🎓 Video 15 Summary: Data Profiling and Cleaning

🧾 Analogy: Inventory in a Store

Just like store managers **count and verify items** during inventory, data analysts **profile and clean data** to ensure it’s accurate and usable.

🔍 What Is Data Profiling?

Data profiling is the process of **exploring data** to identify **quality issues**. It helps analysts understand:

- **Structure** (e.g., columns and data types)
- **Format** (e.g., string, number)
- **Values** (e.g., missing or duplicate entries)
- **Relationships** between fields

🧹 What Is Data Cleaning?

Data cleaning is the process of **fixing or removing** the issues identified during profiling. It ensures the data is:

- **Accurate**
- **Consistent**
- **Complete**

👩‍💼 Example: Arpa, a Retail Data Analyst

- Arpa receives monthly inventory data from multiple stores.
- She uses a **data profiling tool** to:
  - Identify columns (e.g., item name, price, quantity)
  - Determine data types
  - Detect missing or duplicate values
- After profiling, she **cleans the data** by:
  - Fixing missing prices
  - Removing duplicate item names

⚙️ Tools and Efficiency

- Data profiling and cleaning can be done **manually**, but tools are more **efficient** and **less error-prone**.
- Tools help automate the detection and correction of data issues.

🧠 Takeaway

Regular **profiling and cleaning** ensures high-quality data, which leads to **better analysis** and **more informed decisions**.