ALLIANCE UNIVERSITY

Private University established in Karnataka State by Act No.34 of year 2010
Recognized by the University Grants Commission (UGC), New Delhi

Final Report on

InSighto - An AI-Powered Data Analysis and Reporting System

Bachelor of Computer Applications

SEMESTER – IV

Generative AI & Agentic AI Winter Program

Submitted By:

Anuska Ghosh – 2411021240012

Likitha S - 2411021240067

Faculty Signature

HOD Signature

Department of Computer Applications

Alliance University

Chandapura – Anekal Main Road, Anekal

Bengaluru – 562 106

December 2025

Final Report on

InSighto – An AI-Powered Data Analysis and Reporting System

GitHub Repository Link:

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

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

1. Abstract

InSighto is a web-based, AI-assisted data analysis and reporting system designed to simplify exploratory data analysis for users without advanced technical expertise. The system enables users to upload structured datasets (CSV or Excel), preview dataset structure, perform automated profiling, and generate AI-assisted analytical insights using a locally hosted large language model. InSighto also supports exporting analysis reports as PDF, making it suitable for documentation and academic or professional use. The project emphasizes privacy, cost-effectiveness, and explainable AI-driven insights.

2. Introduction

In today's data-driven world, organizations and individuals increasingly rely on data to make informed decisions. However, analyzing raw datasets often requires proficiency in programming, statistics, and specialized tools. Many existing platforms are either complex, expensive, or require cloud-based processing that raises data privacy concerns.

InSighto addresses these challenges by providing a simple web-based interface that automates essential data analysis steps and augments them with AI-generated explanations. The system is designed to be accessible, privacy-conscious, and suitable for both academic and practical use cases.

3. Problem Statement

- Manual data analysis is time-consuming and technically demanding
- Many data analysis tools are costly or difficult for beginners
- Cloud-based AI solutions pose data privacy risks
- Non-technical users struggle to interpret statistical outputs

4. Objectives

- To design a user-friendly data analysis platform
- To automate dataset profiling and basic exploratory analysis
- To integrate Generative AI for insight and report generation
- To ensure data privacy through local processing
- To enable report export in a commonly used format (PDF)

- To build a cost-effective and extensible system

5. System Architecture

5.1 High-Level Architecture

InSighto follows a modular client-server architecture:

- **Frontend:** HTML, CSS, JavaScript (server-side rendered using Flask templates)
- **Backend:** Flask (Python)
- **Database:** SQLite for temporary, session-based storage
- **AI Layer:** Ollama with llama3:8b (local LLM inference)

5.2 Workflow

1. User uploads a dataset
2. Backend stores the file with a unique session identifier
3. Dataset preview and metadata are displayed
4. Analysis agent performs dataset profiling
5. AI model generates insights and summaries
6. A structured report is generated
7. Report is displayed and can be exported as PDF

6. Frontend Implementation

The frontend guides users through a structured workflow:

- **Landing Page:** Introduces the system and its capabilities
- **Upload Page:** Allows CSV/Excel file upload with validation
- **Overview Page:** Displays dataset metadata and preview rows
- **Analysis Page:** Shows profiling results and AI-generated insights
- **Report Page:** Presents a complete, structured analysis report

The report page supports printing and PDF export, allowing users to save reports for documentation, sharing, or submission purposes.

