1.Data Handling:

Q. How would you handle missing values in a dataset? Describe at least two methods?

Ans: Deletion Method: Remove rows with missing values (useful when missing data is minimal and random)

Imputation Method: Replace missing values with: a) Mean/median/mode for numerical columns b) Most frequent value for categorical columns c) Predicted values using statistical techniques like regression

Q. Explain why it might be necessary to convert data types before performing an analysis?

Ans: It is necessary to convert data types before performing an analysis as Using correct data types like integers instead of string for numeric data, optimizes memory usage and processing speed. Proper data types improve accuracy, performance, and prevent errors during analysis.

2.Statistical Analysis:

Q. What is a T-test, and in what scenarios would you use it? Provide an example based on sales data.

Ans: T-test is a statistical test used to compare the means of two groups to determine if the differences between them are statistically significant. It is commonly used when the sample size is small, and the data is approximately normally distributed.

Q. Describe the Chi-square test for independence and explain when it should be used. How would you apply it to test the relationship between shipping mode and customer segment?

Ans: The Chi-square test for independence is a statistical method used to determine whether there is a significant association between two categorical variables. It evaluates whether the observed frequencies in a contingency table differ from the expected frequencies under the assumption that the variables are independent.

3. Univariate and Bivariate Analysis:

Q. What is univariate analysis, and what are its key purposes?

Ans: Univariate analysis examines and summarizes a single variable in a dataset. Its key purposes include understanding the distribution, central tendency (mean, median, mode), dispersion (variance, standard deviation), and identifying outliers

Q. Explain the difference between univariate and bivariate analysis. Provide an example of each.

Ans: The difference Between Univariate and Bivariate Analysis

1. **Number of Variables**:
    - **Univariate Analysis**: Focuses on one variable at a time.
    - **Bivariate Analysis**: Examines the relationship between two variables.

2. **Purpose**:
    - **Univariate Analysis**: Describes the characteristics of a single variable
    - **Bivariate Analysis**: Identifies correlations, patterns, or dependencies between two variables.

4.Data Visualization:

Q. What are the benefits of using a correlation matrix in data analysis? How would you interpret the results?

Ans: Benefits of using a Correlation Matrix in Data Analysis are as follows

- Provides a snapshot of relationships in a dataset, which can guide hypothesis generation and further analysis.
- Informs regression models by highlighting variables that are strongly correlated with target variable
- It helps to detect and quantify the linear relationships in a dataset, which can guide hypothesis generation and further Analysis.

Q. How would you plot sales trends over time using a dataset? Describe the steps and tools you would use?

Ans: We need make sure our dataset contains at least contains two columns for data and time and one for sales figure.

Use data analysis tools like python libraries Pandas and Matplotlib /seaborn

We can make a Line plot for for showing trends over time using a dataset using seaborn

5. Sales and Profit Analysis:

Q. How can you identify top-performing categories based on total sales and profile? Describe the process?

Ans: We need to group the product category and aggregate the total sales and profit for each Python and sort the aggregate data by total sales and profit to identify the top performers, calculate the profit margin for each product category which is the ratio of profit to sales. Plot bar charts to visualize the top performing bases on Sales and profit. The product categories with the highest total sales and profits indicate the most profitable area.

Q. Explain how you would analyse seasonal sales trends using historical sales data?

Ans: we need to Group sales by time and calculate the average for each unit across years to identify patterns, then we can visualize the trends using bar charts or any other chart.

6.Grouped Statistics:

Q . Why is it important to calculate grouped statistics for key variables? Provide an example using regional sales data.

Ans: Grouped statistics are essential in data analysis because they allow us to summarize and understand the underlying patterns within different categories or groups of data. By calculating

grouped statistics, such as mean, sum, or count, we can gain insights that are not obvious when looking at the entire dataset. This helps to break down complex data into manageable segments, allowing for better decision-making.

For example, in sales Data, we calculate the average sales or profit for each product category rather than the entire dataset can reveal which categories are performing the best. This can guide inventory management and promotional strategies.