will guide the analysis:

1. How does the total bill affect the tip amount given by customers?
2. Is there a significant difference in tipping behavior between males and females?
3. Do smokers tend to give different tips compared to non-smokers?
4. Are tipping patterns consistent across different days of the week?
5. Does the time of day (lunch vs. dinner) impact the amount customers tip?
6. How does party size influence the tipping behavior in restaurant settings?

Literature Review

Previous studies showed that the study of tipping behavior has garnered significant attention in the fields of economics, psychology, and hospitality management. Tipping is not only a social norm but also a critical component of income for many service employees, particularly in the restaurant industry. Understanding the determinants of tipping behavior can provide valuable insights for improving service quality and customer satisfaction. The literature review in this study contains only research articles on tipping behavior in the service industry, particularly in the restaurant industry. Factors such as total bills, sex and smoking status of the individual and how they affect their tipping behavior are reviewed.

Review of Literature

Azar (2005) conducted research on the economic theory of tipping. From his study Azar found that tipping serves as an incentive for service quality. He justifies his study by using the efficiency wage theory, which states that customers tip to reward good service and incentivize future performance. Other studies have shown that higher tips are often associated with better service ratings (Lynn & McCall, 2000). Additionally, tipping can be viewed as a way to reduce

the principal-agent problem by aligning the interests of customers (principals) and service providers (agents).

On the psychological factors in tipping, research indicates that tipping behavior is influenced by various cognitive and emotional factors. Social norms and expectations play a significant role; individuals often tip because they believe it is the socially appropriate action (Lynn, 2016). Emotional factors such as empathy and gratitude also impact tipping decisions. For instance, customers are likely to tip more when they feel a personal connection with the server or when they perceive the server to be in a difficult financial situation (Seligman, Finegan, & Hazard, 1987).

Demographic factors also have a profound impact on tipping behavior. Factors such as gender, age, and income level have been found to influence tipping behavior. Research suggests that men and women may tip differently, with men generally leaving larger tips than women (Lynn, et al. 1993). Furthermore, younger customers and those with higher incomes tend to tip more generously (Bodvarsson and Gibson, 1997).

The context and situation in which tipping occurs also affect tipping behavior. Studies have shown that tips vary based on the day of the week, time of day, and the size of the dining party. For example, tips are often higher during weekends and dinner hours when social dining is more common (Lynn and Grassman, 1990). Larger parties tend to leave higher total tips, although the tip percentage may decrease as party size increases (Freeman, et al. 1975).

The quality and attributes of service provided are critical determinants of tipping. Elements such as server friendliness, promptness, and attentiveness significantly impact tipping

behavior (Parrett, 2006). Experimental studies have demonstrated that small gestures by servers, such as introducing themselves by name or touching the customer lightly on the shoulder, can increase tips (Hornik, 1992).

With the advent of digital payment systems, the way tips are collected and perceived has also evolved. Research indicates that customers may tip differently when using digital payment methods compared to cash (Garrity, 2020). Moreover, environmental factors such as restaurant ambiance and cleanliness can indirectly influence tipping by affecting overall customer satisfaction (Lynn and McCall, 2016).

Analysis of Literature

The literature on tipping behavior highlights a complex interplay of economic, psychological, demographic, contextual, and service-related factors. This study aims to build on this body of knowledge by using the `tips` dataset from Seaborn to explore how these various factors influence tipping in a restaurant setting. By employing statistical tests and regression analysis, this study seeks to provide a comprehensive understanding of the determinants of tipping behavior, offering valuable insights for restaurant managers and staff to optimize their service strategies and enhance customer satisfaction.

Methodology

This section provides an overview of the data and the research methodology that will be used in analyzing the tips dataset. It includes the data source, estimation techniques, variable description and the different methods that will be used in the analysis. The chapter is broken down into different sections to capture all these in the analysis.

Data Collection

To perform the analysis on tipping behavior, the study utilizes the `tips` dataset from Seaborn, which includes variables such as total bill, tip, sex, smoker, day, time, and size. The data represents the tipping behavior of different categories of individuals, including gender (male or female) and smoker status (smoker or non-smoker), and examines how the total amount of tip given depends on factors such as total bill, day of the week, time of day, and party size. The data was chosen based on its completeness and the important information that can be deduced from it. The data can be model as follow:

$$\boldsymbol{tip = \beta_0 + \beta_1 total_{bill} + \beta_2 sex_i + \beta_3 smoker_i + \beta_4 day + \beta_5 time + \beta_6 size + \epsilon}$$

Where:

β_0 is the intercept

β_1, β_2, β_3, β_4, β_5, β_6 are the coefficients for each predictor variable

ϵ is the error term

The model aims to explain the variation in tips based on the given predictors.