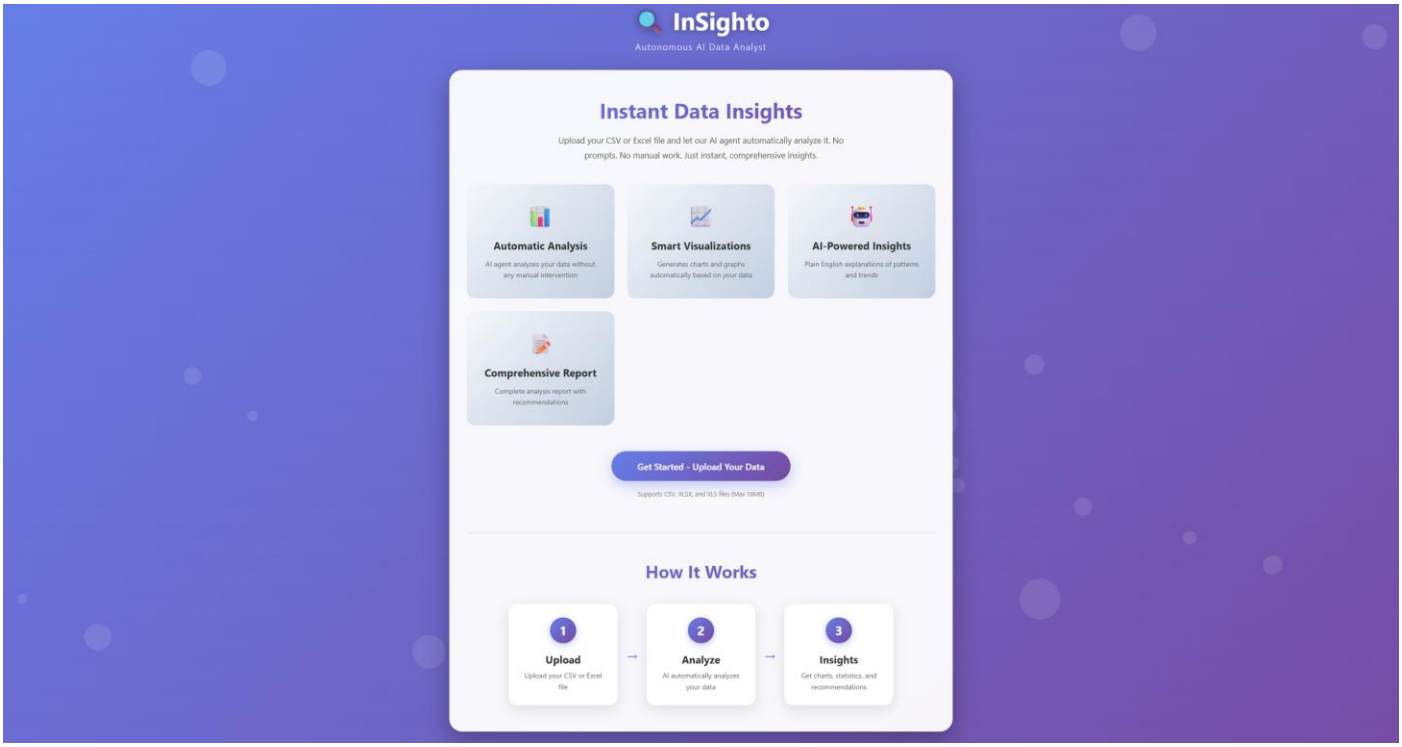


Figure 1: Landing page of InSighto introducing the AI-powered data analysis system

