

Q4. On the basis of which attributes, data will be called good or bad?

Ans:

Good Data	Bad Data
❑ Information is well structured ❑ It is accurate ❑ It is consistent ❑ It is cleanly presented ❑ Contains information which is relevant to our requirement	❑ Information is scattered ❑ Contains a lot of incorrect values ❑ Contains missing and duplicate values ❑ It is poorly presented ❑ Contains information which is not relevant to our requirement

Q5. Which are the associated terms with data acquisition from websites?

Ans:

Data acquisition from websites



Q6. What are the Ethical concerns in data acquisition?

Ans: While gathering data and choosing datasets, certain ethical issues can be addressed before they occur:

Bias	Take steps to understand and avoid any preferences or partiality in data
Consent	Take necessary permissions before collecting or using an individual's data
Transparency	Explain how you intend to use the collected data and do not hide intentions
Anonymity	Protect the identity of the person who is the source of data
Accountability	Take responsibility for your actions in case of misuse of data

Q7. Which factors determine the usability of data?

Ans: There are three primary factors determining the usability of data:

1. **Structure-** Defines how data is stored.
2. **Cleanliness-** Clean data is free from duplicates, missing values, outliers, and other anomalies that may affect its reliability and usefulness for analysis. In this particular example, duplicate values are removed after cleaning the data.
3. **Accuracy-** Accuracy indicates how well the data matches real-world values, ensuring reliability. Accurate data closely reflects actual values without errors, enhancing the quality and trustworthiness of the dataset.