# Data Cleaning and Quality Assurance - SpendHound Take-Home Assessment

## Overview

This Document contains the data cleaning and quality assurance process for financial transactions related to software vendor payments. The goal was to accurately determine valid transactions, assign correct service periods, and ensure data integrity for financial analysis.

## Assessment Objectives

1. Transaction Selection – Identifying and including only valid expense records.  
2. Service Period Assignment – Determining the correct start and end dates for each included transaction.  
3. Justification and Documentation – Explaining decisions, assumptions, and handling uncertain transactions.

## Challenges Faced and Uncertain Transactions

During the data validation process, there were two transactions where assumptions had to be made due to missing or unclear information:

## 1. Ashby Transaction

- Transaction Date: 2024-03-10  
- Amount: $3,600  
- Memo:"Annual Subscription" (No specific period mentioned)  
  
The memo field did not provide explicit service dates. However, after reviewing historical transaction patterns, it was observed that Ashby payments were consistently made in March each year. Given this, the best assumption was that this transaction covered an annual subscription from March 10, 2024, to March 9, 2025.

## 2. Datadog Transaction

- Transaction Date: 2024-06-05  
- Amount: $500  
- Memo: "Usage-Based Charges - May 2024"  
  
This transaction was part of a postpaid billing model, meaning the charge was for May but was processed in June. To ensure correct financial period allocation, the service period was assigned from May 1, 2024, to May 31, 2024. This aligns with Datadog's billing structure and maintains accuracy in the time-series expense tracking.

## Technical Approach

Data Processing Workflow :  
1. Load Data – Import transaction records from Excel files.  
2. Duplicate Removal – Identify and eliminate redundant records.  
3. Exclusion of Non-Spending Transactions – Remove journal entries, amortizations, and accruals.  
4. Service Period Assignment – Extract or infer service dates based on available information.  
5. Handling Missing Values – Fill blank memos with placeholders and estimate missing service periods.  
6. Validation and Final Export – Ensure completeness before saving the cleaned dataset.

## Technology Stack

- Python (`pandas`, `datetime`) – Data manipulation and validation.  
- Jupyter Notebook – Interactive development and debugging.  
- Excel (`pandas.ExcelFile`, `ExcelWriter`)– Data input/output handling.  
- Matplotlib – Visualization for insights on transaction trends.

## Outcome

- Successfully cleaned and structured the dataset for financial analysis.  
- Removed duplicate and non-spending transactions, ensuring data accuracy.  
- Assigned service periods logically, filling in missing values where necessary.  
- Documented all decisions for transparency and future reference.

## Final Thoughts

This project highlights the importance of structured data validation in financial analysis. The ability to clean, interpret, and structure transaction data accurately ensures that financial insights are reliable, actionable, and aligned with business needs.

- Thank you.

- Raghul Rajkumar.R