Unified Mentor, Gurugram, Haryana

Dissertation Project Report

On

Daily Transactions

Presented by:

Swapnil Prabhakar Chavan

Introduction

- **Objective:** Analyze household transactions to find patterns in spending and income.
- **Importance:** Supports budgeting, cost optimization, and financial planning.
- **Scope:** Daily transactions from a household finance dataset.

Dataset Overview

- **CSV File:** Daily Household Transactions.csv
- **Rows:** ~2,461
- **Columns:** Date, Mode, Category, Subcategory, Amount, Income/Expense, Currency
- Mix of income and expense transactions
- Missing values in Subcategory and Note columns

Columns found: ['Date', 'Mode', 'Category', 'Subcategory', 'Note', 'Amount', 'Income/Expense', 'Currency']

Sample rows:

Date	Mode	Category	Subcategory	Note	Amount	Income/Expense	Currency
20/09/2018 12:04:08	Cash	Transportation	Train	2 Place 5 to Place 0	30.0	Expense	INR
20/09/2018 12:03:15	Cash	Food	snacks	Idli medu Vada mix 2 plates	60.0	Expense	INR
19/09/2018	Saving Bank account 1	subscription	Netflix	1 month subscription	199.0	Expense	INR
17/09/2018 23:41:17	Saving Bank account 1	subscription	Mobile Service Provider	Data booster pack	19.0	Expense	INR
16/09/2018 17:15:08	Cash	Festivals	Ganesh Pujan	Ganesh idol	251.0	Expense	INR

Tools & Libraries

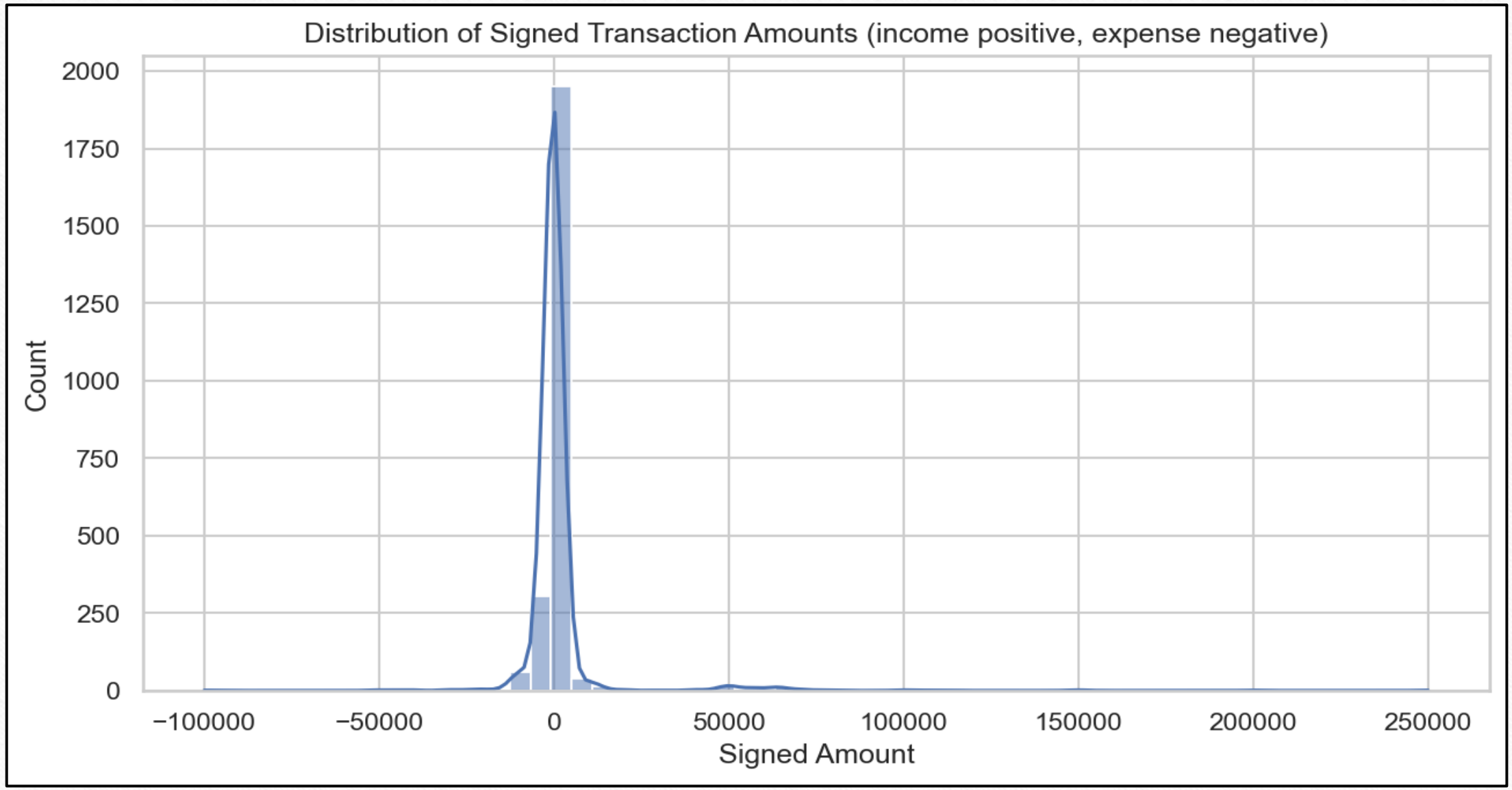
- **Tools:**
- **Python 3** with Jupyter Notebook
- **Libraries:**
- **Pandas** – Used for data cleaning & manipulation
- **Numpy** – Used for numerical calculations
- **Matplotlib** – Used for plotting and graphs
- **Seaborn** – Used for advanced data visualization

