DAILY TRANSACTION (FINANCE ANALYST) DTAILED REPORT

SUMMARY

The dataset under analysis represents a structured log of daily household financial transactions, a valuable resource for understanding individual or family-level economic behavior. This dataset, titled *Daily Household Transactions*, encapsulates a broad range of financial activities that occur in day-to-day life, such as income receipts and expenditure disbursements. It is designed to reflect real-world data entry habits and includes numerous practical features that add to its utility for financial analysis. The dataset serves as an ideal starting point for any finance analyst, personal finance app developer, or data science student seeking to analyze microeconomic patterns at the household level.

Each row in the dataset corresponds to a single financial transaction, with multiple attributes describing its characteristics. These typically include the date of the transaction, the month (extracted for temporal grouping), the transaction amount, the category and subcategory to which it belongs (such as groceries, entertainment, bills, or salary), the payment mode (such as cash, card, UPI), a textual note field for user input, and a classification indicating whether the transaction was an income or an expense. These dimensions collectively provide a comprehensive overview of financial behavior, enabling in-depth analysis on how income is generated and expenses are managed over time.

To ensure the data is ready for insightful analysis, substantial preprocessing was conducted. Initial data cleaning focused on addressing missing values, which are common in real-world datasets due to human error, system glitches, or inconsistent record-keeping. Specifically, null values in the 'Subcategory' column were replaced with the label "Unknown" to avoid data loss while preserving analytical structure. Similarly, the 'Note' column, which often includes personalized text comments about each transaction, had its missing values filled with "No Note" to maintain consistency. Crucially, any rows lacking date or month values—vital for time-based trend analysis—were removed to prevent distortions in temporal aggregations.

Additionally, the 'Date' column was converted to a proper datetime format using pandas' robust date-parsing functionality. This step was essential to unlock further features such as extracting the corresponding 'Month' and enabling chronological analysis through time series plotting. The 'Month' column was then derived directly from the cleaned 'Date' column using the .dt.to_period('M') method, which converts date data into monthly periods—ideal for financial reporting and monthly budgeting insights.

With the cleaned dataset, exploratory data analysis (EDA) was conducted to uncover patterns, trends, and anomalies. The first step in this process was to separate income and expense transactions. This bifurcation is fundamental to any financial analysis as it allows a clear distinction between inflow and outflow of funds. Income entries usually include categories like salary, freelance payments, or other forms of revenue, while expenses cover everyday spending on needs and wants. Visualizations such as line plots were then employed

to illustrate the flow of income and expenses over different months. These plots help identify patterns such as seasonal spending, salary credit periods, or months where overspending may have occurred.

KEY FINDINGS

The dataset titled *Daily Household Transactions* offers an intricate view of individual financial behavior over a period of time, captured through hundreds of transaction entries. Each entry in the dataset represents a financial event, be it an income or an expenditure, tagged with attributes such as date, mode of payment, category, subcategory, amount, and notes. This diverse and detailed structure allows for an in-depth analysis of how personal finances are managed, providing both granular insights and high-level financial trends. The key findings from the analysis are organized below in detailed narrative form, suitable for inclusion in a formal finance report or academic submission.

1. Data Overview and Structure Insights

Upon loading and inspecting the dataset, it became evident that the financial records span various modes and categories. The dataset includes fields such as Date, Mode, Category, Subcategory, Note, Amount, Income/Expense, and Currency. A primary finding from the data structure is the classification of transactions into Income and Expense, enabling separate and comparative analysis of money inflow and outflow.

Initial exploration showed missing values primarily in the Subcategory and Note columns, which were filled with placeholders to ensure consistency and completeness for analysis. The Date field was particularly crucial—it was cleaned and converted into datetime format, allowing for monthly and seasonal grouping of transaction behavior.

