

Data Wrangling of Bank Transaction Records

Problem Statement:

Organizing and categorizing transaction records from multiple bank accounts into a single, easily viewable file is crucial for effective financial management. The main challenge lies in the varying header styles and unclear descriptions provided by different banks. This project aims to normalize the data into a single data file for easy viewing.

OSEMN Process:

Obtain: Downloaded CSV files from various banks' online webpages.

Scrub:

Data Cleaning: Removed unnecessary columns and rows with missing values. Added new headers for clarity and descriptiveness.

Column Splitting: Divided columns into more specific parts for better detail.

Description Cleaning: Standardized and cleaned transaction descriptions. Automated mapping of odd descriptions to simple descriptions using keyphrases.

Categorization: Added categories to transactions for improved organization.

Explore:

Data Mapping: Used JSON configs to manage different bank formats from one location.

Mapping: Employed text files to simplify complex descriptions and assign categories.

Model: Converted the final CSV to JSON and back to CSV for compatibility and flexibility.

Interpret: The cleaned and categorized data offers a clear view of all transactions across banks and accounts, aiding in tracking spending, identifying trends, and managing finances effectively.

Description of data:

The data came in multiple CSV files. Some accounts provided a single file with all transactions, while others had separate files for each statement period, which needed to be merged. All files had similar headers (Description, Amount, Date), though the names and order varied. Extra columns that were not needed were removed. Some Description fields contained comments from the payee with punctuation that disrupted typical CSV file reading, requiring manual handling. The approximated complexity of this data is 5.

Examples of Raw Data:

Description,,Summary Amt.

Beginning balance as of 12/06/2024,, "1,566.21"

Total credits,, "2,547.84"

Total debits,, "-3,100.00"

Ending balance as of 01/07/2025,, "1,014.05"

Date,Description,Amount,Running Bal.

12/06/2024,Beginning balance as of 12/06/2024,, "1,566.21"

12/11/2024,"Zelle payment from for "COSCO"; Conf# T0YBGYKWQ", "52.38", "1,618.59"

12/13/2024,"Zelle payment from for "RENT: OCTOBER";", "781.86", "2,400.45"

12/13/2024,"DES:EDI PYMNTS ID:7D404498E4 CO ID:XXXXX67207 WEB", "-1,000.00", "1,400.45"

12/18/2024,"BKOFAMERICA AMERICA TO NEW YORK NY", "1,680.00", "3,080.45"

12/23/2024," DES:EDI PYMNTS ID:9E6566D914CO ID:XXXXX67207 WEB", "-2,000.00", "1,080.45"

01/02/2025,"Bank of America 0060XXXXX84XXXXX0 CO ID:XXXXX90310 PPD", "33.59", "1,114.04"

01/02/2025,"Zelle payment 2025 vehicle storage"; Conf# aclm8hvbd", "-100.00", "1,014.04"

01/07/2025,"Interest Earned", "0.01", "1,014.05"

Transaction Date,Post Date,Description,Category,Type,Amount,Memo

01/24/2025,01/26/2025,Big Fish Games Inc.,Entertainment,Sale,-2.99,

01/24/2025,01/26/2025,McDonalds 3732,Food & Drink,Sale,-1.00,

01/23/2025,01/24/2025,MCDONALD'S F3732,Food & Drink,Sale,-1.49,

"01/25/2025", "-2.98", "*", "", "APPLE.COM/BILL 866-712-7753 CA"

"01/25/2025", "-43.00", "*", "", "COMCAST CABLE COMM 800-COMCAST OR"

"01/21/2025", "-9.99", "*", "", "APPLE.COM/BILL 866-712-7753 CA"

"01/14/2025", "2472.93", "*", "", "Bill Pay Payment"

"01/11/2025", "-5.99", "*", "", "APPLE.COM/BILL 866-712-7753 CA"

