

Mydigibill: Receipt and Invoice Digitizer



Our Team



Mayank Kumar
Team Lead



**Pranitha
Piduru**
System
Architect



Tisha Gurjar
Frontend Dev.



**Goutham
Mourya**
Backend
Developer



Viabhavi D.
Presentation
Designer



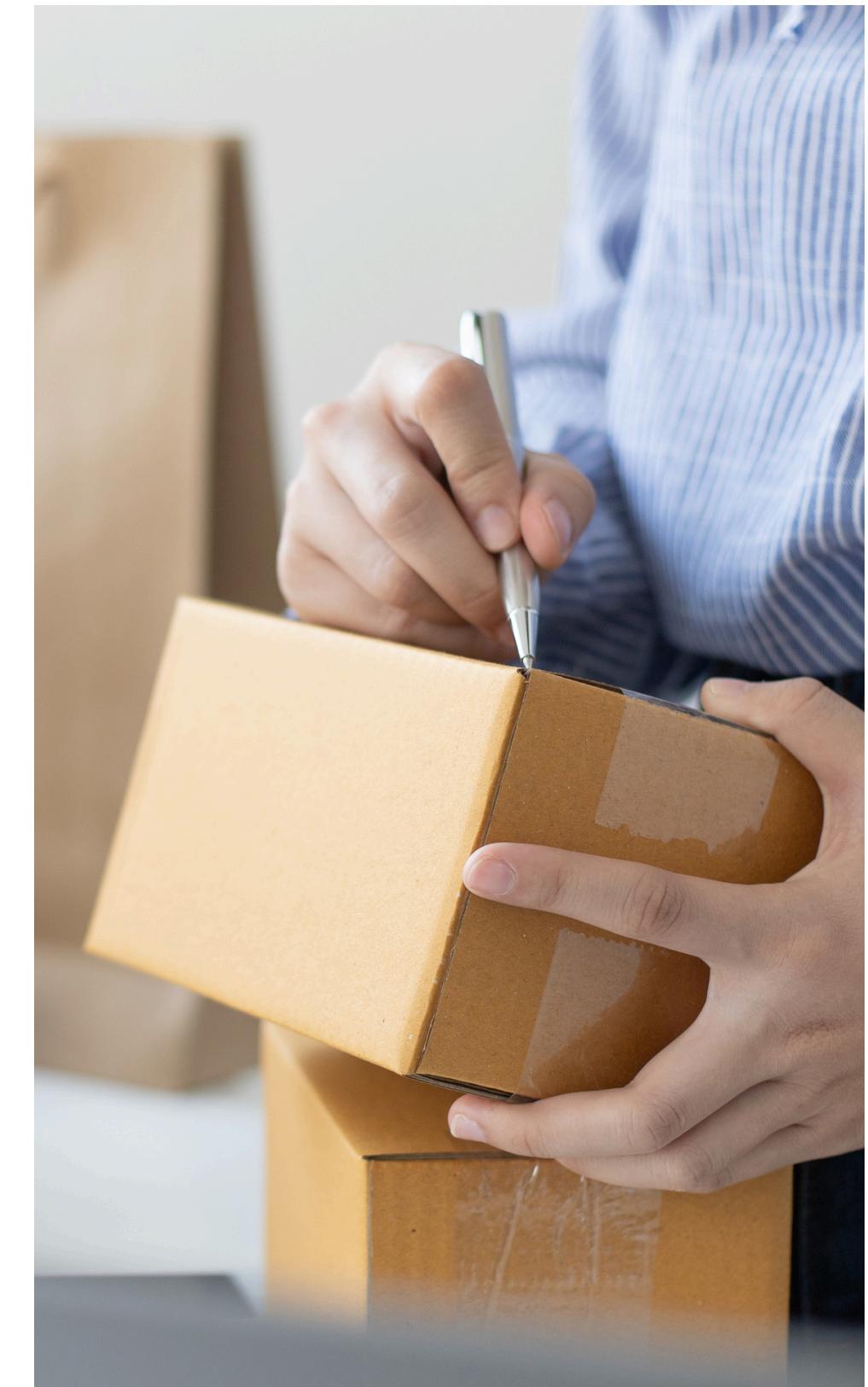
**Niharika
Rangu**
Technical
Writer

Problem Statement

Build a AI driven system that automatically scans, extracts, and digitizes information from receipts and invoices using :

- OCR (Optical Character Recognition)
- NLP-based field extraction.

The digitized data is stored in a structured format, making it easy to search, analyze, and integrate with accounting or ERP systems.



Objectives & Challenges

Objectives	Challenges
1. Prone to Loss, errors, and manual entry delays.	-----
2. OCR Accuracy: ≥85% correct text extraction on test receipts	Noise in Receipts/Invoices. Poor Image Quality, Damaged Text inputs.
3. Field Extraction: ≥90% accuracy on date, vendor, and total fields	Variation of Templates, missing subtotal, Different date formats
4. Validation: Detects duplicates with ≥95% reliability.	-----

Proposed Solution

Mydigibill automates the digitization and processing of physical and digital receipts and invoices using OCR, re and NLP, reducing manual effort and improving accuracy in financial operations.

