International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802046

Automated Data Analytics and Visualization System using Prompts

S Balaji[1], Geethan[2], Vijay Chidhambaranathan[3], Sam Salvius[4]

Associate Professor, Dept. of ADS, Kamaraj College of Engineering & Technology, Virudhunagar, India[1]

UG Scholar, Dept. of ADS, Kamaraj College of Engineering & Technology, Virudhunagar, India[2]

UG Scholar, Dept. of EEE, Kamaraj College of Engineering & Technology, Virudhunagar, India[3]

UG Scholar, Dept. of CSE, Kamaraj College of Engineering & Technology, Virudhunagar, India[4]

Publication History: Received: 25.02.2026; Revised: 20.03.2026; Accepted: 25.03. 2026; Published: 28.03.2026.

ABSTRACT: This project presents a **data analytics–based system for automated dataset analysis and visualization using natural language prompts**. The system focuses on simplifying data exploration and generating meaningful insights from structured datasets at an early stage of analysis. A **Pandas-based data processing approach** is used to analyze datasets by handling operations such as data cleaning, filtering, aggregation, and statistical computation based on user prompts. A **data visualization module using Matplotlib** is used to generate graphical representations such as line charts, bar graphs, and scatter plots to help users understand patterns and trends in the data. Users can upload datasets in CSV format and enter natural language queries to generate automated insights and visualizations instantly. This system helps in faster data understanding and supports analysts and decision-makers in making **better data-driven decisions**.

KEYWORDS: Natural Language Prompt Processing, Automated Data Visualization, Pandas Data Analysis, Matplotlib Visualization

I. INTRODUCTION

Artificial Intelligence and Data Analytics play a significant role in modern organizations by enabling efficient analysis and interpretation of large volumes of data. In many industries such as business, education, research, and finance, vast datasets are generated daily and require timely analysis to extract meaningful insights. Traditional data analysis methods often depend on manual coding and technical tools to process datasets and generate visualizations, which can be time-consuming and sometimes difficult for non-technical users. To address this challenge, this project develops an **automated data analytics and visualization system that operates using natural language prompts**. The system allows users to upload structured datasets and perform analysis without writing complex code. The **data analysis module uses Python-based libraries such as Pandas and NumPy** to process datasets by performing operations such as data cleaning, aggregation, filtering, and statistical evaluation. The **visualization module uses Matplotlib** to generate graphical representations such as line charts, bar graphs, and scatter plots that help users understand patterns and trends within the dataset.

Additionally, the system performs **data preprocessing operations** to improve the quality and consistency of the dataset before performing analysis. The application provides a **simple and user-friendly interface** so that even users with limited technical knowledge can interact with the system easily. It also displays generated visualizations along with basic analytical summaries to help users interpret the results effectively. The integration of **data analytics libraries with an interactive web application** makes the system accessible for real-time data exploration. Overall, the project demonstrates how **artificial intelligence and automated analytics tools can improve the efficiency and accessibility of data analysis and visualization**.

1.1 METHODS AUTOMATED DATA VISUALIZATION SYSTEM

Data Collection and Preprocessing: – Structured datasets such as CSV or Excel files are collected from public data repositories and sample datasets used for data analysis tasks. The collected data is cleaned by removing missing, duplicate, or inconsistent values. Categorical attributes are converted into numerical or structured formats where required, and numerical values are normalized for consistent analysis. Dataset columns are validated and organized to make them suitable for automated data processing and visualization generation.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802046

Data Analysis using Pandas: – The Pandas data analysis library is used to process and analyze the uploaded dataset based on user queries. It performs operations such as filtering, grouping, aggregation, and statistical analysis on different columns of the dataset. When a user provides a prompt requesting a specific analysis, the system identifies the relevant dataset attributes and applies the required data manipulation functions. This approach enables efficient analysis of structured datasets and helps generate meaningful insights.

Data Visualization using Matplotlib: – A Matplotlib-based visualization module is used to generate graphical representations from the analyzed dataset. The visualization component automatically creates charts such as line graphs, bar charts, and scatter plots depending on the nature of the dataset and the user query. Visualization functions help convert numerical data into clear graphical formats, making it easier for users to understand patterns, relationships, and trends within the data.

System Integration using Streamlit: – All data analysis and visualization components are integrated into an interactive web application using the Streamlit framework. The application allows users to upload datasets and enter natural language queries for analysis. The system processes the dataset and prompt, performs the required operations, and displays the generated visualizations instantly through a simple and user-friendly interface.