Explanation of Variables

The variables from the equation above are explain under this section

tip: The dependent variable representing the amount of tip given by a customer.

total_bill: The independent variable representing the total amount of the bill. It is expected that higher bills might result in higher tips, although the exact relationship can vary.

sex_i: A categorical independent variable representing the gender of the customer (e.g., male or female). This variable can be encoded as a binary indicator (e.g., 0 for male and 1 for female).

smoker_i: A categorical independent variable representing whether the customer is a smoker (e.g., smoker or non-smoker). This variable can also be encoded as a binary indicator (e.g., 0 for non-smoker and 1 for smoker).

day: A categorical independent variable representing the day of the week (e.g., Monday, Tuesday, etc.). This variable can capture variations in tipping behavior across different days.

time: A categorical independent variable representing the time of day when the meal took place (e.g., lunch or dinner). This variable can show how tipping behavior differs between lunch and dinner.

size: An independent variable representing the size of the party (number of people at the table). Larger parties might tip differently compared to smaller parties.

Exploratory Data Analysis (EDA):

Descriptive Statistics for numerical variables (total bill and tip).

Descriptive statistics provide a comprehensive summary of the central tendency, dispersion, and shape of a dataset's distribution. For the tips dataset, which includes variables such as total bill, tip, sex, smoker, day, time, and size, descriptive statistics can offer valuable

insights into the data. Below is an explanation of the descriptive statistics for each key variable in the tips dataset:

Table 1 Descriptive Statistics

Variable	Mean	Std Dev	Minimum	Maximum	N
Total_bill	19.7859426	8.9024120	3.0700000	50.8100000	244
Tip	2.9982787	1.3836382	1.0000000	10.0000000	244

The descriptive statistics provided by the MEANS Procedure give a clear overview of the distribution and spread of the total bill and tip amounts in the tips dataset. Key takeaways include:

- **Total Bill:** The average total bill is around $19.79, with a considerable spread (standard deviation of $8.90) indicating variability in spending. The bills range from as low as $3.07 to as high as $50.81, showing a wide range of customer spending behavior.
- **Tip:** The average tip is approximately $2.998, with less variability (standard deviation of $1.38) compared to the total bill. Tips range from $1.00 to $10.00, suggesting that while tips do vary, they are generally within a narrower range compared to total bills.

Understanding these descriptive statistics helps in gaining initial insights into the data, forming hypotheses, and guiding further analysis to explore relationships between variables and tipping behavior in more detail.

Result Presentation and Analysis

This section presents and analyzes the empirical findings of the tips dataset. It is a crucial stage in the study, involving the summarization of findings in a clear and interpretable manner.

Visualization tools such as histograms, bar charts, box plots, and scatter plots are employed to illustrate key patterns, distributions, and relationships in the data. These visualizations make the data more accessible and highlight trends and anomalies that might be overlooked in the raw data.

Additionally, this section explores the relationships between variables using correlation coefficients and regression analysis, discussing how variables interact and influence each other. Significant findings that emerge from the analysis are highlighted. Statistical tests, including t-tests and ANOVA, are implemented to determine if observed differences are statistically significant.

Statistical Analysis

Statistical analysis is a fundamental aspect of data analysis that involves applying statistical methods to interpret, summarize, and draw conclusions from data. In the context of the tips dataset, statistical analysis helps to uncover patterns, relationships, and differences among various variables. Here's a brief explanation of the key statistical techniques used

1. **Total bills paid**

The total bill represents the amount spent by customers on their meals, excluding tips. In the tips dataset, the total bill amounts range from as low as $3.07 to as high as $50.81. The average total bill is approximately $19.79, with a standard deviation of $8.90, indicating significant variability in customer spending. This variability suggests diverse spending patterns, with some customers opting for modest meals while others spend considerably more. Analyzing

the total bill helps in understanding customer behavior, spending trends, and potentially, how the total bill amount influences tipping practices.

The MEANS Procedure

Analysis Variable: Total_bill

Sum
4827.77

The MEANS Procedure output indicates that the sum of the total_bill variable in the tips dataset is $4827.77. This statistic represents the total amount of money spent by all customers in the dataset combined excluding tips. By understanding the total bill sum, stakeholders can make informed decisions regarding pricing strategies, marketing efforts, and resource allocation to optimize revenue and improve customer satisfaction.

2. **Relationship between Total Bill and Tip**

The relationship between the total bill and the tip amount is a critical aspect of understanding customer tipping behavior. Typically, there is a positive correlation between the total bill and the tip, meaning that as the total bill increases, the tip amount also tends to increase. The relationship can be explained by implementing both correlation coefficient and regression analysis. The results are presented below

Correlation coefficient

The correlation coefficient quantifies the strength and direction of the relationship. A value close to +1 indicates a strong positive correlation. In the tips dataset, the correlation coefficient between total bill and tip is around 0.67, confirming a moderate to strong positive relationship.