*Data Ingestion and
Preprocessing*

*Data Extraction and
Validation*

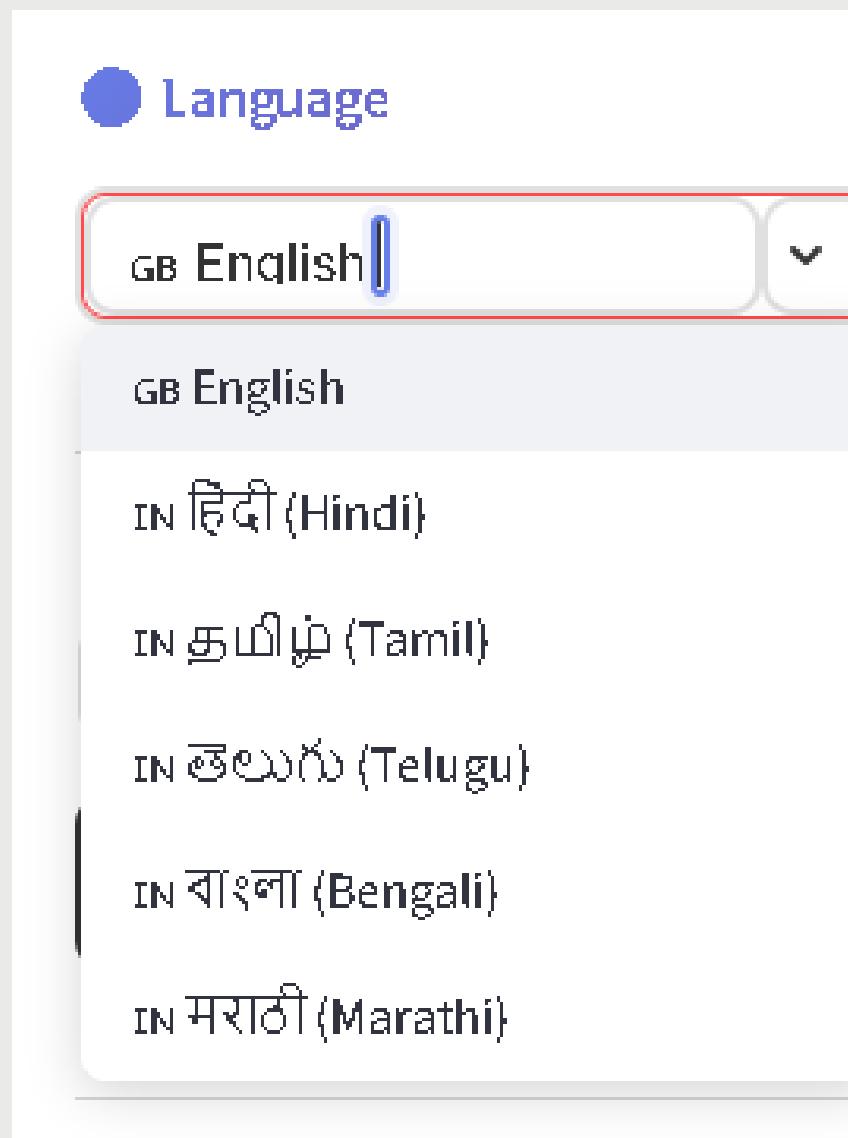
Database Storage

Dashboard & Analytics

Integration Module

Additional Features

- Secure login & authentication system
- Multi-language support for accessibility.
- Chat with the database

