Project Progress Report

Project Title:

InSighto – Privacy-First Agentic Data Assistant

Date: 17th December 2025

Prepared By:

Anuska Ghosh - https://github.com/anuskaGHS/GenAI_DataAnalysis_Assistant

Likitha S - https://github.com/LikithaSrinivas100/GenAI_DataAnalysis_Assistant

BCA (AI & ML) – 3rd Semester
Alliance University

1. Introduction

This report presents the progress made on the project **InSighto**, a privacy-first, AI-assisted data analysis application. The objective of the project is to build an intelligent yet simple system that helps users understand datasets through guided analysis, visible AI reasoning, and meaningful insights, without relying on heavy dashboards or persistent data storage.

The focus during this phase was on building a strong technical foundation, ensuring privacy-first handling of data, integrating AI responsibly, and resolving practical implementation challenges.

2. Application Architecture and Design

A clear and modular architecture was finalized to ensure scalability, clarity, and industry relevance.

2.1 Technology Stack

- **Frontend & Application Framework:** Streamlit
- **Programming Language:** Python 3
- **AI Integration:** Google Gemini / Alternative LLMs (evaluated)
- **Data Handling:** Pandas, NumPy
- **Visualization:** Plotly
- **Storage:** Temporary in-memory storage (no persistence)

This stack was chosen to balance ease of development with real-world applicability.

2.2 Application Flow

The application follows a structured flow:

1. Upload Dataset
2. Overview of Dataset
3. AI-guided Analysis
4. Final Report Generation

This guided flow ensures that users do not skip important steps and receive meaningful explanations at every stage.

3. User Interface Development

3.1 Core Pages Implemented

The following pages were successfully implemented:

- **Upload Page:**
  Allows users to upload CSV or Excel datasets securely.
- **Overview Page:**
  Displays dataset size, column information, data quality indicators, and AI planning.
- **Analysis Page:**
  Shows selected charts and AI-generated reasoning behind the analysis.
- **Report Page:**
  Generates a summarized report with insights and reflections.

Each page maintains a consistent layout and navigation experience.

3.2 Navigation and Reset Behavior

- Implemented a navigation bar with clearly defined steps.
- Clicking on the application title resets the application state safely.
- Session state management ensures clean transitions between pages.

4. Privacy-First Data Handling

A major emphasis was placed on privacy and security.

- Datasets are processed using **temporary files only**.
- No user data is stored permanently on disk or database.
- Each new upload resets previous data and insights.
- The system ensures isolation between different user sessions.

This approach aligns with modern privacy-aware data handling practices.

5. AI Integration and Agent Design

5.1 Agentic AI Structure

The AI component was designed using an agent-based approach with three key stages:

- **Planning:**
  The AI explains what analysis it intends to perform and why.

- **Reasoning:**
  The AI justifies the choice of charts and analysis methods.
- **Reflection:**
  The AI updates its explanations based on user feedback.

This ensures transparency and avoids black-box behavior.

5.2 Model Access and Compatibility Handling

During integration, several practical challenges were identified:

- API model access is account-specific and rate-limited.
- Not all advertised models are available for every API key.
- Free-tier limits were quickly exhausted during testing.

To address this:

- Model access was verified using Google AI Studio.
- The application was aligned to models actually available to the project.
- Alternative AI deployment strategies were evaluated.

6. Error Handling and Debugging

Significant effort was spent resolving real-world development issues, including:

- Python environment and dependency conflicts.
- Streamlit rerun and session state issues.
- Temporary file permission errors on Windows.
- AI SDK deprecations and version mismatches.
- API quota exhaustion and model availability errors.

Each issue was resolved using structured debugging, ensuring stability and correctness.

7. Dark and Light Mode Enhancement

A Dark/Light mode toggle was added to improve user experience.

- Identified limitations of Streamlit’s default theming.
- Implemented manual theme switching using CSS injection.
- Ensured theme state persists across navigation.
- Improved accessibility and visual comfort.

This enhancement contributes to a more polished and user-friendly interface.

8. Evaluation of Deployment Strategies

Multiple deployment approaches were studied:

- **Cloud-only deployment using Streamlit Cloud**
- **Local deployment using open-source LLMs**
- **Hybrid approach combining local and cloud AI**

It was observed that:

- Cloud LLMs have strict rate limits.
- Local LLMs provide unlimited usage and better privacy.
- A hybrid design offers the best flexibility for future expansion.

These insights will guide future development phases.

9. Current Project Status

At the end of this phase:

- Core application flow is complete.
- AI reasoning and explanation features are functional.
- Privacy-first design goals are achieved.
- UI is stable and user-friendly.
- Known limitations are identified and documented.

The project is now at a **working prototype stage** with strong industry relevance.

10. Next Planned Steps

The upcoming work will focus on:

- Improving AI reliability with fallback mechanisms.
- Refining user experience and explanations.
- Preparing deployment-ready configurations.
- Enhancing documentation and reports.
- Final testing with multiple datasets.

11. Conclusion

Today’s progress focused on strengthening the foundation of the InSighto application by combining responsible AI usage, privacy-first data handling, and real-world problem solving. The challenges encountered and resolved during this phase closely reflect industry development scenarios, making the project both academically valuable and professionally relevant.