Bespoke Financial Tracking Tool: Master Workflow

This document outlines the definitive workflow for processing financial statements and maintaining the master transaction ledger. The system uses a two-sided accounting model where all inter-account transfers are explicitly linked with a shared ReconciliationID.

I. Script Inventory

The project folder should be organized into three categories: Core Workflow, Utilities, and an _archive folder for deprecated scripts.

A. Core & Specialized Workflow Scripts

These are the essential scripts for the day-to-day process of importing new statements.

- step2_processor.py: The universal script that cleans and standardizes raw CSV files, enforcing correct transaction polarity.
- step3_categorizer.py: The main importer. It performs a two-sided reconciliation, keeping both sides of a transfer and linking them with a ReconciliationID.
- step4_review.py: The interactive tool for auditing transactions, correcting data, and managing categorization rules.
- **PDF Extractors**: Custom scripts for parsing PDF statements (e.g., extract_target_pdf_v2.py).
- **Amazon Reconciliation Suite**: (step5_..., step6_..., step9_final_merge.py) A specialized workflow for reconciling Amazon orders.
- reconcile_venmo_flow.py: A specialized tool for linking Venmo pass-through payments to their funding source.

B. Utility & Maintenance Scripts

These are powerful tools for auditing data, managing rules, and generating reports.

- step1_inspector.py: A diagnostic tool to inspect the columns and content of a new, unknown CSV or XLSX file before processing.
- generate_excel_report.py: Creates the multi-sheet Excel financial dashboard with yearly summaries and category drill-downs.
- generate_rules.py: Intelligently suggests new categorization rules based on your transaction history.
- combine_csv.py: A utility for merging multiple CSV files into one, useful for combining multiple months of PDF extracts.
- purge_account_data.py: A tool to safely and completely remove all data for a specific account from the master file before a clean re-import.

C. Data Repair & Verification Scripts

These scripts are used to diagnose and fix data integrity issues.

- data_integrity_audit.py: A comprehensive, read-only tool that runs multiple checks on the master file and reports on miscategorized transfers, polarity errors, and unmatched payments.
- verify_reconciliation.py: Audits all ReconciliationID links to ensure every linked pair is correctly balanced (one debit, one credit, summing to zero).
- backfill_reconciliation_ids.py: A powerful repair tool that finds all un-linked transfer pairs in the master file and creates the ReconciliationID link. Essential for reconciling data added by legacy workflows (like the Amazon suite).
- debug_unmatched_pairs.py: A diagnostic tool that explains why specific transfers failed to be automatically reconciled (e.g., dates are too far apart).

II. Standard Monthly Workflow

A. For Standard CSV & PDF Statements

(Applies to: Amex, Discover, US Bank, Target, Wells Fargo, etc.)

1. **(For PDFs Only)** Run the appropriate PDF extractor to generate a raw CSV.
2. **Process**: Run step2_processor.py on the raw CSV to create a processed_... version.
3. **Import**: Run step3_categorizer.py on the processed_... file. It will apply rules and automatically reconcile and link payments.
4. **Review**: Run step4_review.py to categorize any remaining new items.

B. Special Workflow: Chase & Amazon Reconciliation

1. **Process & Merge**: Follow steps 1-4 of the original Chase/Amazon workflow (step2, step5, step6, step9).
2. **Review**: Use step4_review.py to approve the newly added non-Amazon transactions.
3. **Reconcile Payments**: After merging, run backfill_reconciliation_ids.py to find the Chase payments that were just added and link them to the corresponding withdrawals from the checking account.

C. Special Workflow: Venmo Pass-Through Payments

1. **Import Bank Data First**: Process and import your US Bank Checking statement for the month using the standard workflow.
2. **Process Venmo Data**: Run step2_processor.py on your raw Venmo CSV.
3. **Run Venmo Reconciler**: Run reconcile_venmo_flow.py. This script will find payments from your checking account, match them to Venmo expenses, and create the SourceTransactionID link.