Customer Data Cleaning Report

This report describes the data cleaning steps applied to the customer dataset to ensure accuracy, consistency, and usability.

1. Duplicated Rows

The dataset was checked for duplicated rows. Any duplicates found were removed to ensure that each customer entry is unique.

2. Standardizing Text Columns

Text columns, including First Name, Last Name, Company, City, and Country, were standardized to **Title Case**. This ensures uniform capitalization across all entries, improving readability and consistency.

3. Normalizing Phone Numbers

Phone numbers in both Phone 1 and Phone 2 columns were cleaned and standardized. Extensions were separated where applicable, and all numbers were formatted into a consistent national format. This ensures accurate contact information and easier processing for any downstream applications.

4. Subscription Date Formatting

The Subscription Date column was converted to a consistent **DD-MM-YYYY** format. Any invalid dates were identified and corrected where possible. This standardization ensures uniformity across all date entries.

5. Handling Missing Values

Missing values were handled as follows:

- **Country:** Any missing country values were filled with `"Unknown"` to indicate missing information while retaining the row.
- **Website:** Missing website values were filled with `"Not Available"` to clearly indicate absence of data.
- **Customer Id:** Rows missing a Customer Id were dropped, as this is a critical identifier and necessary for maintaining dataset integrity.

Summary

After cleaning, the dataset is now:

- Free of duplicate rows
- Text columns are uniformly capitalized
- Phone numbers are standardized and extensions are separated
- Subscription dates are consistently formatted
- Missing values have been appropriately addressed

The dataset is now ready for analysis, reporting, or export to other systems.