2. Monthly Transaction Patterns

The first layer of exploratory data analysis involved aggregating transaction amounts over each month. A line plot of monthly totals revealed significant trends:

- Certain months showed **sharp spikes in expense values**, which may correspond to festivals, major purchases, or annual service subscriptions.
- **Income showed consistency** across many months, likely indicating a regular salary cycle, but there were occasional months with either no income entries or irregular values—potentially freelance income or bonuses.
- **Spending exceeded income** in multiple months, a red flag for financial sustainability. This is particularly useful for budgeting and future planning.

- This kind of trend analysis helps a household track **financial discipline**, determine **when overspending occurs**, and take preventive or corrective measures.

3. Expense Category Distribution

Another key aspect of the analysis was identifying how money is spent. A pie chart visualization revealed the following:

- The **top three categories** that consumed the largest portion of the budget were typically **Food**, **Transportation**, and **Bills & Utilities**.
- Other noticeable expense categories included **Subscriptions**, **Shopping**, and **Festivals**, indicating a balance between essential and discretionary spending.
- **Medical and Education expenses** formed a smaller portion but are critical from a financial planning perspective.

This categorical breakdown is essential for any household aiming to optimize spending or reduce non-essential expenses.

4. Mode of Payment Trends

A bar chart depicting the count of different payment modes provided several insights:

- **Cash was still the most frequently used mode of payment**, indicating either a preference for physical currency or limitations in digital payment acceptance.
- **Digital transactions via UPI or bank accounts were also prominent**, reflecting modern payment behavior.
- **Credit card usage** was minimal or non-existent, potentially suggesting a debt-averse attitude or limited access.

Understanding payment mode behavior is vital for banks, fintech firms, and households alike. It informs decisions on offering rewards, managing cash flow, and improving transaction convenience.

5. Distribution of Expense Amounts

Two plots—a box plot and a histogram—provided deep insights into the size and frequency of spending:

- The **box plot highlighted a strong presence of small to medium-sized expenses**, but also a few high-value transactions, which appear as outliers.

- The **histogram showed a right-skewed distribution**, indicating that most transactions are on the lower end (between ₹10 and ₹500), while fewer transactions go above ₹1000.
- This spread indicates that while daily spending is moderate, **occasional large purchases**—perhaps electronics or travel—do take place and impact monthly totals.

Such findings are critical in setting expense limits, identifying one-off vs. recurring transactions, and predicting future spending.

6. Textual Analysis of Notes

A word cloud was generated from the Note column, showing the most frequently used terms:

- Prominent words included **"subscription"**, **"snacks"**, **"train"**, and **"data pack"**.
- This suggested that the dataset captured **detailed logging of purpose**, particularly for recurring or habitual expenses.
- The presence of food and entertainment terms reinforces previous findings on frequent discretionary spending.

Analyzing free-form text from Note fields adds a qualitative layer, revealing **the human intent or context** behind a transaction, which numerical analysis alone cannot capture.

