

Project Report: Balance Sheet Analyzer AI

Objective:

The goal of this project is to simplify the financial analysis of corporate balance sheets by leveraging AI models. The system allows users with different roles—**Analyst**, **CEO**, and **Top Management**—to securely log in, upload annual reports in PDF format, extract structured financial data using AI, visualize financial trends, and interact with an intelligent assistant for actionable insights.

Approach and Methodology

The entire project is built using **Streamlit** for the frontend, **SQLite** for the backend database, and **LLM APIs (Groq)** for AI-powered data extraction and analysis. Here's a step-by-step breakdown of the system's functioning:

1. Authentication and Role-Based Access

- Users must log in to access the application.
- The system checks credentials via the SQLite database (`users` table).
- Based on the user's role, the application grants access to certain features:
 - **Analyst**: Can upload new PDF reports and view all companies' data.
 - **CEO**: Can view data and AI insights for **only their assigned company**.
 - **Top Management**: Can view **all companies** under the group and use AI assistance for deeper analysis.

This is enforced using helper functions like `check_login()` and `check_role_access()` in `auth.py`.

2. PDF Processing (Step-by-Step Extraction)

Uploaded annual reports (PDF files) are processed **page by page** using the `pdfplumber` library:

- **Step 1:** Text is extracted from each page and saved as a list of strings.
- **Step 2:** Each page's text is passed through an AI model for financial data extraction.

3. AI-Based Financial Data Extraction

To extract structured financial data from the unstructured PDF text, the system uses the **Groq API** with **LLaMA 3 models** (specifically `llama3-8b-8192` and `llama3-70b-8192`):

- Each page (and sometimes two consecutive pages) is analyzed.
- The AI is instructed to return a valid JSON object containing only the **required financial metrics** (like Revenue, Net Profit, Total Assets, etc.).
- If a value is missing, the AI sets it as `null`—ensuring transparency and no hallucinations.
- Metrics are then **cleaned, converted into float**, and saved to the SQLite database.

Why Groq?

- Initially, OpenAI's GPT and Gemini 1.5 Flash were tested.
- **OpenAI** had latency issues and a very limited number of free messages.
- **Gemini 1.5 Flash** provided fast responses but also had message limits.
- **Groq + LLaMA 3** was chosen for its **structured response capability**, even though latency is slightly slower.

4. Data Visualization & Insights

Once data is saved, users can interact with visual charts and AI insights:

- **Plotly** is used to create:
 - Line charts (e.g., Net Profit over years)
 - Bar charts (e.g., Revenue comparison)
 - Assets vs. Liabilities comparisons

- YoY Growth Charts
- Users can ask the AI assistant questions like:
 - *"How is the financial health of Reliance Jio?"*
 - *"Show me the trend of Net Profit."*

The assistant (powered by Groq) understands context, interprets trends, and gives **concise and actionable summaries**, especially tailored for busy decision-makers like CEOs and executives.

Features Summary

Feature	Description
Role-Based Access Control	Analyst, CEO, Top Management—each with different access permissions
Secure Login	Password authentication via SQLite
PDF Upload (Analyst only)	Only Analysts can upload and extract data from new annual reports
AI-Powered Data Extraction	LLMs extract structured JSON data from financial PDFs
Visual Insights	Charts for trends, growth, comparisons
Chat-Based AI Analyst	Users can ask financial questions and receive summarized insights

