

---

## **Capstone-project**

---

### **About this dataset**

The dataset represents E-Commerce transactions processed via PayPal during the 2025 financial year, covering the period from November to March. While the vast majority of transactions are successfully completed through an automated system, approximately 0.1% fail due to various issues. My role involves investigating these failures, identifying their root causes, and ensuring they are resolved with the appropriate actions.

The dataset includes 38 columns and 4,559 rows, with analyses conducted on a biweekly basis. This report aims to address several key questions:

- What are the main reasons for transaction failures?
- How long does it take to resolve failed transactions?
- Does the total transaction value have a significant impact on the business?
- Which currencies are most associated with failed transactions, and is there a correlation between currency type and the actions taken to resolve them?

### **Let's start this project!**

#### **Inspecting and Cleaning the Dataframe Structure**

This dataset contains 4,552 rows and 38 columns. As a first step, I removed columns that contained no data. I also dropped rows with null values in the AR Comments and Action columns, since these are the key focus areas for our analysis.

After cleaning, we are left with 32 columns, consisting of the following data types: float64 - 5 columns, int64 - 6 columns, object - 22 columns.

Next, the Action column needs to be standardized to ensure data consistency. We'll use a boxplot to help detect and address any outliers. The valid categories for the Action column should be limited to the following:

--> **clear together** - Transactions that result in a zero balance when cleared in their document currency.

--> **write off** - Typically includes:

*Manual refund*: Processed upon a request from customer service.

*Settlement reversed/chargeback*: When a customer reverses the payment directly through PayPal.

*Cancelled invoice/no rebilling*: The operations team cancels the invoice, but no replacement is issued.

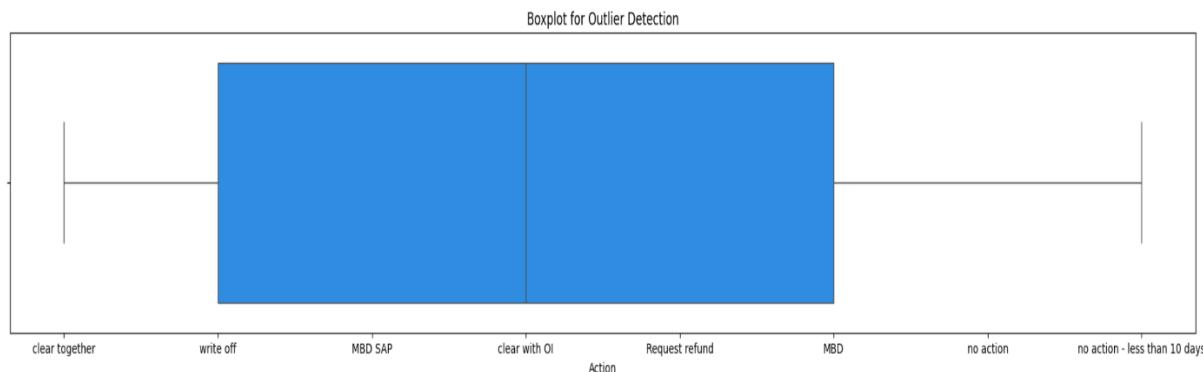
--> **MDB SAP** - Occurs when payment is collected and the order is created in ECC, but no invoice is generated.

--> **MBD** - Payment is collected, but no order is replicated in ECC.

--> **clear with OI** (open items) - Payment exists and can be matched against an open invoice in Accounts Receivable.

