**1. What is EDA and why is it important?**  
EDA (Exploratory Data Analysis) is the process of analyzing datasets using statistical summaries and visualizations to uncover patterns, relationships, and anomalies.

* Importance: Helps understand the structure of data, detect missing/outlier values, check assumptions, and decide on modeling approaches.

**2. Which plots do you use to check correlation?**

* **Heatmap (correlation matrix)** → shows correlation coefficients between numeric variables.
* **Pairplot (scatterplot matrix)** → visualizes relationships between pairs of variables.
* **Scatterplots** → useful for two variables.

**3. How do you handle skewed data?**

* Apply transformations: log, square root, or Box-Cox.
* Use robust models less sensitive to skewness.
* Detect and treat outliers (capping, removal).
* Sometimes keep as is, if skewness is meaningful (e.g., income distribution).

**4. How to detect multicollinearity?**

* Correlation matrix (high correlations >0.8).
* VIF (Variance Inflation Factor) → VIF > 10 suggests multicollinearity.
* Condition Index method.

**5. What are univariate, bivariate, and multivariate analyses?**

* **Univariate:** Analyze one variable (histogram, boxplot, mean, median).
* **Bivariate:** Analyze relationship between two variables (scatterplot, correlation, t-test).
* **Multivariate:** Analyze relationships among 3 or more variables (pairplot, regression, PCA).

**6. Difference between heatmap and pairplot?**

* **Heatmap:** Shows numerical correlation values as colored matrix → quick overview.
* **Pairplot:** Plots scatterplots + histograms for each pair of variables → more detailed visualization.

**7. How do you summarize your insights?**

* State **main findings** clearly (missing data, skewness, outliers, correlations).
* Highlight **key relationships** (e.g., survival linked to class and gender).
* Provide **business or problem context** insights.
* End with **recommendations** (data cleaning steps, features to use, transformations).