

SECTION A — MCQs (10 Questions)

Q1. (AI Basics)

Which statement best defines Artificial Intelligence?

- A. Writing rule-based programs
- B. Making computers behave intelligently by learning and decision-making
- C. Only building robots
- D. Only deep learning models

Answer:

Q2. (Types of AI)

ChatGPT, YouTube recommendations, and face unlock systems are examples of:

- A. AGI
- B. ASI
- C. Narrow AI
- D. Biological Intelligence

Answer:

Q3. (ML Concept)

What is the key difference between **traditional programming** and **machine learning**?

- A. ML uses Python
- B. ML learns patterns from data instead of fixed rules
- C. ML does not need data
- D. ML always uses neural networks

Answer:

Q4. (Learning Types)

Which learning type works **without labels**?

- A. Supervised
- B. Reinforcement
- C. Unsupervised
- D. Deep Learning

Answer:

Q5. (Loss Function)

Which loss function is most suitable for **regression**?

- A. Cross-entropy
- B. Accuracy
- C. Mean Squared Error
- D. Hinge loss

Answer:

Q6. (Data Types)

Which of the following is an **ordinal categorical variable**?

- A. Gender
- B. City
- C. Education level (10th < 12th < B.Tech < M.Tech)
- D. Color

Answer:

Q7. (EDA – Distribution Insight)

If a dataset has **mean > median**, the distribution is most likely:

- A. Symmetric
- B. Left-skewed
- C. Right-skewed
- D. Uniform

Answer:

Q8. (Outliers)

Why is the **IQR method** preferred over Z-score for skewed data?

- A. Faster computation
- B. Uses mean and standard deviation
- C. Robust to extreme values
- D. Assumes normal distribution

Answer:

Q9. (Coding MCQ – Data Leakage)

```
scaler.fit_transform(X)  
X_train, X_test = train_test_split(X)
```

What is the main problem?

- A. Wrong scaler
- B. Scaling should be after splitting
- C. Test size missing
- D. No target variable

Answer:

Q10. (Time Series)

Which is the **correct split strategy** for time-series data?

- A. Random train-test split
- B. K-Fold cross-validation
- C. Train on future, test on past
- D. Train on past, validate on future

Answer:

SECTION B — Short Descriptive (5 Questions)

Q11.

Why did **AI Winters** occur in the history of AI?

Expected points:

Q12.

Explain why **dropping rows with missing values** can be harmful for small datasets.

Expected points:

Q13.

What does “**EDA is a conversation with data**” mean?

Expected points:

Q14.

Why can **accuracy be misleading** for imbalanced datasets?

Expected points:

Q15.

Why is **random shuffling dangerous in time-series problems**?

Expected points:

SECTION C — Descriptive + Coding Logic (3 Questions)

Q16. (EDA → Model Choice)

EDA shows:

- Price is right-skewed
- Area strongly correlates with price
- City changes the price-area relationship

What preprocessing and models would you try, and why?

Expected:

Q17. (Coding Logic – Data Leakage)

Explain step-by-step how you would:

1. Split data
2. Scale features
3. Avoid data leakage

(code needed.)

Q18. (Big Picture Understanding)

Explain the relationship between **AI** → **ML** → **Deep Learning** → **Generative AI** using a real-world example.

