

# PROJECT REPORT

## Insight Genie

---

## AI-Powered Data Analysis Assistant

---

**Team Name: Prime**

**Group Members:**

**Ebaad Khan - DT-22045**

**Ezaan Khan - DT-22046**

**Syed Ahmed Ali - DT-22301**

**Khuzaima Hassan - DT-22302**

**Institution: NED University of Engineering and Technology**

**SUBMITTED TO: DATA INSIGHTS**

# PROJECT REPORT

## Executive Summary

Insight Genie is a Streamlit-based web application developed to automate and enhance the data analysis experience for non-technical users. It enables users to upload datasets and receive insightful visualizations, AI-generated summaries, and detailed reports. Using tools like pandas, seaborn, plotly, and Gemini (Google Generative AI), the application facilitates interactive exploration and conversational data insights.

## Introduction

With the increasing availability of datasets, there is a growing need for intelligent systems that can assist users in understanding their data. Insight Genie is created to bridge this gap by offering a no-code interface where users can upload datasets and get an AI-powered breakdown including cleaning, profiling, visualization, and insights.

**Problem Statement:** Users often lack tools or skills to analyze datasets effectively.

**Scope:** The application supports CSV and Excel files, basic data cleaning, automatic visualizations, and AI-generated insights.

**Limitations:** Requires internet access for AI features, and advanced analytics may be limited by model capabilities.

## Methodology/Approach

- Tools: Python, Streamlit, Pandas, Plotly, Seaborn, Google Generative AI.
- Frameworks: Streamlit frontend, Gemini LLM for AI-driven answers and summaries.
- Steps:
  1. Data Upload and Preview.
  2. Data Cleaning (Missing values, Outliers).
  3. Automatic and custom visualizations.
  4. AI Insight generation using Gemini API.
  5. Report generation in HTML format.

# PROJECT REPORT

## Analysis & Results

Insight Genie was tested with datasets to validate its capabilities. Users were able to see statistical summaries, handle missing/outlier values, and receive interactive visualizations. The AI generated high-quality summaries and recommendations, enhancing interpretability. Below is an example of typical insights produced:

- Feature X is strongly correlated with Y ( $r = 0.89$ ).
- Column A contains 12% missing values; filled with median.
- The most frequent category in Column B is "Retail" (35%).

These insights help drive decisions quickly without requiring advanced data science expertise.

## Conclusion & Recommendations

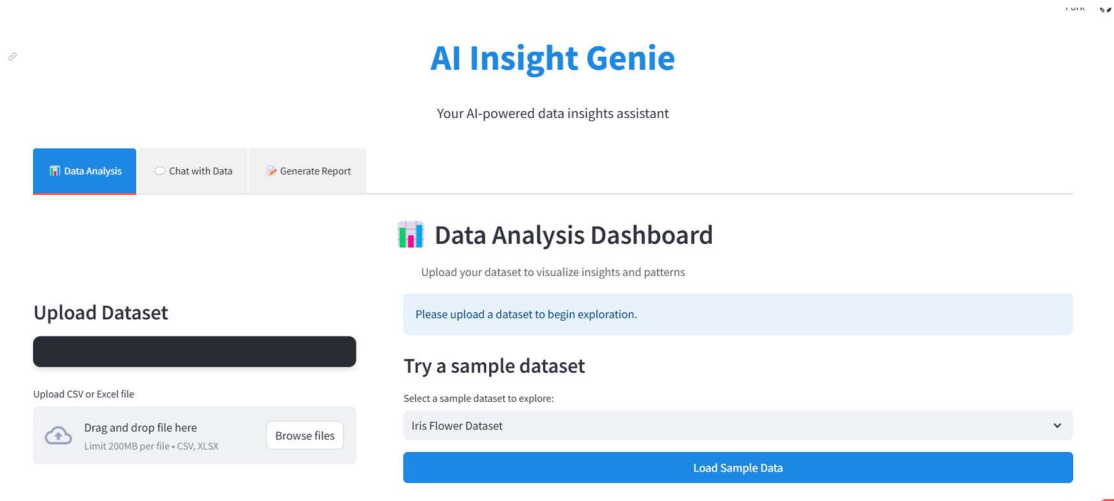
Insight Genie effectively demonstrates how AI can be integrated into data analysis workflows to empower non-technical users. Its ability to produce interactive, interpretable outputs with minimal input is a major strength.

Recommended future enhancements include support for multi-file analysis, integration with cloud storage, UI enhancements and PDF report generation.

