

DATA SCIENCE COURSE TUTORIAL # 09

2.5 Metadata and Data Documentation

1: Metadata

Metadata means "data about data." It gives extra information that explains the actual dataset. Metadata helps us understand what the data represents, how it was collected, and how it should be used. Without metadata, data can be confusing and hard to interpret.

Examples of Metadata:

- Column names and their meanings
- Units of measurement (kg, cm, USD)
- Time and location of data collection
- Data source and author information

2: Data Documentation

Data documentation is a complete description of a dataset. It not only includes metadata but also provides additional details like a data dictionary, usage rules, and notes on cleaning or preprocessing. The purpose of data documentation is to make sure others (or even your future self) can understand and use the dataset correctly.

Well-documented data is easier to analyze, share, and maintain. It increases trust, transparency, and reproducibility in data science projects.

Examples of Data Documentation Includes:

- Metadata details
- Data dictionary (list of variables and their explanations)
- Usage guidelines (how to use the dataset correctly)
- Notes about data cleaning or transformations