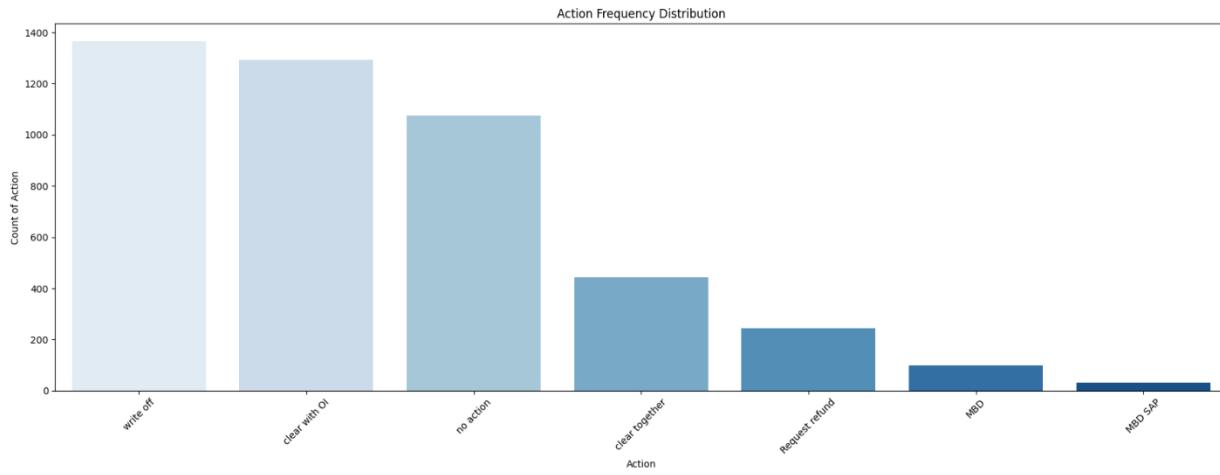
--> **request refund** - Happens when payment is collected twice through different processors, requiring a refund via PayPal.

--> **no action** - Transactions that are 10 days old or less. These usually don't require immediate action unless flagged by the standard formula checker.



From the boxplot above, we can observe that the actions "no action" and "no action - less than 10 days" should be standardized, as they refer to the same status. After applying the necessary replacements, we now have a clean set of seven consistent Action categories.

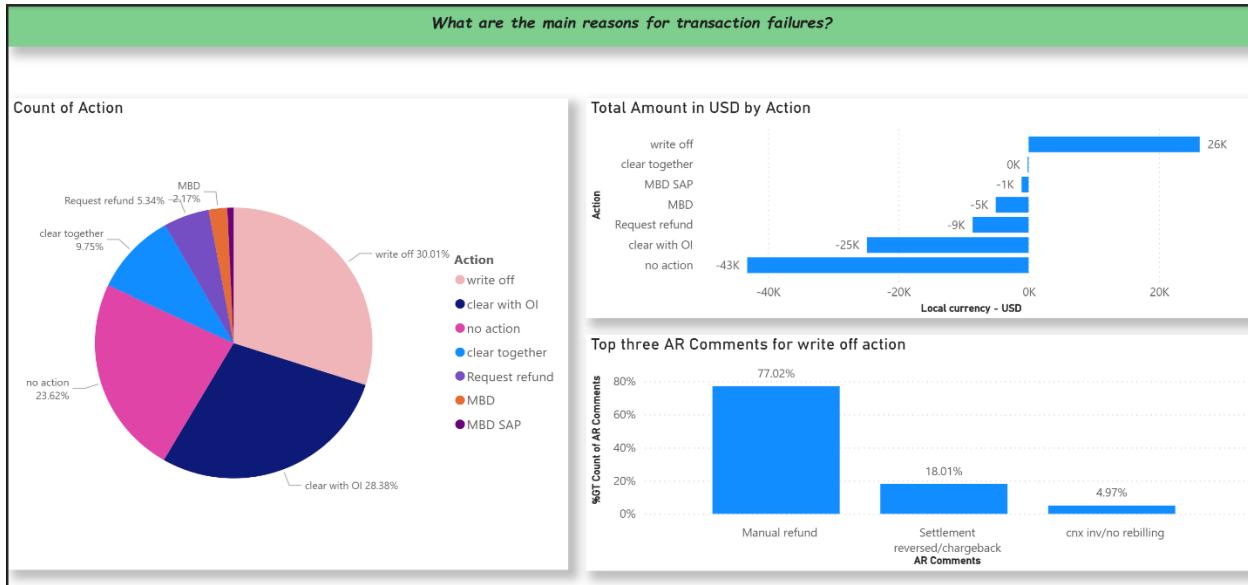
Given that "write off" is the most frequently occurring action, I also standardized the entries in the AR Comments column. This will be essential for addressing our key business questions in the following sections.



The last step of data preparation is to convert categorical features to numerical to enable us to prepare the heatmap.

## Analysis 🔎

The first question we aim to answer is: What are the main reasons for transaction failures?



