30 Student Dataset Discovery Practice Questions

Notes for SHS | Topic 84 | Updated 25 November 2026 | Original educational resource

Question 1: Define dataset

What is a dataset rather than a website?

Difficulty: Foundation. Skill: dataset research and analysis.

Your answer: _____

Question 2: Records and variables

In a rectangular table, what do rows and columns usually mean?

Difficulty: Foundation. Skill: dataset research and analysis.

Your answer: _____

Question 3: Features and targets

Which variable is the classification target in Iris?

Difficulty: Foundation. Skill: dataset research and analysis.

Your answer: _____

Question 4: Structured data

Is a CSV of country-year population counts structured?

Difficulty: Foundation. Skill: dataset research and analysis.

Your answer: _____

Question 5: Unstructured data

Give one example of an image dataset.

Difficulty: Foundation. Skill: dataset research and analysis.

Your answer: _____

Question 6: CSV

What does CSV usually store and what can go wrong?

Difficulty: Foundation. Skill: dataset research and analysis.

Your answer: _____

Question 7: JSON

Why do web APIs often use JSON?

Difficulty: Foundation. Skill: dataset research and analysis.

Your answer: _____

Question 8: Authoritative source

Which site should you choose for UCI Iris?

Difficulty: Foundation. Skill: dataset research and analysis.

Your answer: _____

Question 9: Metadata

What information does a data dictionary provide?

Difficulty: Foundation. Skill: dataset research and analysis.

Your answer: _____

Question 10: Dataset versus portal

Is Kaggle as a whole one dataset?

Difficulty: Foundation. Skill: dataset research and analysis.

Your answer: _____

Question 11: Classification choice

Choose a task from the Digits dataset.

Difficulty: Intermediate. Skill: dataset research and analysis.

Your answer: _____

Question 12: Regression choice

Which provided benchmark uses a numeric prediction outcome?

Difficulty: Intermediate. Skill: dataset research and analysis.

Your answer: _____

Question 13: Time series choice

Which source includes daily station weather measurements?

Difficulty: Intermediate. Skill: dataset research and analysis.

Your answer: _____

Question 14: Visualization choice

What chart suits country population by year?

Difficulty: Intermediate. Skill: dataset research and analysis.

Your answer: _____

Question 15: Download CSV safely

What should happen before opening a CSV in Excel?

Difficulty: Intermediate. Skill: dataset research and analysis.

Your answer: _____

Question 16: Missing marker

Can a -200 air-quality sensor value be a real concentration?

Difficulty: Intermediate. Skill: dataset research and analysis.

Your answer: _____

Question 17: Train/test split

Why keep a held-out test set?

Difficulty: Intermediate. Skill: dataset research and analysis.

Your answer: _____

Question 18: Documentation

Why is metadata necessary before computing revenue?

Difficulty: Intermediate. Skill: dataset research and analysis.

Your answer: _____

Question 19: Licence interpretation

Does a free public download prove free redistribution?

Difficulty: Intermediate. Skill: dataset research and analysis.

Your answer: _____

Question 20: Citation

Write a correct generic dataset citation template.

Difficulty: Intermediate. Skill: dataset research and analysis.

Your answer: _____

Question 21: Outdated data

Can 1994 Adult income data be assumed current?

Difficulty: Applied judgement. Skill: dataset research and analysis.

Your answer: _____

Question 22: Selection bias

Why may Portuguese school outcomes fail to represent all students?

Difficulty: Applied judgement. Skill: dataset research and analysis.

Your answer: _____

Question 23: Sensitive fields

What is risky about demographic predictors?

Difficulty: Applied judgement. Skill: dataset research and analysis.

Your answer: _____

Question 24: Leakage

Which bike-sharing fields directly compose cnt?

Difficulty: Applied judgement. Skill: dataset research and analysis.

Your answer: _____

Question 25: Prediction and causation

Does association between acidity and wine quality prove causation?

Difficulty: Applied judgement. Skill: dataset research and analysis.

Your answer: _____

Question 26: Unsafe sharing

Should a student publish raw restricted taxi records with person-level guesses?

Difficulty: Applied judgement. Skill: dataset research and analysis.

Your answer: _____

Question 27: Portfolio choice

Which beginner dataset gives compact numeric flower records?

Difficulty: Applied judgement. Skill: dataset research and analysis.

Your answer: _____

Question 28: CSV vs Parquet

Why might NYC taxis be in Parquet?

Difficulty: Applied judgement. Skill: dataset research and analysis.

Your answer: _____

Question 29: Ethical project

Design a safe health benchmark study.

Difficulty: Applied judgement. Skill: dataset research and analysis.

Your answer: _____

Question 30: Complete report

Name key sections of a dataset selection report.

Difficulty: Applied judgement. Skill: dataset research and analysis.

Your answer: _____

Verify third-party licences and dataset access conditions at their authoritative source before reuse. Third-party datasets are NOT included.