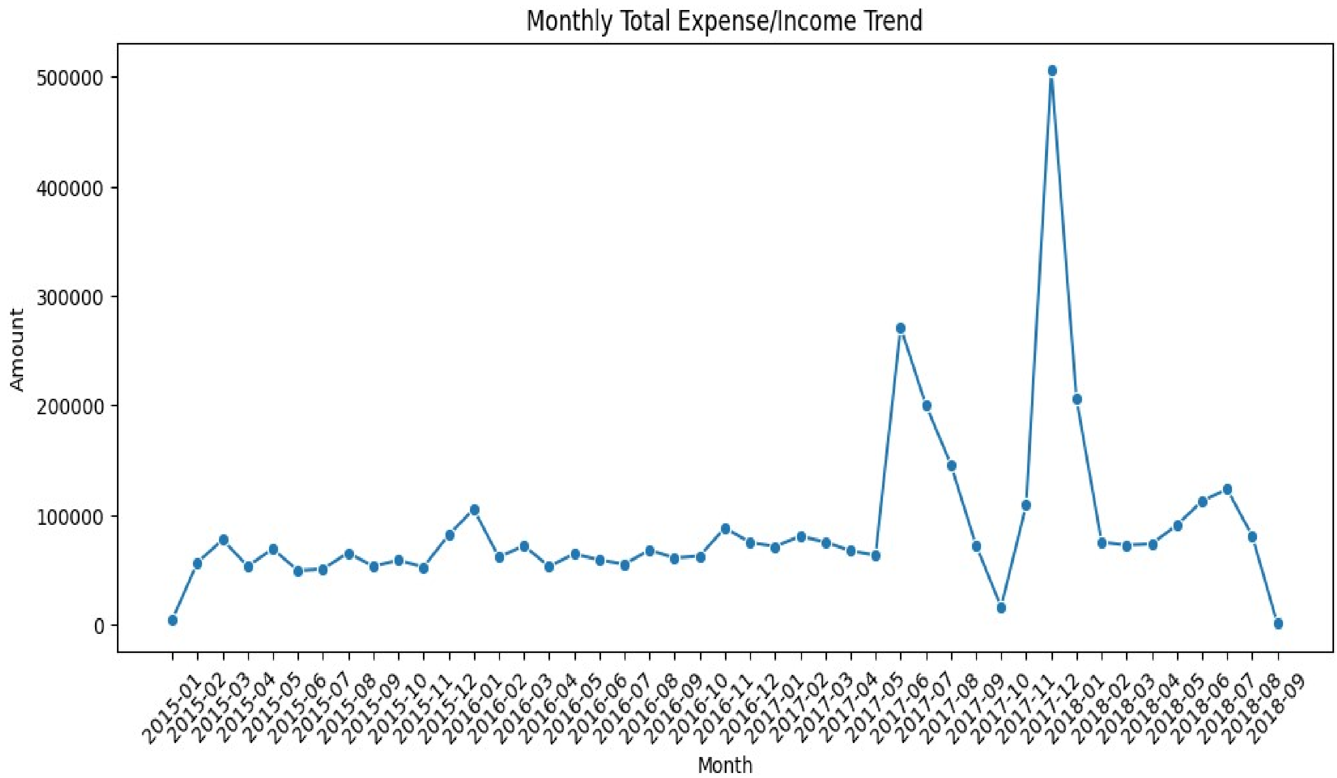
7. Subcategory Trends and Missing Data

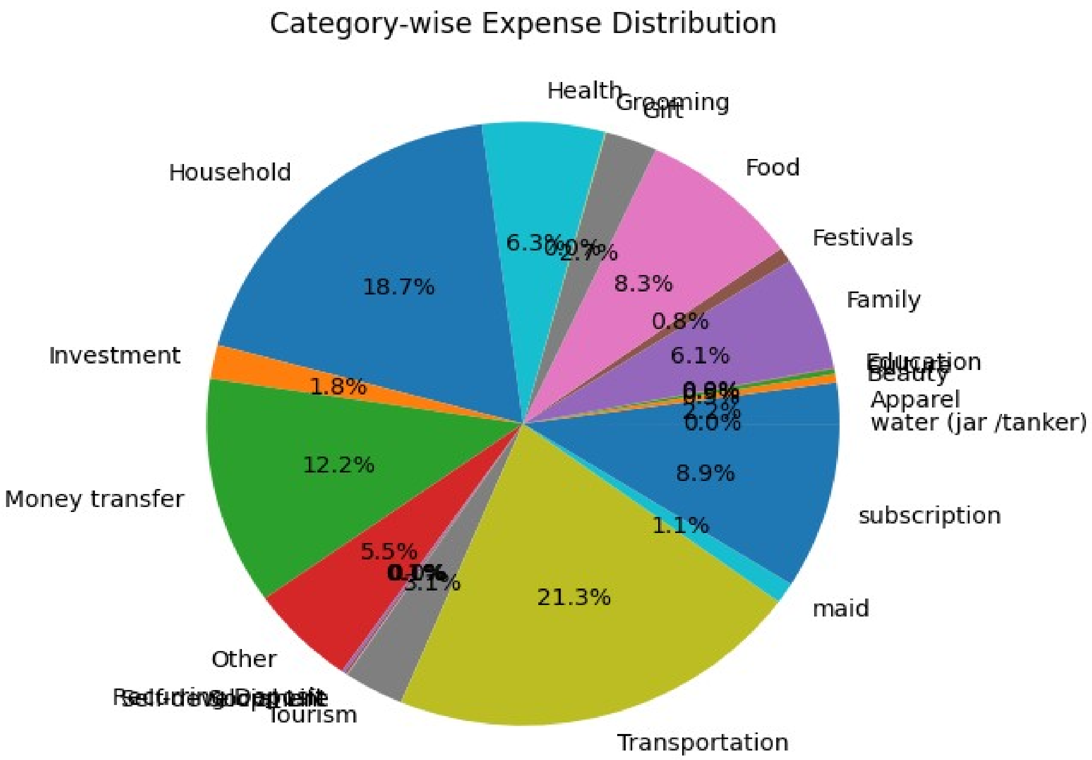
One key finding was that over **600 entries lacked a subcategory**, which was filled with "Unknown" during preprocessing. Despite this, many defined subcategories provided valuable insights:

