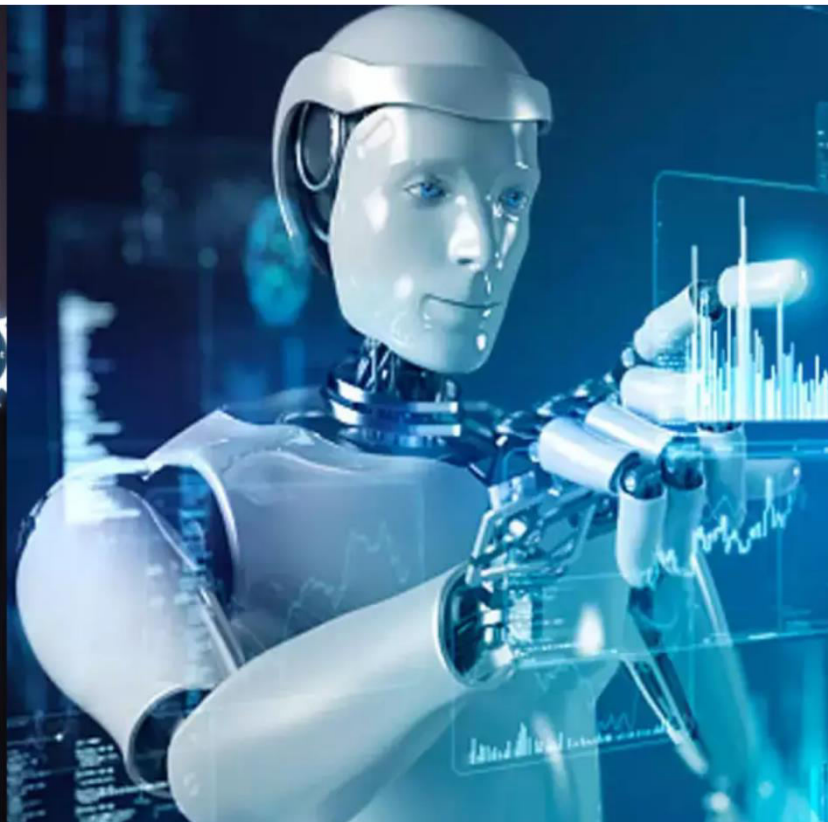




# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 14, Issue 4, April 2025**

# Dataquery AI Assistant: A Natural Language Interface for Intelligent Data Interaction and Analysis

Ms. Sachdev J. C., Akhil Prabhakar Karampuri, Tanisha Shridatta Kadam,

Om Suraj Pardeshi, Dhruv Dattatray Jadhav

Project Guide, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India

**ABSTRACT:** DataQuery AI Assistant is a revolutionary platform on which users are able to query structured data in natural language leveraging the power of advanced large language models (LLMs). By importing Excel or CSV files via the user interface, the data is automatically persisted in an SQLite database. Users are able to execute smart queries, see summarized paragraph results, create data visualizations such as bar charts, line graphs, and pie charts, and lastly export findings in a downloadable PDF report. The design, architecture, and implementation of the assistant are discussed in this paper with examples of how it can be utilized in educational, business, and research.

**KEYWORDS:** NLP, Data Analysis, Visualization, LLM, SQLite, PDF Report Generation, CSV to SQL, AI Assistant.

## I. INTRODUCTION

In the era of big data, making sense of tabular data remains a challenge for non-technical users. Traditional data querying requires knowledge of SQL or programming, creating a gap between users and data insights. To bridge this, we present DataQuery AI Assistant, a platform that enables intuitive interaction with datasets through natural language queries powered by LLMs.

The system automates the end-to-end pipeline—from accepting an Excel or CSV file, storing it in an SQL database, to dynamically generating human-readable insights and visualizations. Unlike conventional dashboards, it allows question-answering over datasets in plain English, helping users make informed decisions quickly.

## II. LITERATURE SURVEY

The union of Natural Language Processing (NLP) and data analytics has grown exponentially in the recent years. Many systems have tried to make data exploration easier by enabling users to query structured data using natural language. Natural Language Interfaces to Databases (NLIDB) first emerged in the 1970s but saw a significant jump with the evolution of modern language models. The initial systems required very large rule-based grammars and templates, which were not domain-scalable. Recent transformer-based LLMs have offered robust query generation without the need for explicit grammar rules.

In [1], Introduced Text-to-SQL systems that can translate user queries into SQL with the assistance of neural networks. Such systems often utilize datasets like Spider or WikiSQL for training. However, most models cannot dynamically adapt to unseen schemas, limiting real-world deployment.

In [2] contrasted different Text-to-SQL models over a range of domains, pointing out generalizability and schema consistency issues. They indicated that model grounding with explicit schema representation is better in terms of accuracy.

Open-source packages like Text2SQLBench and NL2SQL libraries give building blocks for NLIDB systems but don't give end-to-end pipelines with file ingestion, charting, and PDF report generation.

