Project Title: Exploratory Data Analysis of Daily Household Transactions

Author: Abhiraj Bajpai

Date: September 7, 2025

---

1. Introduction

This report presents a comprehensive Exploratory Data Analysis (EDA) on a dataset of daily household transactions. The primary objective of this project is to analyze the provided transaction data to identify spending patterns, track income sources, and uncover behavioral trends in personal finance.

The scope of this analysis includes data cleaning, preprocessing, and the visualization of key metrics through univariate, bivariate, and time-series analysis. The insights derived can be used to forecast expenses and develop more effective budgeting and savings strategies. The project was completed using Python with the Pandas, Matplotlib, and Seaborn libraries in a Google Colab environment.

---

2. Dataset Description

The analysis was performed on the "Daily Household Transactions" dataset. It contains 2,461 individual transaction records, each with 8 initial features.

The columns in the dataset are as follows:

- **Date:** The timestamp when the transaction was made.
- **Mode:** The payment method used (e.g., Cash, Credit Card).
- **Category:** The general category of the transaction (e.g., Food, Transportation).
- **Subcategory:** A more detailed breakdown of the transaction category.
- **Note:** A brief description of the transaction.
- **Amount:** The monetary value of the transaction.
- **Income/Expense:** An indicator classifying the record as income, an expense, or a transfer.
- **Currency:** The currency of the transaction.

---

3. Data Cleaning and Preprocessing

To ensure the integrity and accuracy of the analysis, a rigorous data cleaning and preprocessing stage was conducted. The following steps were performed:

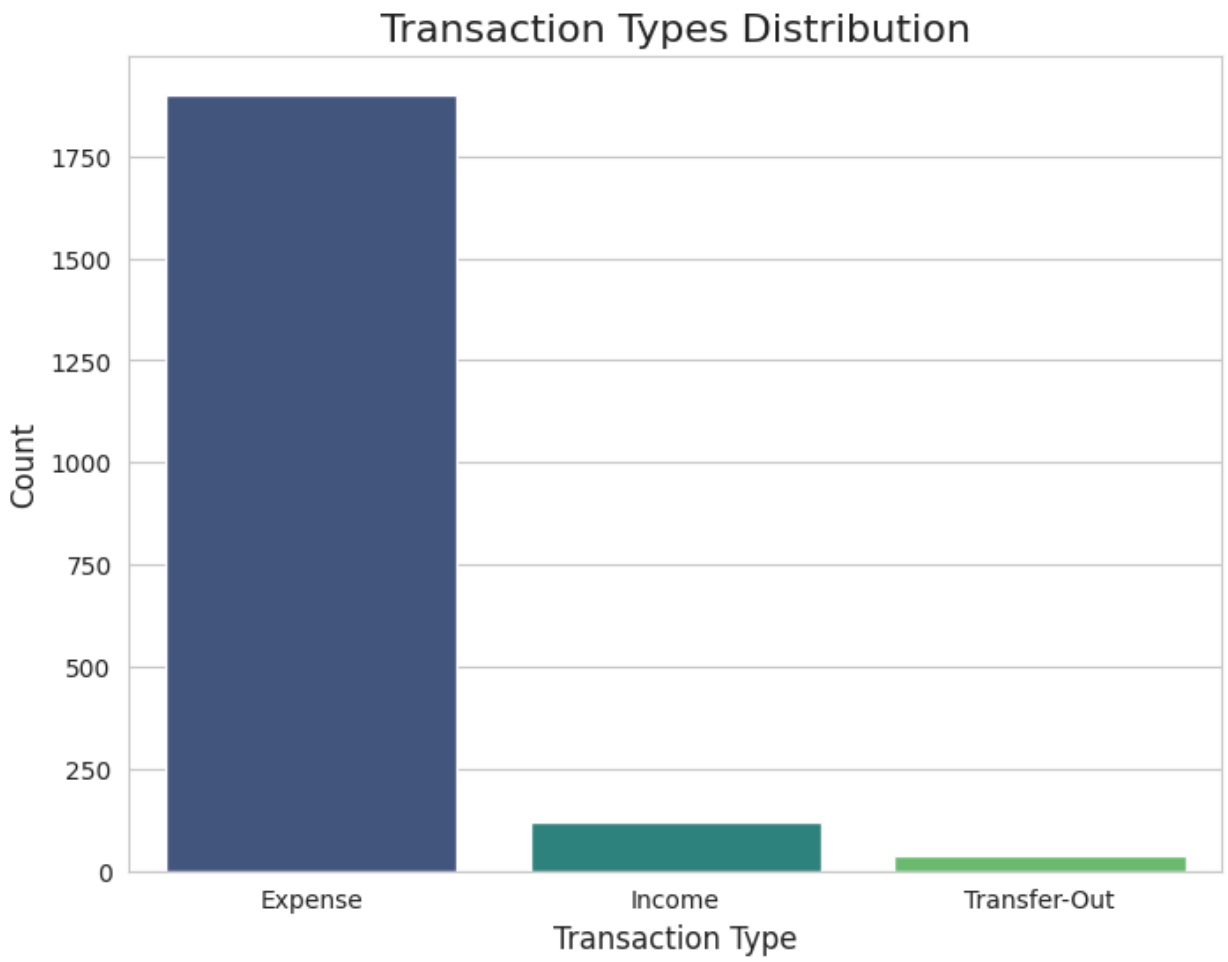
1. **Date Type Conversion:** The `Date` column was initially stored as a text format (object). It was converted into a proper `datetime` format to enable accurate time-series analysis and feature extraction.
2. **Handling Missing Values:** The `Subcategory` and `Note` columns contained 635 and 521 missing values, respectively. These were filled with placeholder text ('None' and 'No Note') to maintain data integrity without discarding valuable rows.
3. **Redundancy Removal:** The `Currency` column was dropped from the dataset, as all 2,461 transactions were recorded in INR, making the column redundant for this analysis.
4. **Duplicate Removal:** The dataset was checked for any duplicate rows, and none were found. This step ensures that each transaction is counted only once.