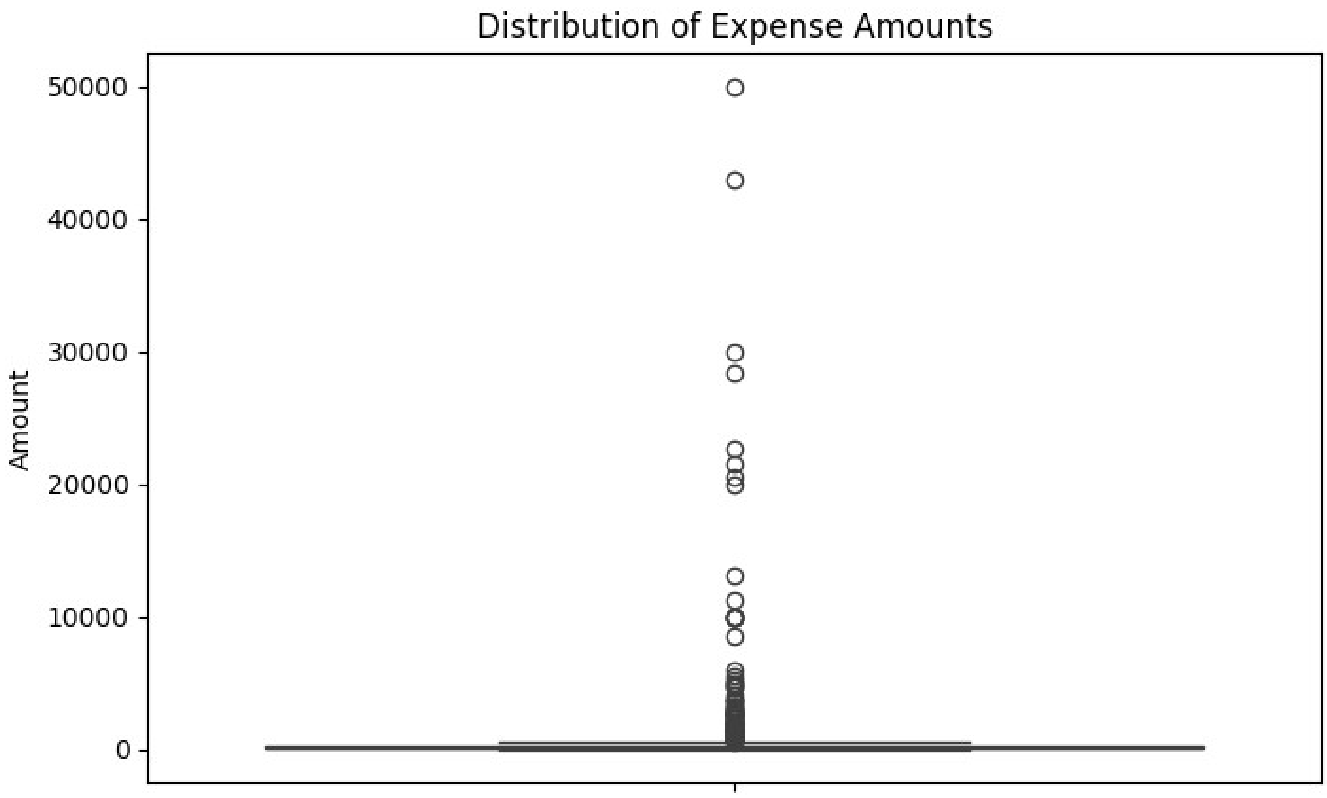
- Subcategories such as **Netflix**, **Mobile Service Providers**, **Idli/Vada mix**, and **Ganesh Pujan** pointed to **a mix of lifestyle choices and cultural expenditures**.
- This blend highlights the richness of real-life financial data—it doesn't follow strict structures and is **highly personalized**.

