Compliance Case Study

NorthAm Compliance Data Analyst
Stone Leiker | March 2025

Case Study Overview

Goal:
Address two regulatory requests (R1 & R2) using customer transaction data and provide clean, auditable outputs.

Regulatory Requests:

- **R1:**
  - Total cross-currency volume in GBP under UK Entity between 01/04/2022 and 01/08/2023
- **R2:**
  - Total cross-currency volume in GBP under US Entity between 01/04/2022 and 01/08/2023
  - Total same-currency volume in GBP under US Entity between 01/04/2022 and 01/08/2023

My Approach

Step 1: Excel Preparation

- Converted the PDF to Excel manually
- Used formulas to split combined text into clean columns: customer_id, customer_type, current_address_county, and customer_since_date
- Save as .csv for clean import

Step 2: SQL cleaning & view creation

- Imported .csv into SQL server
- Created view vw_CleanedWiseStudyCase
- Handled all NULLS and blank fields using Coalesce and NULLIF

Step 3: Power BI Visualization

- Built a full dashboard with slicers, chart trends, and category, breakdowns
- Verified totals and joined dates across customer types.

R1 – UK Entity Results

Regulatory Request R1:

Goal: Total cross-currency volume under the UK Entity**,** where the customer is based in the UK and the route includes GBP.

Approach:

- Filtered by:
  - ○ Current_address_country = UK
  - ○ Customer_since_date between 01/04/2022 and 01/08/2023
- Used Power BI filters to simulate UK-based customer activity

Recommendation for Accurate Currency Reporting:

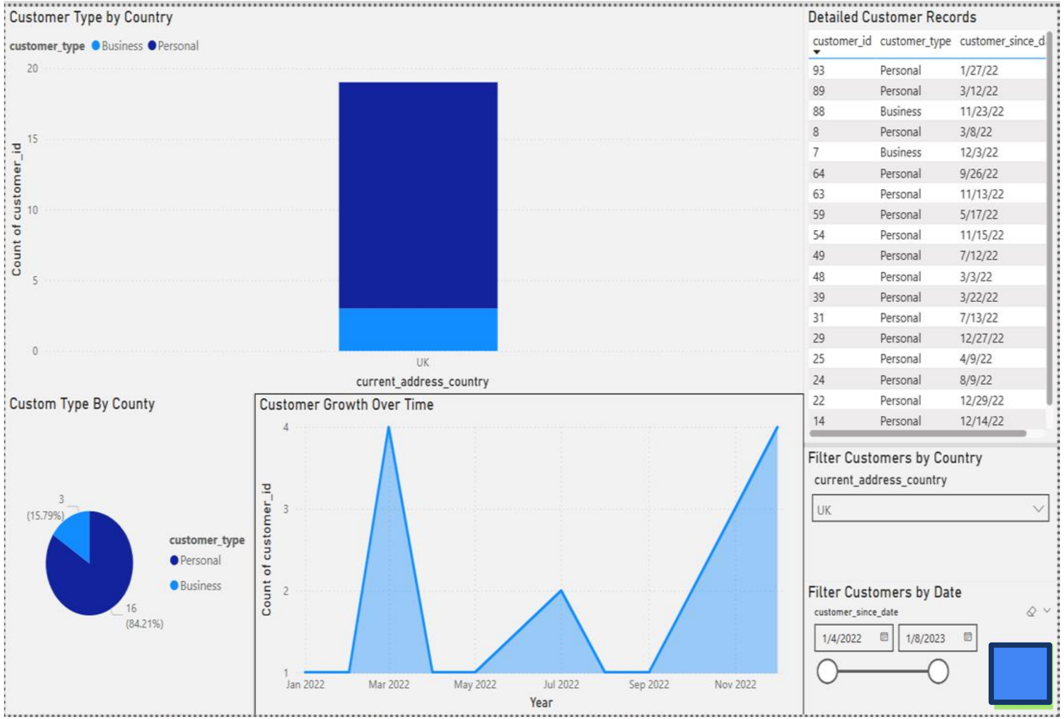
The current dataset lacks currency direction fields to reliably distinguish between cross-currency and same-currency transactions.