Examples of Clean CSV Data:

```
Year,Month,Date,Description,Category,Amount,Type,Bank,Card
2025,January,01/25/2025,apple store,shopping,-2.98,Credit,Bilt,Bilt
2025,January,01/25/2025,xfinity internet,utility,-43.0,Credit,Bilt,Bilt
2025,January,01/21/2025,apple store,shopping,-9.99,Credit,Bilt,Bilt
2025,January,01/14/2025,transfer,transfer,2472.93,Credit,Bilt,Bilt
2025,January,01/11/2025,apple store,shopping,-5.99,Credit,Bilt,Bilt
2025,January,01/07/2025,visible,utility,-25.0,Credit,Bilt,Bilt
2025,January,01/07/2025,visible,utility,-25.0,Credit,Bilt,Bilt
2025,January,01/04/2025,rent,utility,-2261.78,Credit,Bilt,Bilt
2025,January,01/02/2025,pge,utility,-88.59,Credit,Bilt,Bilt
2024,December,12/31/2024,walgreens,shopping,-8.6,Credit,Bilt,Bilt
2024,December,12/25/2024,xfinity internet,utility,-40.0,Credit,Bilt,Bilt
2024,December,12/25/2024,apple store,shopping,-1.99,Credit,Bilt,Bilt
2024,December,12/21/2024,apple store,shopping,-9.99,Credit,Bilt,Bilt
2024,December,12/12/2024,apple store,shopping,-5.99,Credit,Bilt,Bilt
2024,December,12/10/2024,transfer,transfer,2389.67,Credit,Bilt,Bilt
2024,December,12/07/2024,visible,utility,-25.0,Credit,Bilt,Bilt
2024,December,12/07/2024,visible,utility,-25.0,Credit,Bilt,Bilt
2024,December,12/03/2024,rent,utility,-2272.21,Credit,Bilt,Bilt
2024,December,12/02/2024,transfer,transfer,2485.32,Credit,Bilt,Bilt
2024,December,12/02/2024,pge,utility,-67.46,Credit,Bilt,Bilt
2024,November,11/25/2024,apple store,shopping,-1.99,Credit,Bilt,Bilt
2024,November,11/25/2024,xfinity internet,utility,-48.53,Credit,Bilt,Bilt
2024,November,11/21/2024,apple store,shopping,-9.99,Credit,Bilt,Bilt
2024,November,11/16/2024,apple store,shopping,-5.99,Credit,Bilt,Bilt
2024,November,11/13/2024,transfer,transfer,25.0,Credit,Bilt,Bilt
2024,November,11/11/2024,apple store,shopping,-5.99,Credit,Bilt,Bilt
2024,November,11/11/2024,xfinity mobile,utility,-40.26,Credit,Bilt,Bilt
2024,November,11/07/2024,visible,utility,-25.0,Credit,Bilt,Bilt
2024,November,11/07/2024,visible,utility,-5.0,Credit,Bilt,Bilt
2024,November,11/05/2024,rent,utility,-2240.55,Credit,Bilt,Bilt
2024,November,11/02/2024,bilt store,shopping,-40.0,Credit,Bilt,Bilt
2024,November,11/01/2024,pge,utility,-75.04,Credit,Bilt,Bilt
2024,October,10/23/2024,apple store,shopping,-1.99,Credit,Bilt,Bilt
2024,October,10/21/2024,apple store,shopping,-9.99,Credit,Bilt,Bilt
2024,October,10/18/2024,transfer,transfer,412.68,Credit,Bilt,Bilt
2024,October,10/16/2024,apple store,shopping,-5.99,Credit,Bilt,Bilt
2024,October,10/12/2024,xfinity mobile,utility,-68.32,Credit,Bilt,Bilt
2024,October,10/11/2024,apple store,shopping,-5.99,Credit,Bilt,Bilt
2024,October,10/07/2024,rent,utility,-201.65,Credit,Bilt,Bilt
2024,October,10/03/2024,apple store,shopping,-1.99,Credit,Bilt,Bilt
2024,October,10/01/2024,pge,utility,-77.25,Credit,Bilt,Bilt
2024,October,10/01/2024,rent,utility,-50.0,Credit,Bilt,Bilt
2024,September,09/29/2024,mcdonald's,dining,-1.49,Credit,Bilt,Bilt
```

Examples of Clean JSON Data:

```
[
  {
    "Year": "2025",
    "Month": "January",
    "Date": "01/25/2025",
    "Description": "apple store",
    "Category": "shopping",
    "Amount": "-2.98",
    "Type": "Credit",
    "Bank": "Bilt",
    "Card": "Bilt"
  },
  {
    "Year": "2025",
    "Month": "January",
    "Date": "01/25/2025",
    "Description": "xfinity internet",
    "Category": "utility",
    "Amount": "-43.0",
    "Type": "Credit",
    "Bank": "Bilt",
    "Card": "Bilt"
  },
  {
    "Year": "2025",
    "Month": "January",
    "Date": "01/21/2025",
    "Description": "apple store",
    "Category": "shopping",
    "Amount": "-9.99",
    "Type": "Credit",
    "Bank": "Bilt",
    "Card": "Bilt"
  },
  {
    "Year": "2025",
    "Month": "January",
    "Date": "01/14/2025",
    "Description": "transfer",
    "Category": "transfer",
    "Amount": "2472.93",
    "Type": "Credit",
    "Bank": "Bilt",
    "Card": "Bilt"
  },
]
```

Relevant Scripts:

File_importer.py - allows the user to drag in data which will be automatically sorted into specific folders

File_merger.py - does basic fixes to headers (renames unwanted to "*", adds headers, removes headers) then merges same bank files into one per account, handles bad lines

File_cleaner.py - further cleans the data, adding in fields, merging into a single csv file

Bad_lines_cleaner.py - prompts the user to manually fix and input recorded bad lines

Desc_cleaner.py - clean the descriptions to be more uniform and descriptive

Cat_cleaner.py - maps categories based on the description

CSV_to_JSON.py - converts the clean.csv to JSON

JSON_to_CSV.py - converts the clean.json to CSV

Most of these python scripts could be merged together, however, I decided to keep them separate for readability and to better keep track of function process.

All relevant scripts and folders are attached in the zip folder called Transaction Pipeline