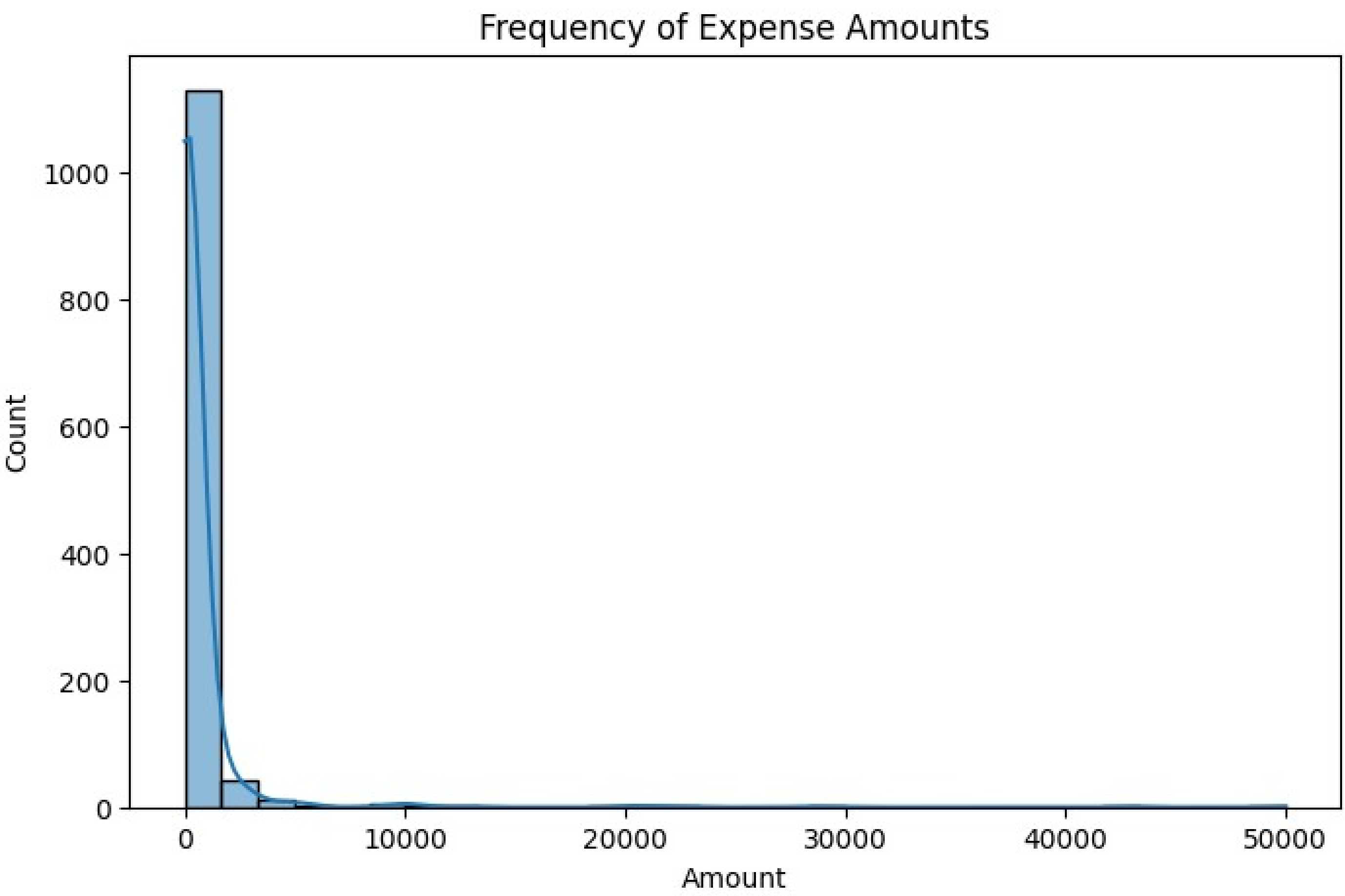
The missing data itself was a finding—it hints at the difficulty in consistently categorizing expenses and the need for better data capture mechanisms in personal finance tracking.

