**Interview Questions**

**1. What is EDA and why is it important?**  
EDA (Exploratory Data Analysis) is the process of looking at your data using statistics and visuals to understand it better. It helps you find patterns, trends, mistakes, and missing values so you know how to clean the data and what questions to ask next.

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

* **Heatmap:** shows correlation values between all numerical columns.
* **Scatterplot:** shows how two variables move together — if they form a clear line, they’re correlated.
* **Pairplot:** shows scatterplots for all pairs of variables together.

**3. How do you handle skewed data?**  
You can:

* **Transform it** (log, square root) to make it more normal.
* **Remove outliers** if they don’t make sense.
* **Use robust methods** (like median) that are less affected by skewness.

**4. How to detect multicollinearity?**  
Check the **correlation matrix** — if two columns are highly correlated (close to +1 or –1), they might cause multicollinearity. For more precise checking, use **VIF (Variance Inflation Factor)** — a high VIF means multicollinearity.

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

* **Univariate:** Look at **one variable** at a time (e.g., histogram of age).
* **Bivariate:** Look at **two variables** together (e.g., scatterplot of age vs. salary).
* **Multivariate:** Look at **more than two variables** (e.g., pairplot, multiple regression).

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

* **Heatmap:** Shows numbers — it’s a colored table of correlation values.
* **Pairplot:** Shows **scatterplots** for all variable pairs and histograms on the diagonal — more visual detail.

7. **How do you summarize your insights?**  
Write clear points about what you found:

* Are there trends?
* Any strong relationships?
* Any outliers?
* What does this mean for the next step?  
  Keep it short and simple — focus on what’s **useful and interesting** in your data.