The most frequent action is "write off", accounting for 30.01% of all cases, followed by "clear with OI" and "no action". The top drivers behind write-offs include:

Customer dissatisfaction → resulting in manual refunds

Disputed payments → often processed as chargebacks

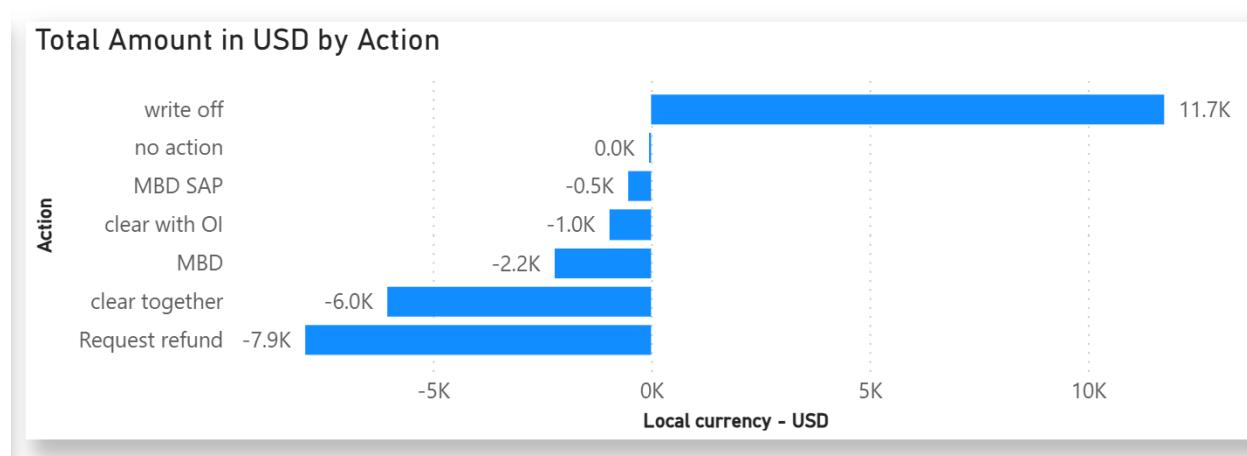
Internal process gaps → such as canceled invoices without rebilling

These trends highlight key areas for improvement in customer support, billing accuracy, and operational coordination.

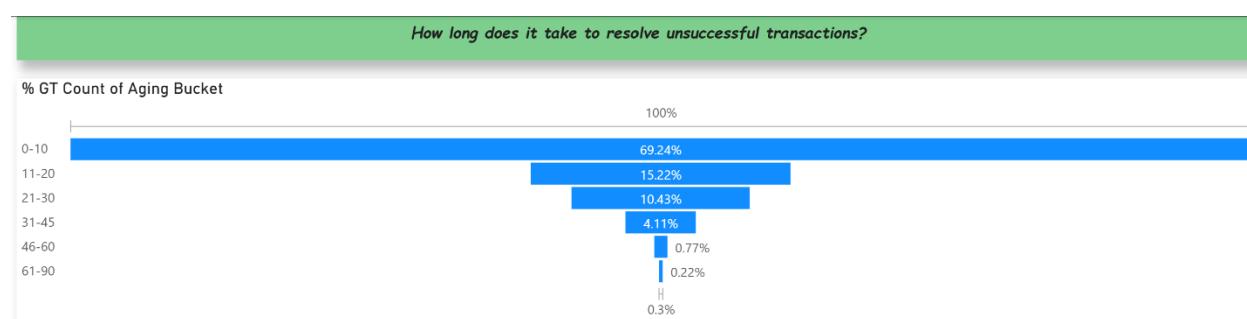
In terms of financial impact:

- Write-offs amount to \$26K
- No action and clear with OI contribute the highest negative balances, with "no action" alone accounting for - \$43K

However, it's important to note that most "no action" items are transactions less than 10 days old, and are typically processed automatically by the system. As analysts, we are not required to address these unless flagged during the standard formula check. Once those recent items are excluded, the true impact of "no action" drops significantly to - \$7.9K, indicating a minimal short-term risk and allowing us to defer further investigation until the next analysis cycle.



Moving on to the next part of our analysis, we explore the question: How long does it take to resolve unsuccessful transactions? 🕒



The data shows that nearly 70% of transactions are under 10 days old. While analysts are not required to take immediate action on these items, standard practice allows for early intervention, especially when transactions fall into predefined automated classifications. Taking proactive steps at this stage helps prevent unnecessary backlog in future analysis cycles.