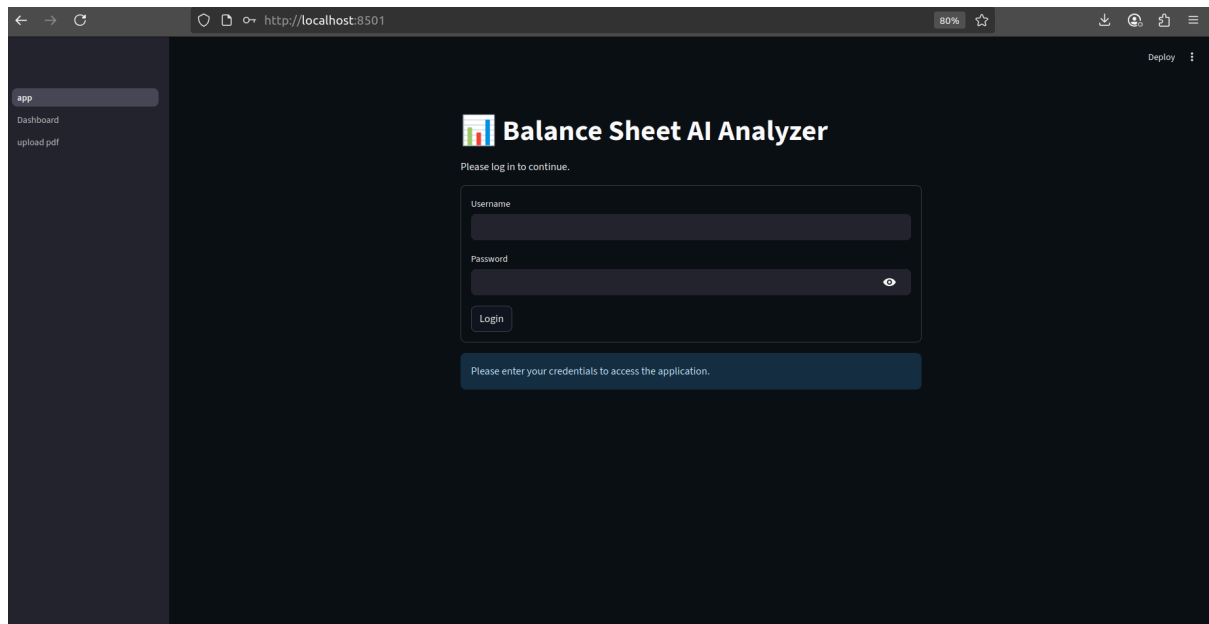
Findings & Key Learnings

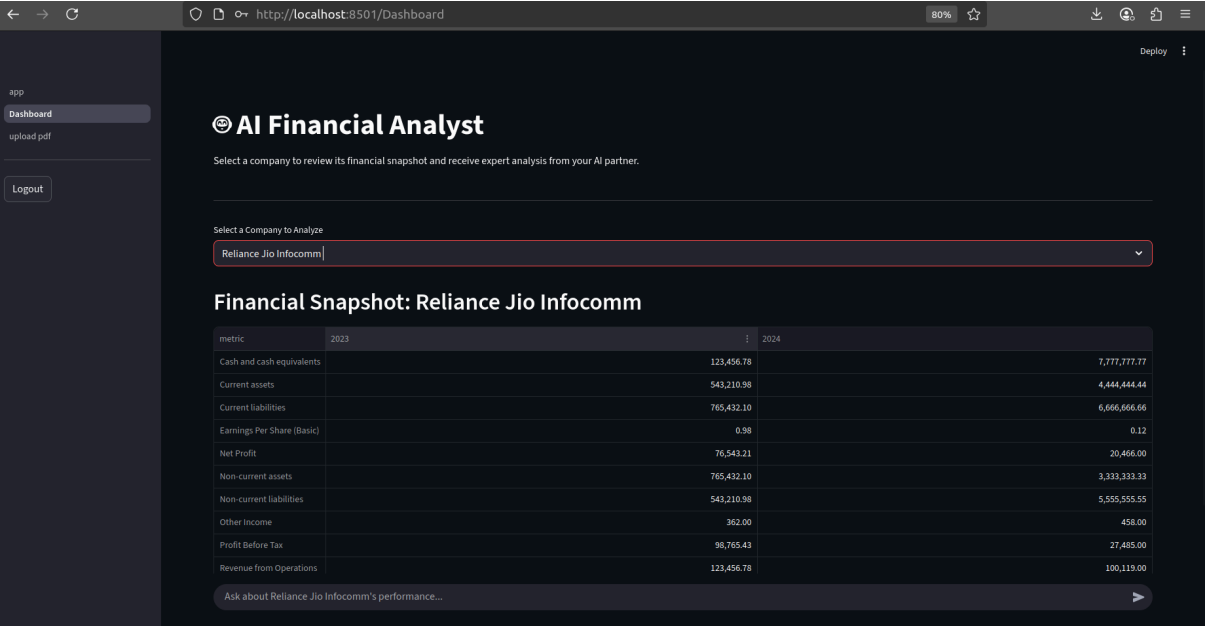
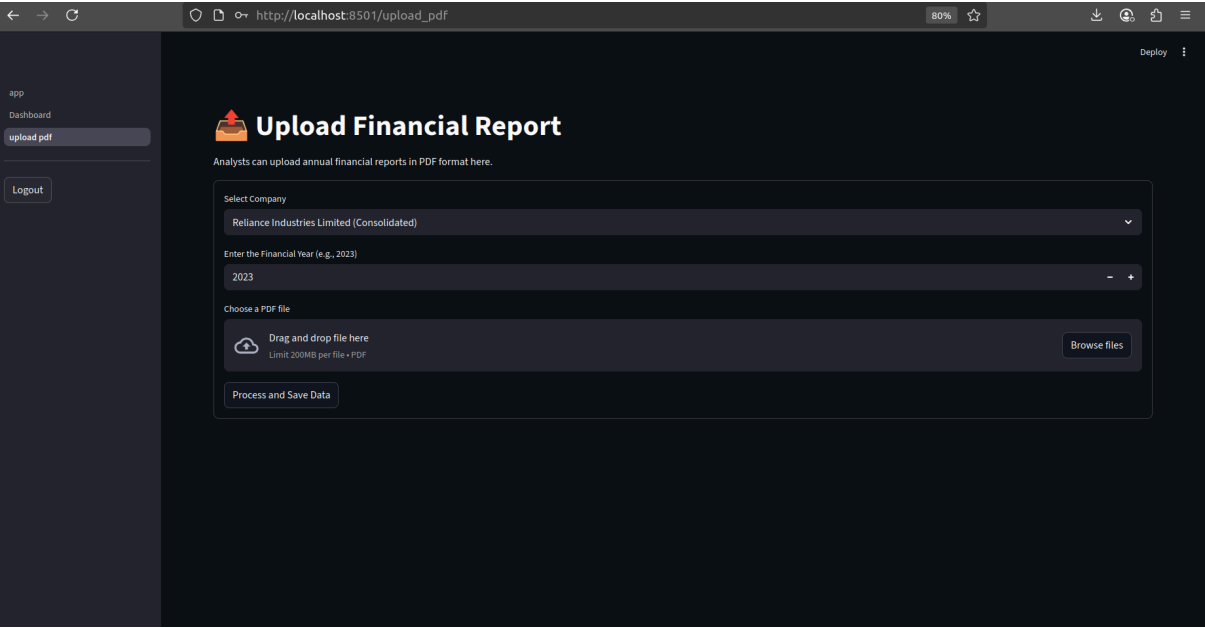
1. **AI + PDF Integration is powerful:** AI can extract meaningful data even from complex financial tables when prompted correctly.
2. **Groq API is ideal for structured JSON tasks:** Despite slightly slower responses than Gemini, it returned more consistent structured outputs.
3. **Role-Based Security is essential:** Separating access ensures confidentiality and relevance—CEOs only see their data, while analysts can operate at a wider scale.
4. **Streamlit enables rapid development:** It offers a fast, interactive UI with minimal setup.

