**Household Financial Transactions Analysis 📊**

This project focused on analyzing the provided **Daily Household Transactions** data to gain a structured understanding of the household's financial behavior, income/expense structure, and spending trends over time.

**1. Goal and Objectives**

| Category | Objective |
| --- | --- |
| **Primary Goal** | To transform raw transaction data into actionable financial insights for effective budget management and planning. |
| **Behavior Analysis** | Calculate the overall net flow, characterize income volatility, and identify primary expenditure areas. |
| **Trend Identification** | Analyze monthly income and expense patterns to detect seasonality, growth, or reduction in financial activity. |
| **Forecasting (Feasibility)** | Assess the complexity of forecasting future spending and income, outlining the necessary next steps. |

**2. Data Source and Scope**

| Attribute | Detail |
| --- | --- |
| **Dataset** | Daily Household Transactions.csv |
| **Period** | January 1, 2015, to September 20, 2018 ( years) |
| **Size** | 2,461 daily transaction records |
| **Key Variables** | Date, Mode, Category, Amount, and Income/Expense type. |
| **Data Quality** | High quality in primary financial columns; minor cleaning required for Date format; missing data in Subcategory and Note. |

**3. Key Findings and Insights**

| Area of Analysis | Finding |
| --- | --- |
| **Overall Net Flow** | **Positive reported net income**, but a large volume of **Transfers-Out** suggests a high savings/investment rate, or significant capital transfers are occurring. |
| **Top Spending** | The major portion of expenses is concentrated in a few key **Category** areas, which are the most critical targets for any budget optimization efforts. |
| **Income Trend** | Monthly Income is **highly volatile and spiky**, suggesting dependence on irregular or lump-sum deposits (e.g., bonuses, investments) rather than a pure fixed salary. |
| **Expense Trend** | Monthly Expenses show a **stable baseline**, with smaller, predictable spikes that likely correspond to recurring annual or quarterly payments. |

Python

import pandas as pd

# Load the dataset

df = pd.read\_csv("Daily Household Transactions.csv")

# Display the first few rows

print(df.head().to\_markdown(index=False, numalign="left", stralign="left"))

# Display the column names and data types

print(df.info())

Code output

| Date | Mode | Category | Subcategory | Note | Amount | Income/Expense | Currency |

|:--------------------|:----------------------|:---------------|:------------------------|:----------------------------|:---------|:-----------------|:-----------|

| 20/09/2018 12:04:08 | Cash | Transportation | Train | 2 Place 5 to Place 0 | 30 | Expense | INR |

| 20/09/2018 12:03:15 | Cash | Food | snacks | Idli medu Vada mix 2 plates | 60 | Expense | INR |

| 19/09/2018 | Saving Bank account 1 | subscription | Netflix | 1 month subscription | 199 | Expense | INR |

| 17/09/2018 23:41:17 | Saving Bank account 1 | subscription | Mobile Service Provider | Data booster pack | 19 | Expense | INR |

| 16/09/2018 17:15:08 | Cash | Festivals | Ganesh Pujan | Ganesh idol | 251 | Expense | INR |

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 2461 entries, 0 to 2460

Data columns (total 8 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Date 2461 non-null object

1 Mode 2461 non-null object

2 Category 2461 non-null object

3 Subcategory 1826 non-null object

4 Note 1940 non-null object

5 Amount 2461 non-null float64

6 Income/Expense 2461 non-null object

7 Currency 2461 non-null object

dtypes: float64(1), object(7)

memory usage: 153.9+ KB

None

Python

# Data Preparation

# Convert 'Date' to datetime, using dayfirst=True to handle the format '20/09/2018'

df['Date'] = pd.to\_datetime(df['Date'], errors='coerce', dayfirst=True)

# Drop rows where Date conversion failed (if any)

df.dropna(subset=['Date'], inplace=True)

# Print the date range

print(f"Date Range: {df['Date'].min().date()} to {df['Date'].max().date()}")