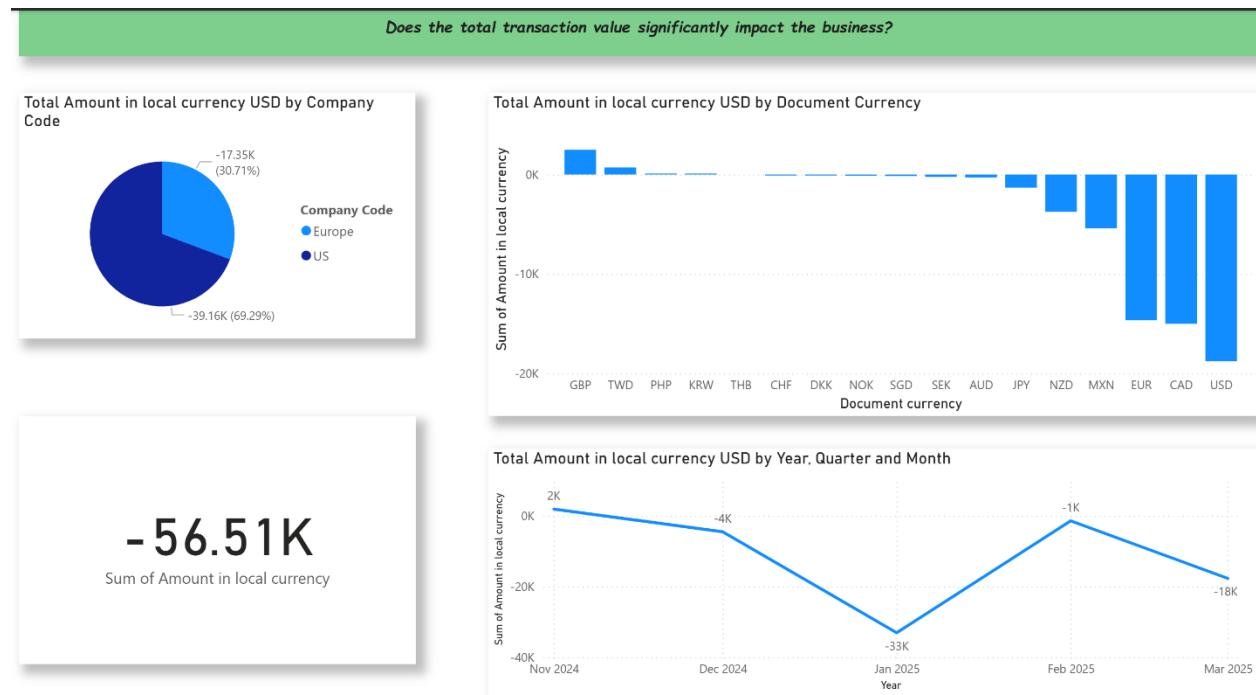
The most common actions taken on these newer transactions are:

- Clear with OI / Write off the difference
- Clear with OI

In many cases, these will be resolved automatically by the system. The analyst's role is to assess whether manual intervention is necessary or if the transaction should be left to process naturally.

Approximately 25% of transactions are resolved between 11 and 30 days, indicating that while the majority are handled quickly, a substantial portion still requires follow-up within a 1-month window.

For most companies, the important question is does the total transaction value significantly impacts the business? 💰

