

Data Science Foundations

Data Science · Answer Key · 8 Questions

1. Which programming language is most widely used in data science for its extensive library support and simple syntax?

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

2. What is the primary purpose of a 'bar chart' in data visualization?

- A) To show the relationship between two continuous variables
- B) To represent parts of a whole as percentages
- C) To compare quantities across different categories**
- D) To display trends over a specific time period

3. In a dataset, what is the term for the numerical value that appears most frequently?

- A) Mean
- B) Median
- C) Mode**
- D) Range

4. Which library is the standard tool for data manipulation and analysis in Python?

- A) Pandas**
- B) Pygame
- C) Tkinter
- D) Requests

5. What is the definition of a 'supervised' machine learning model?

- A) The model learns from unlabeled data without instructions
- B) The model is trained on a dataset containing input-output pairs**
- C) The model requires no data to function
- D) The model is used only for image compression

6. What is the correct definition of the 'median' in a set of numbers?

- A) The average of all values
- B) The difference between the highest and lowest value
- C) The middle value when the numbers are ordered**
- D) The total sum of all values

7. Which file format is most commonly used for storing structured data tables to be processed by data science applications?

A) CSV

B) EXE

C) ISO

D) DLL

8. In machine learning, what do we call the process of cleaning and preparing raw data before analysis?

A) Data mining

B) Data wrangling

C) Data visualization

D) Data caching