📘 Advanced SQL – Lesson 3: String Functions

🎯 What You'll Learn

In this lesson, you'll learn how to:

- Clean, manipulate, and format text data in SQL using **string functions**
- Apply `TRIM`, `REPLACE`, `SUBSTRING`, `UPPER`, and `LOWER` to fix messy or inconsistent values
- Perform fuzzy matching across tables for approximate joins

🧠 Why Use String Functions?

String functions are essential for:

- **Cleaning dirty data** (e.g., extra spaces, typos, inconsistent capitalization)
- **Preparing data** for analysis, joins, or reports
- **Standardizing** entries from multiple sources

🛠️ Key Functions Covered

🔹 `TRIM`, `LTRIM`, `RTRIM`

Remove blank spaces:

- `TRIM` removes from both sides

- `LTRIM` removes from the **left**
- `RTRIM` removes from the **right**

Useful for fixing IDs or names with leading/trailing spaces.

🔹 REPLACE

Replace specific substrings with new values:

- Replace errors like `"Flenderson-Fired"` with `"Flenderson"`

🔹 SUBSTRING

Extract parts of a string:

- Select a portion of text by specifying a **start position** and **length**
- Often used for **fuzzy matching** (e.g., matching “Jimbo” to “Jim”)

Fuzzy Matching Tip:
Use `SUBSTRING` on multiple fields (e.g., first name, last name, age) to join messy datasets with no exact match.

🔹 UPPER and LOWER

Convert text to **uppercase** or **lowercase**:

- Normalize inconsistent casing before comparison or matching
- Great for cleaning entries like “tObY” to “toby” or “TOBY”

💡 Real-World Applications

- **Fixing human data entry errors**
- **Standardizing records** across systems before joins
- **Improving the accuracy** of fuzzy joins without unique identifiers
- **Cleaning inputs** for dashboards, exports, or validation

📌 Recap

✅ Use `TRIM`, `REPLACE`, and `SUBSTRING` to clean and fix broken or inconsistent text
✅ Use `UPPER`/`LOWER` to standardize casing
✅ Combine string functions for **fuzzy matching** when IDs are missing
✅ These tools are foundational for working with real-world, imperfect data