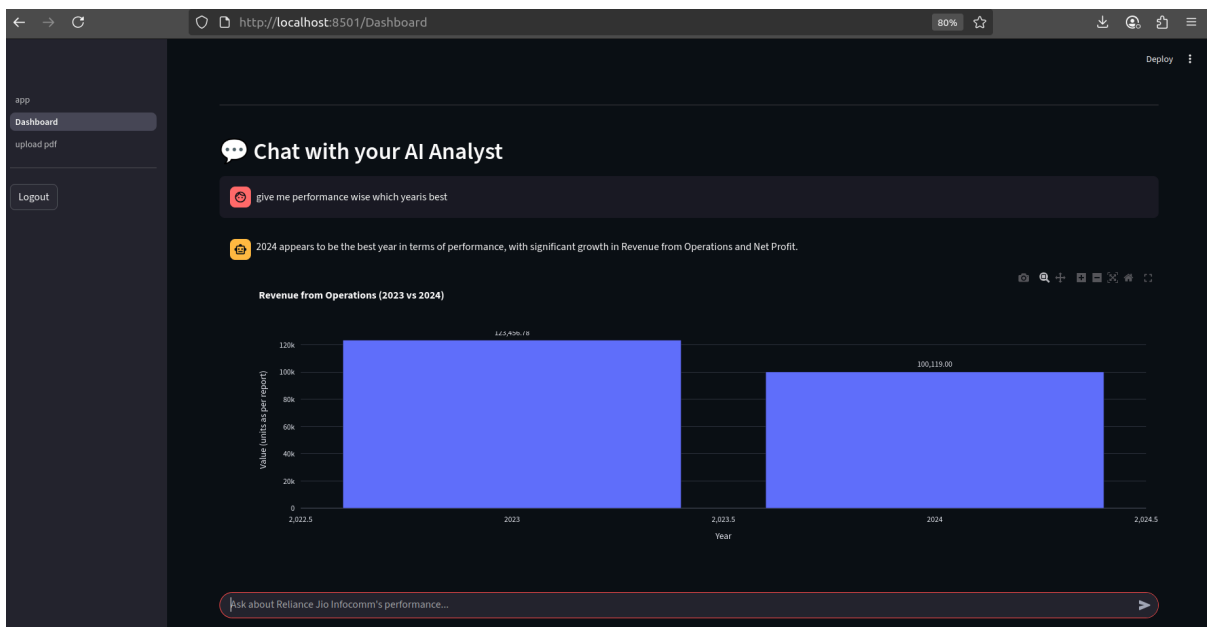
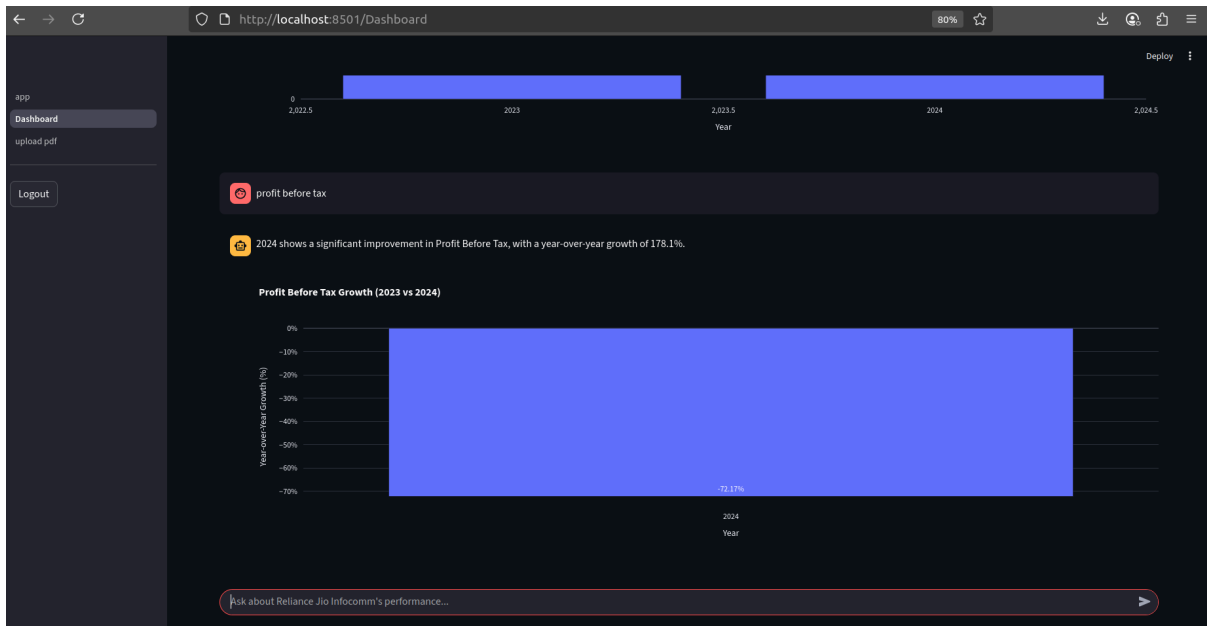
Limitations & Next Steps

- **LLM Cost & Message Limits:** Groq is more reliable for free-tier usage, but scaling to multiple users would require API quota management.
 - **OCR Support (Future):** Currently, only digital PDFs are supported. Adding OCR support for scanned PDFs (e.g., with Tesseract) is a potential improvement.
 - **Error Handling Enhancements:** Better user feedback during failed parsing or AI timeout can improve UX.
-

Frontend :







Network URL: <http://192.168.31.67:8501>

Conclusion

This project demonstrates the effective use of AI to simplify financial analysis for diverse stakeholders. From data extraction to dynamic insights, the system transforms complex financial reports into interactive, insightful dashboards. With further refinement, this could serve as a robust tool for financial analysts and C-level executives.