Pearson Correlation Coefficients, N = 244 Prob > |r| under H0: Rho=0		
	total_bill	**tip**
total_bill	1.00000	0.67573 <.0001
tip	0.67573 <.0001	1.00000

The result from the table above shows that there is a significant positive correlation between the total bill and the tip amount, suggesting that higher bills tend to be associated with higher tips. This aligns with the common practice of tipping a percentage of the total bill. The p-value (< .0001) confirms that the correlation is statistically significant. This means we can confidently say that the relationship observed is real and not due to random chance.

Regression Analysis

A simple linear regression can model the relationship, with the total bill as the independent variable and the tip as the dependent variable. The resulting equation can predict the expected tip based on the total bill amount, helping in understanding and anticipating tipping patterns.

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t-Value**	**Pr > |t|**
Intercept	1	0.92027	0.15973	5.76	<.0001

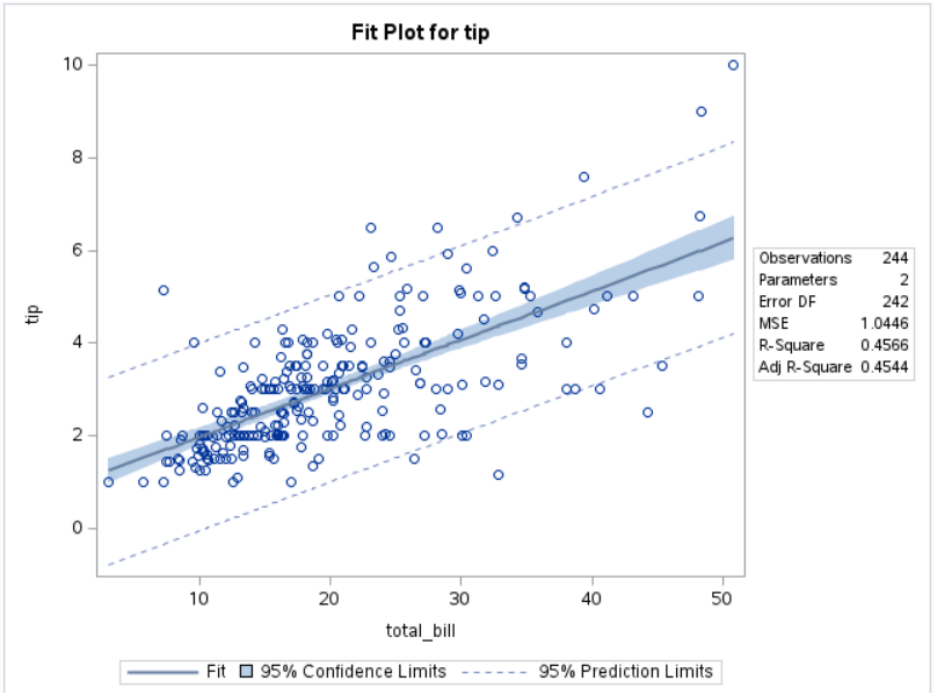
total_bill	1	0.10502	0.00736	14.26	<.0001
Adj R-Sq		0.4544			

From the table, the intercept (0.92027) and slope (0.10502) are both statistically significant ($p < .0001$). This means both parameters are essential for the model. The positive slope indicates a positive relationship between total bill and tip. For every dollar increase in the total bill, the tip increases by about $0.11 on average. The Adj R-Square value of 0.4544 means that 45.44% of the variability in tip amounts can be explained by the total bill amounts indicating a good fit with a slight adjustment for degrees of freedom.

Result Visualization

A scatter plot can be used to visualize the relationship between the total bill and the tip. This visualization provides a clear picture of the trend and direction between these two variables. By plotting each data point, we can observe how tips tend to increase with higher total bills, giving us insight into the tipping behavior of customers.

To better understand the relationship, we can also add a regression line to the scatter plot. This line represents the best fit through the data points and helps in visualizing the overall trend. Moreover, the correlation coefficient can be calculated to quantify the strength and direction of the relationship between total bill and tip. The scatterplot is depicted below



In summary, the scatter plot effectively visualizes the positive correlation between the total bill and the tip amount, while also highlighting the variability and potential outliers in tipping behavior. This visual representation helps in understanding the general tipping pattern and identifying areas for further analysis, such as examining the impact of other variables like service quality or customer demographics.

3. Tipping Behavior by Gender

Analyzing tipping behavior by gender involves comparing the average tip amounts given by male and female customers. By examining this, we can gain insights into whether there are any significant differences in tipping patterns between genders.

