

# Chapter 5: Cleaning and Resolving Data Issues

## Part II: Profiling, Cleaning, and Transforming Data

### → Key Objectives

- Understand data profiling and quality assessment
- Identify common data issues and their impact
- Explore techniques for data cleaning and standardization
- Learn strategies for data transformation and integration

### Data Workflow



### Why It Matters

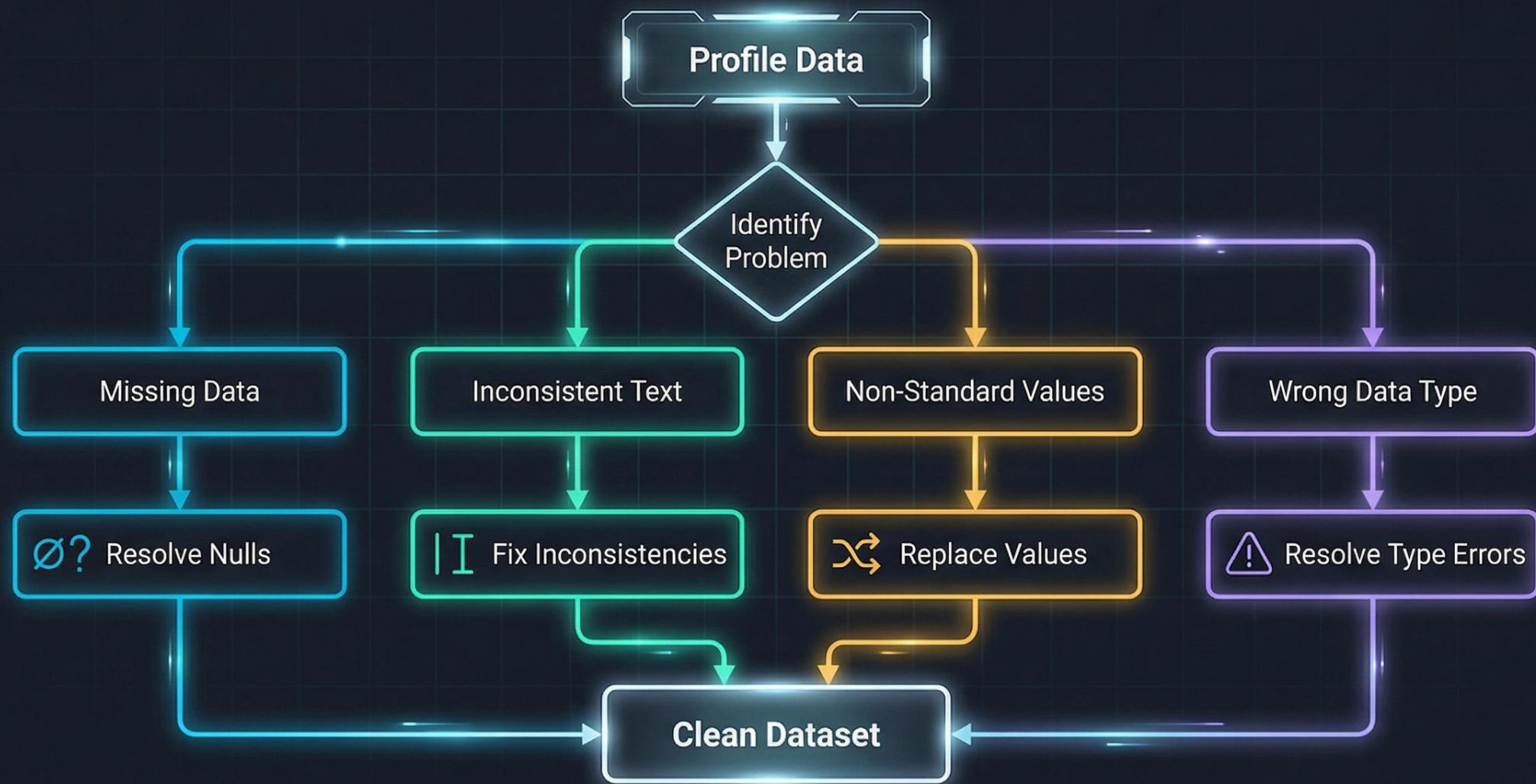
Ensuring data accuracy and consistency is crucial for reliable insights, operational efficiency, and informed decision-making.

# Why It Matters: The Messy Reality

- Real-world data arrives **messy**. Every single time.
- **Power BI** is not forgiving like **Excel**.
- **Inconsistencies** lead to **incorrect** reports.
- **Data cleaning** is the **professional** standard.



# The Cleaning Decision Tree



# Problem 1: Missing Data (Nulls)

The “Empty Chair” of Data



## → What is a Null?

- It is the ABSENCE of value.
- NOT Zero (0)
- NOT a Blank Space (" ")

## Context is King:

- The meaning depends on the situation.

## The 3 Fix Strategies



### 1. Remove Rows

When the data is useless or irrelevant.



### 2. Replace Values

When we know what the missing value *should* be (e.g., Null tip = \$0).