---

4. Exploratory Data Analysis & Key Findings

4.1 Transaction Type Distribution

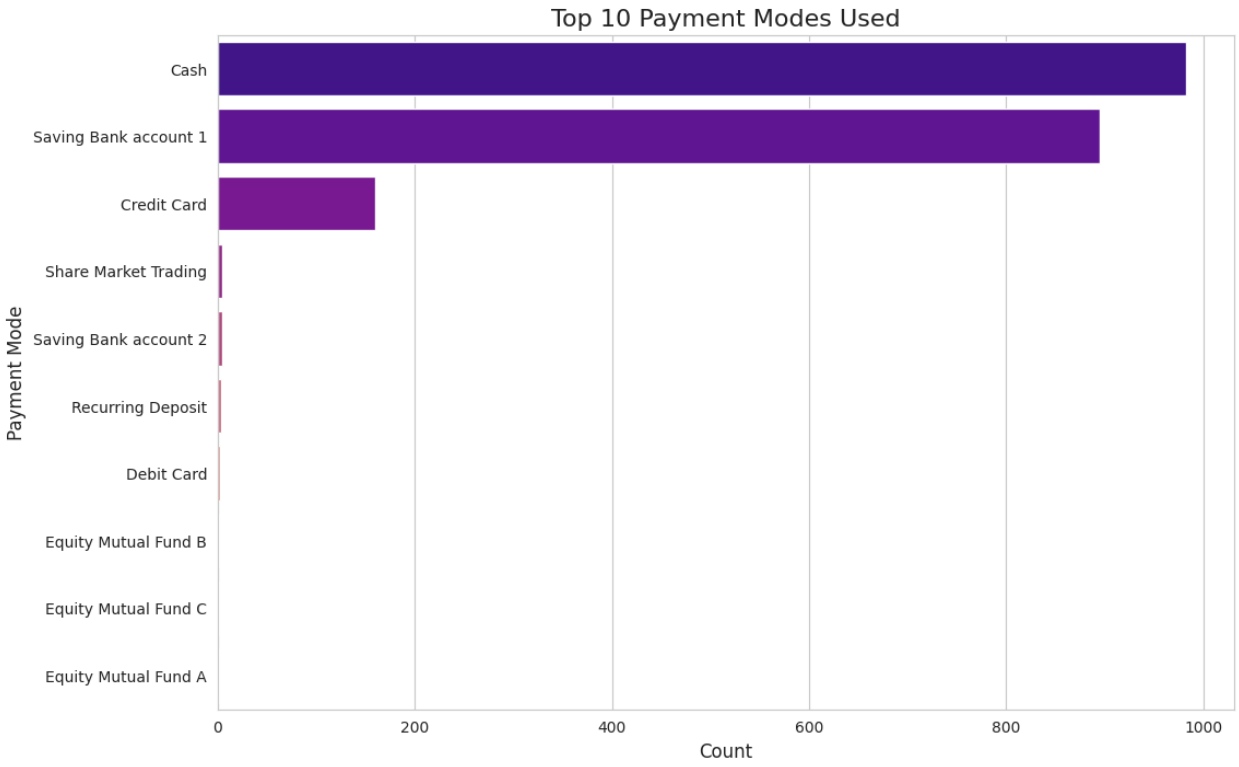
The initial analysis focused on the overall distribution of transaction types. Expenses account for the vast majority of records, indicating that the dataset is primarily used for tracking spending.