Analysis Variable: tip

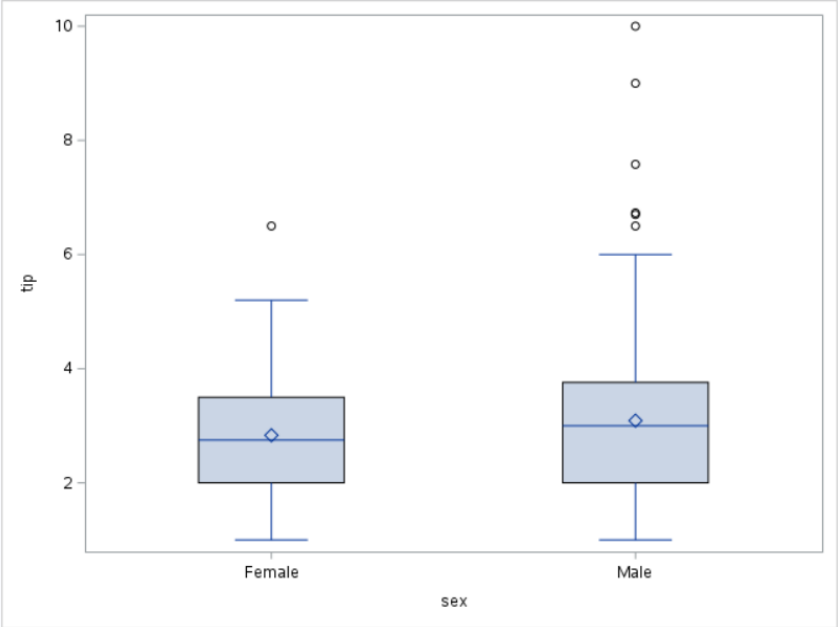
sex	N Obs	Mean	Std Dev	Pr > |t|
Female	87	2.8334483	1.1594945	0.1665

Male	157	3.0896178	1.4891018	0.1378

The data indicates that males, on average, tip slightly more than females. The mean tip for males is higher by about 0.26 dollars compared to females. Additionally, the variation in tipping amounts (as measured by the standard deviation) is greater among males than females, suggesting that male tipping behavior is less consistent. However, the t-tests yield p-values greater than 0.05 (0.1665 and 0.1378, respectively), indicating that the observed difference in mean tips between genders is not statistically significant at the 0.05 level. Overall, gender does not appear to have a significant impact on tipping behavior in this dataset and the assumption of equal variances between genders is violated. Further analysis could involve exploring other factors and interactions to gain deeper insights into tipping behavior.

Result Visualization

A box plot can be used to visualize the relationship between tipping behavior by gender. This visualization provides a clear depiction of the distribution, central tendency, and variability of tips given by females and males.



For females, the box plot shows the median tip amount of approximately $2.5, the interquartile range (IQR) from $2 to $4, and any outliers. The median represents the middle value of the tips given by females, while the IQR displays the range within which the middle 50% of the tips fall. Outliers, if any, indicate tips that are significantly higher or lower than the rest of the data.

For males, the box plot similarly presents the median tip amount of around $3.00, the IQR from about $2.00 to $4.00, and any outliers. This allows for a comparison between the tipping behaviors of males and females.

The box plot reveals a tendency for males to tip slightly more generously than females, as evidenced by the higher median and greater number of higher-value outliers. The spread of tipping amounts is also wider for males, indicating more variability in their tipping behavior. On the other hand, females exhibit a more consistent tipping pattern with fewer high-value outliers.

4. **Tipping behavior by smoker status**

Tipping behavior by smoker status examines whether there are differences in the amount of tips given by smokers compared to non-smokers. The analysis involves comparing the average tip amounts and the variability of tips within these two groups. Here are the key findings from our results.

