

Introduction to Data Science Concepts

Data Science · Answer Key · 8 Questions

1. Which of the following programming languages is commonly used in data science for its extensive libraries like Pandas and NumPy?

- A) Java
- B) Python**
- C) C++
- D) Ruby

2. In data science, what is the primary purpose of data cleaning?

- A) To add more features to the dataset
- B) To remove irrelevant data and correct errors**
- C) To create complex visualizations
- D) To build predictive models

3. What type of data is represented by categories or labels, such as 'red', 'blue', or 'green'?

- A) Numerical data
- B) Categorical data**
- C) Time-series data
- D) Text data

4. Which statistical measure represents the average value of a dataset?

- A) Median
- B) Mode
- C) Standard Deviation
- D) Mean**

5. What is the term for algorithms that learn patterns from data without explicit programming for each task?

- A) Supervised Learning
- B) Unsupervised Learning
- C) Reinforcement Learning
- D) Machine Learning**

6. Which data visualization technique is best suited for showing the distribution of a single numerical variable?

- A) Scatter Plot
- B) Bar Chart
- C) Histogram**
- D) Line Graph

7. In the context of machine learning, what is 'feature engineering'?

- A) Selecting the target variable
- B) Creating new input features from existing data**
- C) Evaluating the model's performance
- D) Choosing the algorithm

8. What does the term 'overfitting' refer to in machine learning?

- A) A model that performs poorly on training data
- B) A model that performs well on training data but poorly on unseen data**
- C) A model that is too simple to capture the underlying patterns
- D) A model that is computationally too expensive