Finding: The dataset is heavily skewed towards expense tracking, with significantly fewer records for income or internal transfers.

4.2 Payment Mode Preference

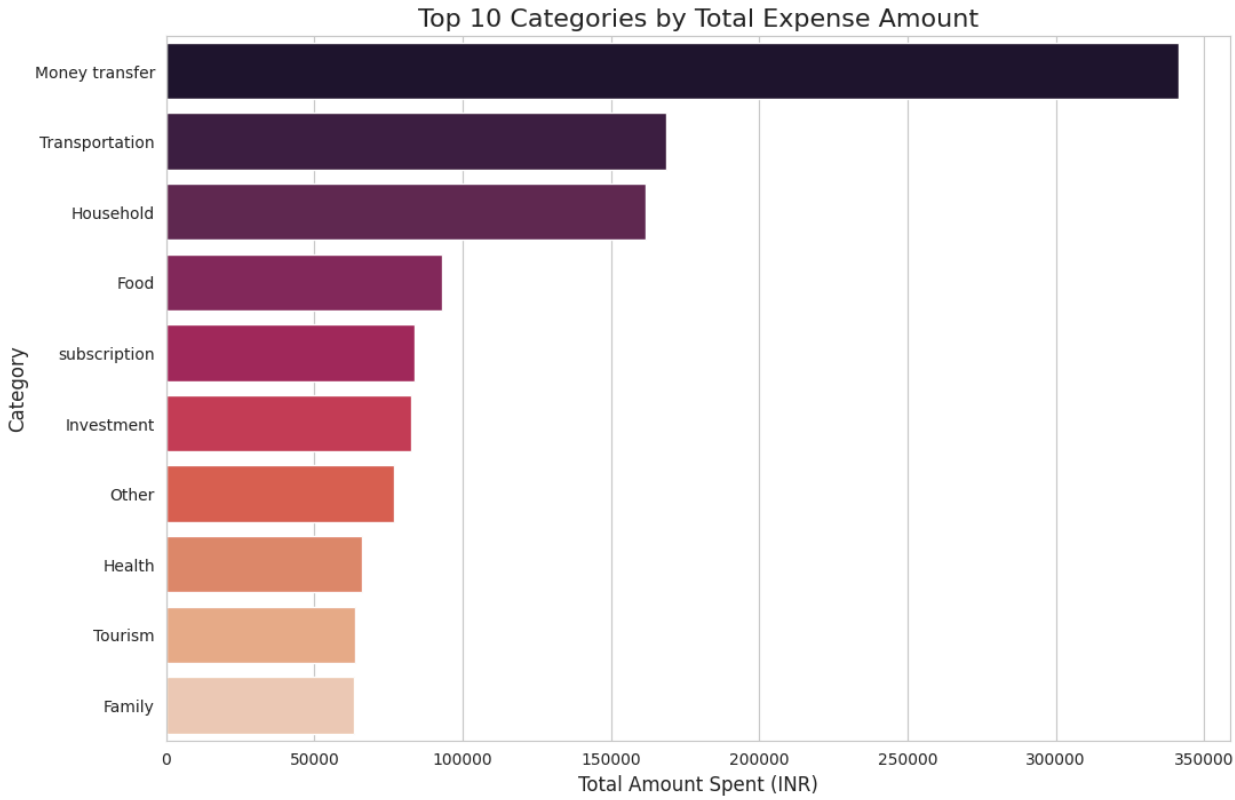
Analysis of the payment methods reveals a clear preference for direct debit and cash transactions.



