Data Manipulation Logic
Stone Leiker | March 2025

Data Manipulation Logic – Global Compliance Case Study

Step 1: Excel Preparation
I started by manually converting the provided customer data from PDF into Excel format. The raw data included combined text columns, so I used formulas to split those into clean and separate fields:

- "customer_id"
- "customer_type"
- "current_address_country"
- "customer_since_date"

I verified that each record had values in the necessary fields and saved the cleaned version as a .csv file for importing into SQL Server.

Step 2: SQL Cleaning and Transformation
I imported the cleaned .csv file into SQL Server and created a view named "vw_CleanedWiseStudyCase" to structure the data for analysis.

To ensure consistency and usability:

- I handled all blank or NULL fields using the SQL functions "NULLIF()" and "COALESCE()" — replacing them with the value "Unknown".
- I manually updated specific missing "customer_type" entries for customer IDs 19, 53, 60, and 67.

I then validated the data by grouping and counting:

- Total number of customers
- Customer distribution by "customer_type"
- Country distribution to confirm proper regional mapping

This helped ensure all 100 rows were complete, clean, and ready for reporting use.

Step 3: Power BI Filtering and Analysis
Next, I loaded the cleaned SQL view into Power BI to simulate the reporting for both R1 and R2.

I applied the following filters:

- Date range: Only included customers with a "customer_since_date" between 01/04/2022 and 01/08/2023
- Region-specific filters:
    - For R1: "current_address_country" = "UK"
    - For R2: "current_address_country" = "USA"

Since the dataset did not include transaction data, this analysis focused only on filtering customers based on location and date.

Assumptions and Limitations

- The dataset did not include "currency_sent", "currency_received", or "transaction_amount", so I was unable to classify transfers as cross-currency or same-currency or calculate totals in GBP.
- The analysis is based only on customer-level metadata, not transaction-level behavior.
- All reporting logic is a simulation, not a full regulatory output.

Recommended Enhancements for Future Reporting
To build an accurate and regulator-ready process, I recommend:

- Including fields for "currency_sent", "currency_received", and transaction "amount"
- Adding "entity_id", "account_open_country", and "IP_address" to improve legal mapping and fraud prevention
- Integrating FX rate data (such as from OANDA) to convert transactions to GBP based on transaction date
- Creating automated validation scripts to flag nulls or outliers at the point of data ingestion

Tools Used

- Excel: For structuring and cleaning raw data
- SQL Server: For transformation, NULL handling, and view creation
- Power BI: For filtering and simulating reporting outputs

Interpretation
I chose this Excel - SQL - Power BI pipeline because it reflects my regular workflow in real-life projects. I use these tools consistently in my current and past roles to solve data issues, automate compliance tasks, and prepare reports that are both audit-ready and scalable. These habits help me stay efficient, organized, and aligned with industry practices for regulatory reporting.