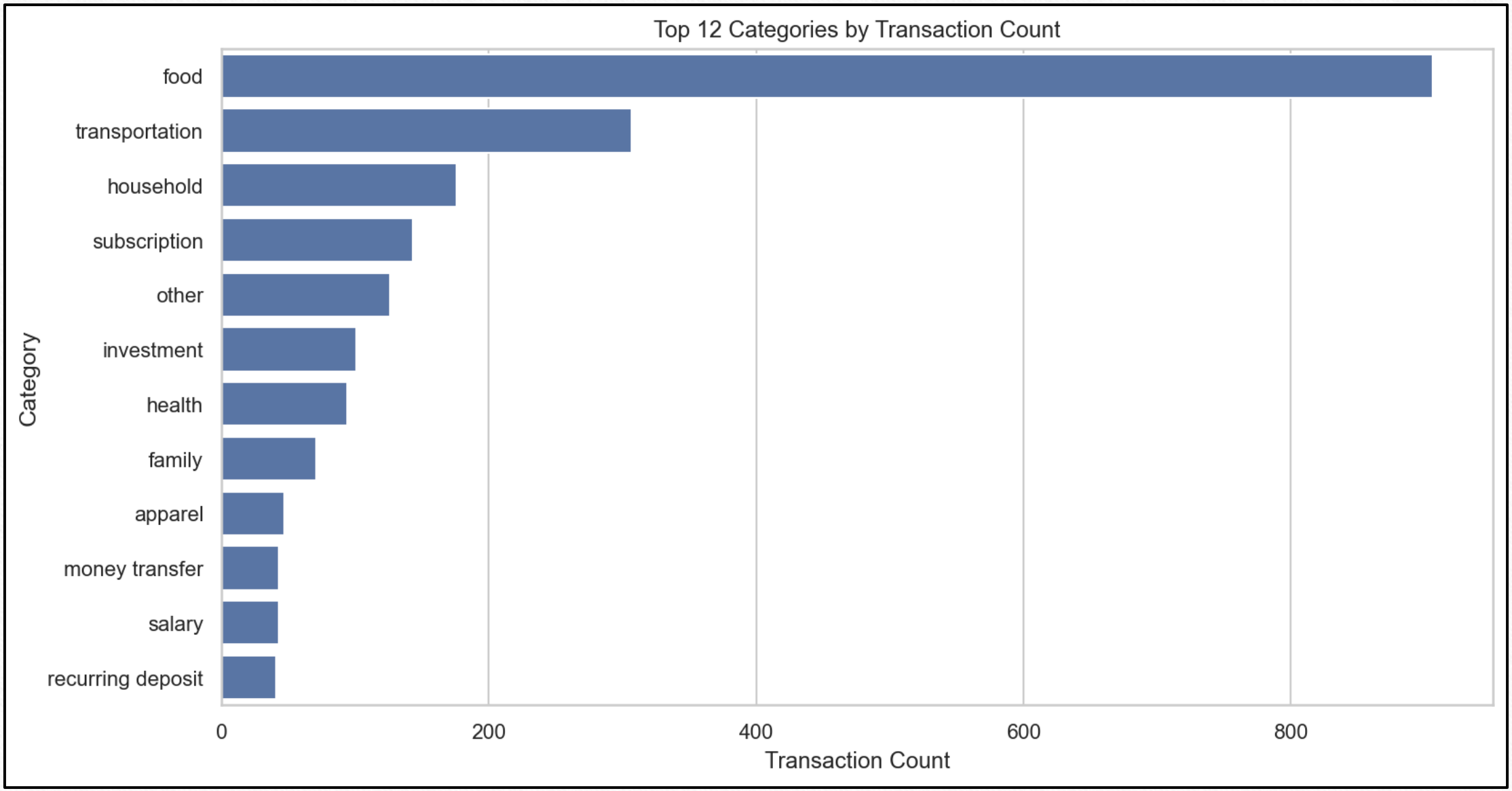
Data Cleaning Steps

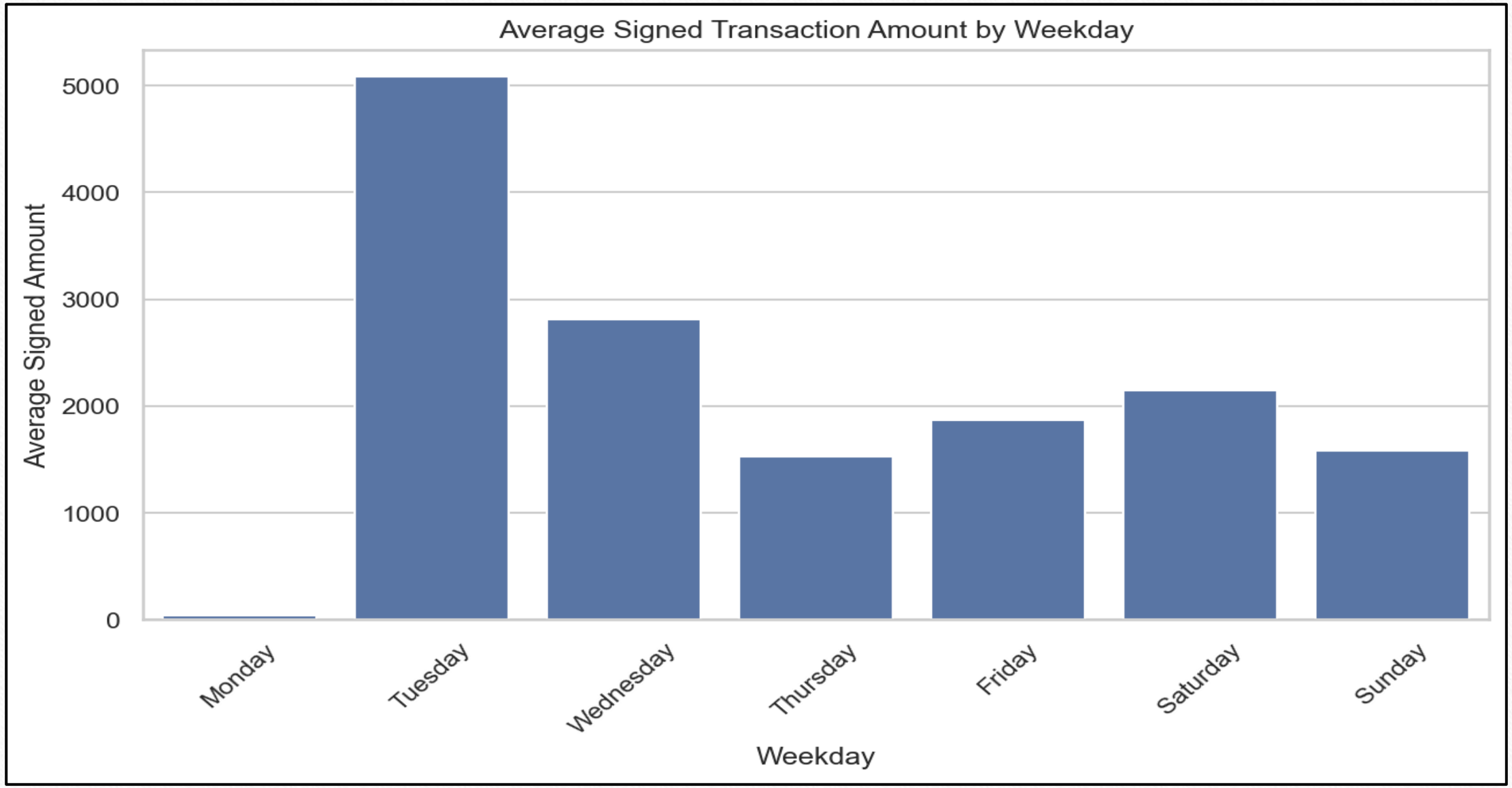
- Removed duplicates
- Handled missing values with fill or drop methods
- Standardized text (lowercase, trimmed spaces)
- Converted Amount to numeric values
- Created new columns: Year, Month, Weekday, Signed Amount

Exploratory Data Analysis (EDA)

- Transaction count by category
- Transaction amount distribution (income positive, expense negative)
- Most used payment modes
- Average transaction size

