Enabling AI-Powered Business Intelligence for Organizations

Course-end Project 1 — InsightForge (LangChain + RAG + LLM)

1. Introduction

Organizations generate large volumes of business data, but many small and mid-sized teams struggle to convert it into actionable insights. InsightForge is an AI-powered Business Intelligence assistant that uses Large Language Models (LLMs) together with Retrieval-Augmented Generation (RAG) to analyze business data, answer natural-language questions grounded in computed KPIs, and visualize trends.

2. Problem Scenario

Many BI solutions are expensive, complex, and require specialist skills. The objective is to deliver an automated BI assistant that can: (a) analyze business datasets for trends and patterns, (b) generate insights and recommendations in natural language, and (c) present visual dashboards.

3. Objectives

- Analyze business data: compute KPIs, trends, and breakdowns.
- Generate insights: answer questions with numeric evidence.
- Visualize insights: interactive charts and summary tables.
- Maintain context: persistent conversation memory.
- Evaluate quality: automated evaluation with QAEvalChain.

4. Data Sources

Structured datasets are ingested from CSV and Excel files:

- sales_data.csv — transactional sales records
- records.xlsx — supporting business records (optional)

The system performs best-effort column detection for dates, sales/revenue, products, regions, and customer demographics.

5. System Architecture (High Level)

Pipeline:

1) Data ingestion (CSV/XLSX)

2) KPI computation (monthly/quarterly sales, top products/regions, segmentation)

3) KPI tables converted into knowledge-base documents

4) Embeddings + FAISS VectorStore for similarity retrieval

5) RAG chain (Retriever → Prompt → LLM)

6) Streamlit UI for chat + dashboards

7) Persistent memory to track user interactions over time

6. Implementation Details

- LangChain orchestration: prompt templates + LLM abstraction (Groq/Ollama)
- RAG: similarity search over KPI ‘documents’ stored in FAISS
- Memory: simple JSON chat memory persisted to storage/chat_memory.json
- Visualization: Matplotlib charts for trends, top products, and top regions
- Evaluation: QAEvalChain compares model outputs to reference answers for a small eval set

7. Key Features

- Natural-language BI assistant grounded in computed KPIs
- Actionable recommendations with supporting numbers
- Modular pipeline for data-heavy workflows
- Works with cloud LLM (Groq) or local LLM (Ollama)
- Streamlit dashboards + evaluation tab

8. Results (What to Demonstrate in Screenshots)

- Chat: ask questions like ‘Which region has highest sales?’ and show numeric evidence.
- Dashboard: monthly sales trend + top products/regions charts.
- Evaluation: run QAEvalChain on 2–3 reference questions and show graded results.

9. Conclusion

InsightForge demonstrates how LLMs combined with RAG can deliver explainable business insights from organizational data without relying on heavy BI tooling. The design is modular and extensible to additional datasets, metrics, and forecasting.

10. Future Enhancements

- Real-time ingestion (APIs/streams)
- Forecasting (Prophet/ARIMA/ML)
- Advanced segmentation (RFM/cohort analysis)
- Role-based dashboards and access control
- Deployment on cloud with scheduled index refresh