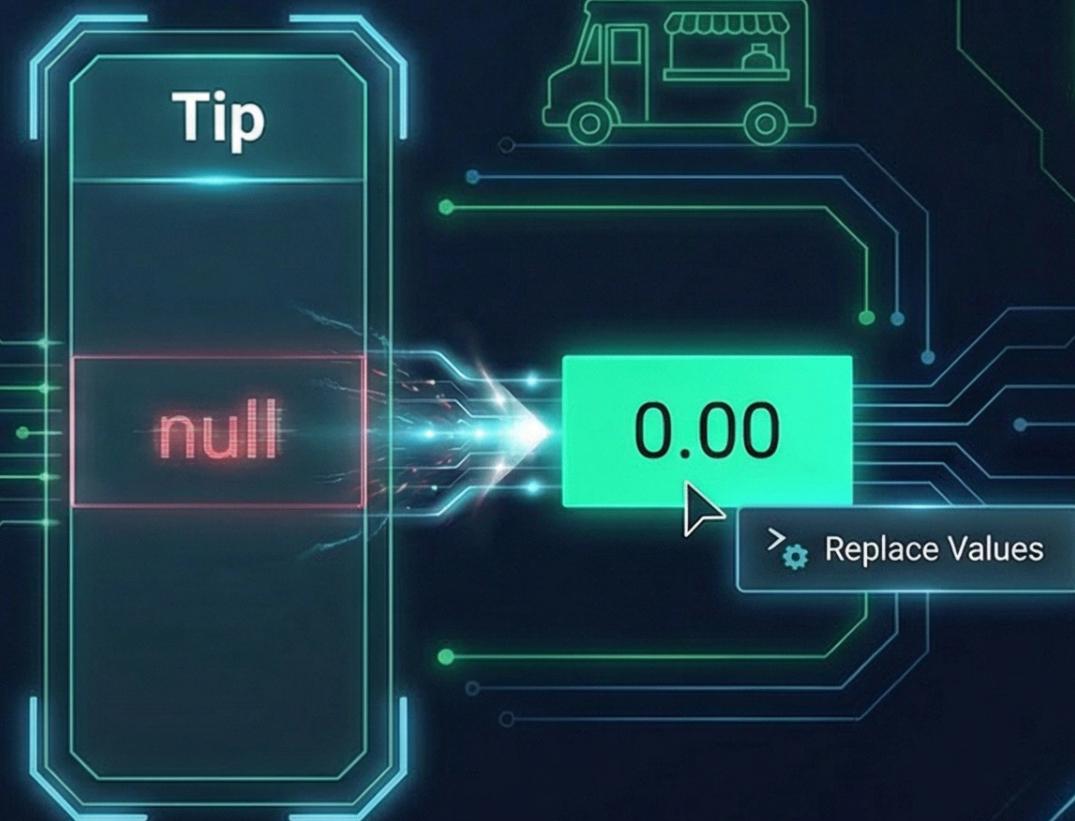
### 3. Fill Down

When the value applies to rows below it.

# DEMO: Resolving Nulls in Sabor Miami

## → Key Points

- **Scenario:** Tip column has nulls
- **Logic:** Null means 'No Tip' (\$0)
- **Action:** Replace nulls with 0
- **Tool:** Right-click > Replace Values



# Problem 2 & 3: Inconsistent & Non-Standard Text

**Power Query is Case-Sensitive!**



3 Different Values

Cash

CASH

cash \_

## The Fix Toolkit



**Replace Values**  
(Standardize specific words)



**Proper Case**  
(Fix capitalization)



**Trim**  
(Remove invisible spaces)

# DEMO: Fixing Inconsistencies

## Scenario 1: Menu Categories

Input (Top):

*Entrée* Entree  
*entree*



Replace  
Values

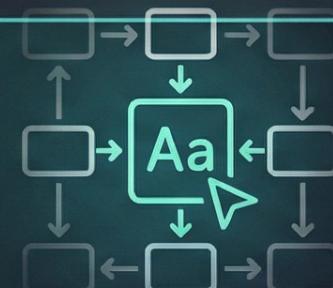
Output (Bottom):

**Entrée**

## Scenario 2: Employee Names

Input:

MARIA GARCIA  
maria garcia



Proper Case  
+ Trim

Output:

**Maria Garcia**

# Problem 4: Data Type Errors

## The 'Golden Rule' of Data Cleaning

### The Problem

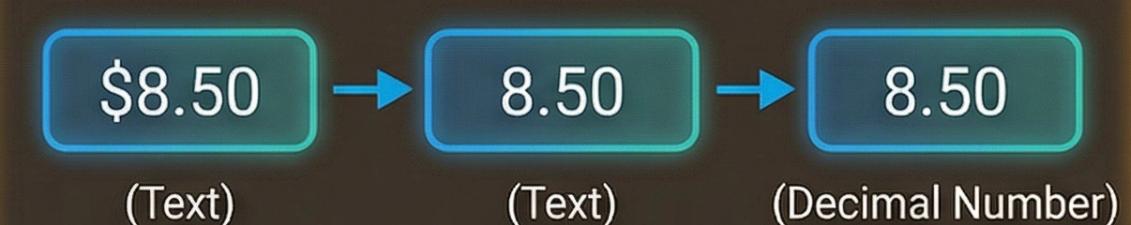


Power Query is **strict**! It sees the '\$' as a letter, not a **number**.