Use of LLMs like LLaMA has revolutionized question-answering and SQL generation work. Qwen's Coder 2.0 model [3] demonstrates outstanding zero-shot and few-shot performance in natural language to structured query representation mapping, making it best suited for dynamic database use. For data visualization, libraries such as Matplotlib and Recharts are widely used. In [4], The use of interactive charting libraries to enable user interaction with data, with support for static and dynamic visual components for exploratory data analysis.

In [5], Systems combining LLMs with tabular data to generate visual summaries but not with capabilities like file-to-database conversion or automated reporting.

### III. METHODOLOGY

The DataQuery AI Assistant follows a modular pipeline that begins with file upload and ends with intelligent data insights and visual outputs. When a user uploads an Excel or CSV file, it is automatically parsed and stored in an SQLite database on the backend. The system extracts the schema and displays it for user reference. A locally hosted open-source LLM processes user queries in natural language, generates corresponding SQL statements, and executes them on the database. The results are converted into human-readable paragraphs. Based on user input or query results, the system generates bar charts, pie charts, and line graphs using Python visualization libraries. All interactions, insights, and visuals can be compiled into a downloadable PDF report, enabling end-to-end automation of data understanding.

### IV. EXPERIMENTAL RESULTS

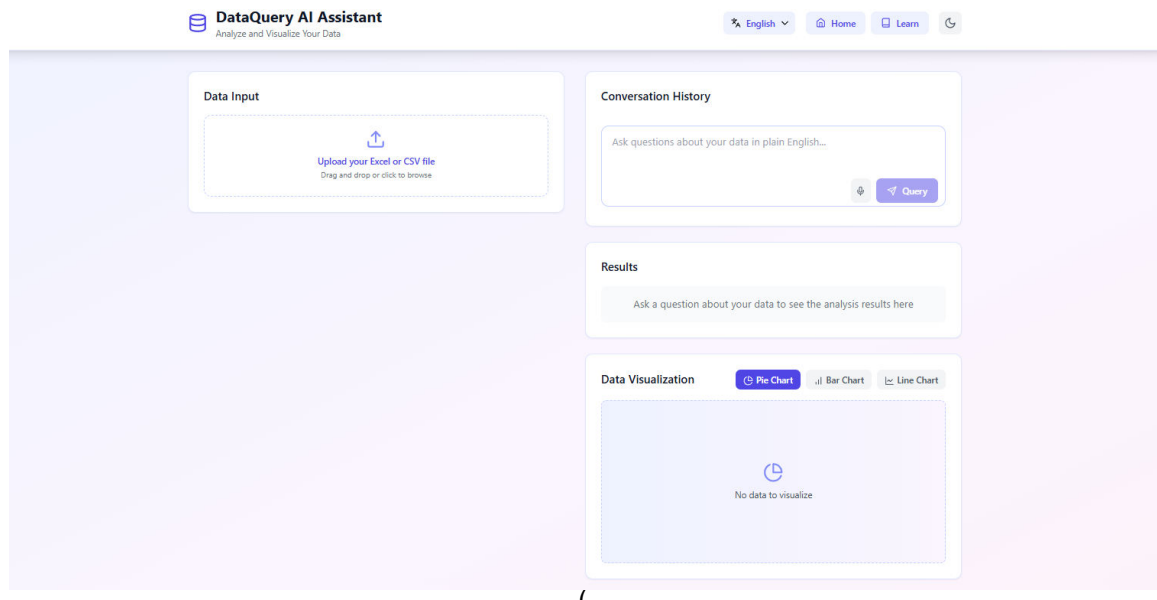


Figure 1(a) shows the initial interface of the system, where users can upload Excel or CSV files for analysis. The layout includes modules for data input, natural language query, result generation, and visualizations.