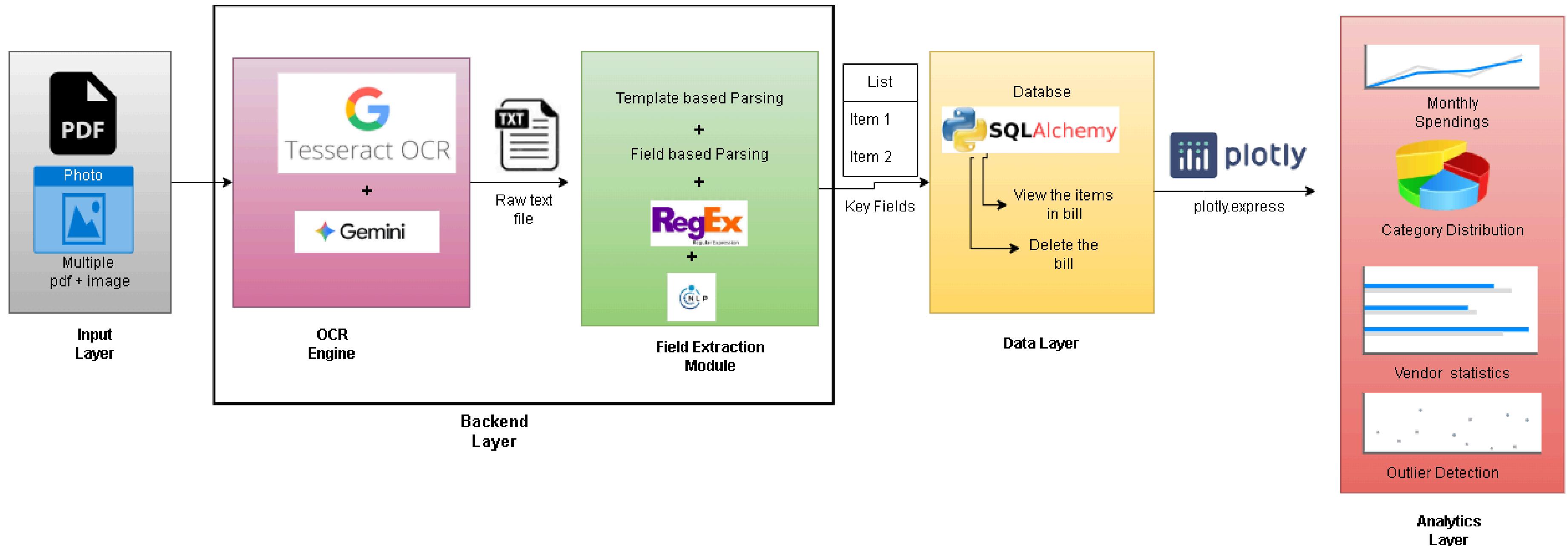


A screenshot of a sign-up form. At the top right is a "Sign in with Google" button with a yellow key icon. Below it is a horizontal line with the text "OR" in the center. The form consists of several input fields:

- Full Name: John Doe
- Email Address: your.email@example.com
- Password: (Redacted)
- Confirm Password: (Redacted)

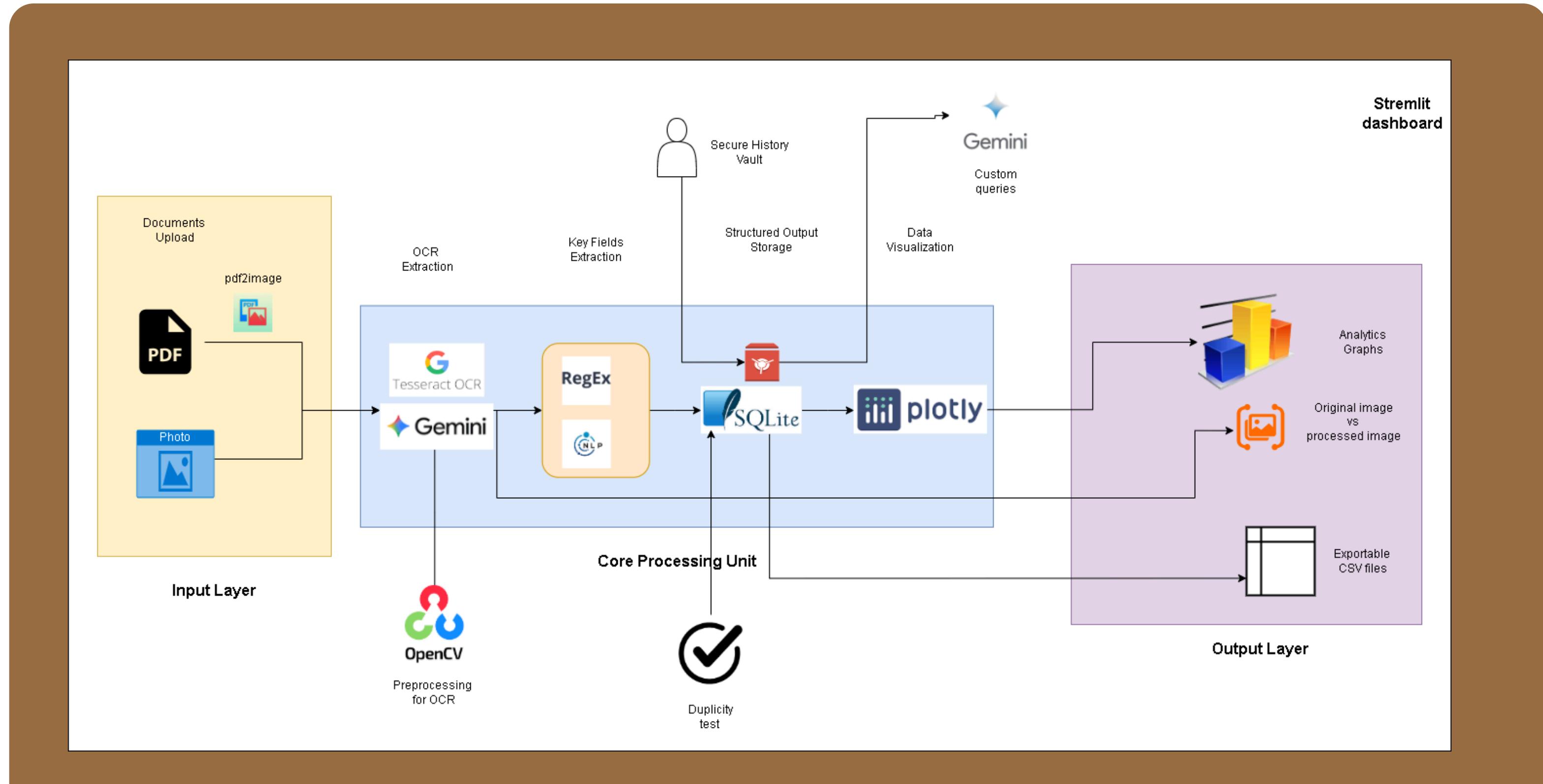
Login" is displayed. At the very bottom is a "Login" button.

