**Data Cleaning Report**

**Steps for Cleaning the Data**

1. **Handle Missing Values:**
   * For numerical columns, use imputation (e.g., mean, median, or interpolation).
   * For categorical columns, impute with the most frequent value or a placeholder like "Unknown."
   * Remove rows with excessive missing values if they do not contribute meaningfully.
2. **Fix Data Types:**
   * Convert dates to datetime format.
   * Ensure numeric columns are stored as integers or floats.
   * Verify and correct data types for categorical variables.
3. **Standardize Categorical Values:**
   * Apply consistent capitalization (e.g., all values in lowercase or title case).
   * Fix typos or merge similar categories (e.g., "Male" vs. "male").
4. **Identify and Treat Outliers:**
   * Use statistical methods (e.g., IQR) or visual tools (e.g., boxplots) to detect outliers.
   * Decide whether to cap, transform, or remove extreme values based on their business impact.
5. **Data Validation:**
   * Check for logical inconsistencies (e.g., negative transaction amounts).
   * Ensure data integrity by validating against expected ranges and patterns.
6. **Documentation:**
   * Keep a log of cleaning steps to ensure transparency and reproducibility.

**Methods and Tools Used**

* **Python Libraries:**
  + pandas for handling missing data and transformations.
  + numpy for statistical imputation and numerical operations.
  + datetime for managing and validating date fields.
* **Key Techniques:**
  + Imputation strategies for missing data.
  + Standardization for categorical and text fields.
  + Outlier detection using IQR and Z-scores.