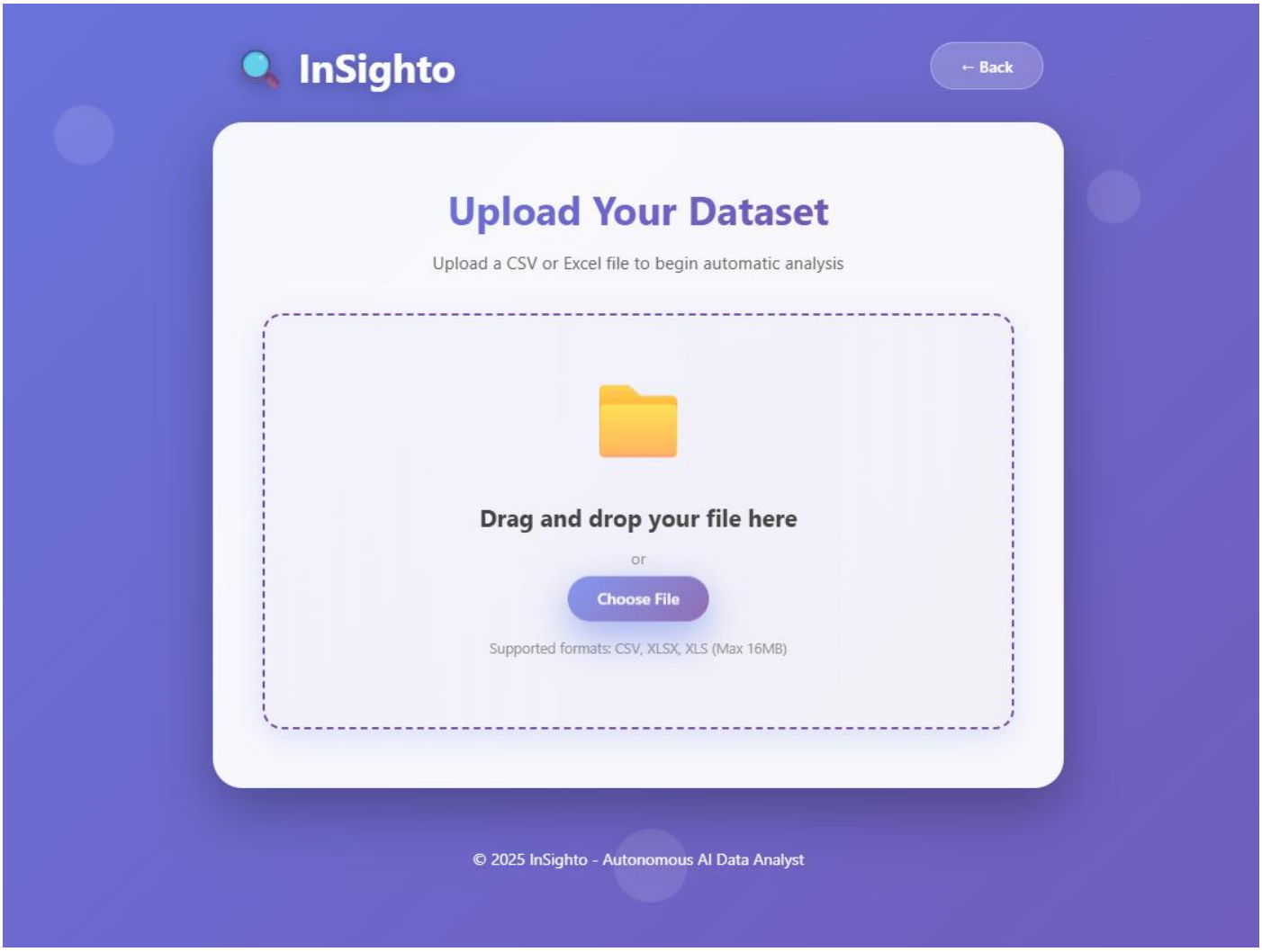


Figure 2: Dataset upload interface supporting CSV and Excel files

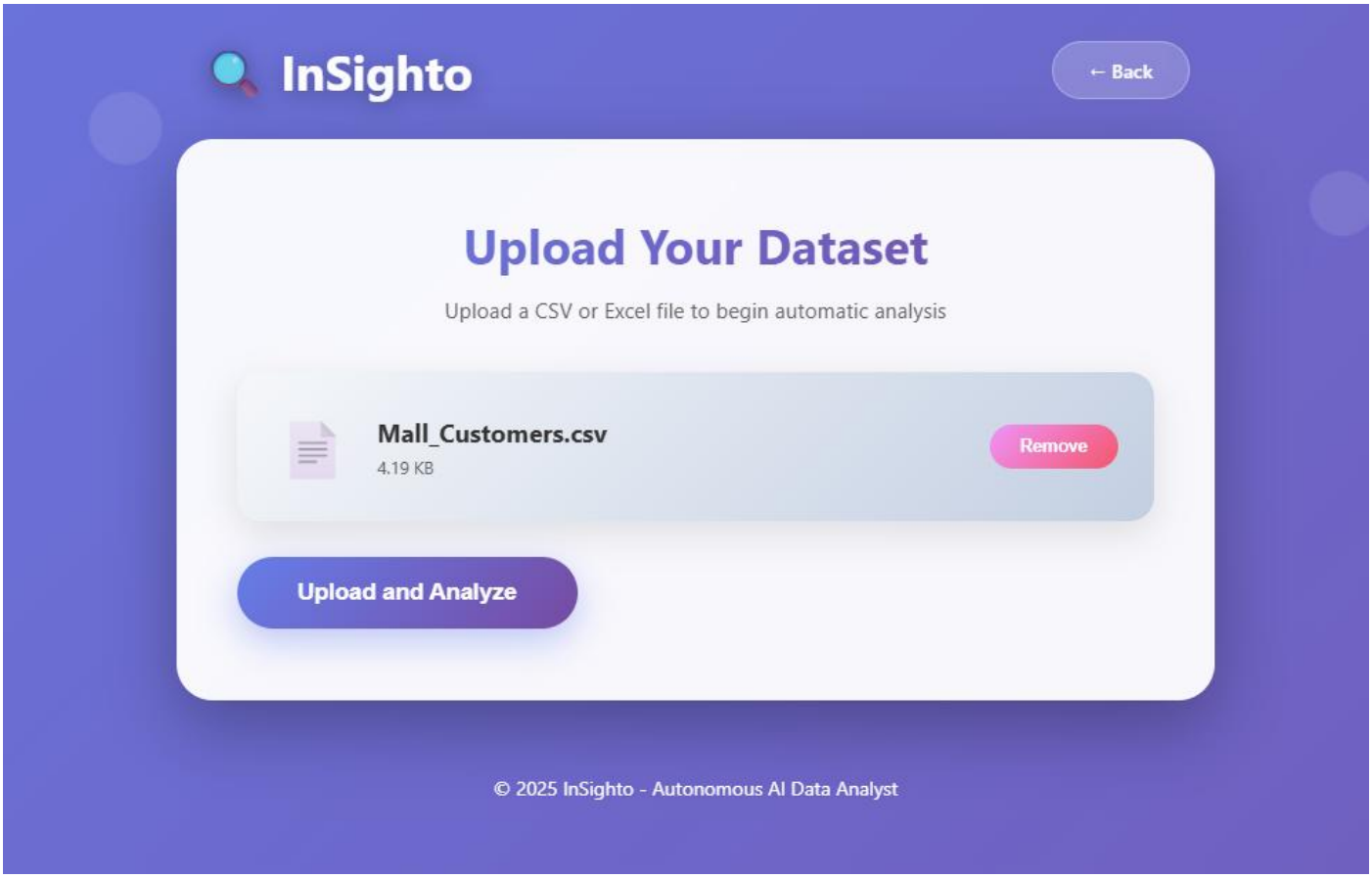


Figure 3: Dataset uploaded interface

7. Backend Implementation

The backend is implemented using Flask with a clear separation of responsibilities:

7.1 Storage Module

- Manages file uploads and session tracking
- Stores datasets and analysis results in SQLite
- Ensures session-level data isolation

7.2 Agent Module

- Loads real user-uploaded datasets
- Performs dataset profiling (rows, columns, missing values)
- Coordinates AI insight generation
- Orchestrates report creation

7.3 Report Generator

- Structures analysis results into logical sections
- Supports extensible report formatting
- Prepares report content for UI display and PDF export

8. Generative AI Integration

8.1 Model Used

- **Model:** llama3:8b
- **Framework:** Ollama (local inference)

8.2 Role of AI

The AI component is used strictly for interpretation and explanation, not for statistical computation.

AI-generated outputs include:

- Key observations
- Patterns and trends
- Data quality remarks
- Executive summary

Strict prompt constraints ensure factual accuracy and prevent hallucinations.

