Optimizing Regulatory Reporting for Cross-Currency and Same-Currency Transactions: A Case Study for R1 and R2 Compliance

-Prakhar Srivastava

1a. How would you summarize the results of the data?

In this case study, I analyzed the transaction data for both UK and US entities, focusing on cross-currency and same-currency volumes in GBP. The key findings are:

- **R1 (UK Entity Cross-Currency Volume):** 676 GBP in cross-currency transactions were recorded under the UK entity for the specified period.
- **R2 (US Entity Cross-Currency Volume):** 2135 GBP in cross-currency transactions were identified under the US entity.
- **R2 (US Entity Same-Currency Volume):** There were no same-currency transactions recorded for the US entity, resulting in 0 GBP.

These results clearly show that the US entity is engaged in more cross-currency activities compared to the UK entity. Additionally, the absence of same-currency transactions for the US entity could be due to the specific nature of the transactions or the absence of domestic transfer activities in GBP within the timeframe.

1b. How would you summarize the main caveats to the available data?

There are a few key caveats that need to be addressed when interpreting the data:

1. **Currency Route Formatting:**
    - The format for currency routes is not standardized across all transactions. For instance, routes such as "GBP --> USD" might be entered differently in some cases, leading to the potential misclassification of cross-currency transactions. If there is inconsistency in the data entry, it could affect the accuracy of the report.
    - **Solution:** Implement a data standardization process to ensure that all currency routes follow the same format before running the analysis.
2. **Customer Data Accuracy:**
    - The field Current_Address_Country in the customer data may not always be up-to-date, especially for long-standing customers who may have moved to a different country. As a result, there's a risk of transactions being attributed to the wrong regulatory entity.
    - **Solution:** Ensure that customer data is periodically updated, especially for fields related to their current country of residence.
3. **Time Range Constraints:**

- The transactions analyzed are limited to those within the date range of 01/04/2022 to 01/08/2023. This may not provide a complete picture of the entity's transaction activity, especially if there are transactions outside this range that are relevant for the report.
- **Solution:** Work with the stakeholders to confirm that the date range aligns with regulatory reporting requirements. If not, adjust the date filter and pull in the necessary data.

By addressing these caveats upfront, we can ensure that the stakeholders understand the limitations of the data and take appropriate actions to enhance future reporting.

1c. Are there any issues you identified with the data? If so, how do you propose to fix these?

Yes, I identified a few data-related issues that could impact the accuracy and reliability of the analysis:

1. **Inconsistent Currency Routes:**
   - As mentioned earlier, the format of currency routes varies across transactions, which can lead to misclassification. If the currency route is misformatted, the query logic may not capture all relevant cross-currency transactions.
   - **Proposed Fix:** Introduce a data pre-processing step where all currency routes are normalized into a consistent format before running the queries. This can be

done through simple string operations or using a Python script for batch processing.

2. **Outdated Customer Information:**
    - The Current_Address_Country field may not be up-to-date for some customers. If customers have relocated but their records haven't been updated, this could result in incorrect categorization under UK or US entities.
    - **Proposed Fix:** Implement regular data updates or enforce data validation checks, especially for customer information fields such as address and country of residence. Ideally, customer data should be refreshed at regular intervals to ensure accuracy.
3. **Potential Data Gaps:**
    - The time range of the transactions analyzed is limited to a specific period, which might exclude relevant transactions that fall outside this range.
    - **Proposed Fix:** Expand the date range in the query if required by the stakeholders. Additionally, work with the data engineering team to ensure that historical data is readily available and correctly integrated into the reporting system.

By implementing these fixes, we can improve the quality of the data and ensure that future reports are more accurate and reliable.

1d. Is there any information you are missing? If so, what is the nature of the missing information?

There are two key pieces of information that would be beneficial to improve the depth of the analysis:

1. **Customer Activity History:**
    - The current dataset only includes the customers' current address. However, having access to historical customer information, particularly when and where they moved, would allow us to track changes in their transaction behavior more accurately. For example, if a customer recently moved from the UK to the US, this information would help in categorizing their transactions more precisely for regulatory purposes.
    - **Requirement:** A historical customer dataset that tracks changes in the customer's country of residence over time. This would be linked to the existing data via Customer_Id.
2. **Transaction Type Metadata:**
    - The dataset currently lacks details on the nature or type of each transaction (e.g., remittance, purchase, or business payment). This metadata would be useful in understanding the context behind each transaction and determining whether certain transactions should be classified differently under regulatory reporting requirements.

- **Requirement:** A column in the transaction data specifying the transaction type. This would provide additional context to the regulatory bodies and help segment the data more effectively.