Analysis Variable: tip

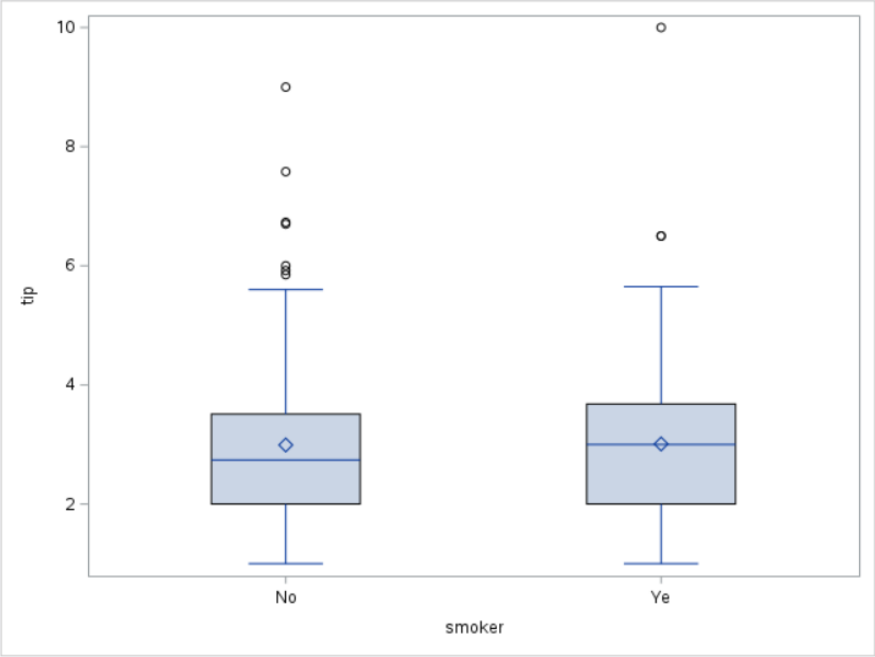
smoker	N Obs	Mean	Std Dev	Pr > |t|
No	151	2.9918543	1.3771901	0.9266
Yes	93	3.0087097	1.4014676	0.9269

The table from the MEANS procedure provides descriptive statistics for the variable tip based on the smoking status of customers. The result from the table shows that there are more observations for non-smokers (151) compared to smokers (93), indicating more non-smokers were surveyed or recorded in the dataset. On average, tipping behavior is nearly identical for smokers ($2.99) and non-smokers ($3.01), with smokers tipping slightly more on average. However, the difference is minimal. The variability or spread of tips is also very similar between the two groups, indicating that both smokers and non-smokers have a comparable range of tipping amounts. The identical behavior between the two groups can be justified from p-value results, with the t-tests yielding p-values greater than 0.05 (0.9266 and 0.9269, respectively), indicating that the observed difference in mean tips between smokers and non-smokers is not statistically significant. In summary, the analysis shows that there is no significant difference in tipping behavior between smokers and non-smokers. Both groups have nearly identical mean tips and similar variability in their tips. The statistical tests confirm that any observed differences are

not statistically significant. Thus, smoking status does not appear to have a significant impact on how much people tip.

Result Visualization

A box plot can be used to visualize the relationship between tipping behavior and smoker status. This visualization provides a clear picture of the distribution, central tendency, and variability of tips given by non-smokers and smokers.



The box plot for non-smokers (No) shows a median tip amount around $2.5 with an interquartile range (IQR) from $2 to $3.5. The mean tip amount is slightly higher than the median, indicating a slightly right-skewed distribution. Several outliers above the upper whisker suggest that some non-smokers gave significantly higher tips than the majority.

For smokers (Yes), the median tip amount is around $3, with an IQR from $2 to $4. The mean tip amount is close to the median, indicating a more symmetric distribution. There are

fewer outliers compared to non-smokers, with one notable outlier significantly higher than the rest.

Overall, the box plot shows that smokers tend to tip slightly more and exhibit more variability in their tipping behavior compared to non-smokers. This visualization effectively highlights the similarities and differences in tipping patterns between the two groups.

5. **Tipping trends by day of the week**

Analyzing tipping trends by day of the week can reveal interesting patterns in customer behavior and spending habits. Generally, tips tend to be higher on weekends (Friday to Sunday) compared to weekdays. This could be due to more leisure dining, celebrations, and relaxed spending habits. Understanding these trends helps businesses in the service industry optimize their operations, improve customer service, and anticipate revenue patterns. The table below reveals the findings

Analysis Variable: tip			
day	**N Obs**	**Mean**	**Std Dev**
Fri	19	2.7347368	1.0195771
Sat	87	2.9931034	1.6310143
Sun	76	3.2551316	1.2348803
Thu	62	2.7714516	1.2402232
Pr > F		0.1736	

From the means, it shows that Sunday has the highest average tip amount ($3.26), followed by Saturday ($2.99), Thursday ($2.77), and Friday ($2.73). The standard deviations

