Data Wrangling Questions for Tips Dataset

1. What are the first five and last five rows of the dataset? (Use .head() and .tail().)

2. How many rows and columns are in the dataset?

3. What are the data types of each column, and are there any missing values?

4. What are the summary statistics (mean, median, min, max, etc.) for the numeric columns?

5. How many unique values are there in each categorical column?

6. Filter the data to include only records where the tip is greater than $5.

7. Select all rows where total\_bill is over $20 but the tip is less than $3.

8. Find all records where the meal was on a weekend (Saturday or Sunday).

9. Filter the dataset for rows where the person is a smoker.

10. Find all rows where the group size (size) is larger than 4.

11. Calculate the average tip by gender.

12. What is the total total\_bill amount for each day of the week?

13. Find the average tip given by smokers vs. non-smokers.

14. Calculate the average tip and total\_bill for each day and each time (Lunch or Dinner).

15. What is the average total\_bill by table size (size)?

16. Create a new column tip\_percentage as (tip / total\_bill) \* 100 and show the first 5 values.

17. Add a new column that categorizes each row as High or Low tip based on a threshold (e.g., tip greater than 15% of total\_bill).

18. Create a column that categorizes the total\_bill as Low, Medium, or High based on defined ranges (e.g., Low: <10, Medium: 10-20, High: >20).

19. Create a new column showing the difference between the total\_bill and the tip.

20. Find the maximum and minimum tip\_percentage by day of the week.

21. Sort the dataset by tip in descending order and display the top 5 rows.

22. Sort the data first by day and then by total\_bill within each day in ascending order.

23. Pivot the data to show the average tip by day and time of day.

24. Melt the data to transform the day and time columns into a single column with the corresponding values.

25. Reset the index of the dataset after filtering to re-index the rows.