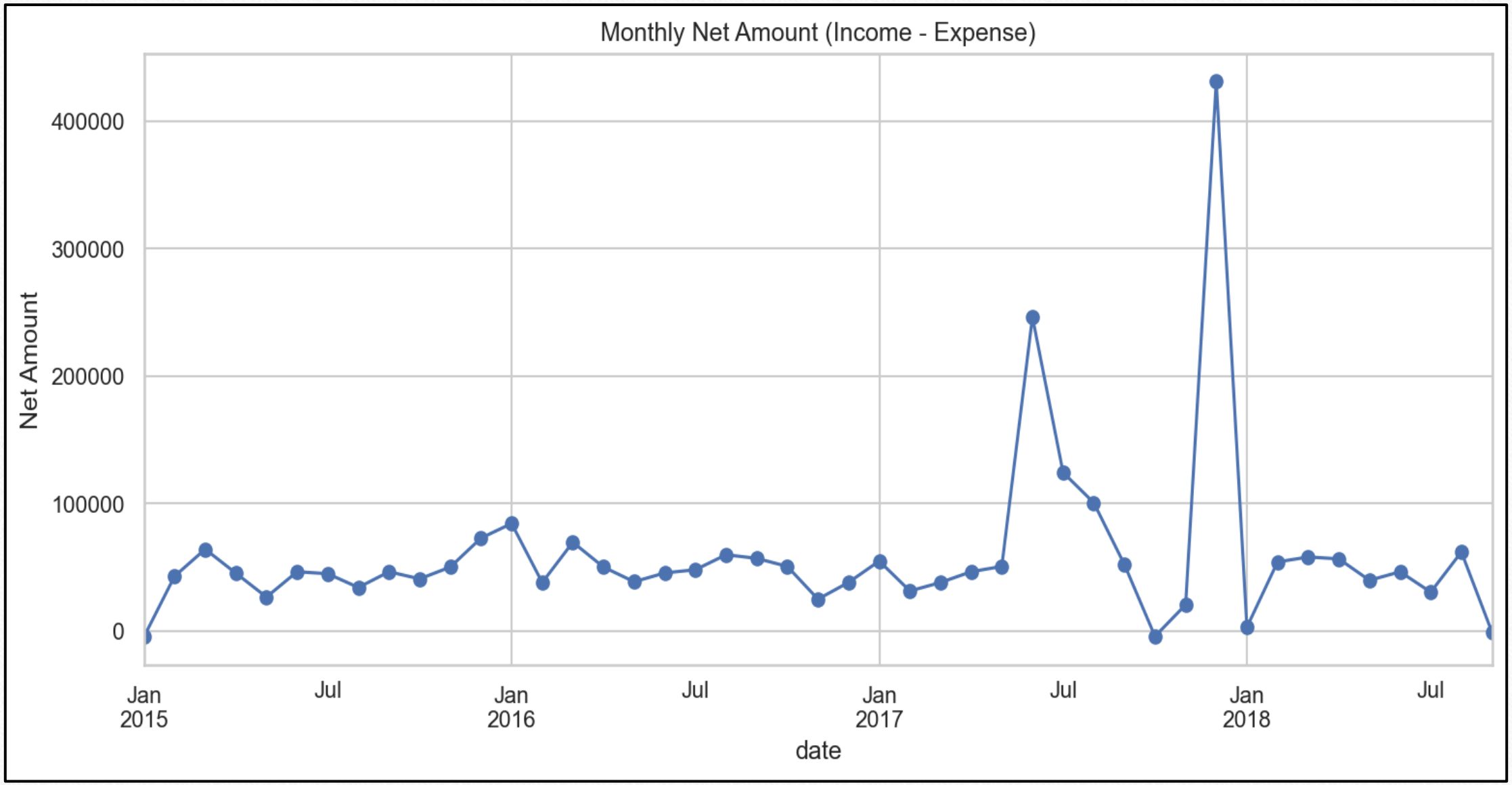

Distribution of Signed Transaction Amounts (income positive, expense negative)
Count
2000
1750
1500
1250
1000
750
500
250
0
−100000
−50000
0
50000
100000
150000
200000
250000
Signed Amount