II. SYSTEM DESIGN

Dataset Collection – Structured datasets used for analysis are collected from reliable data sources such as Kaggle and public data repositories. These datasets contain numerical and categorical attributes representing different types of information used for performing automated data analysis and visualization tasks.

Data Preprocessing – The collected datasets are cleaned by removing missing, duplicate, or incorrect values to improve the quality and reliability of the data. Normalization and formatting techniques are applied so that the dataset becomes suitable for data processing and analytical operations.

Feature Processing – Dataset attributes such as categorical labels or textual fields are converted into appropriate formats so that they can be efficiently processed by the analysis system. This step ensures that the data structure is compatible with data processing libraries.

Dataset Validation – Uploaded datasets are verified to ensure that the column names, data types, and structure are suitable for performing automated analysis. Validation helps prevent errors and improves the reliability of the generated results.

Module Selection – The Pandas library is selected for data analysis because of its powerful capabilities in handling structured datasets. The Matplotlib library is chosen for visualization due to its ability to generate multiple types of charts and graphical representations effectively.

Data Processing – The data processing module performs operations such as filtering, grouping, aggregation, and statistical analysis on the dataset. These operations help extract meaningful information from the dataset before visualization.

Visualization Generation – After processing the dataset, the system generates visualizations such as bar charts, line graphs, and scatter plots to represent the data insights in graphical form.

Module Integration – The data processing and visualization modules are integrated into a web-based system using the Streamlit framework, allowing smooth interaction between the analysis engine and the user interface.

User Input – Users can upload structured datasets such as CSV files and provide natural language prompts describing the type of analysis they want to perform on the dataset.

Analysis Process – The system processes the dataset along with the user prompt, identifies the required analysis operations, and performs the necessary data processing tasks.

Result Display – The generated visualizations and analytical summaries are displayed on the web interface in an easy-to-understand graphical format.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802046

Deployment – The application is deployed using the Streamlit framework to provide real-time accessibility. Users can access the system through a web browser and generate automated data insights and visualizations instantly.

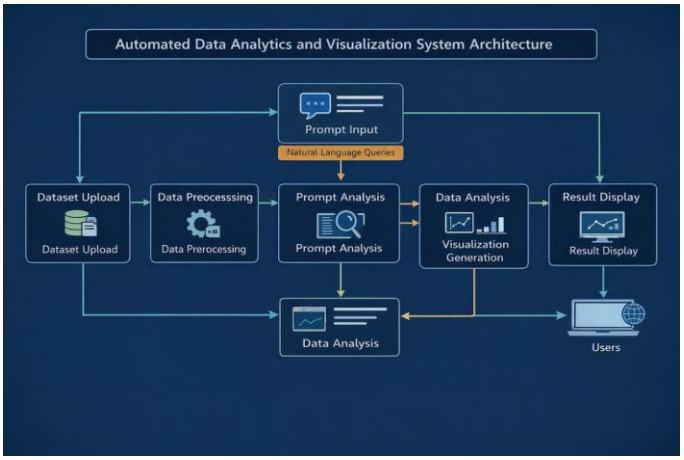


FIGURE.1. System Design

Work Flow

The system workflow begins with collecting **structured datasets such as CSV or Excel files** used for performing automated data analysis and visualization. The collected datasets undergo preprocessing steps such as **data cleaning, normalization, and validation of dataset attributes** to improve data quality and ensure accurate analysis. The processed dataset is then handled by the **Pandas-based data analysis module** to perform operations such as filtering, grouping, aggregation, and statistical evaluation based on the user query. The processed data is further passed to a **Matplotlib visualization module** to generate graphical representations such as line charts, bar graphs, and scatter plots that help interpret the dataset effectively. Finally, the system processes the dataset and user query, performs the required data operations, and displays the generated visualizations and insights through a **user-friendly interactive interface**.

III. RESULTS AND DISCUSSION

RESULTS

The system successfully performs **automated data analysis and visualization** based on user-provided natural language queries. It analyzes structured datasets using the **Pandas data processing module** to perform operations such as filtering, aggregation, and statistical evaluation. It also generates graphical insights from the dataset using the **Matplotlib visualization module**, producing charts such as line graphs, bar charts, and scatter plots. The generated analysis results and visualizations are displayed through a **user-friendly Streamlit-based web interface**. The system demonstrates reliable performance in processing datasets and generating meaningful visual insights. This helps users in **faster data interpretation** and supports analysts and decision-makers in making **better data-driven decisions**.

