

SalesAnalytics Data Cleaning Tasks List

This document outlines the **activities and tasks** to be accomplished in each of the SQL scripts for the SalesAnalytics Data Cleaning Project. This serves as a checklist for building and executing the cleaning scripts.

01_Cleanup_Customers.sql

Goal: Clean and standardize the `Customers` table.

Tasks:

1. Remove leading/trailing spaces from `FirstName`, `LastName`, `Email`, `Phone`, `City`, etc.
 2. Standardize name capitalization (`FirstName` and `LastName`).
 3. Convert all `Email` entries to lowercase.
 4. Remove special characters from `Phone` numbers.
 5. Replace empty strings or NULLs in `City` with 'Unknown'.
 6. Identify and remove duplicate customers (based on `Email`).
 7. Validate cleaned data: check NULLs, distinct cities, and duplicate counts.
-

02_Cleanup_Products.sql

Goal: Clean and standardize the `Products` table.

Tasks:

1. Trim whitespace from `ProductName`, `Category`, and `SubCategory`.
 2. Standardize text case for `ProductName`, `Category`, and `SubCategory`.
 3. Ensure numeric fields (`Price`, `Cost`, `QuantityInStock`) are non-negative.
 4. Handle NULL values for `QuantityInStock` (set to 0 if NULL).
 5. Identify and remove duplicate products (based on `ProductName`).
 6. Validate ranges for `Price`, `Cost`, and `QuantityInStock`.
-

03_Cleanup_Orders.sql

Goal: Clean and standardize the `Orders` table.

Tasks:

1. Trim whitespace from `Status` and other text fields.
 2. Standardize `Status` values to consistent categories ('Shipped', 'Pending', 'Cancelled').
 3. Validate `OrderDate` and `ShipDate`:
 - o Convert string dates to proper `DATE` type if necessary.
 - o Identify invalid dates.
 4. Check for duplicate `OrderID` entries.
 5. Validate `TotalAmount` for non-negative values.
-

04_Cleanup_Reviews.sql

Goal: Clean and standardize the `Reviews` table.

Tasks:

1. Trim whitespace from `Comment` and other text fields.
 2. Standardize `Rating` to ensure values are between 1 and 5.
 3. Replace missing `ReviewDate` with current date (`GETDATE()`).
 4. Remove or correct invalid `ProductID` or `CustomerID` references if needed.
 5. Identify and remove duplicate reviews for the same `ProductID` and `CustomerID`.
-

05_Cleanup_InventoryTransactions.sql

Goal: Clean and standardize the `InventoryTransactions` table.

Tasks:

1. Trim whitespace from `TransactionType`.
 2. Standardize `TransactionType` values to only 'Sale', 'Purchase', or 'Return'.
 3. Ensure `Quantity` is non-negative.
 4. Validate `TransactionDate` is within a reasonable range.
 5. Identify duplicate transactions if necessary.
 6. Validate `ProductID` references exist in the `Products` table.
-

Note:

- Each script should include **validation queries** after cleaning to ensure the dataset is consistent and ready for ETL or analytics tasks.
- All scripts should be **idempotent** to allow safe repeated execution.

This checklist can be used as a **reference while building and executing the cleaning scripts** in the SalesAnalytics Data Cleaning Project.