Data Science (Basic !’ Advanced) Top 25 One-Liner Interview Questions & Answers

1. What is Data Science? – Extracting insights from data using statistics, programming, and domain knowledge.

2. Data Science vs Data Analytics? – Analytics explains what happened; data science often predicts/optimizes what will happen.

3. What is the typical DS workflow? – Collect !’ clean !’ explore !’ model !’ evaluate !’ deploy !’ monitor.

4. What is EDA? – Exploratory Data Analysis to understand patterns, outliers, and relationships.

5. What is data cleaning? – Fixing missing values, duplicates, inconsistent formats, and noisy data.

6. How do you handle missing values? – Drop, impute (mean/median/mode), model-based imputation, or use “missing” category.

7. Mean vs median—when to use which? – Median for skewed data/outliers; mean for symmetric distributions.

8. What is an outlier? – A point far from others that may be error, rare event, or meaningful signal.

9. What is standardization vs normalization? – Standardization scales to mean 0/std 1; normalization scales to a fixed range (e.g., 0–1).

10. What is feature engineering? – Creating better input variables to improve model performance.

11. What is one-hot encoding? – Converting categorical variables into binary indicator columns.

12. What is train/validation/test split? – Train fits the model, validation tunes it, test estimates final generalization.

13.What is cross-validation? – Repeated training/testing on different folds to reduce evaluation variance.

14.What is data leakage? – Using information in training that wouldn’t be available at

prediction time.

15. Bias vs variance? – Bias is underfitting; variance is overfitting due to sensitivity to data.

16. How do you prevent overfitting? – Regularization, simpler models, more data, early stopping, and cross-validation.

17. What is regularization? – Penalizing complexity to improve generalization (e.g., L1/L2).

18. L1 vs L2 regularization? – L1 promotes sparsity (feature selection); L2 shrinks weights smoothly.

19. Classification vs regression? – Classification predicts classes; regression predicts continuous values.

20. Accuracy vs precision vs recall? – Accuracy is overall correctness; precision is “correct positives”; recall is “found positives.”

21. What is F1-score? – Harmonic mean of precision and recall, useful for imbalanced classes.

22. What is ROC-AUC? – Measures ranking quality across thresholds (probability discrimination).

23. What is a confusion matrix? – A table of TP/FP/TN/FN showing classification outcomes.

24. What is the difference between Bagging and Boosting? – Bagging reduces variance via averaging; boosting reduces bias by sequentially correcting errors.

25. How do you deploy and monitor a model? – Serve via API/batch, track drift/latency/accuracy, retrain when performance drops.