



# Intelligent Automated Data Analysis System using Generative AI for Insight Generation

K Naresh<sup>1</sup>, Nuthangi Ysaswini<sup>2</sup>

<sup>1</sup>Assistant Professor, Department of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

<sup>2</sup>Postgraduate, Department of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

## How to Cite this Article:

Yasaswini, N. (2026). Intelligent Automated Data Analysis System using Generative AI for Insight Generation. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(04).  
<https://doi.org/10.55041/ijcope.v2i4.076>

## License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i4.076>

## Abstract

The necessity for effective data analysis tools that may yield significant insights with little human interaction has grown due to the quick expansion of data across numerous disciplines. In order to streamline the process of analyzing data, this study proposes an intelligent automated data analysis system that combines conventional data analytics methods with generative artificial intelligence. The suggested solution enables users to submit datasets in a variety of formats, including CSV and Excel, and uses a Generative AI model to automatically carry out preprocessing, statistical analysis, and visualization in addition to producing analytical insights that like those of a human.

The Django framework is used in the development of the system, which uses Python modules for data handling and visualization, including Pandas, Matplotlib, Seaborn, and Plotly. Furthermore, based on dataset properties, the Google Gemini API is leveraged to produce contextual insights, suggestions, and possible machine learning strategies. The method makes data interpretation accessible to non-technical users and drastically cuts down on the time and skill needed for data analysis. The technology successfully generates significant insights and visual representations, which improves decision-making processes, according to experimental results.

## Keywords

Generative AI, Automated Data Analysis, Data Visualization, Machine Learning, Django Framework, Pandas, Insight Generation



## I. Introduction

Organizations produce enormous volumes of data in today's data-driven environment, necessitating effective processing and analysis. Conventional data analysis techniques frequently require substantial time commitment, programming skills, and domain experience. For non-technical users who require quick and insightful data analysis, this creates a gap. Automated data analysis systems have drawn interest as a solution to this problem. By automating data preprocessing, visualization, and interpretation, these systems seek to minimize manual labor. Nevertheless, the majority of current technologies just offer charts and statistical results without providing a human-readable explanation of the insights.

It is now feasible to produce intelligent explanations and suggestions from unprocessed data thanks to the development of generative AI. In order to provide both textual and visual insights, this study suggests an automated data analyzer that blends traditional data analytics methods with generative AI.

In addition to analyzing datasets, the system offers appropriate machine learning methods, explains patterns, and makes preprocessing recommendations. For novices, analysts, and researchers, this makes it an effective tool.

## II. Objective

This project's main goal is to create an intelligent, user-friendly automated data analysis system that makes it easier to extract valuable insights from unprocessed datasets. Without requiring sophisticated technical knowledge, the system seeks to allow users to submit data in a variety of formats, including CSV and Excel, and automatically carry out crucial functions including data pretreatment, statistical analysis, and visualization. Integrating generative AI into the analysis pipeline is another important goal. This will enable the system to produce human-like explanations, spot trends, and offer practical suggestions based on the information. The project also aims to improve accuracy and accessibility while cutting down on the time and effort required for manual data processing.

## III. Methodology

The suggested system's methodology adheres to a methodical pipeline that guarantees effective data processing and analysis. Data upload is the first step in the process, when users submit datasets via a web interface created with the Django framework. After being

submitted, the dataset is processed using Python tools like Pandas and kept safely. Finding the file format, dealing with missing values, and extracting important data like dataset dimensions and data types are all part of the preprocessing step. Following preprocessing, statistical analysis is carried out to produce descriptive summaries that include metrics like standard deviation, mean, and median.

In order to help users comprehend relationships among the data, the system then generates visuals utilizing libraries like Matplotlib, Seaborn, and Plotly, such as distribution graphs, pie charts, and correlation heatmaps. The integration of Generative AI via the Gemini API, which evaluates the processed data and produces insights, spots trends, and recommends possible machine learning techniques, is a crucial part of the methodology. An interactive dashboard that incorporates statistical data, visuals, and AI-generated explanations displays the finished product.

## IV. Dataset

The suggested system supports several file formats, including CSV, Excel, and TSV, and is made to operate with user-provided datasets. The system can handle structured data with both numerical and category features because the datasets are uploaded dynamically. The inability of users to comprehend and analyze datasets because of a lack of technical expertise and the complexity of current tools is one of the main issues this work attempts to address. Preprocessing data, choosing suitable analysis techniques, and interpreting findings are challenges that many users face. The development of a system that automates these procedures while offering understandable and significant insights is the main goal of the issue statement.