DISCUSSION

The developed system demonstrates how **automated data analytics and visualization techniques can assist in efficient data exploration and interpretation**. The **Pandas-based data analysis module** effectively processes structured datasets and performs operations such as filtering, grouping, and statistical analysis to generate meaningful insights. The **Matplotlib visualization module** successfully creates graphical representations such as line charts, bar graphs, and scatter plots to illustrate patterns and trends within the dataset.The interactive interface allows users to upload datasets, enter **natural language prompts**, and obtain visual insights quickly. The system reduces the effort required for manual data analysis and supports analysts and decision-makers in interpreting data efficiently.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802046

FIGURE 3.1 Results

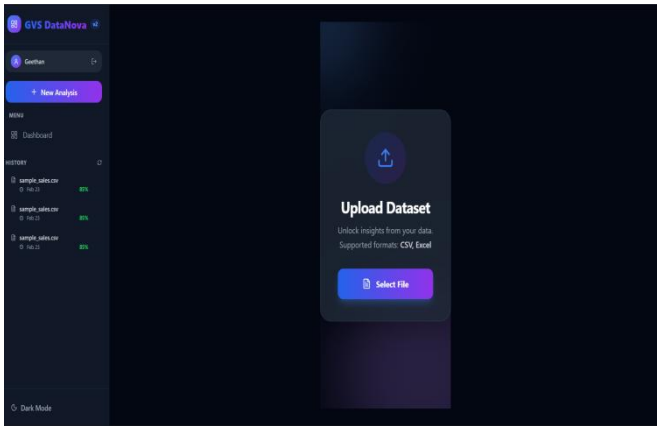


FIGURE 3.2 Results

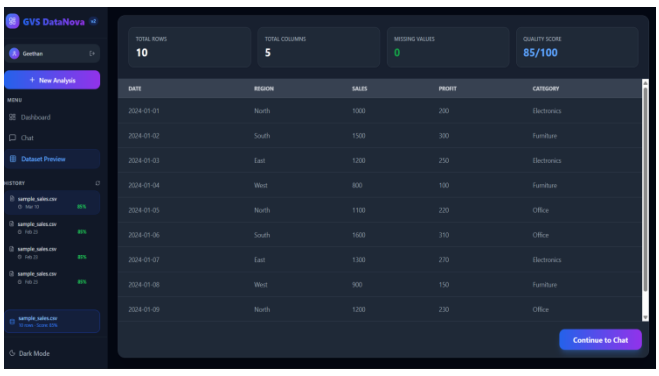


FIGURE 3.3 Results

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802046

FIGURE 3.4 Results

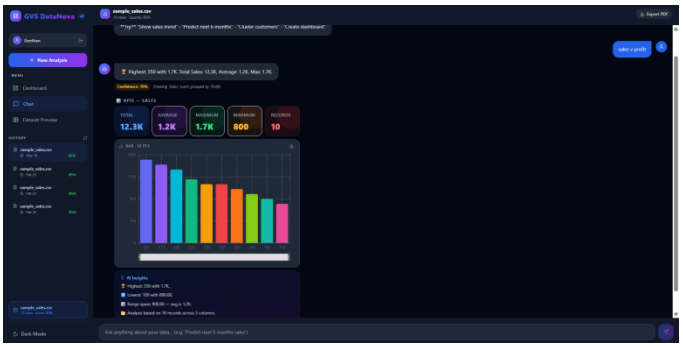
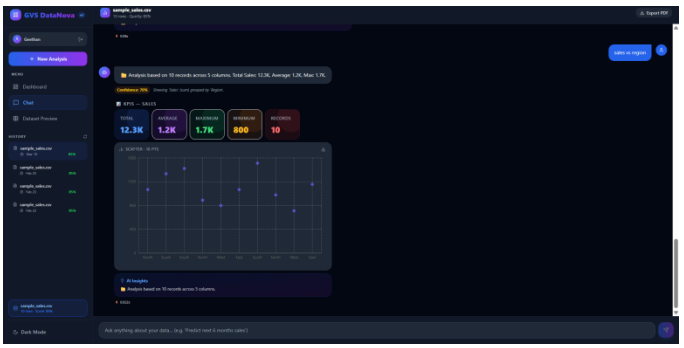


FIGURE 3.5 Results

FIGURE 3.6 Results



|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802046

IV. CONCLUSION

This project presents a **data analytics–based system for automated dataset analysis and visualization using natural language prompts**. The **Pandas data processing library** is used to analyze structured datasets for generating meaningful insights and statistical summaries. A **Matplotlib visualization module** is used to create graphical representations such as charts and plots from processed datasets. The system helps in faster data understanding and supports analysts and decision-makers in interpreting information accurately. Overall, the project demonstrates how **artificial intelligence and data analytics tools can improve data exploration and analytical efficiency**.

V. ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to our **project guide** for their valuable guidance and continuous support throughout this project. I also thank the **faculty members of the Department of Artificial Intelligence and Data Science** for their encouragement and suggestions. My sincere thanks to our **institution** for providing the necessary resources and facilities to complete this project successfully. Finally, I thank my **friends and family** for their constant motivation and support.

REFERENCES

Patel et al. (2021) proposed a data analytics–based approach for automated data visualization using structured datasets. The study focuses on analyzing tabular datasets to generate meaningful insights automatically. Data preprocessing techniques are applied to improve the quality of datasets before performing analysis. Data processing libraries are used to extract important patterns and relationships from the dataset. The system is implemented using structured datasets containing multiple attributes and numerical values. The analysis module can generate graphical representations such as charts and plots from a given dataset. The proposed method helps in improving data interpretation and reducing the manual effort required for analysis. This research demonstrates the effectiveness of data analytics techniques in automated data analysis and visualization systems.

JOURNAL REFERENCE STYLE:

1) A. Smith and R. Brown, “Automated Data Visualization Using Python Libraries,” 2021.
2) M. Johnson, P. Lee, and K. Walker, “Natural Language Interfaces for Data Analytics Systems,” 2021.
3) S. Kumar, A. Gupta, and R. Singh, “Automated Data Analysis Using Python and Visualization Tools,” 2019.
4) C.Nagarajan and M.Madheswaran - ‘Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques’- Taylor &Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011. DOI: 10.1080/15325008.2010.541746
5) C.Nagarajan and M.Madheswaran - ‘Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter’ - Journal of Electrical Engineering, Vol.63 (6), pp.365-372, Dec.2012. DOI: 10.2478/v10187-012-0054-2
6) C.Nagarajan and M.Madheswaran - ‘Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis’- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011. DOI 10.1007/s00202-011-0203-9
7) S.Tamilselvi, R.Prakash, C.Nagarajan,“Solar System Integrated Smart Grid Utilizing Hybrid Coot-Genetic Algorithm Optimized ANN Controller” Iranian Journal Of Science And Technology-Transactions Of Electrical Engineering, DOI10.1007/s40998-025-00917-z,2025
8) S.Tamilselvi, R.Prakash, C.Nagarajan,“ Adaptive sliding mode control of multilevel grid-connected inverters using reinforcement learning for enhanced LVRT performance” Electric Power Systems Research 253 (2026) 112428, doi.org/10.1016/j.epsr.2025.112428
9) S.Thirunavukkarasu, C. Nagarajan, 2024, “Performance Investigation on OCF and SCF study in BLDC machine using FTANN Controller," Journal of Electrical Engineering And Technology, Volume 20, pages 2675–2688, (2025), doi.org/10.1007/s42835-024-02126-w
10) C. Nagarajan, M.Madheswaran and D.Ramasubramanian- ‘Development of DSP based Robust Control Method for General Resonant Converter Topologies using Transfer Function Model’- Acta Electrotechnica et Informatica Journal , Vol.13 (2), pp.18-31,April-June.2013, DOI: 10.2478/aeei-2013-0025.
11) C.Nagarajan and M.Madheswaran - ‘DSP Based Fuzzy Controller for Series Parallel Resonant converter’- Springer, Frontiers of Electrical and Electronic Engineering, Vol. 7(4), pp. 438-446, Dec.12. DOI 10.1007/s11460-012-0212-0.
12) C.Nagarajan and M.Madheswaran - ‘Experimental Study and steady state stability analysis of CLL-T Series

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802046

Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.

13) C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
14) Suganthi Mullainathan, Ramesh Natarajan, "An SPSS and CNN modelling based quality assessment using ceramic materials and membrane filtration techniques", Revista Materia (Rio J.) Vol. 30, 2025, DOI: https://doi.org/10.1590/1517-7076-RMAT-2024-0721
15) M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022
16) L. Zhang, Y. Chen, and H. Zhao, "Interactive Data Visualization for Large Datasets," 2020.
17) R. Patel and N. Shah, "AI-Based Data Exploration and Visualization Framework," 2020.
18) T. Anderson and D. Clark, "Overview of Data Visualization Techniques in Data Science," 2020.
19) P. Williams, J. Thomas, and K. Scott, "Visualization Methods for Analytical Data Interpretation," 2019.