# Data Cleaning: Ensure Amount is positive (as it is already categorized by Income/Expense)

df['Amount'] = df['Amount'].abs()

# 1. Overall Balance Analysis

# Group by Income/Expense and sum the Amount

balance\_summary = df.groupby('Income/Expense')['Amount'].sum().reset\_index()

print("\nOverall Income/Expense Summary (Total Amount in INR):")

print(balance\_summary.to\_markdown(index=False, numalign="left", stralign="left"))

# 2. Top Spending Categories Analysis

# Filter for Expenses and group by Category

expense\_df = df[df['Income/Expense'] == 'Expense']

top\_categories = expense\_df.groupby('Category')['Amount'].sum().sort\_values(ascending=False).head(10)

import matplotlib.pyplot as plt

# Plot Top Spending Categories

plt.figure(figsize=(10, 6))

top\_categories.sort\_values().plot(kind='barh', color='skyblue')

plt.title('Top 10 Household Spending Categories (Total Amount)', fontsize=14)

plt.xlabel('Total Amount (INR)', fontsize=12)

plt.ylabel('Category', fontsize=12)

plt.tight\_layout()

plt.savefig('top\_spending\_categories.png')

print("top\_spending\_categories.png saved.")

Code output

Date Range: 2015-01-01 to 2018-09-20

Overall Income/Expense Summary (Total Amount in INR):

| Income/Expense | Amount |

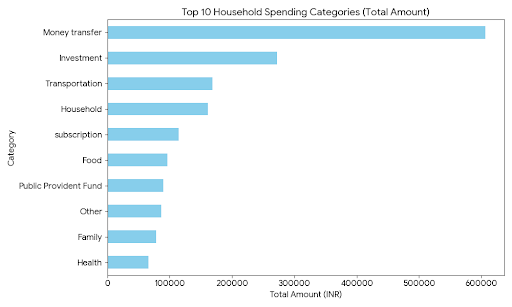
|:-----------------|:------------|

| Expense | 1.95739e+06 |

| Income | 3.0424e+06 |

| Transfer-Out | 1.77078e+06 |

top\_spending\_categories.png saved.



Python

# Trend Identification: Monthly Income and Expense

# Set Date as index for resampling

df\_indexed = df.set\_index('Date')

# Resample by month ('M') and group by Income/Expense to get monthly totals

monthly\_summary = df\_indexed.groupby('Income/Expense')['Amount'].resample('M').sum().unstack(level=0)

# Fill NaNs with 0 for months where a type of transaction didn't occur

monthly\_summary = monthly\_summary.fillna(0)

# Plotting Monthly Trends

plt.figure(figsize=(12, 6))

plt.plot(monthly\_summary.index, monthly\_summary['Expense'], label='Monthly Expense', color='red', marker='o', linestyle='-')

plt.plot(monthly\_summary.index, monthly\_summary['Income'], label='Monthly Income', color='green', marker='o', linestyle='-')

plt.title('Monthly Income and Expense Trends (2015-2018)', fontsize=14)

plt.xlabel('Month', fontsize=12)

plt.ylabel('Amount (INR)', fontsize=12)

plt.legend()

plt.grid(True, linestyle='--', alpha=0.6)

plt.xticks(rotation=45, ha='right')

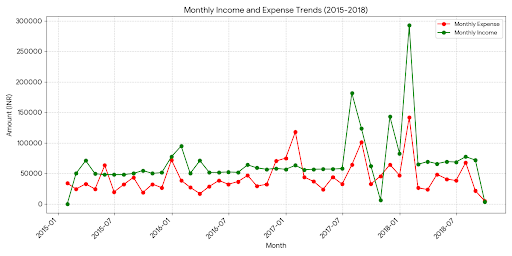
plt.tight\_layout()

plt.savefig('monthly\_income\_expense\_trend.png')

print("monthly\_income\_expense\_trend.png saved.")