9. Data Privacy & Cost Effectiveness

9.1 Data Privacy

- All processing occurs locally
- No dataset is sent to external cloud services
- Session-based isolation ensures user data separation
- Temporary storage is used to minimize data retention

9.2 Cost Effectiveness

- Open-source technologies only
- No paid API usage
- Local AI inference eliminates recurring costs
- Suitable for students, educators, and small organizations

10. Results & Output

Based on system execution:

- Dataset upload and preview function correctly
- Automated profiling executes on real datasets
- AI-generated insights are displayed on the analysis page
- A structured final report is generated
- Reports can be exported and saved as PDF for offline use

- Users can restart analysis with a new dataset

The system successfully demonstrates an end-to-end AI-assisted data analysis workflow.

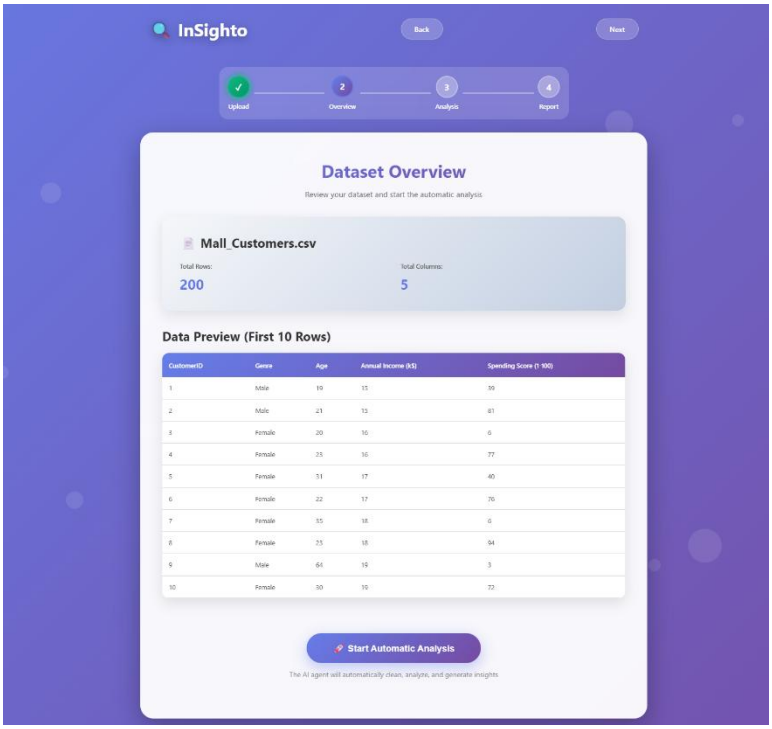


Figure 3: Dataset overview showing column structure and preview rows

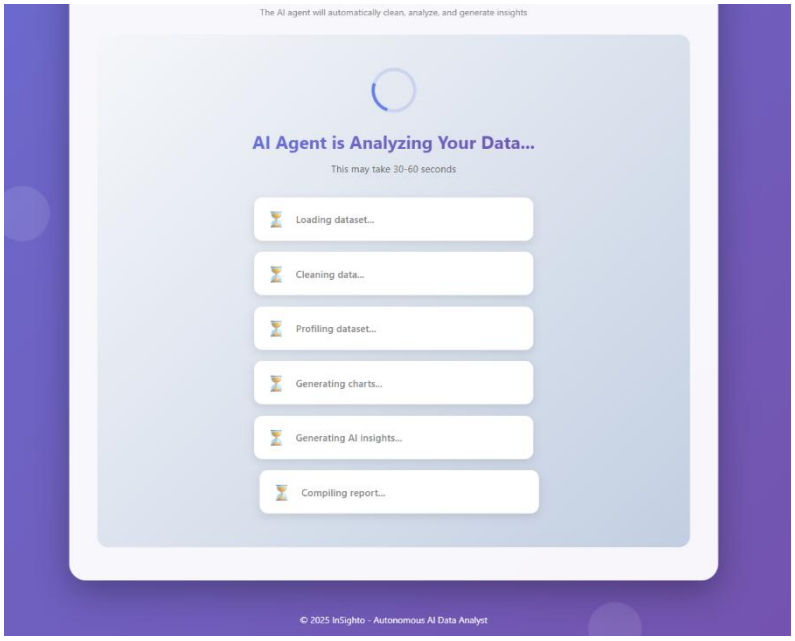


Figure 4: Overview page showing real-time analysis progress while the AI agent processes the uploaded dataset

