Joshlynn Hilburn

Exploring my financial activity
Data Cleaning Log

• Opened raw .CSV file in Microsoft Excel
• Saved a copy as Spending.XLSX for manipulation

• Created a header row with appropriate column names (Date, Description, Value, Balance) and froze it
• Expanded all column names to fit content
• Used conditional formatting to highlight all blank cells and none existed
• Changed the Date column to date type, Description to text, Value to currency, and Balance to currency

• Sorted the Date column by ascending and descending to verify all the dates are within the specified range (01/01/2021 to 02/03/2022)
• Calculated the sum of the Value column to verify that it added up to my current bank account balance
    > This value also matched with the Balance field for the latest entry ($8683.17)
• Created a column called Category to assign each transaction type to a specific category
    > Categories included are described in the table below with examples for each

Category Name	Example Transactions
Beauty	Haircuts, esthetician services, etc.
Clothes	In-store and online clothes shopping
Deposit	Initial deposit to open bank account
Education	Coursera subscription
Entertainment	Concert tickets, video games, Spotify subscription, arcades, etc.
Food	Dine-in restaurants, delivery food, to-go and pickup food
Groceries	Food and household items from HEB, 99 Ranch, etc.
Housing	Rent, electric service, laundry machines
Income	Earnings from job
Item	Books, gifts, computer and electronic goods, shipping and tax filing services, etc.
Medical	Doctor’s appointments and prescriptions
Transportation	Gas, bus tickets
Verification	Small deposits for verification purposes

    > Created a pivot table from the data to find repeated transactions and mass-assigned appropriate Category labels; this accounted for 92 of 162 transactions so manual input was only required for the remainder
    > For unspecified transactions (Venmo, transfers, etc.) I referenced outside records such as receipts for the appropriate categories they belonged to
• Created a column named ID to function as a Primary Key for querying if needed
    > With the data sorted in ascending order by date, I assigned the first row the ID of 1 with the following rows being +1 of the one above, so that there was one unique ID for each row
    > Range of IDs is 1 to 162

• No need to check for duplicate rows as duplicates are allowed to exist; however, there were none
• After all cleaning was done, column widths were adjusted appropriately and the cleaned dataset was saved as Spending.CSV
• Locked sheet to prevent accidental edits