System Workflow

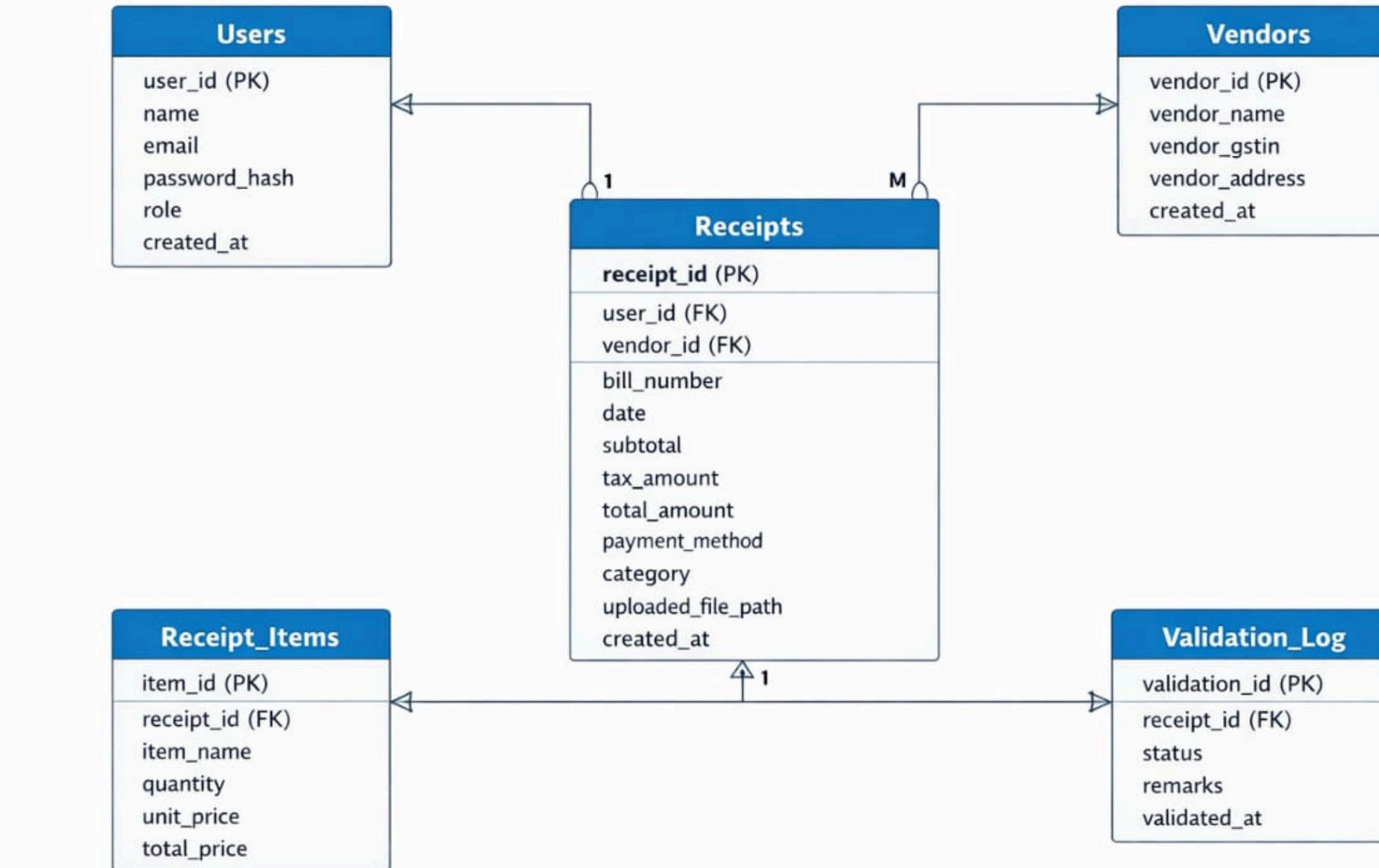


Architecture Diagram

Modular Diagram

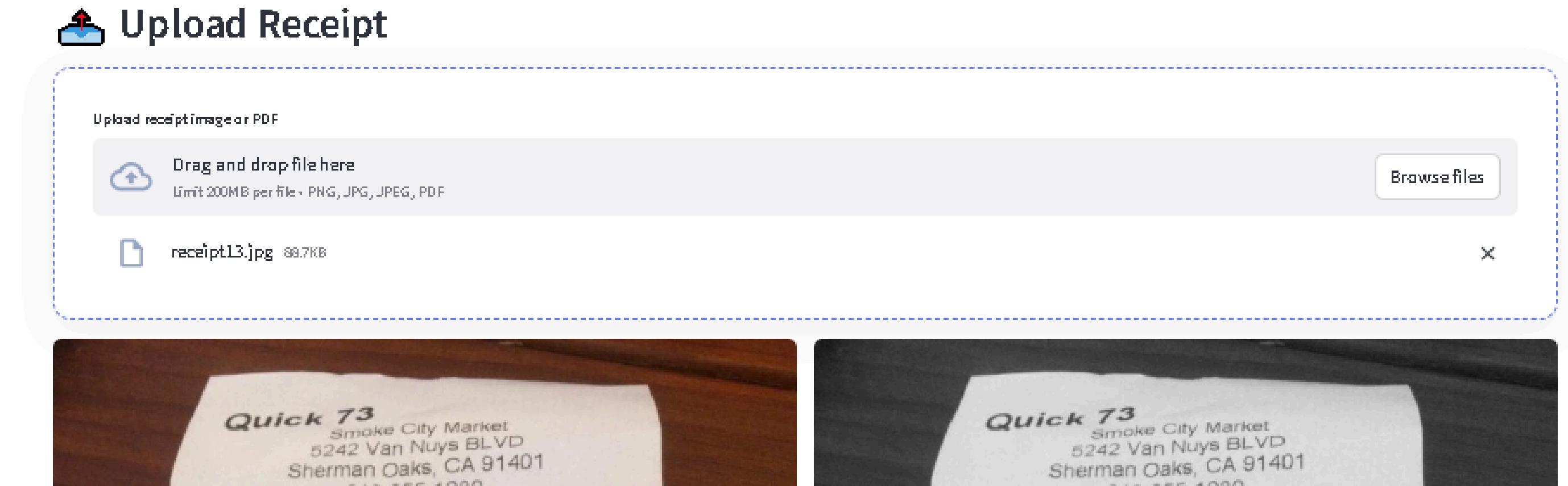


Schema Diagram



Document Ingestion & OCR

- Supports multiple receipts upload (PDF, image)
- Detects duplicates via hash(**hashlib**).
- Automated OCR-based text extraction from scanned receipts and invoices.
- Uses both Pytesseract + Gemini to extract raw text.



Field Extraction & Validation

- NLP/Regex-based extraction for date, vendor, invoice ID, tax, total amount, line items.
- Field-level parsing for date, vendor, amount, tax, and line items.
- Validate extracted totals (subtotal + tax = total).
- Capable of Handling multiple layouts. (template based parsing)

The image shows a screenshot of a receipt processing interface. At the top, there's a section titled "Receipt Summary" with a table showing bill details. Below it is a section titled "Item-wise Details" with a table showing individual item purchases.

Receipt Summary

	Bill ID	Vendor	Category	Date	Subtotal (₹)	Tax (₹)	Amount (₹)