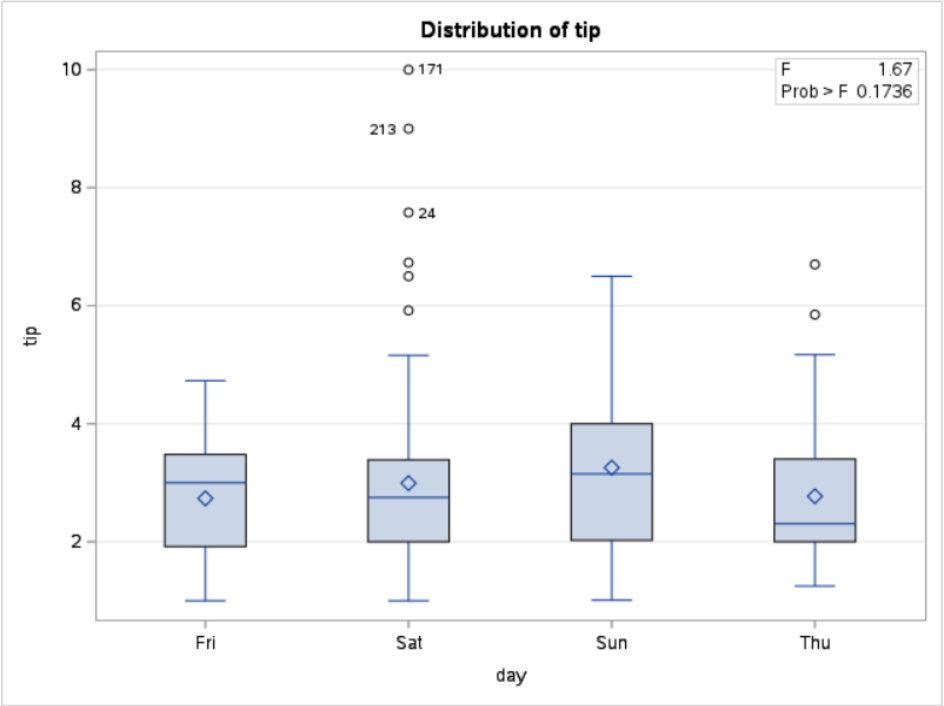
suggest that tipping behavior is most variable on Saturdays ($1.63) and least variable on Fridays ($1.02).

The F-value (1.67) and the associated p-value (0.1736) indicate that there is no statistically significant difference in tipping amounts between the days of the week at the 0.05 significance level. This means that while there are differences in mean tipping amounts, these differences are not large enough to be considered statistically significant given the sample data.

In summary, the analysis shows that while there are observable differences in average tips between different days of the week, these differences are not statistically significant according to the ANOVA test. The highest average tips occur on Sundays, and the variability in tipping is highest on Saturdays. However, the day of the week does not appear to be a strong predictor of tipping behavior overall.

Result Visualization

A box plot can be used to visualize the relationship between tipping trends by day of the week. This visualization provides a clear picture of the distribution, central tendency, and variability of tips given by non-smokers and smokers. The boxplot is depicted below:



The box plot visualizes the distribution of tips given on different days of the week: Friday, Saturday, Sunday, and Thursday. The median (represented by the line inside each box) appears to be slightly higher on Sunday compared to other days. Friday and Saturday have similar median tip amounts, slightly lower than Sunday. Thursday shows the lowest median tip amount among the days analyzed. The boxes represent the interquartile range, where the middle 50% of the data lies. The IQR is larger for Sunday, indicating greater variability in tips given on that day compared to other days. In general, while there are some differences in tipping behavior across different days of the week, with Sunday having slightly higher median tips and more variability, these differences are not statistically significant. This suggests that factors other than the day of the week may have a more substantial impact on tipping behavior.

6. **Tipping behavior by mealtime**

Tipping behavior by mealtime examines whether there are differences in the amount of tips given at various times of the day, such as breakfast, lunch, and dinner. Analyzing tipping behavior by mealtime provides valuable insights into customer tipping patterns, allowing for better resource allocation and service improvements to maximize tip earnings across different periods of the day. The table below depicts our findings:

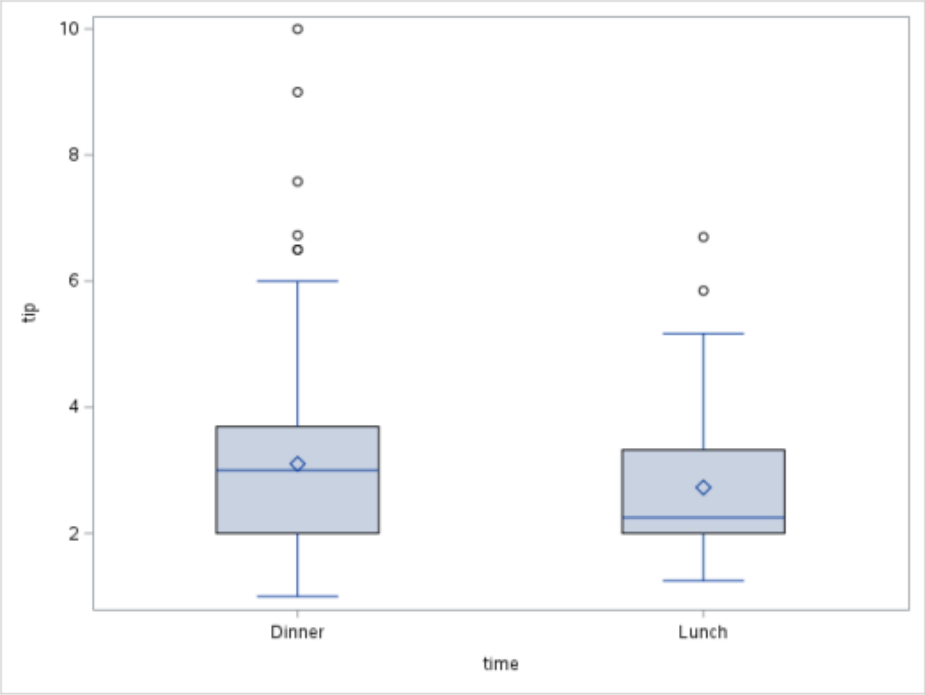
Analysis Variable: tip			
time	**N Obs**	**Mean**	**Std Dev**
Dinner	176	3.1026705	1.4362428
Lunch	68	2.728xxxx0882	1.2053454
Pr > F		0.0413	