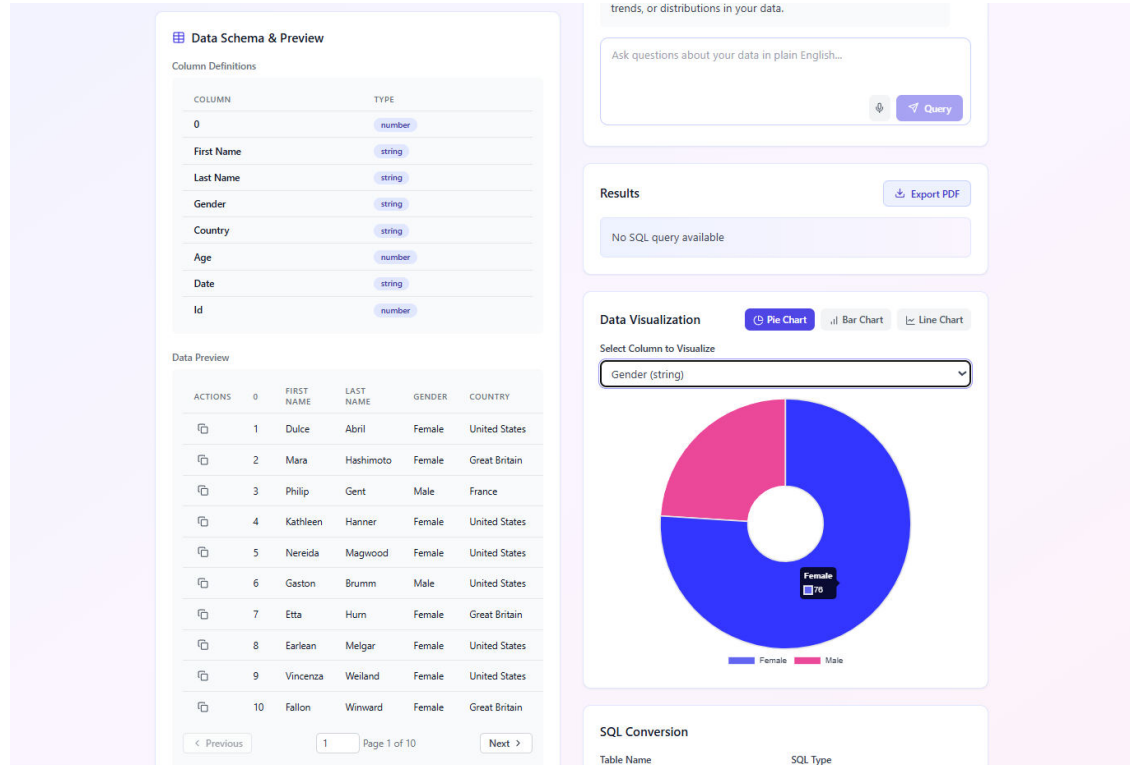


Figure 2(b) represents the result after uploading a sample dataset containing personal information including names, gender, country, age, and date. The system successfully extracts the data schema and displays a data preview with column definitions and data types.

In the Data Visualization section (Figure 2), the user selected the Gender column for visualization. The system generated a Pie Chart that clearly illustrates the distribution of gender data, identifying 78 entries as Female and the rest as Male. This demonstrates the tool's ability to quickly and effectively translate raw data into meaningful visual insights.

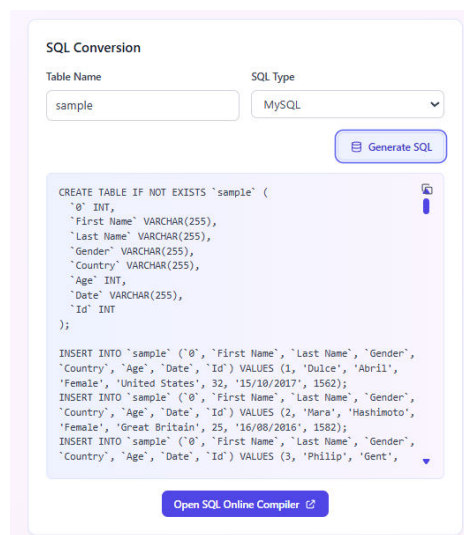


Figure 3(c) illustrates the SQL Conversion feature of the DataQuery AI Assistant. Once a user uploads an Excel or CSV file, the system automatically converts the structured data into SQL queries based on the selected database type—MySQL, PostgreSQL, or SQLite.

In this example, the selected SQL type is MySQL, and the system successfully generated both the CREATE TABLE and INSERT INTO statements.



Figure 4(d) illustrates the PDF Report Generation feature of the DataQuery AI Assistant. Once the user engages with the assistant through data-related queries and visualizations, they can download a comprehensive PDF report that captures the entire interaction.

## V. CONCLUSION

We have implemented an AI-powered data analysis assistant capable of automatically generating insights, visualizations, and SQL queries from structured data files. Our system successfully interacts with users through natural language, processes Excel or CSV files, and converts them into meaningful responses and visual representations. We have tested our assistant on various datasets and found that it accurately understands user queries, generates correct SQL conversions, and provides intuitive visual analytics.

## REFERENCES

- [1] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al., "Language Models are Few-Shot Learners," *Advances in Neural Information Processing Systems*, NeurIPS, 2020. OpenAI, "GPT-4 Technical Report," 2023.
- [2] Rajkumar, R., He, B., Liu, Z., et al., "Natural Language to SQL: A Benchmark Evaluation," *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021.
- [3] Plotly Technologies Inc., "Interactive Visualizations in Python," <https://plotly.com>
- [4] SQLite.org, "SQLite Documentation," <https://sqlite.org>
- [5] Meta AI, "LLaMA 3.2 Language Model," <https://ai.meta.com/llama>
- [6] Python Software Foundation, "Pandas Library," <https://pandas.pydata.org>
- [7] Streamlit Inc., "Streamlit: The fastest way to build and share data apps," <https://streamlit.io>
- [8] PyFPDF, "FPDF for Python," <https://pyfpdf.github.io>
- [9] H. V. Jagadish, J. Gehrke, A. Labrinidis, et al., "Big Data and Its Technical Challenges," *Communications of the ACM*, vol. 57, no. 7, pp. 86–94, 2014.
- [10] Y. Li, Y. Zhang, and D. Jiang, "Text-to-SQL: Recent Progress and Research Challenges," *IEEE Transactions on Knowledge and Data Engineering*, 2022.
- [11] T. Chen, Z. Xu, and Y. Ma, "Building a Natural Language Interface to Databases for Non-Technical Users," *ACM Transactions on Interactive Intelligent Systems*, vol. 11, no. 2, 2021.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details