0	234573 #1	Smoke City Market	Food	2015-06-26	59.11	5.32	64.43

Item-wise Details

Item	Price
Beef Ribs	36.86
Pulled Pork Sand	8.5
*Reg	0
Creamed Corn[Small]	3.35
Longhorn Mac & Cheese[sm]	4.2
Chocolate Cake[Slice]	3.95
Fountain Soda	2.25

Database Storage

- Digitized data is stored in a structured format in a SQL database (SQLAlchemy).
- Support search by vendor, date, or amount.
- Export to CSV/Excel.
- Support Items wise Details info
- Allows Entries deletion for better Database Management.

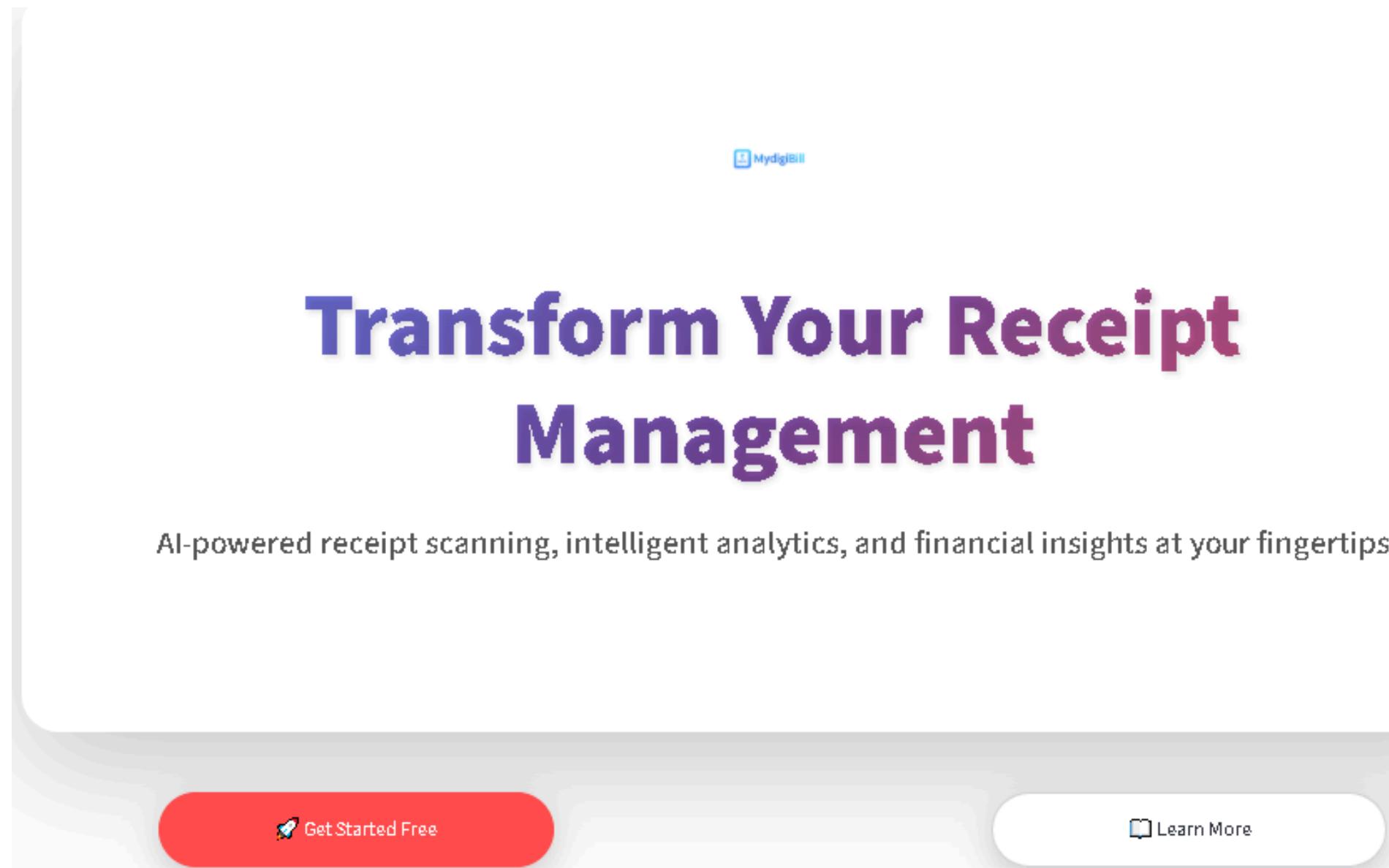
The screenshot shows a user interface titled "Stored Receipts". At the top, there are four filter sections: "Bill ID" (with a "Filter by ID..." button), "Vendor" (with a "Filter by vendor..." button), "Category" (with a dropdown menu set to "All"), "Subtotal (₹)" (with a "Filter by subtotal..." button), "Tax (₹)" (with a "Filter by tax..." button), and "Total (₹)" (with a "Filter by total..." button). Below the filters is a table with the following data:

Delete?	Bill ID	Vendor	Date	Total (₹)	Tax (₹)	Subtotal (₹)	Category
<input type="checkbox"/>	1/A-297424	Subway	2019-05-01	₹13.22	₹1.23	₹11.99	Food
<input type="checkbox"/>	2763	Loaded Cafe	2019-02-23	₹48.53	₹4.21	₹44.32	Food
<input type="checkbox"/>	118888	Taco Maria	2018-01-23	₹45.79	₹3.29	₹42.50	Food
<input type="checkbox"/>	GUEST-CHECK-20170616-	Thai Gusto Restaurant	2017-06-16	₹30.15	₹2.67	₹27.48	Food
<input type="checkbox"/>	128	Grotto Pizzeria & Tavern	2017-05-12	₹45.58	₹2.58	₹43.00	Food
<input type="checkbox"/>	234673 #1	Smoke City Market	2015-06-26	₹64.43	₹5.32	₹59.11	Food
<input type="checkbox"/>	177437	Panda Express	2013-03-16	₹164.89	₹13.89	₹151.00	Food

At the bottom, there are three buttons: "Delete Selected Receipts" (with a trash icon), "Filtered Results" (with a bar chart icon), and "Filtered Total" (with a money bag icon). The total value is displayed as ₹412.59.

Dashboard & Analytics

- Uses KPI cards to showcase the top results.
- Streamlit based Web dashboard.
- Clean, Modern and Responsive Interface
- Uses Multiple Charts to showcase the time dependent and other stats.
- Provides AI insights along corresponding to charts.



The image shows a screenshot of the Mydigibill website. At the top, there's a navigation bar with links like 'Home', 'About', 'Services', 'Contact', and 'Log In'. Below the navigation is a large banner with the text 'Transform Your Receipt Management' in bold purple letters. Underneath the banner, it says 'AI-powered receipt scanning, intelligent analytics, and financial insights at your fingertips'. At the bottom of the page are two buttons: a red one labeled 'Get Started Free' and a white one labeled 'Learn More'.

