

Data Mastery: Conquer the Information Age in 11th Grade

Analyze complex visual patterns and identify algorithmic bias to make smarter decisions in policy, tech, and marketing careers.

1. A public health researcher tracks a disease outbreak using 'proxy data' such as an increase in searches for 'fever' rather than clinical reports. What is the primary risk of using this data source?

- A. The data is too localized to provide national trends.
- B. Correlation does not imply causation; searches may spike due to media coverage.
- C. Search engine data is strictly private and cannot be used by scientists.
- D. Proxy data is inherently more accurate than clinical lab tests.

2. When building a machine learning model, using a 'representative sample' means that the demographics of the training data should match the demographics of the population it affects.

- A. True
- B. False

3. In the context of data ethics and privacy, the process of removing personally identifiable information (PII) from a dataset is known as _____.

- A. Data Mining
- B. Anonymization
- C. Encryption
- D. Indexing

4. You are examining a 'Heat Map' used by a city council to show crime rates. If the map uses very large geographic blocks, it might mask high-crime 'hotspots' or exaggerate safety in others. This is an example of:

- A. The Modifiable Areal Unit Problem (MAUP)
- B. Data Redundancy
- C. The Hawthorne Effect
- D. Data Normalization

5. Data 'veracity' refers specifically to the speed at which new data is generated and transmitted across a network.

- A. True
- B. False

6. A specialized software tool used to store, manipulate, and analyze geographic or spatial data is called a _____.

- A. CMS (Content Management System)
- B. GIS (Geographic Information System)

Name: _____ Date: _____

- C. CRM (Customer Relationship Manager)
- D. ERP (Enterprise Resource Planner)

7. Which of the following describes 'Survivor Bias' in data interpretation?

- A. Over-representing data from failures to avoid making the same mistakes.
- B. Focusing only on the people or things that made it past a selection process.
- C. The tendency to prefer data that confirms our existing personal beliefs.
- D. An error in data entry where 'zero' values are recorded as 'null' values.

8. Open Data initiatives are projects where governments and organizations make their datasets freely available for anyone to use and republish.

- A. True
- B. False

9. The 'unit of analysis' in a study refers to the _____ unit that is being investigated (e.g., an individual, a city, or a country).

- A. Most complex
- B. Most expensive
- C. Smallest
- D. Randomized

10. When evaluating a data visualization, which feature is most likely used to 'lie' or mislead the viewer without changing the raw data values?

- A. Using a descriptive title.
- B. Truncating the y-axis (not starting it at zero).
- C. Including a link to the data source.
- D. Labeling all units of measurement clearly.