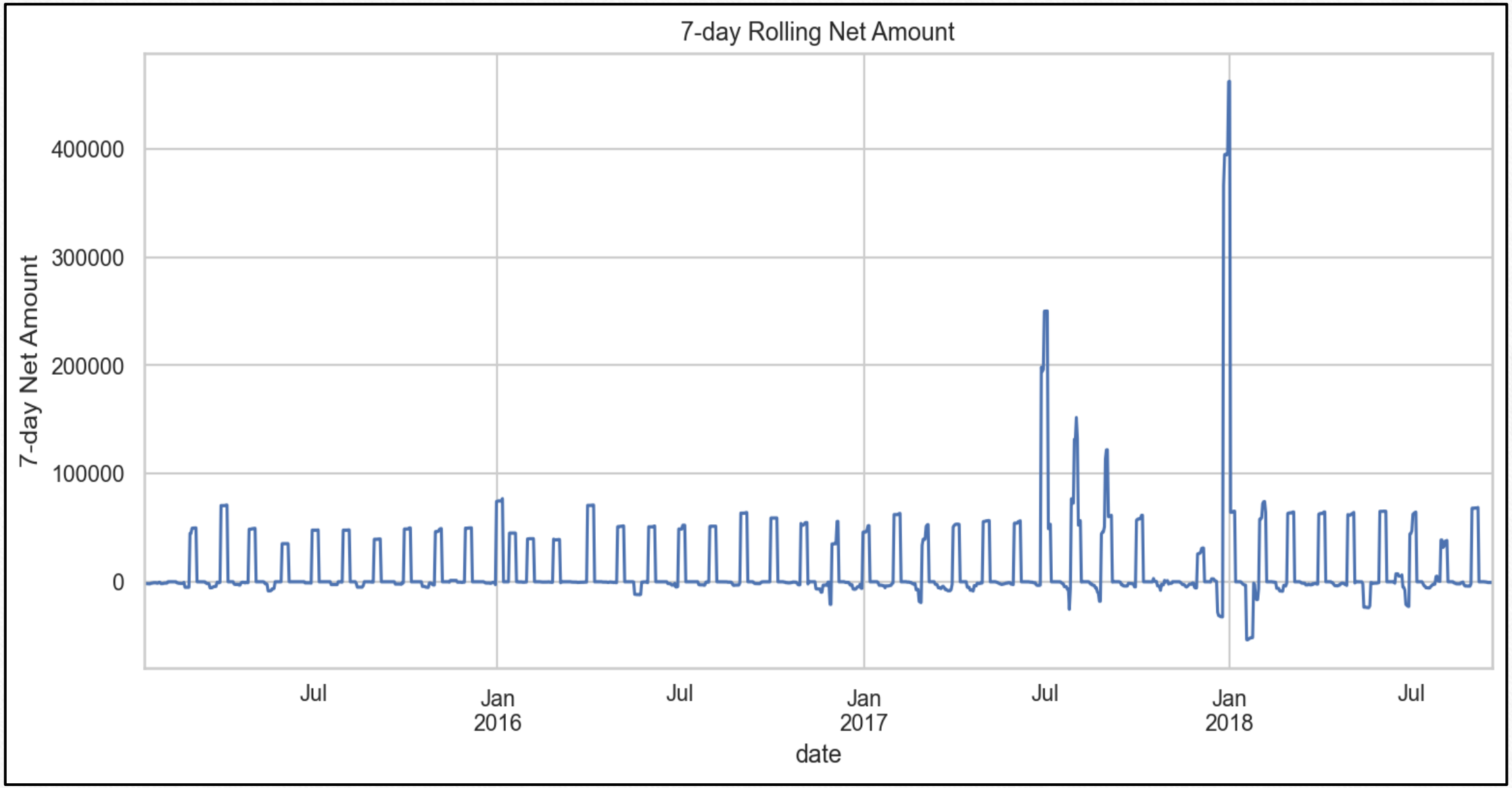

Top 12 Categories by Transaction Count
food
transportation
household
subscription
other
investment
health
family
apparel
money transfer
salary
recurring deposit
Category
0
200
400
600
800
Transaction Count


Average Signed Transaction Amount by Weekday