Mydigibill

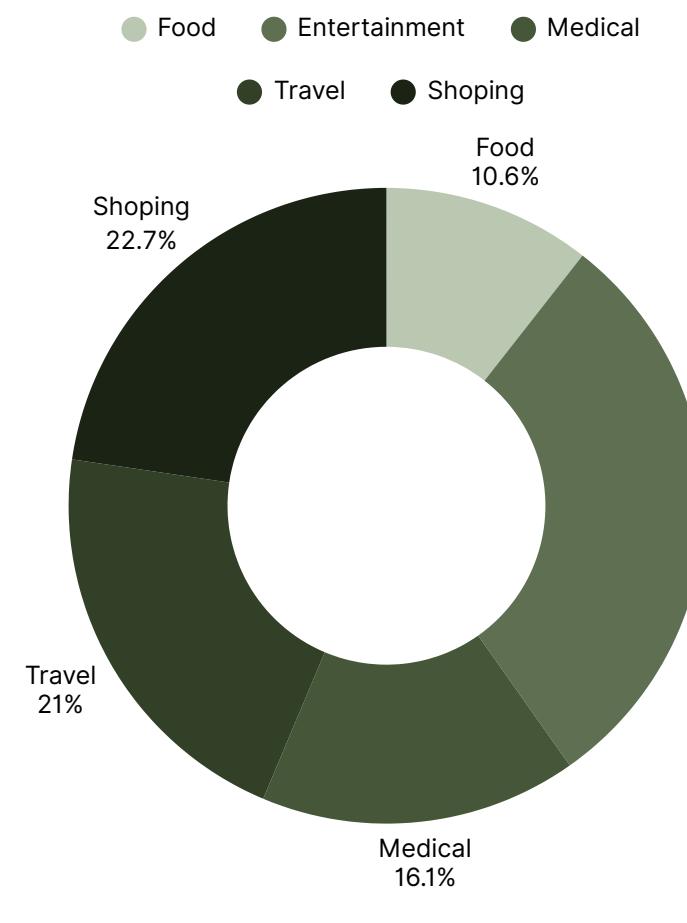
Transform Your Receipt Management

AI-powered receipt scanning, intelligent analytics, and financial insights at your fingertips