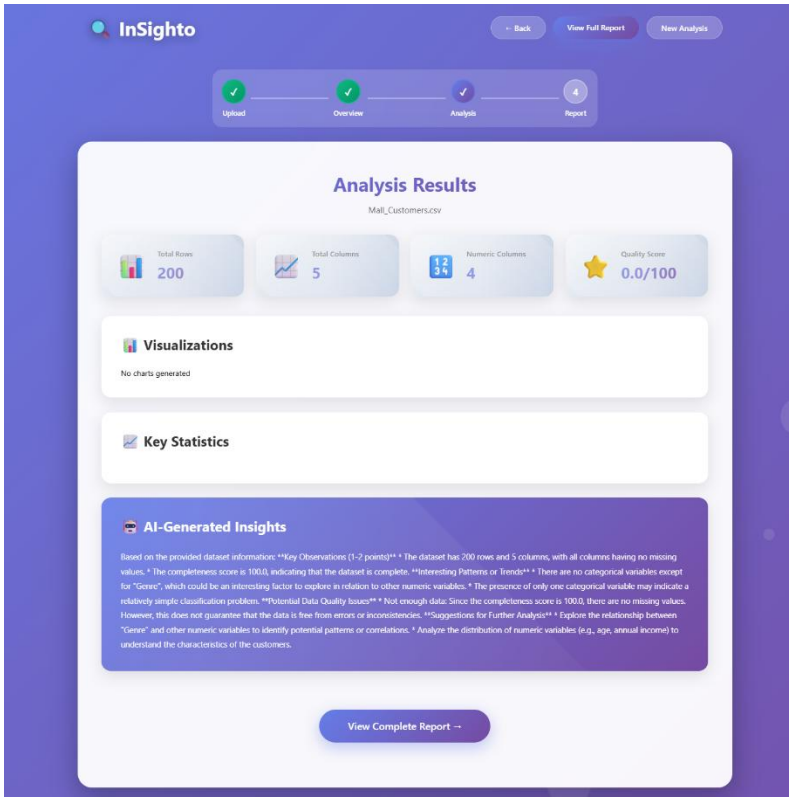
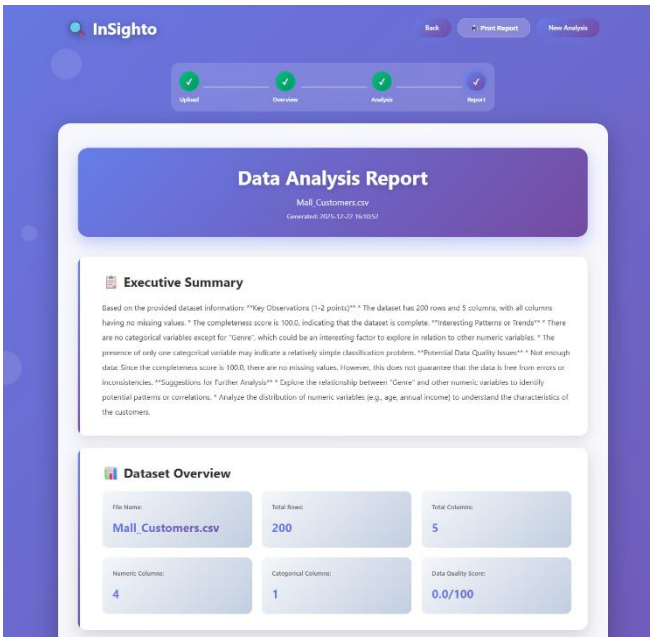


Figure 5: Analysis results page displaying dataset summary and AI-generated insights