VISUALIZATION

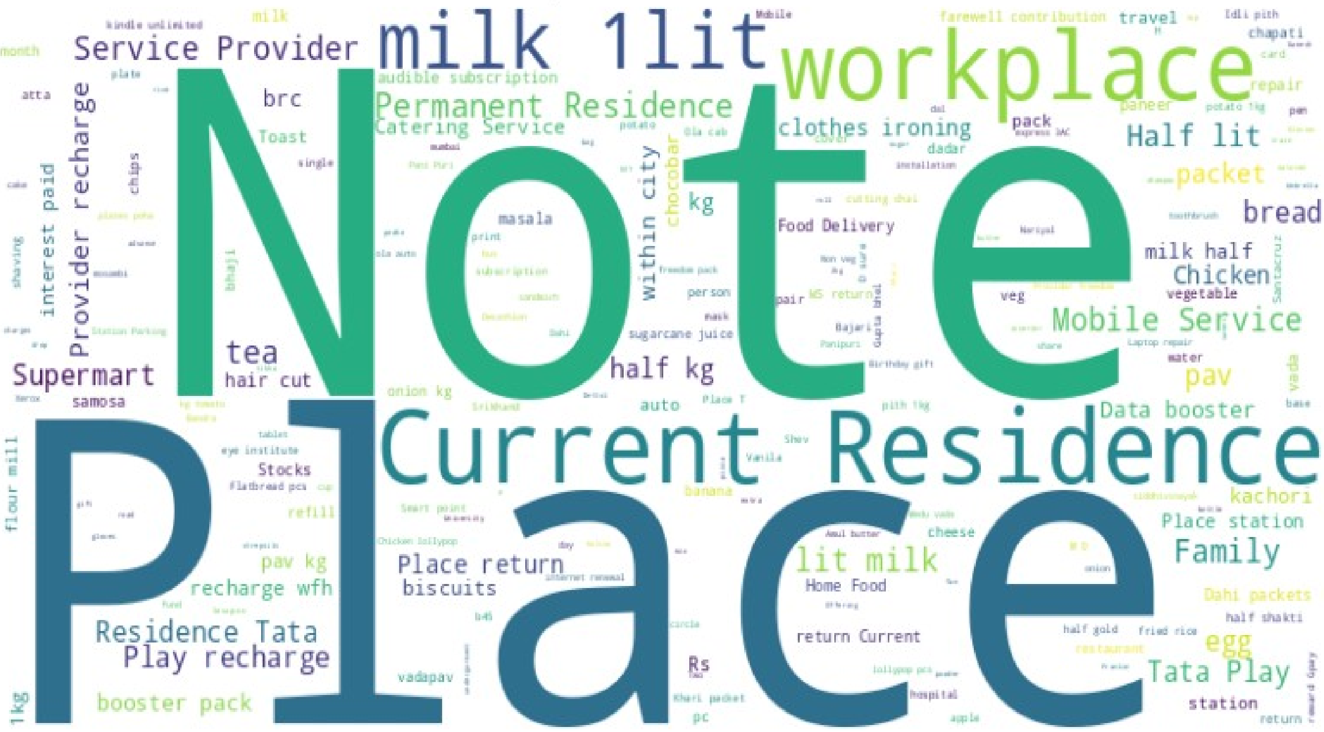




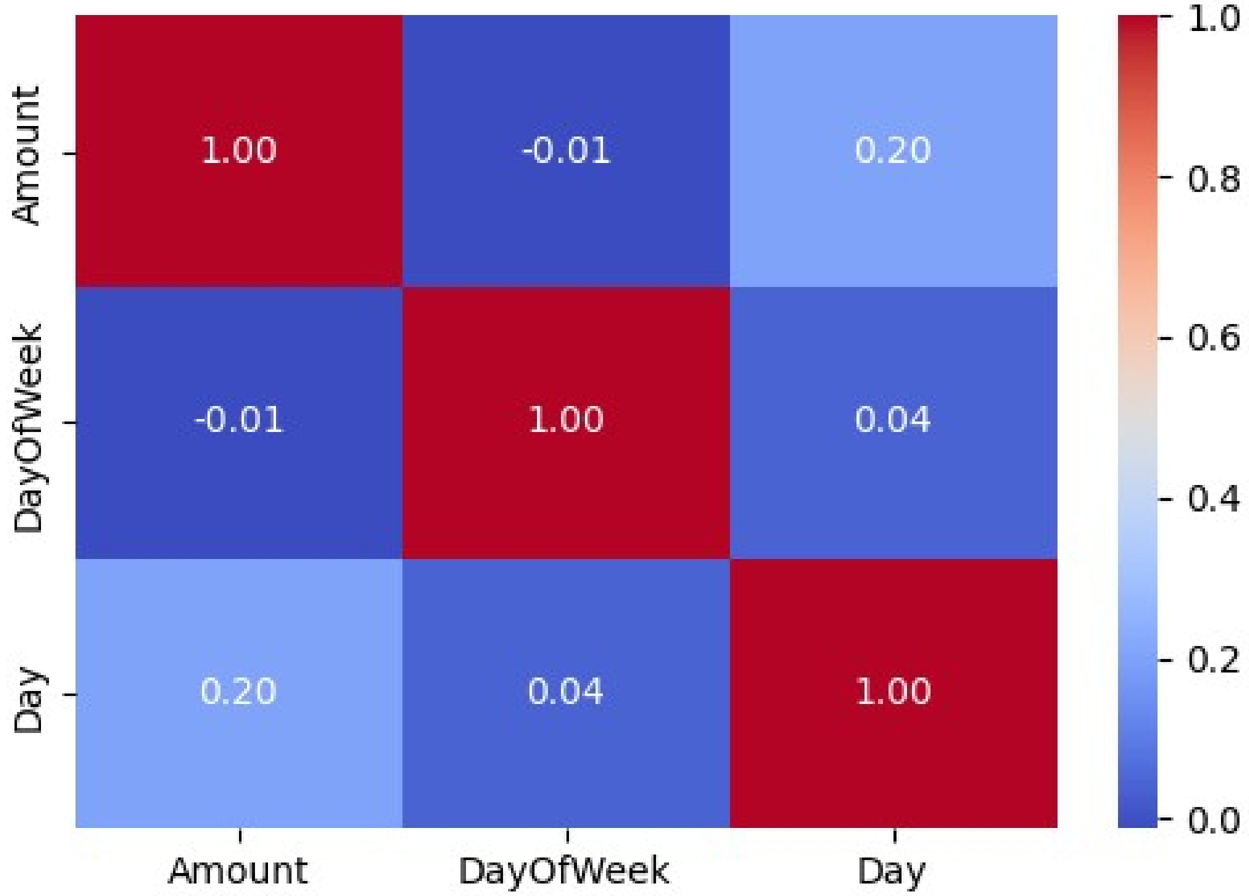

Distribution of Expense Amounts
Amount
50000
40000
30000
20000
10000
0