From the table, the mean tip is higher for dinner ($3.10) compared to lunch ($2.73), suggesting that customers tend to tip more generously during dinner. The standard deviation for dinner is also higher, indicating more variability in tipping amounts. The t-test result shows a p-value of 0.0413, which is less than 0.05, indicating a statistically significant difference in tipping behavior between lunch and dinner.

In general, the analysis indicates that tipping behavior does vary by mealtime, with dinner generally attracting higher tips than lunch. This difference is statistically significant according to the Satterthwaite method. The variability in tipping amounts is similar across both mealtimes, as indicated by the equality of variances test. Understanding these patterns can help restaurants optimize their service and staffing strategies to potentially enhance tipping outcomes.

Result visualization

A box plot can be used to visualize the relationship between tipping trends by day of the week. This visualization provides a clear picture of the distribution, central tendency, and variability of tips given by non-smokers and smokers. The boxplot is depicted below:



The box plot displays the distribution of tips given during two different mealtimes: dinner and lunch. The median tip amount (represented by the line within each box) is higher for dinner compared to lunch. This suggests that on average, customers tend to tip more during dinner than lunch. The box plot suggests a tendency for customers to leave higher tips during dinner compared to lunch. The variability in tips is also greater during dinner, as evidenced by the larger interquartile range and the presence of more and higher outliers. This pattern could be influenced by several factors, such as different dining experiences, meal prices, or customer demographics at different times of the day.

7. Party Size on Tipping

Analyzing tipping trends by party size can provide valuable insights into customer behavior and spending habits. Larger parties might lead to higher total tips, but the tip percentage per person could decrease as the party size grows. Conversely, smaller parties may tip a higher percentage of the bill, reflecting a different approach to tipping. Understanding these patterns can help businesses tailor their service and improve customer satisfaction, potentially influencing tipping behavior positively. This analysis can also guide staff allocation, ensuring optimal service levels across different party sizes. We can gain more insight into this by analyzing the party size tipping data:

Analysis Variable: tip			
Size	**N Obs**	**Mean**	**Std Dev**
1	4	1.4375000	0.5065159
2	156	2.5823077	0.9855012
3	38	3.3931579	1.5573437
4	37	4.1354054	1.6406682
5	5	4.0280000	1.4401111
6	4	5.2250000	1.0531698
Pr > F		<.0001	

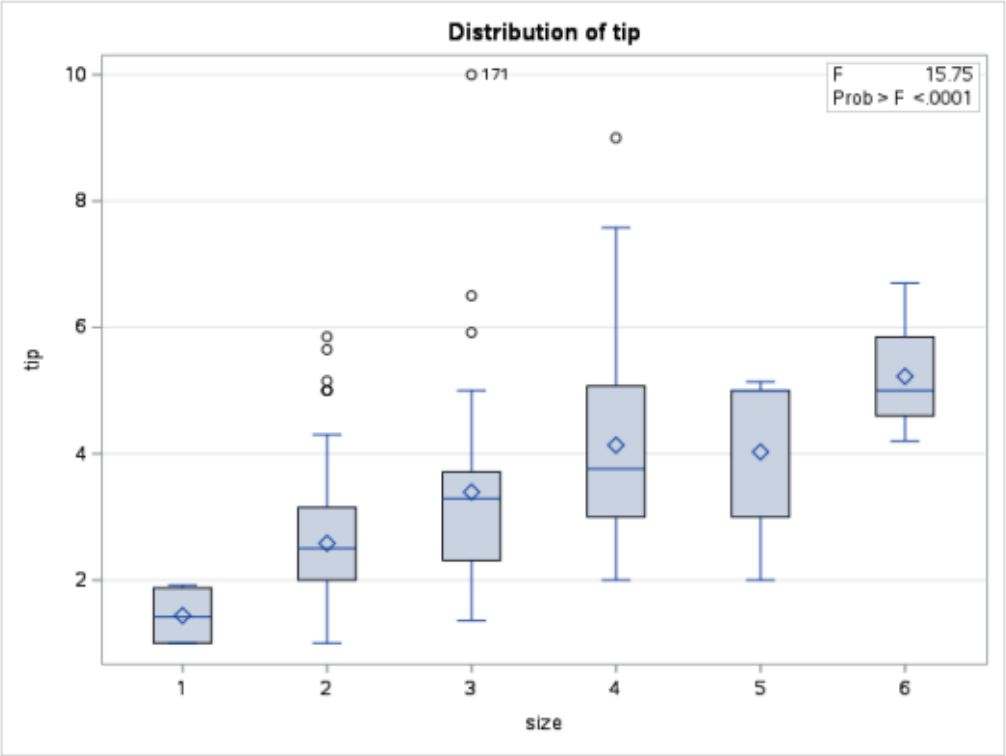
The tables provide an analysis of tipping behavior based on party size using both descriptive statistics and ANOVA. The mean tip amount tends to increase with party size. The