### The Golden Rule

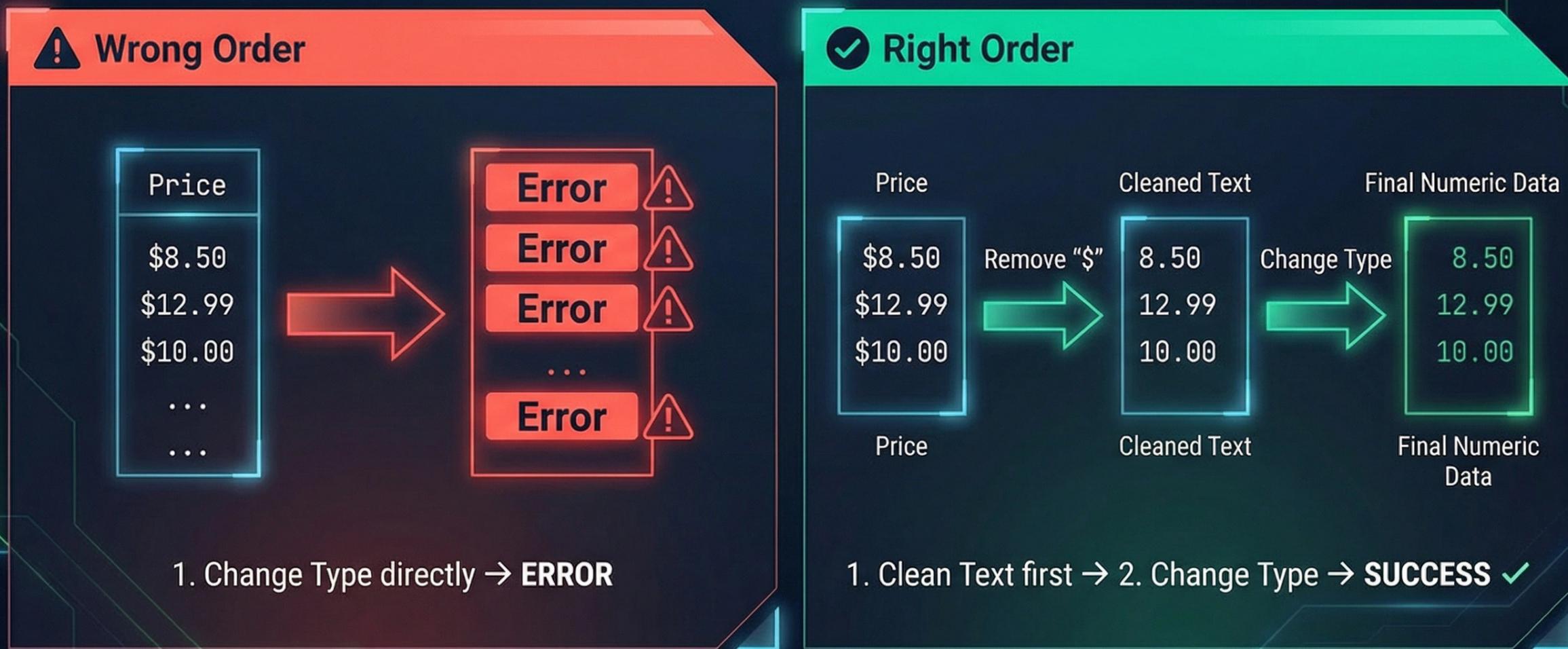
#### The 'Golden Rule' of Data Cleaning

- 1. Clean the text **FIRST** (Remove '\$')
- 2. **THEN** change the data type



# DEMO: The Deliberate Error

## Order of Operations Matters!



# Problem 5: Poor Naming Conventions

Making Data Human-Readable

Cryptic & Confusing



Truck\_ID

EmpFN

T\_Amt

Clear & Descriptive

Truck ID

Employee First Name

Transaction Amount



Convention: Use spaces. No abbreviations. Be descriptive.

# Practice & Verify

Your Turn to Clean the Kitchen

## Key Tasks

1. Clean the full Sabor Miami dataset (all 5 tables).
2. Use your Chapter 4 Profiling Report as your checklist.

## Verification Checklist (The “Definition of Done”)

- [] **Column Quality:** 100% Valid (Green Bars)
- [] **Applied Steps:** Clean, logical history
- [] **Column Distribution:** No duplicate categories



# You've Cleaned the Kitchen!

Next Up: Chapter 6 - Transforming Data

## Key Achievements

- ✓ **Accurate:** Nulls resolved, typos fixed.
- ✓ **Consistent:** Standardized text and categories.
- ✓ **Usable:** Correct data types and clear naming.