

Technical Challenge!

Duration: 1-2 days

Objective: Build an MVP (minimal viable product) demo of an LLM-powered data analysis tool that can ingest, query, and visualize Excel/CSV data.

Context: Organizations want to be able to visualize their data but has no easy way of doing this and they have hired an entire research development team to do this task manually. It takes several days to easily create new tables to visualize the current data. They are handling the data in excel sheets and would like to explore the possibility of an AI agent that can help the organization to "talk-to-your-excel-sheets" and be able to streamline their processes for data visualization and be able to derive insights more quickly.

Required Deliverables

Submit **both** of the following by the deadline:

1. Live Deployment URL

- Must be accessible online (Vercel, Railway, Render, Azure, etc.)
 - Include test credentials if authentication is implemented
-

Core Features (Must Have)

1. Data Ingestion

- Upload CSV or Excel files (.csv, .xlsx)
 - There will be 3 CSV/Excel sheets provided (based on year, 2023-2025)
 - sample data will be provided
- Parse and store data in a queryable format (database or in-memory)
- Handle basic data validation (e.g., detect empty files, malformed data)

2. Master Data Checking

- View the merged data in a single large table after all the data has been ingested and aggregated into the master table
- Edit any information here for any last minute changes

- Export this master table into a .csv / .xlsx

3. LLM-Powered Querying

Users should be able to ask natural language questions and receive accurate responses. The system must support:

- **Descriptive Analytics:**
 - Minimum, maximum, average, mean values
 - Count entries, summary statistics
 - Example: *"What is the average age of participants?"*
- **Cross-Referencing Between Tables:**
 - Query relationships across multiple uploaded files
 - Example: *"Show me all participants from Table A who appear in Table B"*
- **Table Generation:**
 - Generate new subset tables based on query results
 - Display results in a structured table view within the chat interface
- **Export Functionality:**
 - Allow users to download master dataset (after ingestion and validation step)
 - Allow users to download generated subset tables (CSV or Excel format)

Bonus Features (Nice to Have)

4. Data Visualization

- Generate simple charts/graphs from query results (bar chart, line chart, pie chart, etc.)
- Use any charting library (Recharts, Chart.js, D3.js, etc.)

Tech Stack Flexibility

You may use **any** tech stack, but we recommend:

- **Frontend:** Next.js, React, or similar

- **Backend:** Python (FastAPI/Flask), or similar
- **Database:** PostgreSQL, or in-memory storage
- **LLM Integration:** OpenRouter (API key will be provided if needed)
- **Deployment:** Vercel, Railway, Render, Azure, AWS, or any platform of your choice

Evaluation Criteria

Criteria	Weight	What We're Looking For
Functionality	40%	Does it work end-to-end? Can we upload data, query it, and get accurate results?
Speed	25%	Completed within 1-2 days
Problem-Solving	20%	Creative solutions to technical challenges (e.g., LLM prompt engineering, data parsing edge cases)
Communication	10%	If unsure, there needs to be clear communication to align and review
Code Quality	5%	Clean, readable, well-structured code with proper error handling

Submission Checklist

Before submitting, ensure:

- ☐ Live deployment is accessible and functional
- ☐ Core features (data ingestion + LLM querying) are fully functional
- ☐ Export functionality works (can download generated tables)

Submission

Email the following to **[insert email]** by **[insert deadline]**:

1. Live deployment URL
2. Document of how to interact with it best