

Video 6 Summary: Preparing Data for Processing

Context:

You've collected the data needed for a dashboard project. Before moving to the **process** stage, you must **prepare** the data to ensure it's clean, consistent, and usable.

Key Concepts & Explanations

1. Raw Data Issues

- **Incomplete or missing data:** Some fields may be empty or not recorded.
- **Duplicated or incorrect data:** Errors that can distort analysis results.
- **Impact:** These issues must be fixed before analysis to avoid misleading insights.

2. Data Transformation

- **Definition:** Converting raw data into a consistent, usable format.
- **Purpose:** Ensures data is error-free and ready for analysis and visualization.
- **Examples:** Standardizing date formats, correcting typos, removing duplicates.

3. Data Processing vs. Data Transformation

- **Data Processing:** A broad term that includes collecting, cleaning, transforming, analyzing, and visualizing data.
- **Data Transformation:** A **specific** part of processing focused on format and structure conversion.

4. Goals of Data Transformation

- Make data **accessible** and **usable** for the entire team.
- Fix errors, **add missing information**, and **reduce unnecessary detail**.

5. Six Basic Types of Data Transformation

- **Data Smoothing:** Reducing noise or fluctuations.
- **Attribute Construction:** Creating new features from existing data.
- **Data Generalization:** Abstracting data to a higher level.

- **Data Aggregation:** Summarizing data (e.g., totals, averages).
- **Data Discretization:** Converting continuous data into categories.
- **Data Normalization:** Scaling data to a standard range.

6. Transformation Methods

- **Manual:** Using SQL or other programming languages.
- **Automated:** Using data pipelines or tools that handle transformation automatically.
- **Choice depends on:** Project needs and data type.

Takeaway

Before moving to the **store** stage, ensure your data is clean and well-structured. A solid foundation here will lead to **successful analysis and visualizations** later.