Get Started Free

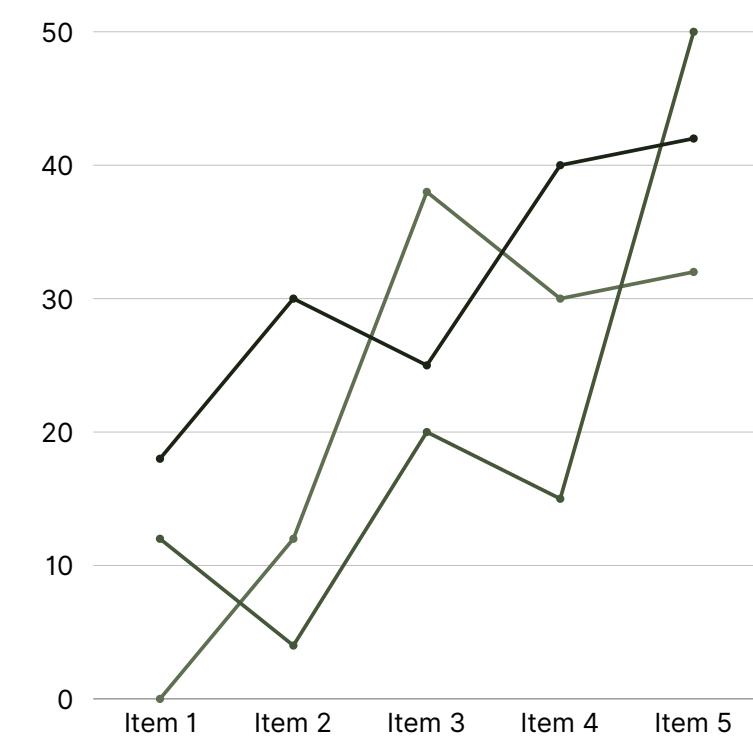
Learn More

Analytics



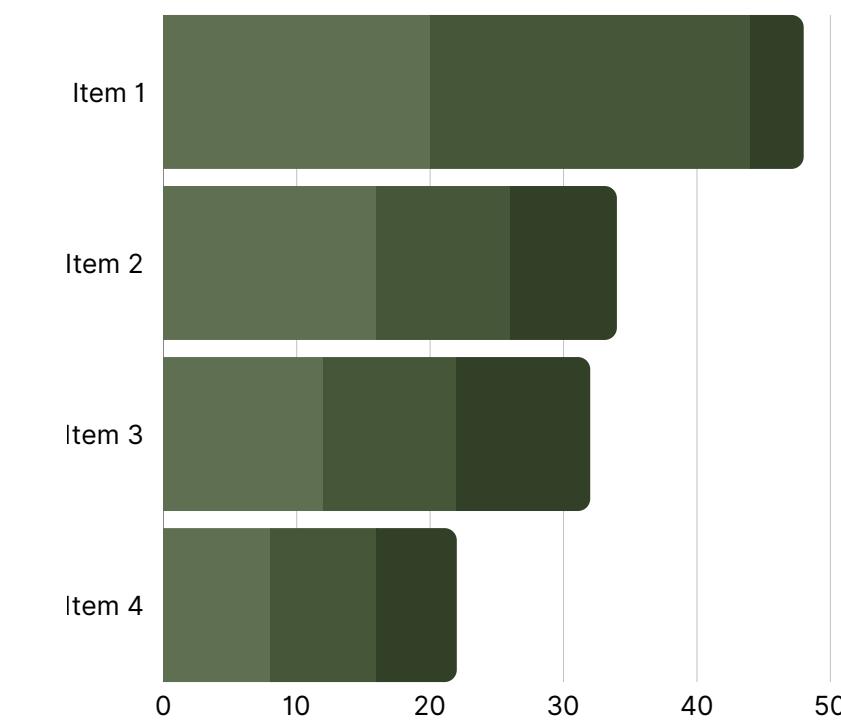
Category Distribution

Help users understand spending distribution



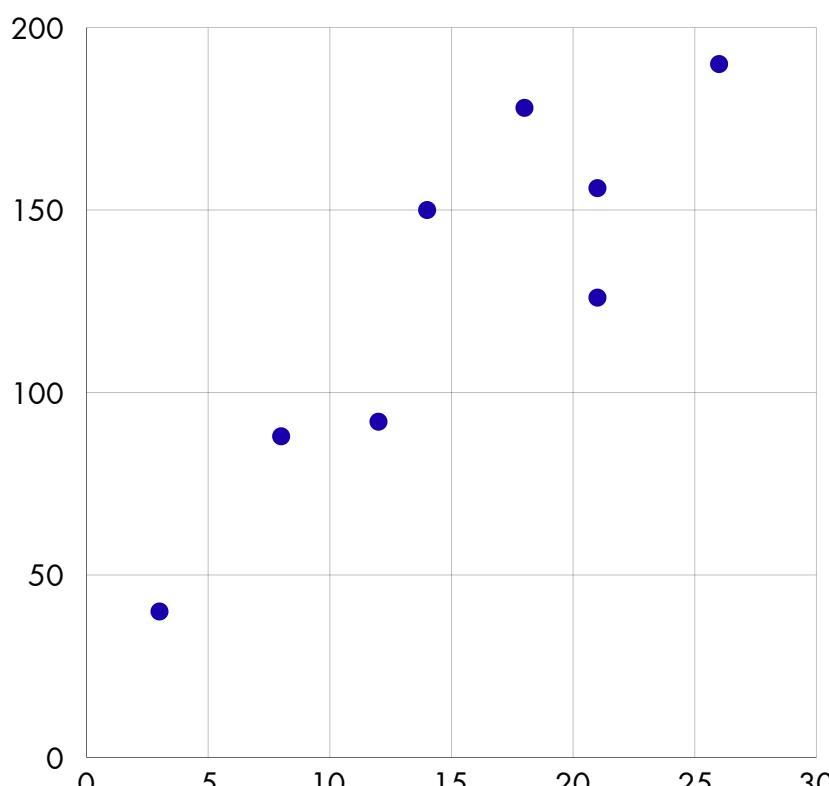
Monthly Spendings

Displays month-wise expense trends.



Vendors Statistics

Compares spending across vendors.



Outlier Detection

Identifies unusual expense values.

Integration with ERP Systems

- Deployment:
<https://mydigibill.streamlit.app>
- Enables communication between the digitizer and external ERP systems.
- Supports **ERPNext**, a open Source ERP system.
- API-based synchronization
- Maintains data Integrity.

Receipt Vault Analyzer API 1.0.0 OAS 3.1

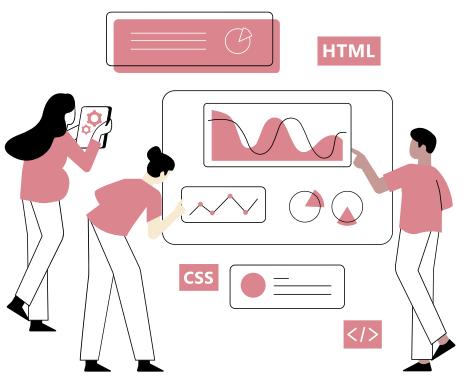
[/openapi.json](#)

REST API for ERP integration and external data access

default

<code>GET</code>	/ Read Root	▼
<code>GET</code>	/api/v1/receipts Get Receipts	▼
<code>GET</code>	/api/v1/receipts/{bill_id} Get Receipt	▼
<code>POST</code>	/api/v1/erp-sync Sync To Erp	▼
<code>POST</code>	/api/v1/ocr/process Process Image	▼

Technology Stack



Frontend
Streamlit

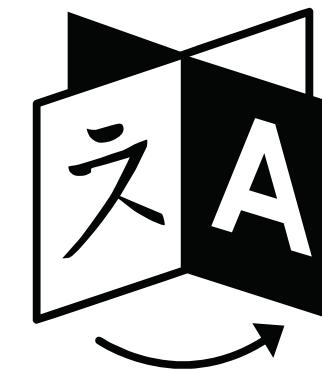


Backend and frameworks

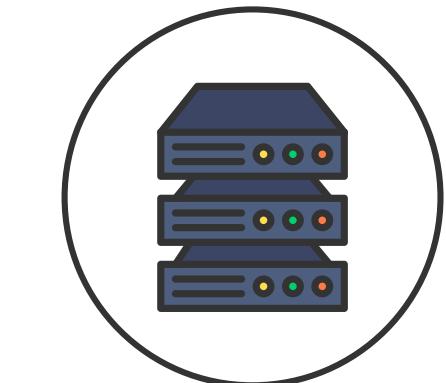
Python,
Numpy, Pandas,
Pytesseract, Re



Data Vizualization
plotly.express



LLM used
Gemini(LLM)



Database:
sqlite3,
SQLAlchemy

References

- 8 hidden costs of outdated accounting systems
- Tesseract OCR Best Practices
- Google's Document AI (Conceptual)
- spaCy: Industrial-Strength NLP
- InvoiceNet (Academic Paper)
- Regex for Financial Data: Reliable patterns for dates, amounts, IDs
- Perceptual Hashing (ImageHash Library) for near-duplicate image detection

**THANK
YOU**