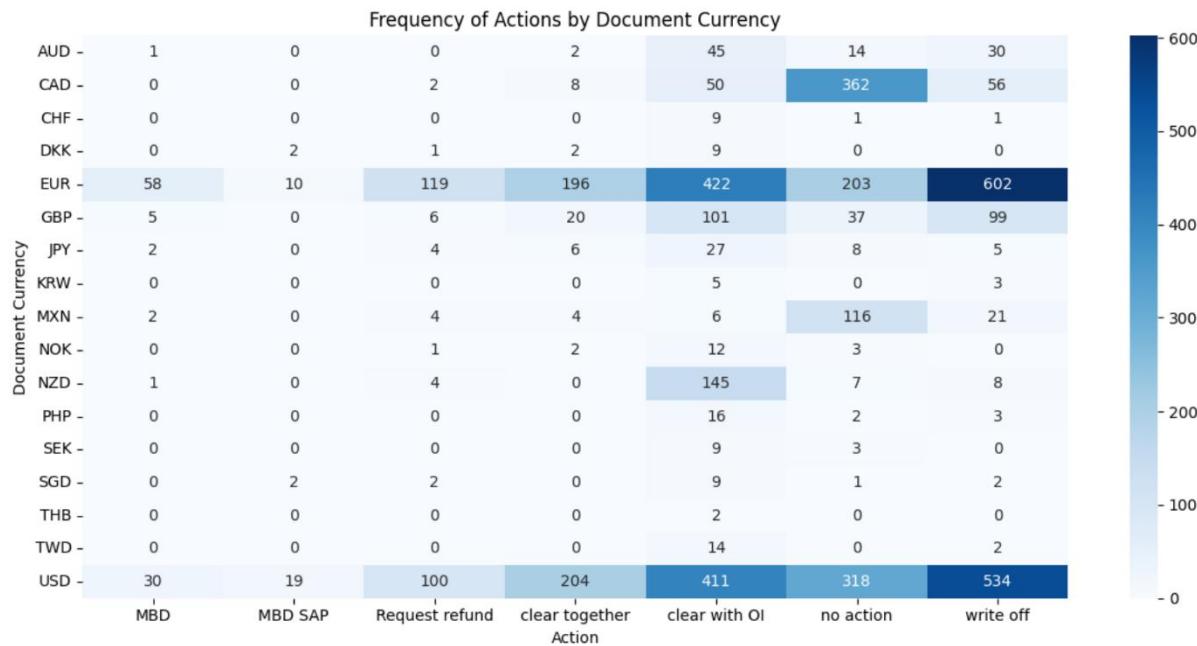


As previously mentioned, the dataset covers transactions from November 2024 to March 2025. According to the chart, the total value of unresolved transactions in local currency (USD) is approximately - \$56.51K.

Given that the company processes over 7 million transactions per week across 14 different payment processors, this PayPal-specific amount is not immediately significant. However, if left unresolved, these smaller discrepancies can accumulate into a much larger financial risk.

Further analysis shows that nearly 70% of these transactions originate from the U.S. subsidiary, which includes operations in the United States, Mexico, and Canada - making it a key area to monitor closely.

Finally, let's investigate what the main currency is that fails and there is any relation between currency and actions taken to resolve them? 



The heatmap reveals a strong correlation between certain currencies — USD, EUR, CAD, GBP, MXN, and NZD — and specific actions such as write off, no action, clear with OI, clear together, and request refund.

Among these, "write off" emerges as the most frequently taken action. Diving deeper into this category, the most common reason for a write-off is manual refund. These refunds are typically processed by the refund team following a request from customer service, often based on prior agreements made between the customer and the company.

This trend suggests a notable volume of customer dissatisfaction, particularly in key markets such as US (USD) and Germany (EUR), raising the question: Why are so many customers requesting refunds? Investigating the root causes behind these cases could provide valuable insights and improvement opportunities.

The second most frequent write-off category is settlement reversed/chargeback, which again points to customer dissatisfaction. In these cases, customers reverse payments directly through PayPal, bypassing internal resolution processes. This further highlights potential issues in the customer journey or post-sale experience that merit closer examination.

	Write-Off Reason (AR Comments)	Count
0	Manual refund	1022
1	Settlement reversed/chargeback	239
2	cnx inv/no rebilling	66

## Conclusion

This analysis of PayPal E-Commerce transaction failures during the 2025 financial year reveals key insights into both operational inefficiencies and customer behavior. Although failed transactions represent a small portion (0.1%) of the total volume, their cumulative financial impact and the patterns behind them offer valuable opportunities for process improvement.

Key findings include:

- "Write off" is the most frequent resolution action, often tied to customer dissatisfaction, chargebacks, or process gaps like canceled invoices without rebilling.
- A significant portion of failed transactions are less than 10 days old and typically auto-resolve, reducing their immediate risk. However, early monitoring helps prevent backlog buildup.
- The total value of unresolved PayPal transactions (~\$56.5K) is relatively small when compared to the company's total processing volume, but consistent tracking is crucial to avoid accumulation.

- Currency analysis highlights strong links between specific currencies (USD, EUR, CAD) and failure patterns, especially in key markets like the U.S. and Germany. High refund rates in these regions suggest customer experience issues that warrant further investigation.

In summary, while the financial risk posed by failed transactions is currently limited, the underlying trends—particularly in write-offs and regional dissatisfaction—highlight areas for strategic improvement. Proactive action on these insights could improve customer satisfaction, reduce manual intervention, and enhance overall efficiency in transaction handling.