Average Signed Amount
5000
4000
3000
2000
1000
0
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Weekday

Time Series Trends

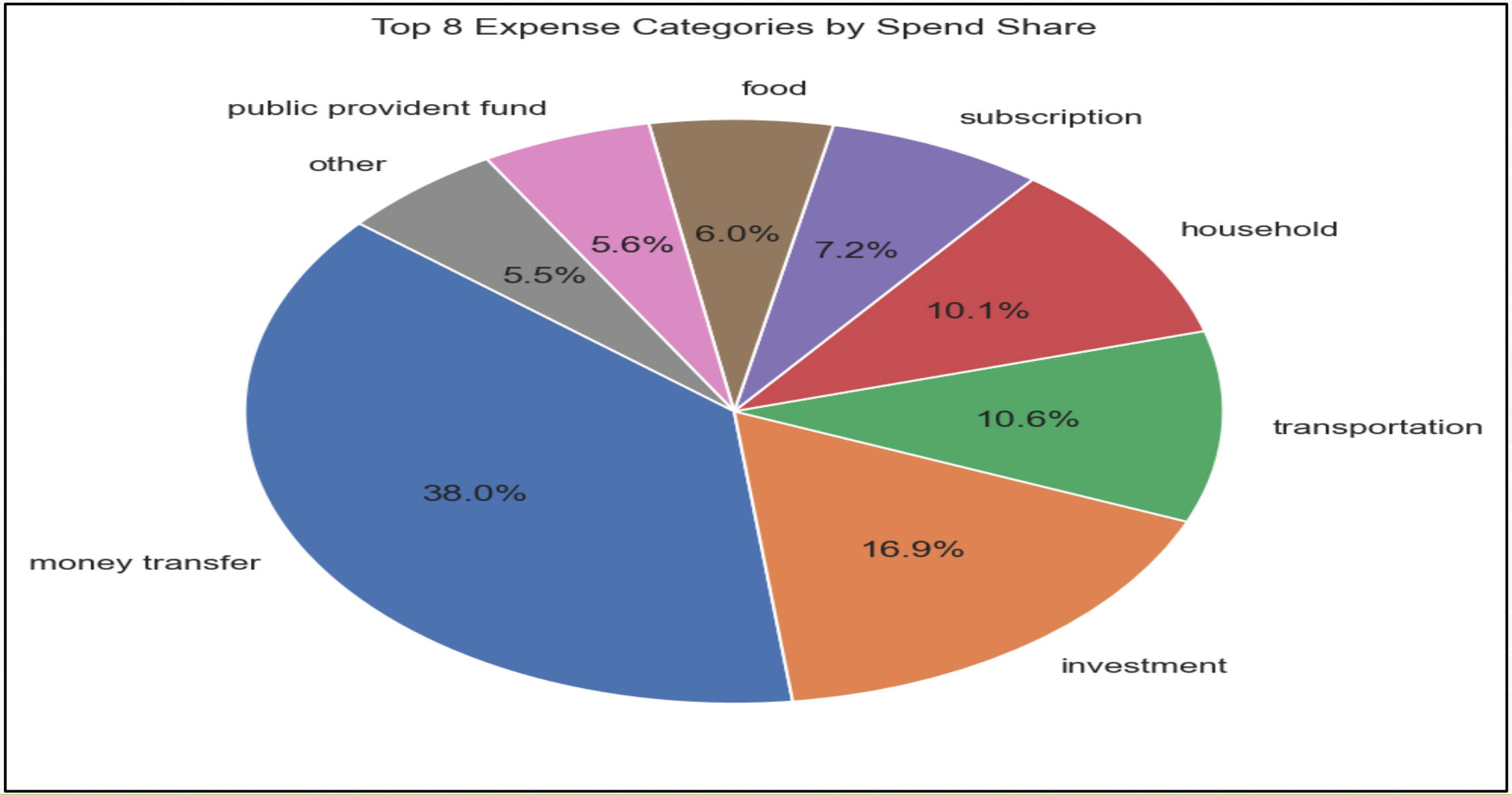
- Monthly net balance trend
- Rolling 7-day spending average
- Seasonal peaks in expenses

