Yuan Tianyi - IAD Data Engineer/Data Analyst, GIC Private Limited

Financial Transactions Dashboard — Project Documentation

1) Overview

An interactive Streamlit app to explore financial-transactions dataset, with insights from data visualisation, interactive components and features, mock-up chatbot and outlier detection.

2) How to run

- Clone / unzip this project from https://github.com/tianyiiiiiii/financial-transactions-dashboard.git
- Install dependencies

  ```bash
  pip install -r requirements.txt
  ```

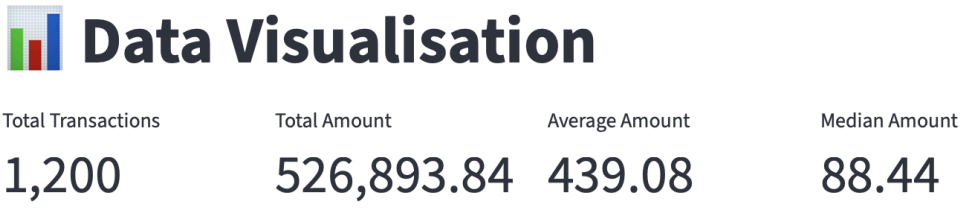
- Run the app

  ```
  streamlit run Home.py
  ```

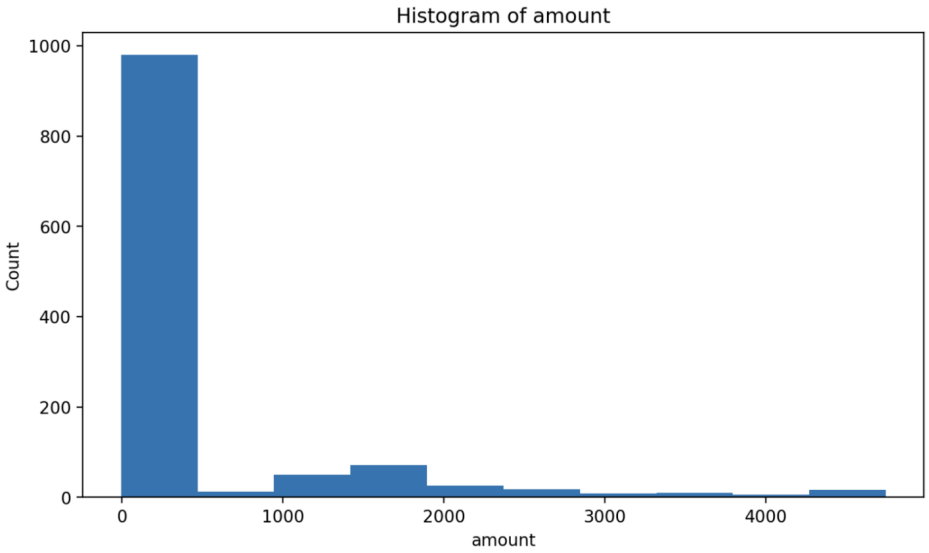
3) Project structure

Home.py
pages/
  1_Dashboard_Analytics.py
  2_Data_Exploration.py
  3_Mockup_Chatbot.py
  4_Outlier_Detection.py
financial_transactions.csv
requirements.txt
README.md

- **Home.py** greets users and page links to navigate to the other pages
- **Dashboard_Analytics.py** with time series, histograms/boxplots, category/merchant bar charts, and heatmaps to visualize trends and patterns.



-



-

- **2_Data_Exploration.py** to preview, filter, and assess the dataset with summary stats, missing values, and duplicate checks.
- **3_Mockup_Chatbot.py** is a chatbot that answers dataset questions on spend, counts, top merchants/categories, payment mix, outliers, and date range
- **4_Outlier_Detection.py** to detect unusual transaction amounts using the robust Interquartile Range (IQR) method
    - Why IQR: IQR is simple, interpretable, and effective for highly skewed financial data, with many small amounts and a few very large ones. IQR is robust against this skew and provides clear thresholds to identify unusually high-value or irregular transactions
- **README.md** provides overview of the project, setup instructions, requirements, and file structure for easy reference

4) Insights

- Transactions fluctuate weekly without a consistent trend, with spending heavily skewed by a few large outliers.
- Shopping, Travel, and Groceries dominate categories, while a handful of merchants (e.g., Mortgage, Netflix) drive a large share of activity.
- Cash is the most common payment method, but cards and mobile payments show strong adoption.
- Heatmaps highlight clear categories–payment method links and seasonal spikes in early 2025.

6) Limitations

- Outlier detection is applied only on the amount column, without multivariate context.
- The mock-up chatbot is rule-based with predefined queries, not a natural language interface.

7) Possible Extensions

- Incorporate multivariate anomaly detection to capture more complex outliers.

- Enhance the chatbot with natural language processing (NLP) or LLM integration for free-text queries.
- Extend with forecasting and predictive analytics (e.g., time-series forecasting of spend trends).

8) Acknowledgements
This project was developed by Yuan Tianyi as part of the IAD Data Engineer/Data Analyst Internship online assessment at GIC Private Limited.
I greatly appreciate GIC the chance to be considered for a potential internship, and I look forward to the possibility of contributing to GIC's data-driven initiatives.