

The quiz is to test your understanding of concepts of intelligent agents and practical problem solving.

Q1: About data for AI [4 + 16 = 20 points]

Instructions: Give your answers in bullet points.

a) What is open data? Given an example of open data that you produce which others can use?

[2 + 2 = 4]

- **Open data:** Data that can be **freely used, modified, and shared by anyone for any purpose** (usually with an open license).
- **ExampleCrime data:** [Here](#), usually anything with “.gov” i use for public data.

b) You are analyzing a dataset and some attributes are missing.

b.1) What could be any 2 reasons why they are missing? [2 + 2 = 4]

- The value **was not provided** or **does not exist / has no practical interpretation**.
- The value is **hidden/redacted for privacy** (or considered **not reliable**).

b.2) What are any 2 ways you can still proceed with data analysis despite the missing values.

For each, mention

what assumption you are making and what are its risks. [(2+2+2) \* 2 = 12]

**Method 1: Omission (drop rows/columns with missing values).**

- **Assumption:** Missingness is **MCAR** (Missing Completely At Random), so dropping cases won't bias relationships.
- **Risk: Less data** (smaller sample) and potential bias if the MCAR assumption doesn't hold.

**Method 2: Simple imputation (e.g., mean/median for numeric; a default/mode for categorical).**

- **Assumption:** Missingness is roughly **MAR** (Missing At Random); the **summary/default value** is a reasonable proxy.
- **Risk: Shrinks variance**, can **dampen correlations** and bias results if the MAR assumption is false (esp. MNAR).