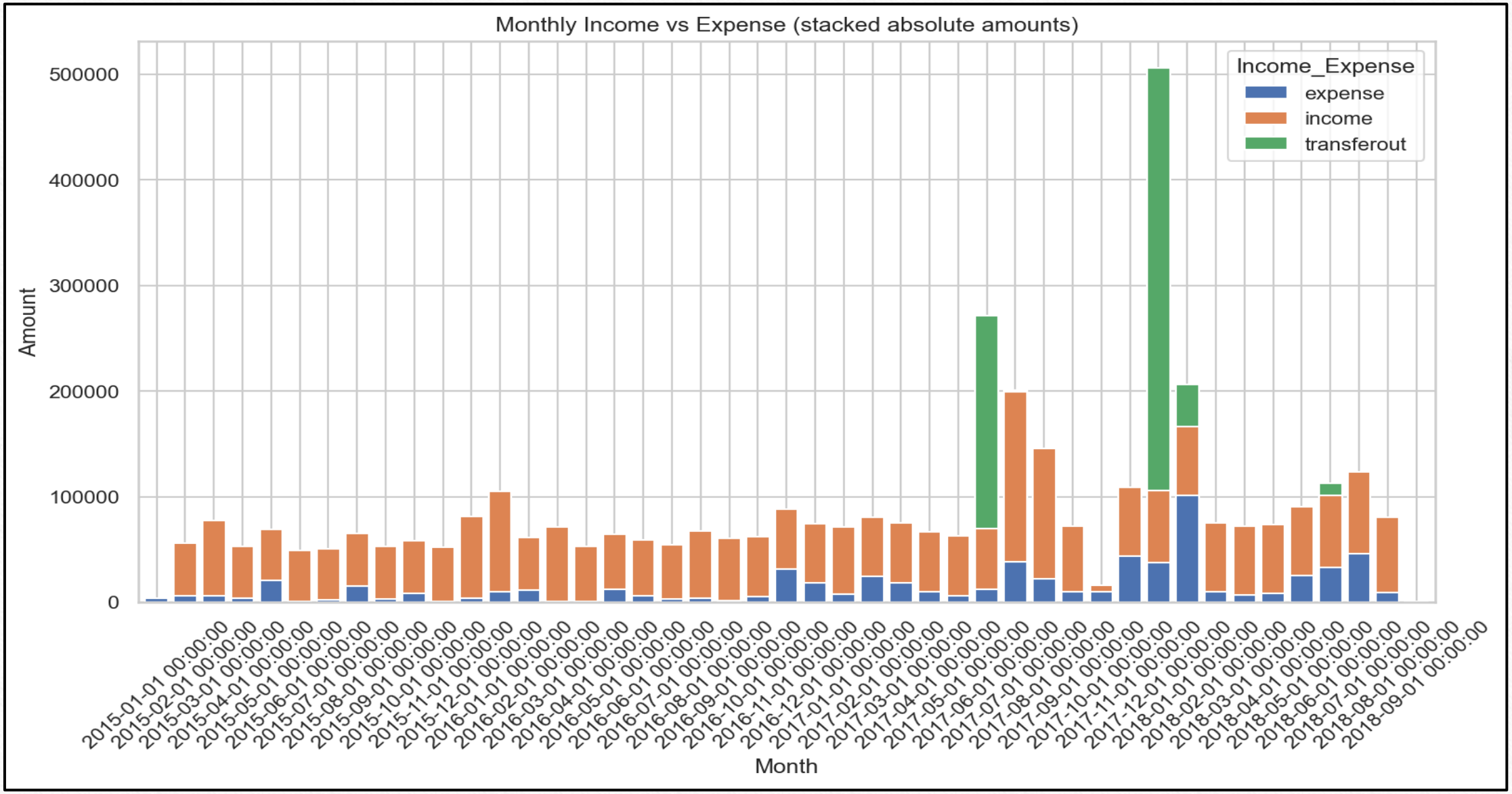

Monthly Net Amount (Income - Expense)
Net Amount
400000
300000
200000
100000
0
Jan
2015
Jul
Jan
2016
Jul
Jan
2017
Jul
Jan
2018
Jul
date