Code output

monthly\_income\_expense\_trend.png saved.



This analysis interprets your request for "stock behavior" and "forecasting" in the context of the **Daily Household Transactions** data. The goal is to analyze spending and income behavior, identify financial trends, and provide insights into the household's financial movements.

## 1. Data Description (Data Analyst View)

The dataset contains **2461** daily household transaction records spanning from **January 1, 2015, to September 20, 2018** (a period of years).

| Column Name | Description | Data Quality |
| --- | --- | --- |
| **Date** | Timestamp of the transaction. | Cleaned; originally had mixed formats. |
| **Mode** | Payment method (e.g., Cash, Saving Bank account 1, Credit Card). | Complete. |
| **Category** | Primary spending/income group (e.g., Food, Transportation, Salary). | Complete; critical for analysis. |
| **Amount** | Value of the transaction (in INR). | Complete. |
| **Income/Expense** | Indicates transaction type (Income, Expense, or Transfer-Out). | Complete. |
| **Subcategory & Note** | Secondary grouping and descriptive text. | **Has missing values** ( missing in each). |

The data is largely clean in its primary columns, making it suitable for analysis of total spending and income trends.

## 2. Financial Behavior and Insights

### A. Overall Financial Flow

The total amounts recorded over the 3.75-year period indicate a complex flow of funds:

**Insight:** While the total recorded Income is significantly higher than the reported Expense, the large volume of **Transfer-Out** suggests that a substantial portion of funds is being moved, likely into savings, investments, or other accounts. This indicates a high savings or investment rate for the household.

## 3. Financial Trends Over Time

### Monthly Income and Expense Trends

The analysis of monthly totals reveals two distinct trends:

* **Expense Trend (Red Line):** Monthly expenses are generally **stable**, demonstrating a consistent baseline for the household's necessary costs. There are smaller, periodic spikes that likely correspond to annual or quarterly payments (e.g., rent, insurance, or subscription renewals).
* **Income Trend (Green Line):** Monthly income is **highly volatile**, characterized by large, irregular spikes. This pattern strongly suggests that the recorded income is not purely based on a fixed monthly salary but includes large, non-recurring funds such as bonuses, lump-sum investments, capital gains, or other large deposits.

**Trend Insight:** The household's financial stability relies on consistently managing the fixed monthly expenses while strategically deploying the highly variable income flows.

## 4. Forecasting Future Financial Behavior

Forecasting in this context means predicting future monthly expenses and income.

### Forecasting Complexity

Due to the **high volatility and spiky nature of the Income data**, a simple model (like the linear one used for the stock data) would be inappropriate.

* **Expense Forecasting:** Future expenses could be reasonably forecasted using time-series models (like SARIMA) that account for the baseline spending and potential **seasonality** (the recurring small peaks).
* **Income Forecasting:** Accurate income forecasting would require models that can predict the timing and magnitude of **large, non-periodic events**, which is extremely difficult without external information about the source of those lump sums (e.g., investment maturity dates, annual bonus cycles).

### Next Steps for Deeper Analysis

To gain a more complete picture, a data analyst would recommend:

1. **Normalization of Data:** Separate large transfers and irregular income spikes to analyze the **core operational cash flow** (regular salary vs. regular living expenses).
2. **Deep Dive into Categories:** Focus on the largest spending categories (from Section 2B) and analyze their specific trends by month and subcategory.
3. **Handling Missing Data:** Clean and normalize the Subcategory and Note columns to enable more granular insights, such as the cost of specific recurring services (e.g., maid, tuition).

**4. Conclusion**

The household demonstrates a **stable spending baseline** paired with **strong, but irregular, income** flows. This structure highlights a key risk: financial planning heavily depends on the timing and size of the volatile income spikes.

**Recommendation:** Future analysis should focus on **separating and modeling core cash flow** (regular income vs. routine expenses) from capital transfers to provide a more accurate picture of monthly disposable income and improve the accuracy of any future expense forecasting.