Statistics and Trends (Applied Data Science 1)

"Analysing Tipping Behavior: Insights from the Tips Dataset"

Student Name: **Aaradhya Begari**

Student ID: **23068805**

Abstract:

One popular dataset for data analysis and visualisation is the Tip dataset, which comes from a restaurant environment. Key characteristics such as total_bill, tip, sex, smoker, day, time, and size are included in its 244 observations. This dataset provides information on how gender, party size, and bill amounts affect the tipping habits of customers. By analysing this dataset, changes in tipping patterns and the correlations between factors can be investigated.
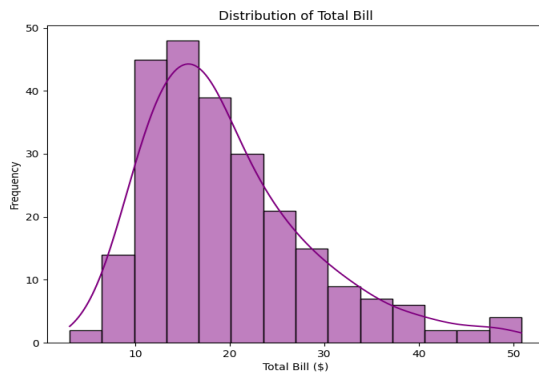
Histogram of Total Bill

Figure 1

The above visualization is the histogram for total_bill, which is the distribution of the customers' total bill amount, and it's overlayed with a Kernel Density Estimate curve that indicates the smooth approximation of the distribution. A smaller percent of the customers has substantially higher bills, while most have relatively lower total bills, so the distribution should be right-skewed. It is expected that the peak of the distribution would fall between 10 and 30 because these amounts are the most frequent total bills in this sample. Very few bills exceed $60, indicating that most dining experiences are likely to be casual rather than fine dining.

Boxplot of Tips by Day

Figure 2

The following boxplot displays the gender specific tip distribution across each day of the week: Each day has a different median tip amount; with Friday and Saturday have a larger median compared with other days. The smallest median tips occur on Sundays, implying that customers are generous at the end of the week. There are some outliers, mainly apparent on Saturday, when a few persons tipped an extraordinarily higher amount compared to others. Males tend to always leave a little bit more tips on most days, indicating the difference in tipping behavior between males and females.
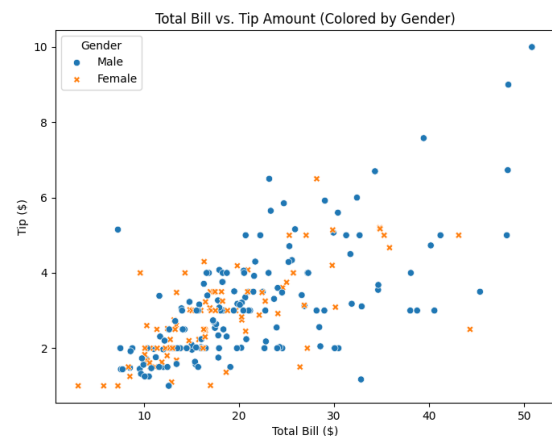
Scatterplot of Total Bill vs. Tip Amount

Figure 3

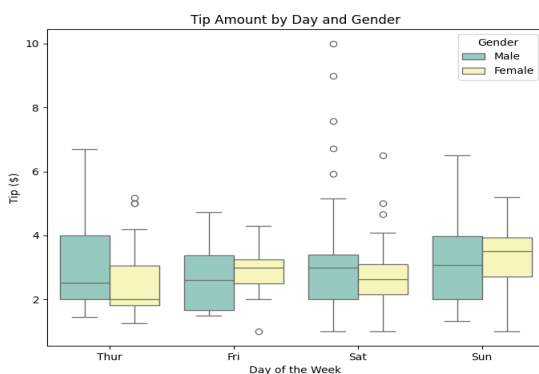The scatterplot below colours the points by gender and displays the relationship between the tip amount and the total bill amount. It can be seen from the graph that the relationship is positive, meaning that as the total bill increases, so does the tip between the total bill and the tip amount. For the same total bills, male consumers have wider variation in the dispersion of tips compared to their female customers. Most of the points fall between the smaller total bills of less than $20, suggesting that for both genders, most of the tipping behaviour falls in the same range for smaller bills while it is more dispersed for larger bills.

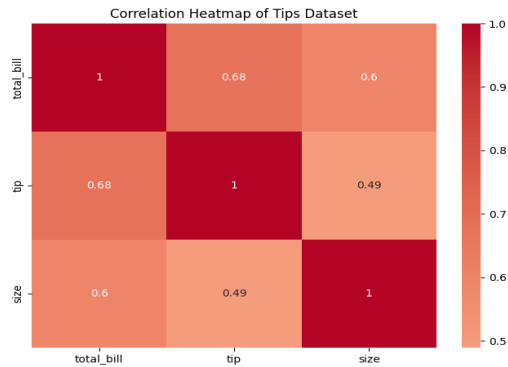**Heatmap of Correlation Between Numerical Features**

Figure 4

The correlation between the dataset's numerical features is shown visually in the heatmap. High positive value of the correlation coefficient ~ 0.68- between total_bill and tip confirms the scatterplot's conclusion that a bigger bill, in general, corresponds to a higher tip. This implies that the total bill goes hand in hand or increases with other characteristics such as size, which means the larger the gathering, the bigger the bills. Numerical properties are generally positively correlated; there is no obvious negative correlation in the data.

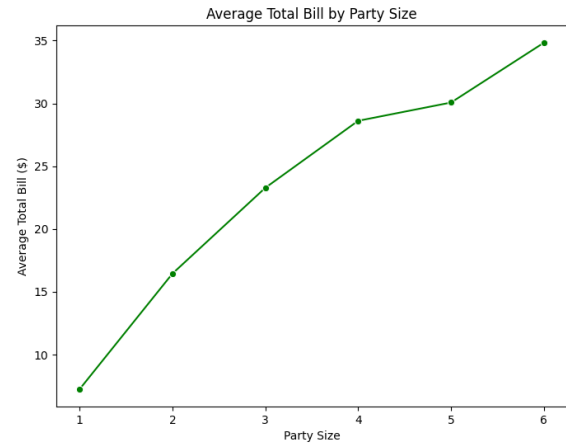**Line Plot of Average Total Bill by Party Size**

Figure 5

This line plot shows how the average total bill varies with party size. Larger groups tend to order more, increasing the cost, shown by mean total bill increased as party size increased in. This increase is not linear for the total bill, so it appears that each additional person contributes more and more to the total amount. This information could be used by restaurants for further modification of their marketing plans or prices in respect to large parties.

Summary:

The visualizations derived from the analysis of the Tips dataset provide an insight into consumer behavior at a dining environment. They show trends in tipping, customer demographics, and how different factors like party size and overall bill amounts affect tipping behavior. By being aware of these trends, a restaurant manager can improve client satisfaction and service tactics.

The following visualizations were created in an attempt to further understand the dynamics of customer interaction in a context of eating and how such analyses might be applied to improve one's business.