

Customer Sales Data Cleaning Report

1. Introduction

This report outlines the data cleaning process performed on the **Customer Sales Dataset** to ensure accuracy, consistency, and completeness. The dataset initially contained missing values, duplicates, incorrect data formats, and inconsistencies that required cleaning for effective analysis.

2. Data Overview

- **Dataset Name:** Customer Sales Data
- **Total Rows:** 104
- **Total Columns:** 6
- **Key Fields:** Customer ID, Transaction Date, Purchase Amount, Payment Method, Email, Country

3. Data Cleaning Process

3.1 Handling Missing Values

- Identified and replaced rows with missing Payment Method and Country.

3.2 Removing Duplicates

- Identified and merged duplicate customer records with slight name variations.

3.3 Correcting Data Formats

- Converted Transaction Date to YYYY-MM-DD format.

3.4 Handling Outliers

- **Issue:** Some numerical fields had extreme values that were unrealistic.
- **Fix:**
 - Used the **IQR method** to detect and cap extreme outliers.

- Verified corrections to ensure data integrity.

4. Cleaned Dataset Overview

- **Final Row Count:** 100
- **Final Column Count:** 6
- **Data Quality Score:** Improved for analysis