**Interview Questions and Answers**

**1. What is EDA, and why is it important?**  
EDA (Exploratory Data Analysis) is the process of analyzing datasets to summarize their main characteristics using statistics and visualizations.  
**Why it's important:**

* Understands data distribution and patterns
* Identifies missing values and outliers
* Helps in feature selection and engineering
* Prevents errors in further analysis or modeling

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

* Heatmap (shows correlation matrix)
* Pairplot (scatter plots between all pairs)
* Scatter plot (for two numeric variables)
* Bubble chart (optional, for 3 variables)

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

* Log Transformation: np.log(x) or np.log1p(x)
* Square Root or Box-Cox transformation
* Winsorization (limits outliers)
* Binning (converts numerical to categorical)
* Use RobustScaler (scikit-learn)

**4. How to detect multicollinearity?**

* Correlation matrix (look for correlation > 0.8 or < -0.8)
* VIF (Variance Inflation Factor):

python

CopyEdit

from statsmodels.stats.outliers\_influence import variance\_inflation\_factor

VIF = [variance\_inflation\_factor(X.values, i) for i in range(X.shape[1])]

* VIF > 5 or 10 indicates multicollinearity

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

* **Univariate**: One variable (e.g., histogram, boxplot)
* **Bivariate**: Two variables (e.g., scatter plot, correlation)
* **Multivariate**: More than two variables (e.g., pairplot, regression, PCA)

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

* **Heatmap**: Color-coded matrix showing correlation between variables
* **Pairplot**: Grid of scatter plots and histograms showing relationships and distributions  
  Use heatmap for correlation; pairplot for overall pairwise visual analysis.

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

* Use simple language to explain findings
* Back insights with charts or tables
* Mention key patterns, trends, or outliers
* Highlight business or analytical implications  
  **Example**:  
  "Students with higher parental education tend to score better in math and reading. Gender differences are noticeable in writing scores."