RECEIPT AND INVOICE DIGITIZER

Milestone 4: Polishing & Integration

By: Team D

TEAM DETAILS

Team Leader

- Name: Gajendra Rajpoot
- Role and Responsibilities:
- Improve extraction accuracy with template-based parsing.

Team Member 1

- Name: Shivangi Bhargava
- Role and Responsibilities: Add search/filter in dashboard.
- Enhance data accessibility by implementing search and filter capabilities.

Team Member 2

- Name: Divya Sri Yaganti
- Role and Responsibilities:
- Optimize DB queries and reports.

PROBLEM STATEMENT

- **Inconsistent Naming:** AI parses the same vendor differently across receipts, breaking spending categorization.
- **Missing Data:** Standard OCR often fails to read faint tax lines, leading to mathematical discrepancies.
- **Lack of Context:** Generic models do not know specific vendor tax rates (e.g., 8.25%), preventing automatic correction of incomplete data.

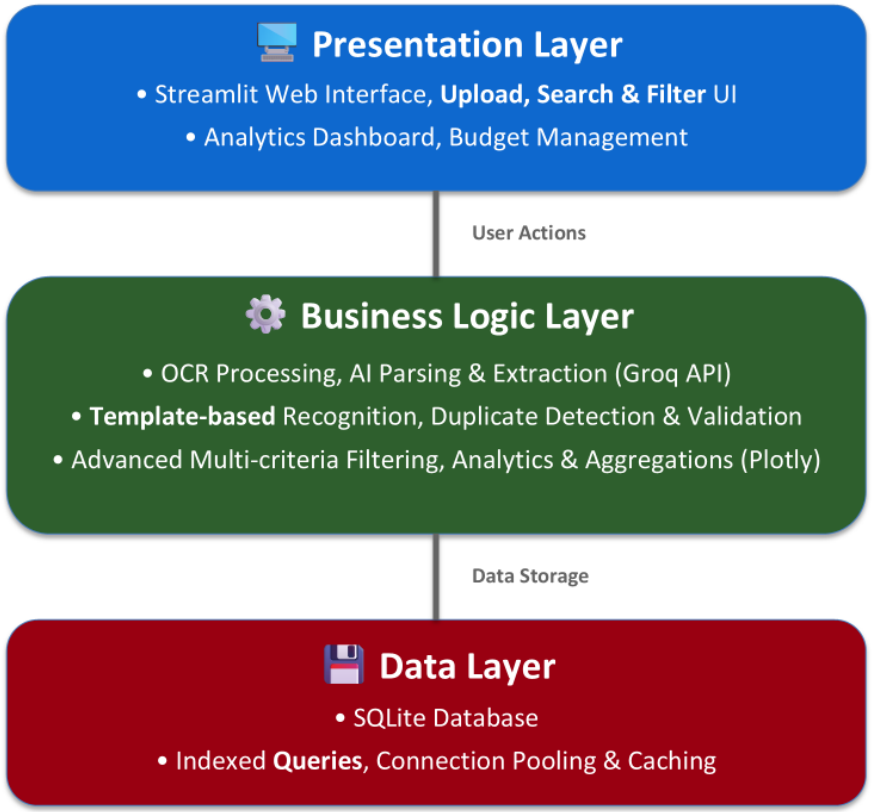
- **Information Overload:** Aggregating all history into one view hides recent spending behaviors.
- **Lack of Actionability:** Users cannot isolate specific cost centers (e.g., "Office Supplies") to verify budget compliance.
- **Static Reporting:** Inability to drill down into custom date ranges makes period-over-period comparison impossible.

- **Inefficient Data Retrieval:** Currently loading 100% of rows just to calculate a single sum (e.g., "Total Spend").
- **Memory Overhead:** Pandas DataFrames consume application RAM proportional to the database size, which is not scalable.
- **Missing Indexes:** Queries for specific dates or vendors are slow because the database must inspect every single record sequentially.

SOLUTION

- Enhanced Extraction (Template-based parsing)
- Smart Search/Filter (Vendor, Date, Amount)
- One-click Export (CSV/Excel)
- Interactive Analytics (Monthly trends)
- Optimized Performance (Indexed queries)

System Architecture Diagram



ADD SEARCH/FILTER IN DASHBOARD

Enhance data accessibility by implementing robust search and filter capabilities within the review dashboard.

Key Features: Enable users to quickly locate specific receipts/invoices using filters like Vendor, Date Range, and Amount.

Technical Implementation: Integrate front-end filter components with backend SQL queries for dynamic and efficient data retrieval.

User Experience: Move from static data viewing to interactive data exploration, allowing users to narrow down results instantly without manual scanning.

Drastically reduce the time spent finding historical records, improving efficiency for audits, reporting, and financial analysis.

IMPROVE EXTRACTION ACCURACY WITH TEMPLATE-BASED PARSING

Objective: Uses vendor-specific extraction rules instead of only generic regex.

Coordinate-Based Extraction: Detects vendor name from OCR text before parsing

Improves date, total, and invoice number accuracy.

Increases overall extraction accuracy from ~85% to 90–95%.

OPTIMIZE DB QUERIES AND REPORTS

- Get invoice data quickly
- Avoid slow loading dashboards
- Make reports smooth even with thousands of receipts

THANKYOU