Finding: 'Saving Bank account 1' and 'Cash' are the two most dominant payment modes, far surpassing the use of credit cards or other methods. This suggests a financial habit centered on using readily available funds.

4.3 Top Spending Categories

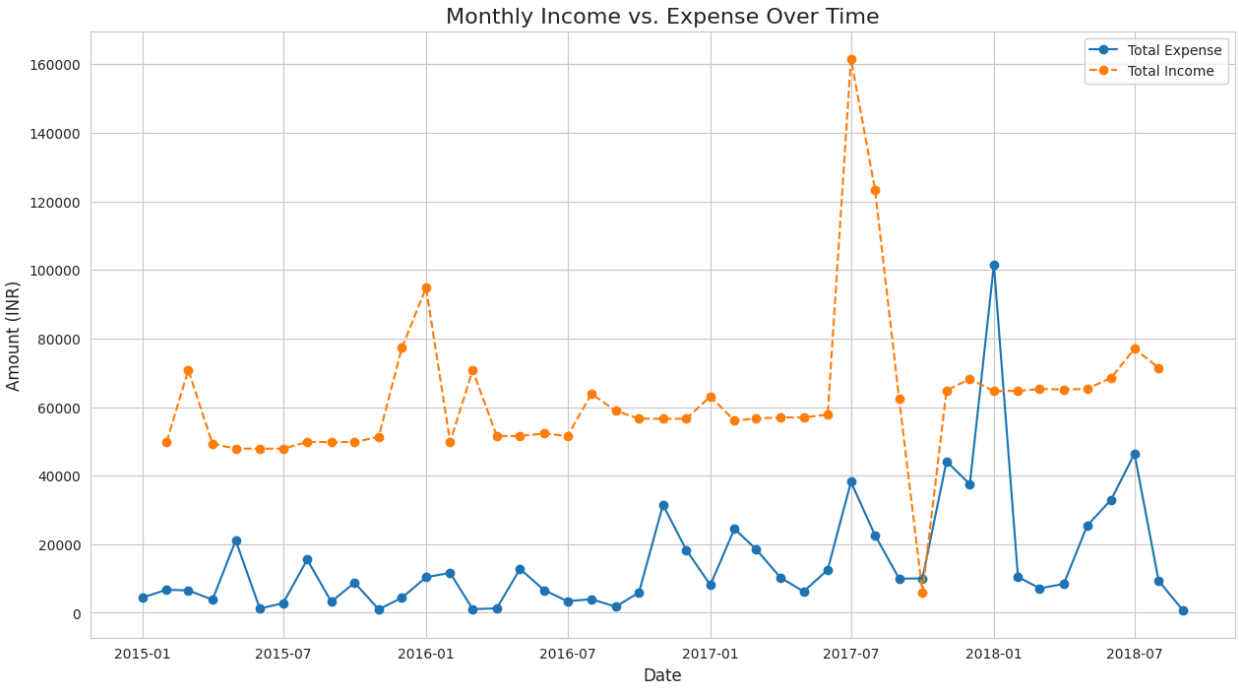
A breakdown of expenses by category highlights where the majority of money is spent.



Finding: 'Food', 'Transportation', and 'Household' are consistently the top three spending categories. This is typical for standard household expenditures and confirms that daily necessities drive the bulk of spending.