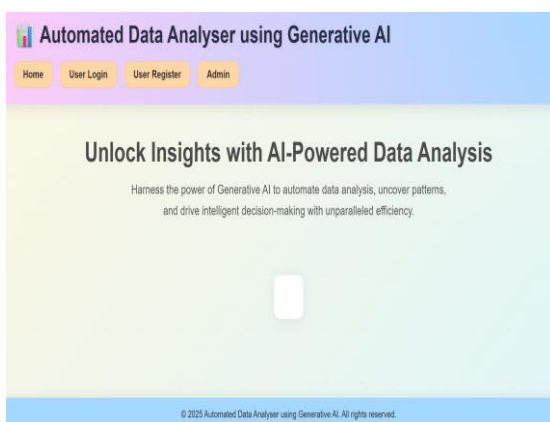
## V. Problem Statement

The project's primary goals are to develop a user-friendly interface for uploading data, automate statistical analysis and preprocessing, create visualizations to show trends, and include generative AI to generate recommendations and explanations that like those of a human. By bridging the gap between unprocessed data and useful insights, the system hopes to empower users to make well-informed decisions without the need for sophisticated technical knowledge.

## VI. Output Screenshots

The system generates a number of outputs that show how well it works. Users can safely access the system using

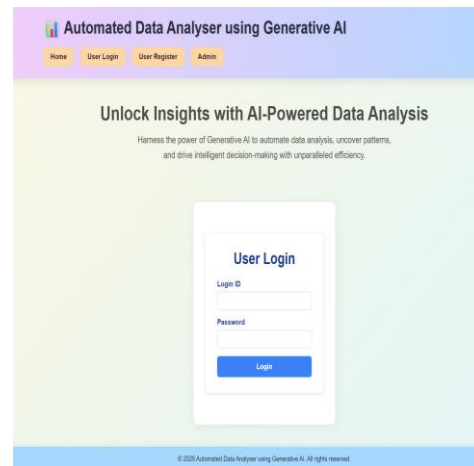
the main interface, which consists of a homepage and user authentication pages. Users can upload datasets via a special upload interface after logging in. The system presents organized tables and statistical summaries that give an overview of the dataset after the data has been processed. To show patterns and relationships within the data, visual outputs like graphical charts and correlation heatmaps are produced. The system also offers a special area for AI-generated insights, where users can see thorough justifications and suggestions drawn from the dataset. Together, these results demonstrate the system's capacity to convert unprocessed data into information that can be understood.



**Figure 1: Automated Data Analyser**

The system automatically starts processing the data after the dataset has been successfully uploaded. It organizes the dataset into an organized manner, finds missing values, and cleans the data. Following processing, the data is shown in the form of well-structured tables and statistical summaries, providing users with an easy-to-understand summary of key descriptive statistics, variable types, and record counts. Without requiring manual analysis, this stage enables users to rapidly comprehend the fundamental structure and quality of their data.

The system produces a range of visual outputs in addition to numerical summaries that greatly improve data interpretation. These include various graphical chart forms like bar graphs, pie charts, and distribution plots, as well as correlation heatmaps that illustrate links between variables. By making it simpler to see trends, patterns, and anomalies in the dataset, these visualizations help users quickly get deeper insights. The whole user experience is further enhanced by the inclusion of interactive and eye-catching charts.



**Figure 2: User Login**

## VII. Conclusion

To improve the overall data analysis experience, the suggested intelligent automated data analysis system effectively combines conventional data processing methods with generative AI. The system lowers the work and knowledge needed to examine datasets by automating complicated processes including preprocessing, statistical evaluation, and visualization. By offering lucid and comprehensible explanations of data patterns and trends, AI-generated insights enhance usability even more. Non-technical users benefit most from the system since it makes it possible for them to comprehend and analyze data without requiring a lot of training. To further boost the system's efficacy and scalability, future improvements may include support for real-time data processing, integration of cutting-edge machine learning models, and enhanced explainability features.

## References

- [1] Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 2017, pp. 5998–6008, A. Vaswani et al., "Attention Is All YouNeed."
- [2] T. Brown et al., "Language Models are Few-Shot Learners," Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Virtual, 2020, pp. 1877–1901.
- [3] Data Analysis in the Era of Generative AI, J. P. Inala, C. Wang, A. Upadhyay, and D. S. Weld, Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI), Honolulu, HI, USA, 2024.
- [4] "Understanding Generative AI-Assisted Data Analysis Workflows," arXiv preprint arXiv:2407.02903, 2024, I. Drosos, A. G. Forbes, and M. Tory.
- [5] M. J. Buehler, "Generative Knowledge Extraction: Principles and Applications," arXiv preprint arXiv:2403.11996, 2024.