🎬 Video 23 – Understanding Joins and Null Values

🔗 What Are Joins?

- **Definition**: Joins combine data from two or more tables based on a related column.
- **Use Case**: Essential for cloud data analysts to bring together datasets for analysis.

⚠️ Common Challenges with Joins

1. **Missing Data**:
    - Occurs when expected rows don't appear in the result.
    - Often caused by the **type of join** used, not necessarily an error.
2. **Null Values**:
    - Represent **missing or unmatched data** in join results.
    - Not the same as zero or blank — they indicate absence of a value.

🔄 Types of Joins

🔶 Inner Join

- Returns only rows with matching values in both tables.
- **Example**: If a shoe ID exists in both inventory and info tables, it appears in the result. If not, it's dropped.

🔶 Outer Joins

- Return matched **and** unmatched rows.
- **Types**:
    - **Left Outer Join**: All rows from the left table + matched rows from the right.
    - **Right Outer Join**: All rows from the right table + matched rows from the left.
    - **Full Outer Join**: All rows from both tables, matched and unmatched.

🧠 Handling Nulls in Outer Joins

- Nulls appear where no match is found.
- Can be confusing if:
  - Nulls existed before the join.
  - Nulls are introduced by the join itself.
- **Pro Tip**: Always consider how nulls affect your analysis before choosing an outer join.

✅ Best Practices

- Choose the **right type of join** based on your data needs.
- Be aware of how **null values** and **missing rows** can impact your results.
- Use joins strategically to **combine data efficiently** and **preserve data integrity**.