smallest parties (size 1) have the lowest mean tip ($1.44), while the largest parties (size 6) have the highest mean tip ($5.23). This trend indicates that larger parties tend to leave higher tips, likely due to larger total bills. Standard deviations increase with party size, suggesting greater variability in tipping behavior among larger groups. The significant F Value (15.75, $p < 0.0001$) indicates that there are statistically significant differences in mean tipping amounts across different party sizes.

In summary, the analysis highlights a clear trend where larger parties tend to leave higher tips, with significant differences observed between different party sizes. The variability in tipping also increases with the size of the party, indicating more diverse tipping behaviors in larger groups.

Result visualization

A box plot can be used to visualize the relationship between party size on tipping. This visualization provides a clear picture of the distribution, central tendency, and variability of tips given by non-smokers and smokers. The boxplot is depicted below:



The box plot illustrates the distribution of tips based on the size of the party (number of people). The median tip amount (represented by the line within each box) increases with party size, indicating that larger parties tend to leave higher tips. The highest median is observed for parties of size 6, while the lowest median is for parties of size 1. The analysis reveals a positive correlation between party size and the amount tipped, with larger parties generally giving higher tips. The increase in the median tip and the broader IQR for larger groups suggest that not only do larger parties tip more on average, but there is also greater variability in the amounts tipped. This information can be valuable for restaurants in planning and optimizing service strategies for different party sizes, potentially enhancing overall tipping income.

Conclusion

This analysis of tipping behavior using the tips dataset will contribute valuable insights into customer behavior in restaurants. By understanding the factors that influence tipping, restaurant managers can implement strategies to enhance service quality, improve customer satisfaction, and potentially increase revenue through higher tips. The findings from this study will serve as a foundation for informed decision-making and further research in the hospitality industry. Here are the key findings from our results. Furthermore, the project underscores the complexity of tipping behavior, influenced by a combination of demographic, situational, and possibly psychological factors. These insights can help service industry businesses better understand customer behavior, allowing for tailored service approaches that enhance customer satisfaction and potentially increase revenue.

References

Seaborn website: https: https://seaborn.pydata.org/.

Fernandez, S. et.al,. (2024). To tip or not to tip? Explaining tipping behavior in restaurants with service-inclusive pricing. *International journal of Hospitality Management.* https://doi.org/10.1016/j.ijhm.2023.103640.

L. Berkowitz (Ed.), Advances in Experimental Social Psychology, 2, Academic Press, New York (1965), pp. 267-299.

Int. J. Hosp. Manag., (2012). The underlying dimensions of tipping behavior: an exploration, confirmation, and predictive model 31 (1) (2012), pp. 247-256, 10.1016/j.ijhm.2011.05.00.

J. Appl. Psychol., Want a tip? Service performance as a function of emotion regulation and extraversion 96 (6) (2011), pp. 1337-1346, 10.1037/a0022884

Lynn, M. (2001). Restaurant tipping and service quality: A tenuous relationship. *Cornell Hotel and Restaurant Administration Quarterly, 42*(1), 14-20.

Koku, P. S., & Savas, S. (2016). Restaurant tipping and customers' susceptibility to emotional contagion. *Journal of Services Marketing, 30*(7), 762-772.

Lynn, M. (2015). Tipping in restaurants and around the globe: An interdisciplinary review. *Handbook of contemporary behavioral Economics*, 648-666.

Ross, J. A. (2018). The Politics of Service: Affectivity, Social Relations, and the Power of Tipping in the Restaurant Industry.