4.4 Financial Trends Over Time

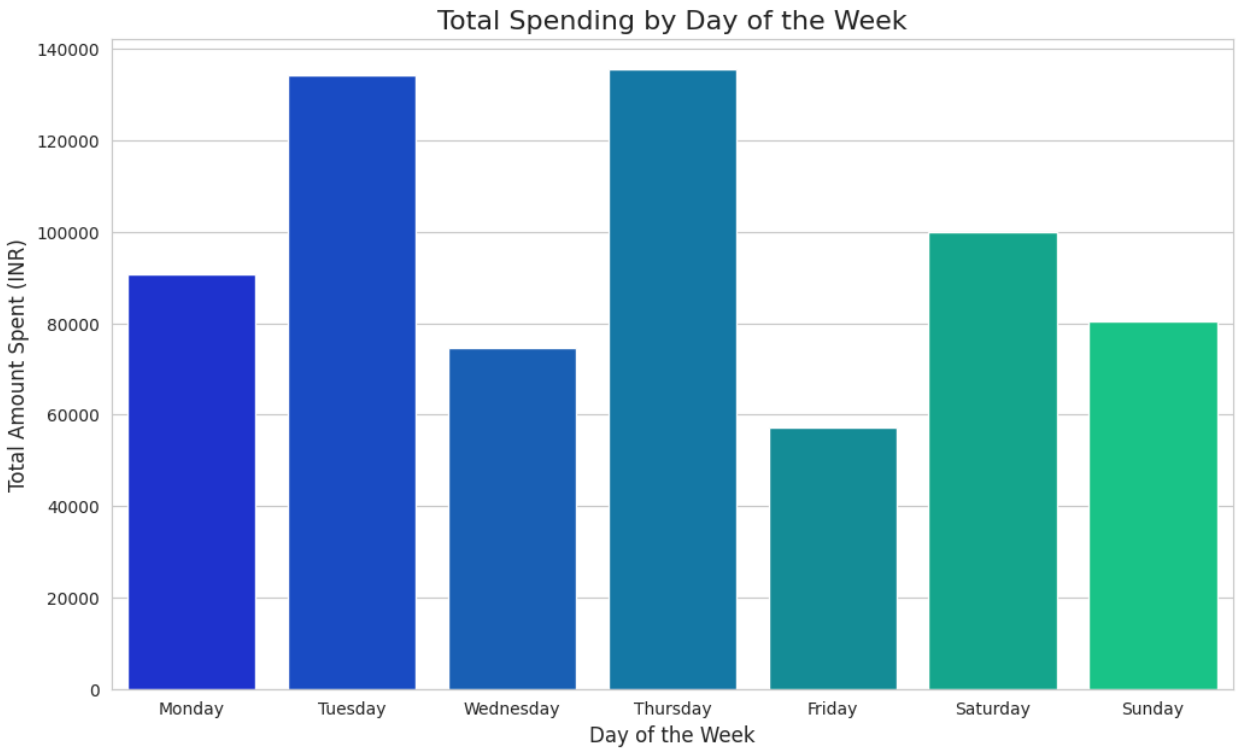
The time-series analysis of monthly income versus expenses provides a clear view of financial health over the years.



Finding: Income spikes are highly regular and occur at the end of the month, corresponding to salary credits. While expenses are more variable, certain months show significantly higher spending, which could be correlated with seasonal events, festivals, or large one-time purchases.

4.5 Weekly Spending Patterns

Analyzing spending habits across the week reveals a distinct behavioral pattern.



Finding: Spending is significantly higher on weekends, particularly on Saturdays and Sundays. This strongly suggests that major shopping, grocery runs, and leisure activities are reserved for the end of the week.

---

5. Conclusion & Recommendations

5.1 Conclusion

The exploratory data analysis successfully provided valuable insights into the financial habits recorded in the dataset. Key patterns were identified, including a reliance on cash and direct bank payments, a high proportion of spending on essential goods, predictable monthly income cycles, and a concentration of purchasing activity during weekends.

5.2 Recommendations

Based on the findings, the following recommendations are proposed to optimize financial management:

1. **Implement Category-Specific Budgets:** Given that 'Food' is the highest and most frequent expense category, creating a detailed weekly or monthly food budget could be the most effective first step in controlling overall spending.

2. **Automate Savings:** Since income arrives in a predictable pattern at the end of each month, setting up an automated transfer to a savings or investment account on the 1st of the following month would be a powerful strategy to build wealth consistently.
3. **Conduct Subcategory Analysis:** For a more granular understanding, further analysis should focus on the `Subcategory` column within high-spend categories like 'Household' and 'Food' to identify specific items or services that can be optimized for cost savings.