

# Data Quality Report

## Dataset Overview

- **Number of Rows:** 1,000
- **Number of Columns:** 11
- **Memory Usage:** ~86.1 KB
- **Data Types:** All columns are `object`

## 1. Missing Values

There are **3 missing values** across 3 columns:

Column	Missing Count	Recommended Action
Customer Id	1	Investigate and drop row if value cannot be recovered
Country	1	Infer value based on corresponding <b>City</b>
Website	1	Fill with <code>"Not Available"</code> or leave blank depending on business rules

## 2. Duplicated Records

- **Total Duplicates:** 4 rows

**Action:** Remove duplicate rows to maintain data integrity.

## 3. Inconsistent Text Formatting

- Columns affected: `First Name`, `Last Name`, `Company`, `City`, `Country`
- Issue: Inconsistent letter casing (mix of uppercase, lowercase, title case)

**Action:** Convert all text columns to **Title Case** for consistency.

## 4. Phone Number Formatting

- Columns affected: `Phone 1`, `Phone 2`
- Issue: Inconsistent formats, mixed separators, and extensions

**Action:**

- Standardize phone numbers into a uniform format:
  - Extract and separate **phone number** and **extension** (if available)
  - Recommended format: `+CountryCode-Number xExtension`






## 5. Subscription Date Format

- Issue: Dates are stored as strings with inconsistent formats

**Action:**

- Convert `Subscription Date` to datetime
- Use a consistent format: **DD-MM-YYYY**

## Summary of Cleaning Steps

-  Handle missing values (infer, impute, or drop)
-  Drop duplicate rows
-  Standardize text columns to title case
-  Normalize phone number formats
-  Convert subscription dates to DD-MM-YYYY format