To improve accuracy, I recommend requesting the data team to include:

- Currency_sent
- Currency_received

This would allow accurate classification of:

- Cross-currency: USD > GBP
- Same-currency: GBP > GBP



Detailed Customer Records

customer_id	customer_type	customer_since_d
93	Personal	1/27/22
89	Personal	3/12/22
88	Business	11/23/22
8	Personal	3/8/22
7	Business	12/3/22
64	Personal	9/26/22
63	Personal	11/13/22
59	Personal	5/17/22
54	Personal	11/15/22
49	Personal	7/12/22
48	Personal	3/3/22
39	Personal	3/22/22
31	Personal	7/13/22
29	Personal	12/27/22
25	Personal	4/9/22
24	Personal	8/9/22
22	Personal	12/29/22
14	Personal	12/14/22

R2 – US Entity Results

Regulatory Request R2:

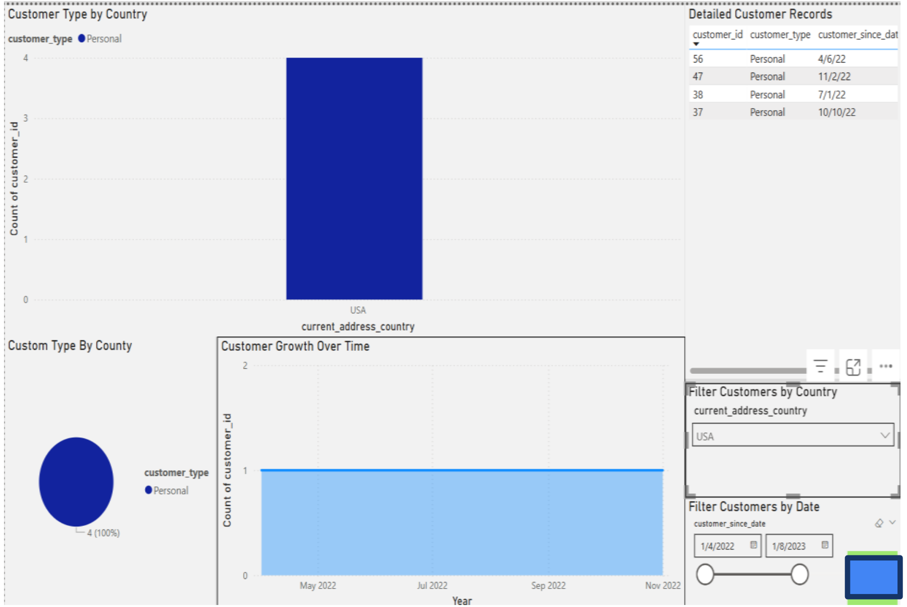
Goal: Total cross-currency volume + same-currency volume under the US Entity, Where customer is based in USA and route includes GBP.

Approach:

- Filtered by:
  - Current_address_country = USA
  - Customer_since_date between 01/04/2022 and 01/08/2023
- Simulated volume breakdown using available customer data

Currency-Level Classification:

Follows the same recommendation outlined in R1: Include currency_sent and currency_received fields to ensure proper classification between cross-currency and same-currency transfers.



Detailed Customer Records

customer_id	customer_type	customer_since_dat
56	Personal	4/6/22
47	Personal	11/2/22
38	Personal	7/1/22
37	Personal	10/10/22

Filter Customers by Country

current_address_country

USA

Filter Customers by Date

customer_since_date

1/4/2022 1/8/2023

SQL Cleaning Logic

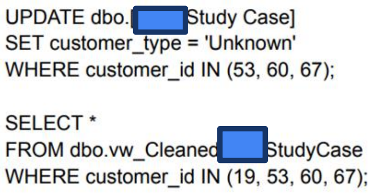
Key View:
`vw_Cleaned[redacted]StudyCase`

Key Objective:

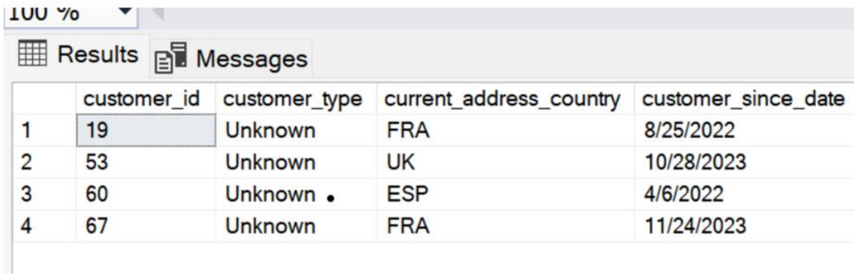
Create a clean, complete, and audit-ready customer dataset to support reporting, filtering, and verify accuracy in downstream analysis.

