# Smart CSV Dashboard with AI-Powered Q&A

## Objective

The goal of this task was to build a Streamlit-based smart dashboard to perform time-based analysis on uploaded CSV files and allow users to ask natural language questions about the data using a local LLM (FLAN-T5). The solution should support year, quarter, month, and day insights, with AI-generated summaries and answers.

## Technologies Used

|  |  |
| --- | --- |
| Tool/Library | Purpose |
| Pandas | CSV parsing and data preprocessing |
| Streamlit | User interface for dashboard and file upload |
| Plotly | Data visualization with time-based trends |
| SentenceTransformers | Embedding model for text representation |
| Transformers (FLAN‑T5) | Lightweight open-source LLM for Q&A |
| Python | Backend logic and data handling |

## Approach

The system is built using the following pipeline:  
  
1. Upload a CSV file via the Streamlit interface  
2. Parse the date column and extract Year, Quarter, Month, and Day  
3. Select a numeric column to analyze  
4. Display interactive time-based line chart using Plotly  
5. Summarize aggregated data  
6. Pass the summary and user question to FLAN-T5 model for Q&A  
7. Return the answer in a chatbot-like text interface

## Sample Q&A Output

|  |  |
| --- | --- |
| Question | Answer |
| Which year had the highest total sales? | The year 2023 had the highest total of ₹42,000. |
| What is the trend of sales over the months? | Sales are increasing from January to June. |
| Which quarter had the lowest average sales? | Q1 2023 had the lowest average sales. |

## Interface Screenshots

Screenshots are included in the screenshots/ folder of the GitHub repository.

## Project Structure

SmartCSV\_QA/  
├── app.py  
├── README.md  
├── report.pdf  
├── requirements.txt  
├── sample\_csv/sample\_sales.csv  
└── screenshots/

## Optional Features Explored

- CSV upload validation  
- Support for multiple time groupings (Year, Quarter, etc.)  
- Model loading using @st.cache\_resource for efficiency  
- Local-only LLM integration with FLAN-T5 (no API keys needed)

## Challenges Faced

- Ensuring date parsing from messy or inconsistent CSVs  
- Handling empty or incorrect metric selections  
- Delays in model loading (especially with FLAN-T5-base)  
- Streamlit responsiveness with large datasets

## Result

- Fully working CSV Dashboard with Q&A engine  
- Uses only open-source libraries  
- Efficient UI for real-time charting  
- Successfully completes Task requirements

## Repository & Demo

- GitHub: [Santhiyagithub/Data\_dashboard\_Q-A: Developed a chat bot and integrate the Question and Answer on a PDF along a Streamlit Smart Dashboard](https://github.com/Santhiyagithub/Data_dashboard_Q-A)  
- Streamlit Demo (Local): Run using `streamlit run app.py`

## Author Info

Name: Santhiya B  
Branch: CSE – AI/ML  
Graduation Year: 2027  
Email: [santhiyabalaji2005@gmail.com](mailto:santhiyabalaji2005@gmail.com)

## Final Note

This task demonstrates my ability to build an AI-powered data analysis platform integrating visualization, semantic search, and natural language understanding in a clean, deployable format. Looking forward to further contributions at eMedlogix.