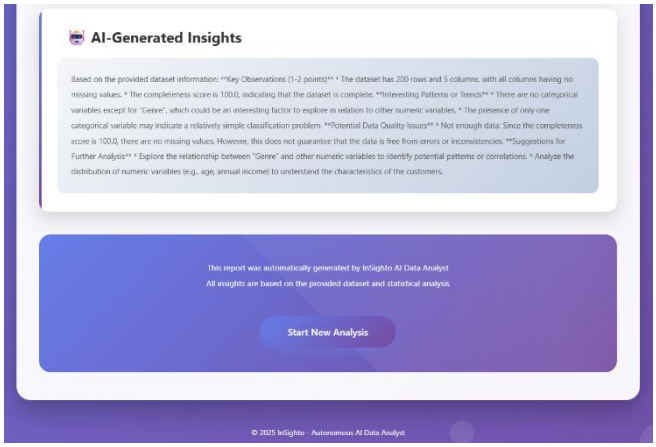


Figure 6: Final structured analysis report generated by the system

11. Real-World Applications

- **Education:** Teaching data analysis concepts interactively
- **Small Businesses:** Analysing sales or customer datasets
- **Research:** Exploratory analysis for academic studies
- **Non-Technical Teams:** Understanding data without coding
- **Privacy-Sensitive Domains:** Local processing without cloud exposure

12. Limitations

- Advanced statistical analysis is limited
- Visual charts are architecturally supported but not yet implemented
- No user authentication system
- No multi-user concurrency management

13. Future Scope

- Automated chart generation
- Advanced statistical summaries
- Data cleaning and preprocessing modules
- AI-generated recommendations integrated into reports
- Export reports in multiple formats (Word, enhanced PDF)
- Support for multiple AI models
- Role-based user access and dashboards

14. Conclusion

InSighto demonstrates how AI-assisted data analysis can be made accessible, privacy-preserving, and cost-effective. By combining automated dataset profiling with local Generative AI and PDF report export capabilities, the system bridges the gap between raw data and actionable insights. The project provides a strong foundation for future enhancements and real-world deployment.

15. Appendix

Appendix A: Application Workflow Reference

Screenshots illustrating the user interface and system outputs have been included within the main body of the report. These figures visually demonstrate the application workflow, analysis progress interface, AI-generated insights, and final report generation. The appendix therefore focuses on supplementary implementation details rather than repeating visual content.

Appendix B: Supported File Formats

The InSighto system supports the following dataset formats for analysis:

- CSV (.csv)
- Microsoft Excel (.xls, .xlsx)

Uploaded files are validated on both the frontend and backend to ensure format correctness and size compliance before processing.

Appendix C: AI Model & Configuration Details

- **LLM Provider:** Ollama
- **Model Used:** llama3:8b
- **Inference Mode:** Local
- **Purpose:**
    - Generating dataset insights
    - Producing executive summaries
    - Assisting in report narrative generation

Strict prompt constraints are enforced to ensure that the AI model generates explanations strictly based on provided dataset information, minimizing hallucinations.

Appendix D: Report Export Capability

The system supports exporting the final generated analysis report as a PDF file. This allows users to save, print, or share the analysis results for academic, professional, or documentation purposes.

16. References

1. Python Software Foundation. *Python Programming Language Documentation*. https://docs.python.org/
2. Pallets Projects. *Flask Web Framework Documentation*. https://flask.palletsprojects.com/
3. McKinney, W. *Data Structures for Statistical Computing in Python*. pandas Documentation. https://pandas.pydata.org/docs/
4. SQLite Consortium. *SQLite Documentation*. https://www.sqlite.org/docs.html
5. Ollama. *Ollama Documentation for Local Large Language Models*. https://ollama.com/
6. Meta AI. *LLaMA 3 Model Overview*. https://ai.meta.com/llama/
7. Mozilla Developer Network (MDN). *HTML, CSS, and JavaScript Documentation*. https://developer.mozilla.org/