Having these two additional pieces of information would enhance the depth of the analysis and provide more granular insights for regulatory reporting.

1e. Based on your answers from (1a) to (1d), what would be your approach given the reports need to be submitted now?

Given that the reports need to be submitted immediately, my approach would be to proceed with the current data but include a disclaimer outlining the known data limitations, such as the potential issues with currency routes, outdated customer address information, and the limited time range.

In parallel, I would start implementing the proposed fixes for future reports:

- Standardize the currency route formats.
- Set up a process for regular updates to customer address data.
- Expand the time range for transactions, if necessary.

The disclaimer ensures transparency with the stakeholders, while the ongoing improvements will result in better data quality and more reliable reporting in subsequent cycles.

2a. What technologies would you use for this task? Why?

For this regulatory reporting task, I would utilize a combination of the following technologies:

- **SQL (Structured Query Language):** SQL is ideal for querying structured datasets, particularly for filtering and aggregating data from the customer and transaction tables. It allows for efficient extraction of specific data points that meet the regulatory reporting criteria.
- **Python:** Python is extremely useful for additional data processing tasks that go beyond basic SQL queries. For instance, I would use Python to perform data validation, normalization of currency routes, and any necessary transformations. Libraries like Pandas and NumPy are well-suited for handling such tasks.
- **ETL (Airflow/DBT):** To automate the data extraction, transformation, and loading (ETL) process, I would use tools like Airflow or DBT. This ensures that the reporting process is scalable and can be run on a regular schedule with minimal manual intervention.
- **Data Visualization (Power BI/Tableau):** For presenting the results and tracking trends, I would use Power BI or Tableau. These tools enable stakeholders to visualize the data in a way that is easy to interpret, especially for non-technical audiences.
- **Version Control (Git):** To ensure that the queries and scripts are properly versioned and can be re-used in future reporting

cycles, I would use Git. This also enables collaboration and tracking of any changes made to the queries.

2b. How would you ensure that you are calculating the right metrics to meet the definitions by regulators?

Ensuring that the metrics align with regulatory definitions requires close collaboration with the regulatory and compliance teams. Here's how I would approach it:

1. **Regular Consultation with Regulatory Teams:**
    - I would work closely with the regulatory teams to verify the definitions of key metrics, such as cross-currency and same-currency volumes, and ensure that the queries accurately capture these metrics.
    - By involving the compliance teams early in the process, I can ensure that any changes in regulatory guidelines are reflected in the queries.
2. **Documentation of Definitions:**
    - I would create and maintain a shared glossary that clearly defines each metric, along with the SQL logic used to calculate it. This document would be regularly updated and reviewed to ensure alignment with regulatory requirements.
3. **Validation and Testing:**
    - Before running the final report, I would conduct validation checks to ensure that the calculations align with previous

reports and regulatory expectations. This might involve comparing the results against historical data or conducting a peer review of the query logic.

2c. Process Adjustments for Weekly Reporting, Annual Reports, and Ad-hoc Requests

1. **Weekly Reporting:**
    - If this data needs to be submitted on a weekly basis, I would automate the process using an ETL tool like **Airflow** or **DBT**. The SQL queries can be scheduled to run automatically on a weekly basis. By building an automated pipeline, the necessary data can be extracted, transformed, validated, and delivered consistently.
    - Additionally, I would ensure **data validation checks** are built into the pipeline to catch inconsistencies or missing data before submission. For example, I could set alerts to notify the team if there are any significant discrepancies compared to the previous week's data.
2. **Annual Reporting:**
    - When preparing annual reports, where the queries from a year ago are being re-used, I would implement **version control** through tools like **Git**. This allows for tracking any changes made to the queries over time and ensures we can revert to a previous version if needed.
    - To ensure that the queries remain relevant and correct for the current dataset, I would also run **test queries** on smaller samples of the current data and compare the

results against previous reports. Any differences or issues identified would prompt a review of the business logic to ensure regulatory requirements are still being met.

3. **Ad-hoc Requests:**
    - In the event of an ad-hoc request, such as one asking for the total number of accounts, I would approach this by leveraging the **current customer dataset**. The query would involve filtering the Customer_Id column based on the specified criteria (e.g., active accounts, accounts opened within a specific time range).
    - To ensure consistency with previous reports, I would re-use the same logic and filtering criteria that were applied in prior regulatory submissions. This ensures that the data aligns with what has been reported in the past, maintaining consistency and accuracy.

By following this process, I can ensure that weekly, annual, and ad-hoc reporting is handled efficiently, with robust data validation checks in place to ensure accuracy. The use of automated ETL pipelines and version control systems like Git ensures that the reporting process is scalable and consistent over time.