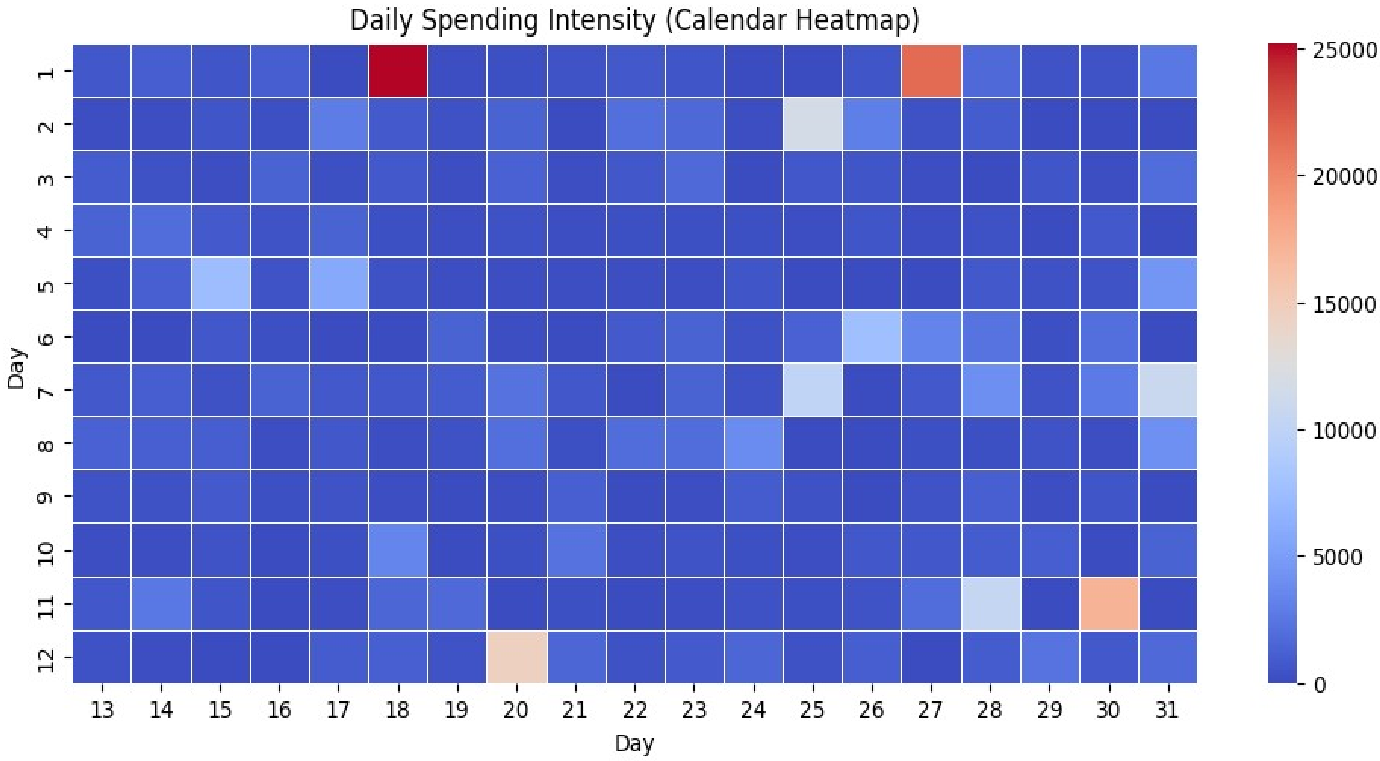

Frequency of Expense Amounts
Count
1000
800
600
400
200
0
0
10000
20000
30000
40000
50000
Amount

Common Keywords in Transaction Notes



Correlation Matrix




Daily Spending Intensity (Calendar Heatmap)
Day
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
Day
25000
20000
15000
10000
5000
0

THANK YOU