Cleaning Steps Taken:

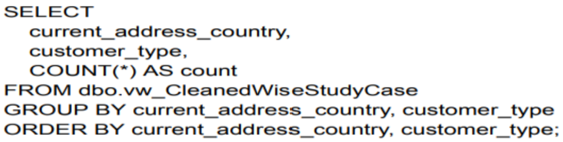
- Created SQL view: vw_Cleaned[redacted]StudyCase
- Replaced all blank or NULL values using:
- COALESCE(NULLIF(...), 'Unknown') for:
  - Customer_type
- Manually updated specific NULL values for:
  - customer_id = 19, 53, 60, 67
- **Created grouped summaries to verify:**
  - All customer types are filled
  - All current address countries are accounted for
  - All 100 rows include valid customer_since_date



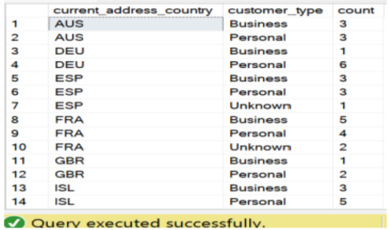
```
UPDATE dbo.[[redacted]Study Case]
SET customer_type = 'Unknown'
WHERE customer_id IN (53, 60, 67);

SELECT *
FROM dbo.vw_Cleaned[redacted]StudyCase
WHERE customer_id IN (19, 53, 60, 67);
```



	customer_id	customer_type	current_address_country	customer_since_date
1	19	Unknown	FRA	8/25/2022
2	53	Unknown	UK	10/28/2023
3	60	Unknown	ESP	4/6/2022
4	67	Unknown	FRA	11/24/2023



```
SELECT
    current_address_country,
    customer_type,
    COUNT(*) AS count
FROM dbo.vw_CleanedWiseStudyCase
GROUP BY current_address_country, customer_type
ORDER BY current_address_country, customer_type;
```



	current_address_country	customer_type	count
1	AUS	Business	3
2	AUS	Personal	3
3	DEU	Business	1
4	DEU	Personal	6
5	ESP	Business	3
6	ESP	Personal	3
7	ESP	Unknown	1
8	FRA	Business	5
9	FRA	Personal	4
10	FRA	Unknown	2
11	GBR	Business	1
12	GBR	Personal	2
13	ISL	Business	3
14	ISL	Personal	5

Query executed successfully.

Data Quality & Validation

Short-Term Actions:

- Identified and fixed 4 rows with missing customer_type
- Trimmed and cleaned all values for readability and consistency
- Validated that all 100 rows had complete and usable data

Long-Term Proposal:

- Build automated pre-ingestion validation scripts
- Flag unexpected or NULL values, including fields such as customer_type
- Standardize logic for:
  - Entity classification (It would be helpful to have entity_id and account_open_country)
  - Transaction routing fields such as sender_country, receiver_country
  - Additional logic to support enhanced compliance (more details in Slide 10)
- Prepare data to be regulator-ready at the source for faster submission and easier auditing

Regulatory Reporting Process (Proposal)

Recommended Reporting Pipeline:

1. Input from standard format (CSV, API)
2. SQL logic applied in staging layer
3. Power BI used for validation & internal review
4. Final output prepared for regulatory submission

Stakeholder Flow:

- Compliance defines rules & flags
- Data team prepares & validates
- Reporting team submits audited output

Tools Used

- **Excel:** For early-stage data structuring and cleanup
- **SQL Server:** To clean and transform raw data
- **Power BI:** To visualize, filter, and validate results
- **Text/Code Output:** PDF, SQL file, dashboard slides for delivery

Deliverables Provided:
Clean SQL Script
CSV Dataset
PBIX Dashboard
This Slide Deck (Final Presentation)

Final Summary & Recommendations

- Addressed both R1 and R2 requirements
- Built a repeatable and scalable data pipeline
- Cleaned and validated full dataset (100 rows)
- Delivered all files, visuals, and audit-ready SQL
- Ready to answer questions or expand this into an automated workflow for real-time regulatory reporting.

Recommended Enhancements:

- Adding currency_sent / currency_received fields
- Logging metadata (file version, data source, row count)
- Including entity_id for legal classification
- Validating country codes and routing fields during ingestion
- Adding account_open_country to support consistent entity mapping
- Capturing IP_address for location validation and fraud detection
- Tracking failed login_attempts (COUNT(*) > 3) to prevent unauthorized access
- Validating transaction amounts to ensure they fall within expected ranges (1 to 10,000 GBP)