

Project Title

- **Personal Budgeting & Expense Forecaster** (Phase 2: Data Categorization and Basic Reports)

Introduction

- This phase takes the raw transaction data entered in Milestone 1 and makes it useful for analysis.
- We need to sort the data so users can understand where their money is going, not just how much they spent.
- Creating clear reports and charts is essential for better financial management and is part of the strategy to avoid cost problems that cause about **11.4% of investment to be wasted**.
- The goal is to move from simple data entry to meaningful financial insights.

Objectives

- **Automate Data Sorting:** Create a system that automatically labels transactions into categories (like 'Groceries' or 'Rent').
- **Generate Core Reports:** Develop functions to summarize spending totals by month and compare income against expenses.
- **Build the First Dashboard:** Create a simple visual screen that displays key financial summaries using charts.

System Overview

- This phase focuses on processing the existing data and showing the first results.
- **Step 1: Get Raw Transactions:** Use the data entered by the user in the previous phase.
- **Step 2: Categorization Engine:** Run the transaction descriptions through the automatic

sorting rules.

- **Step 3: Data Aggregation:** Use the Pandas library to group and calculate totals (sums) for reporting.
- **Step 4: Display Reports and Charts:** Show the calculated summaries and visuals on the user's dashboard.

Methodology / Workflow

- **Categorization Method:** We will use either simple rule-based matching (e.g., checking for keywords like "Starbucks" or "Rent") or a basic text analysis tool like NLTK (Natural Language Toolkit).
- **Data Processing Tool:** The **Pandas** library will be used to handle and manipulate all the data, making it easy to calculate summaries like monthly totals and category spending.
- **Visualization Tools:** To draw the charts, we will use the Matplotlib and Seaborn libraries. Seaborn is chosen because it creates "attractive and informative statistical graphics", which are easy for users to understand.
- **Manual Control:** Even with automation, the user must always be able to manually change any category the system assigns incorrectly.

System Design

- **Categorization Mechanism:** Must include a list of rules that link transaction keywords to standard categories (e.g., 'Groceries', 'Utilities', 'Transport').
- **Reporting Structure:** Must be designed to easily produce three types of summaries:
 - Total spending for each category.
 - A breakdown of spending for each month.
 - A simple total comparison of income versus expense over a defined period.
- **Dashboard Components:** The initial dashboard screen must include:
 - A list showing the most recent transactions.
 - A basic chart (like a pie chart or bar chart) showing what percentage of total money was spent in each category.

Implementation

A. Environment Preparation

- We assume the secure environment (Python, virtual environment) and the basic web interface are already set up from Milestone 1.

B. Development Milestone: (Module 2: Transaction Categorization & Basic Reporting)

- **Automated Categorization:** Implement simple keyword-based rules to assign transactions to categories. Ensure the user can click and change the assigned category.
- **Spending Summary Reports:** Build the code using Pandas to calculate:
 - Category spending totals.
 - Monthly expense summaries.
 - The difference between income and expense.
- **Initial Dashboard View:** Update the user screen (UI) to display:
 - The calculated spending summaries.
 - The initial charts (pie/bar) using Matplotlib/Seaborn to visually show spending distribution.

Testing & Results

- **Categorization Accuracy Testing:** Test the automatic categorization rules on sample data to ensure they correctly assign categories at least 80% of the time. Test the manual override function to confirm users can correct mistakes.
- **Report Accuracy Testing:** Verify that the Pandas calculations correctly add up category totals and monthly expenses. The total income vs. expense calculation must be proven accurate.
- **Visualization Quality Testing:** Check that the charts produced by Seaborn are clear, easy to read, and accurately reflect the calculated data.

Conclusion

- Milestone 2 successfully transforms raw data into actionable financial information.
- By adding categorization and summary reports, the system now provides the user with meaningful insight into their spending habits, preparing the ground for the advanced prediction stage.

Future Scope

- The next critical phase (Milestone 3) is to integrate the Prophet forecasting model.
- This will use the categorized and summarized data to predict future expenses and allow users to set financial goals.