**1. What is EDA and why is it important?**

**EDA (Exploratory Data Analysis)** is the process of analyzing and visualizing datasets to summarize their main characteristics, often using statistical graphics and data visualization tools.

**Importance:**

* Helps detect patterns, anomalies, and relationships.
* Validates assumptions through visual insights.
* Guides feature engineering and model selection.
* Identifies missing or incorrect data.
* Informs further statistical analysis or predictive modeling.

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

To check correlation between variables:

* **Heatmap** (shows correlation coefficients visually in a matrix).
* **Pairplot** (shows scatterplots for each pair of features and histograms for univariate distributions).
* **Scatterplot** (used for two continuous variables).
* **Correlation matrix** (numeric, but not a plot — often used with a heatmap).

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

To handle skewed data:

* **Log transformation** (log(x+1)) for right-skewed data.
* **Square root or cube root transformation**.
* **Box-Cox or Yeo-Johnson transformation**.
* **Winsorizing** (limiting extreme values).
* **Using robust models** that are not sensitive to skewness (e.g., tree-based models).

**4. How to detect multicollinearity?**

Multicollinearity can be detected using:

* **Variance Inflation Factor (VIF)**: VIF > 5 (or 10) indicates multicollinearity.
* **Correlation matrix**: Check for high correlation (>0.8 or < -0.8) between independent variables.
* **Condition number** of the matrix (used in regression diagnostics).

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

* **Univariate analysis**: Analysis of a single variable (e.g., histogram, boxplot, summary statistics).
* **Bivariate analysis**: Analysis between two variables to find relationships (e.g., scatterplot, correlation).
* **Multivariate analysis**: Analysis involving more than two variables (e.g., multiple regression, PCA, clustering).

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

| **Feature** | **Heatmap** | **Pairplot** |
| --- | --- | --- |
| Purpose | Shows correlation matrix visually | Shows pairwise scatterplots |
| Visuals | Colored grid | Multiple scatterplots and histograms |
| Data Type | Numeric correlations | Numeric and categorical (some cases) |
| Insight Type | Strength and direction of relationships | Patterns and distributions |

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

To summarize insights:

* **Highlight key patterns**: trends, correlations, outliers.
* **Use visuals**: plots with captions and titles.
* **Compare groups**: mention differences or similarities between categories.
* **Quantify findings**: e.g., “X is 20% higher in group A than B”.
* **Keep it actionable**: suggest how insights can inform decisions or next steps.