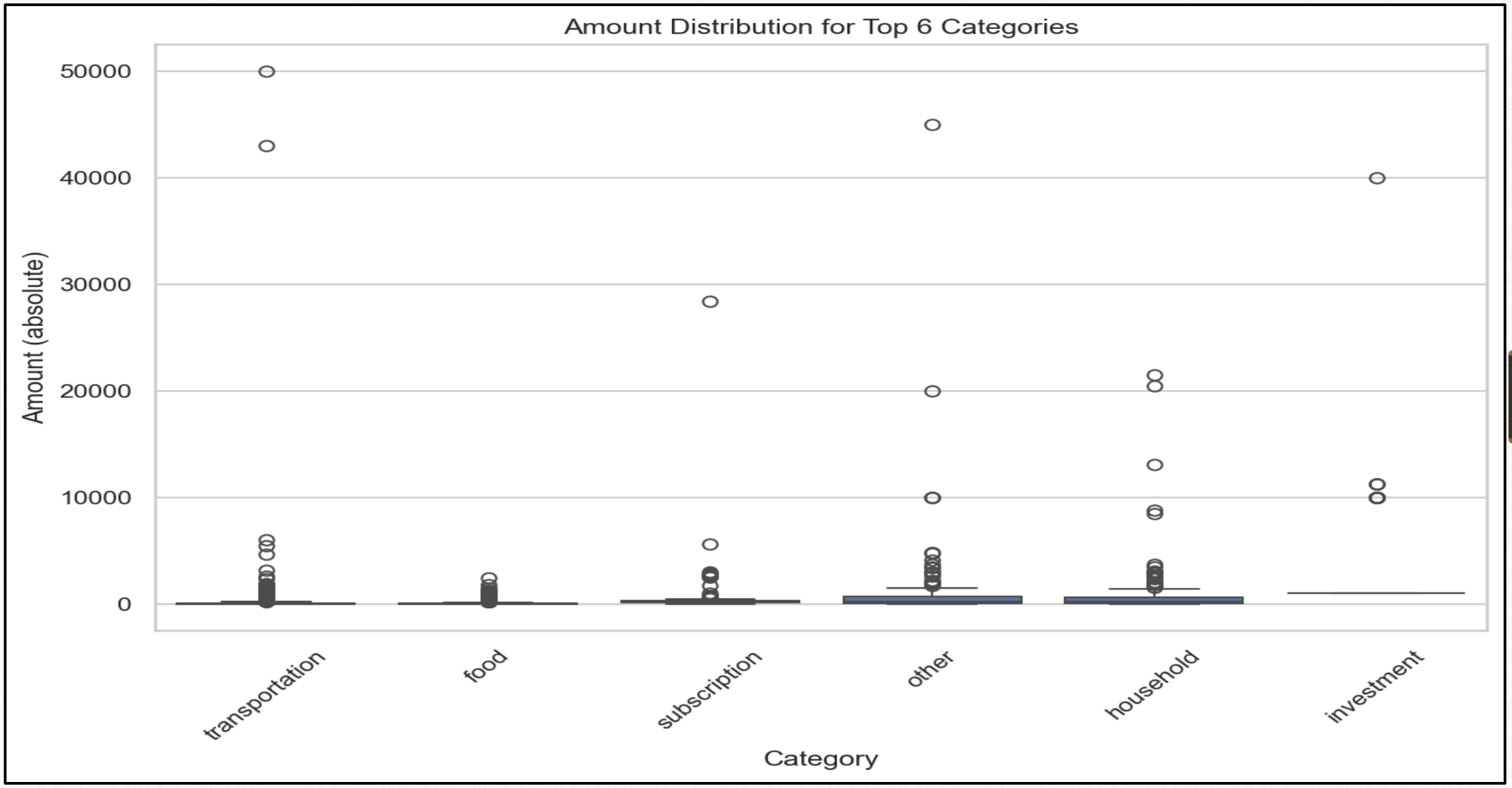

7-day Rolling Net Amount
7-day Net Amount
400000
300000
200000
100000
0
Jul
Jan
2016
Jul
Jan
2017
Jul
Jan
2018
Jul
date

Category & Mode Analysis

- Top expense categories share of spending
- Relationship between payment mode and spending
- Variations in transaction amounts by category

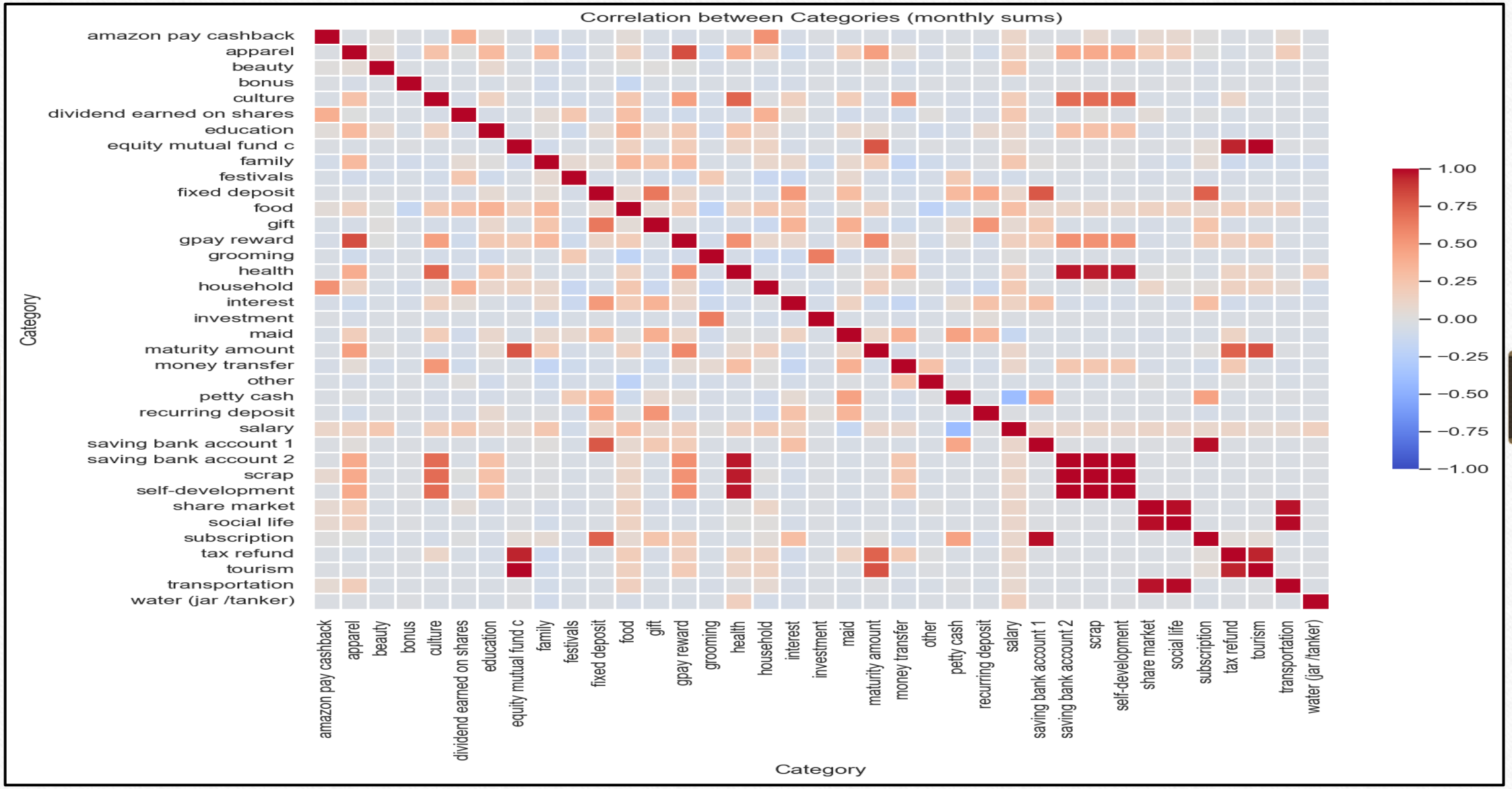

Top 8 Expense Categories by Spend Share
food
public provident fund
subscription
other
household
5.6%
6.0%
7.2%
5.5%
10.1%
10.6%
transportation
38.0%
16.9%
money transfer
investment


Monthly Income vs Expense (stacked absolute amounts)
Income_Expense
expense
income
transferout
Amount
500000
400000
300000
200000
100000
0
Month
2015-01-01 00:00:00
2015-02-01 00:00:00
2015-03-01 00:00:00
2015-04-01 00:00:00
2015-05-01 00:00:00
2015-06-01 00:00:00
2015-07-01 00:00:00
2015-08-01 00:00:00
2015-09-01 00:00:00
2015-10-01 00:00:00
2015-11-01 00:00:00
2015-12-01 00:00:00
2016-01-01 00:00:00
2016-02-01 00:00:00
2016-03-01 00:00:00
2016-04-01 00:00:00
2016-05-01 00:00:00
2016-06-01 00:00:00
2016-07-01 00:00:00
2016-08-01 00:00:00
2016-09-01 00:00:00
2016-10-01 00:00:00
2016-11-01 00:00:00
2016-12-01 00:00:00
2017-01-01 00:00:00
2017-02-01 00:00:00
2017-03-01 00:00:00
2017-04-01 00:00:00
2017-05-01 00:00:00
2017-06-01 00:00:00
2017-07-01 00:00:00
2017-08-01 00:00:00
2017-09-01 00:00:00
2017-10-01 00:00:00
2017-11-01 00:00:00
2017-12-01 00:00:00
2018-01-01 00:00:00
2018-02-01 00:00:00
2018-03-01 00:00:00
2018-04-01 00:00:00
2018-05-01 00:00:00
2018-06-01 00:00:00
2018-07-01 00:00:00
2018-08-01 00:00:00
2018-09-01 00:00:00


Amount Distribution for Top 6 Categories
Amount (absolute)
50000
40000
30000
20000
10000
0
transportation
food
subscription
other
household
investment
Category

Correlation Analysis

- Correlation between monthly category spends
- Strong positive/negative relationships help spot related expense habits


Correlation between Categories (monthly sums)
Category
amazon pay cashback
apparel
beauty
bonus
culture
dividend earned on shares
education
equity mutual fund c
family
festivals
fixed deposit
food
gift
gpay reward
grooming
health
household
interest
investment
maid
maturity amount
money transfer
other
petty cash
recurring deposit
salary
saving bank account 1
saving bank account 2
scrap
self-development
share market
social life
subscription
tax refund
tourism
transportation
water (jar /tanker)
Category
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

Conclusion & Recommendations

- **Summary:** Spending dominated by a few categories, seasonal peaks, clear mode preferences
- **Recommendations:**
  - Reduce discretionary spending in top expense categories
  - Allocate budget limits for high-expense months
  - Consider more cost-efficient payment modes

Thank you!