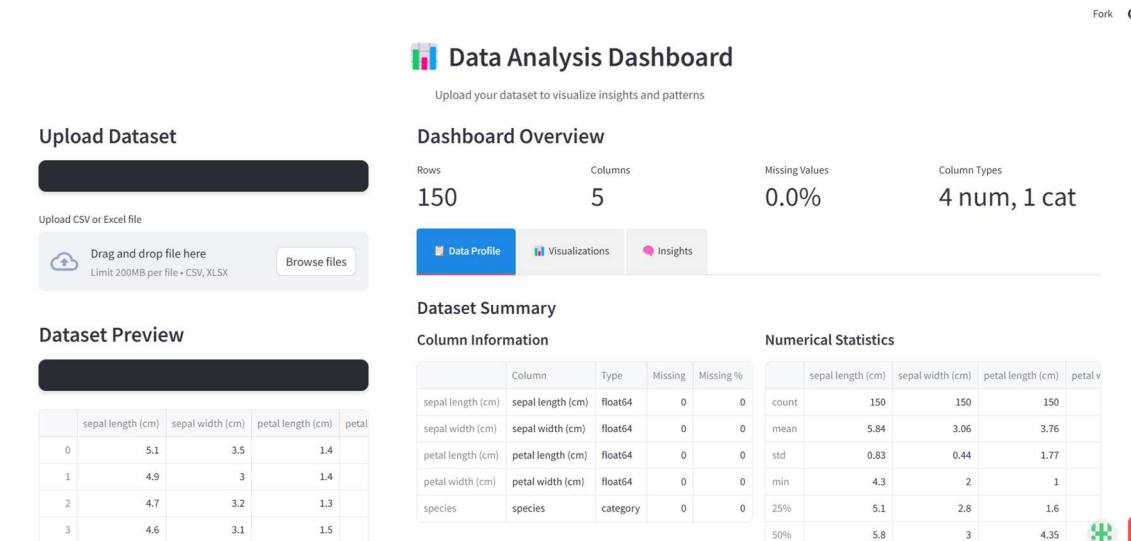
PROJECT REPORT

SCREENSHOTS OF UI

The UI works better with light mode



Dataset loaded



PROJECT REPORT

Visualization tab

Drag and drop file here  
Limit 200MB per file • CSV, XLSX

Browse files

Dataset Preview

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
0	5.1	3.5	1.4	0.2	seto
1	4.9	3	1.4	0.2	seto
2	4.7	3.2	1.3	0.2	seto
3	4.6	3.1	1.5	0.2	seto
4	5	3.6	1.4	0.2	seto

Data Cleaning Tools

Missing Values

Outliers

Handle missing values

Drop missing rows

Apply

Data Profile

Visualizations

Insights

Visualizations

Filter visualizations by type:

All

Distribution of sepal length (cm)

Distribution of sepal width (cm)

Insights tab

Drag and drop file here  
Limit 200MB per file • CSV, XLSX

Browse files

Dataset Preview

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
0	5.1	3.5	1.4	0.2	seto
1	4.9	3	1.4	0.2	seto
2	4.7	3.2	1.3	0.2	seto
3	4.6	3.1	1.5	0.2	seto
4	5	3.6	1.4	0.2	seto

Data Cleaning Tools

Missing Values

Outliers

Handle missing values

Drop missing rows

Apply

Data Profile

Visualizations

Insights

AI-Generated Insights

AI-Generated Insights

Here are four key insights from the provided dataset:

1. \*\*Setosa species is clearly separable based on petal dimensions:\*\* The dataset contains an almost even split between 'setosa' and 'versicolor' (50 instances each). However, a crucial difference lies in petal dimensions. 'Setosa' flowers have significantly smaller petal lengths (mean 1.46 cm, median 1.5cm) and widths (mean 0.24 cm, median 0.2 cm) compared to 'versicolor', suggesting petal measurements could be used to accurately classify these two species. Further analysis with the missing 'virginica' species would determine the generalizability of this finding.

2. \*\*Sepal width shows a higher degree of variation than sepal length:\*\* While both sepal length (mean 5.84 cm, range 4.30-7.90 cm) and width (mean 3.06 cm, range 2.00-4.40 cm) exhibit variation, the range of sepal width is proportionally larger, implying a wider distribution. This suggests sepal width might be a less reliable predictor compared to sepal length when considering species classification or other analyses.

3. \*\*Petal length and width exhibit a strong positive correlation (likely):\*\* The provided summary statistics do not directly show correlation, but the substantial difference in petal dimensions between the two present species strongly implies a positive correlation between petal length and width. A further correlation analysis is needed to confirm this, but such a relationship could be valuable for predictive modeling.

4. \*\*The dataset is balanced (for the two visible species), facilitating fair model training:\*\* The near-equal representation of 'setosa' and 'versicolor' (50 instances each) minimizes class imbalance issues, which often affect the performance of machine learning models. This balance allows for the development of robust classification models without needing to address overfitting or underfitting stemming from an uneven distribution of classes. The addition of the 'virginica' species' data would need to be considered to ensure this remains balanced.

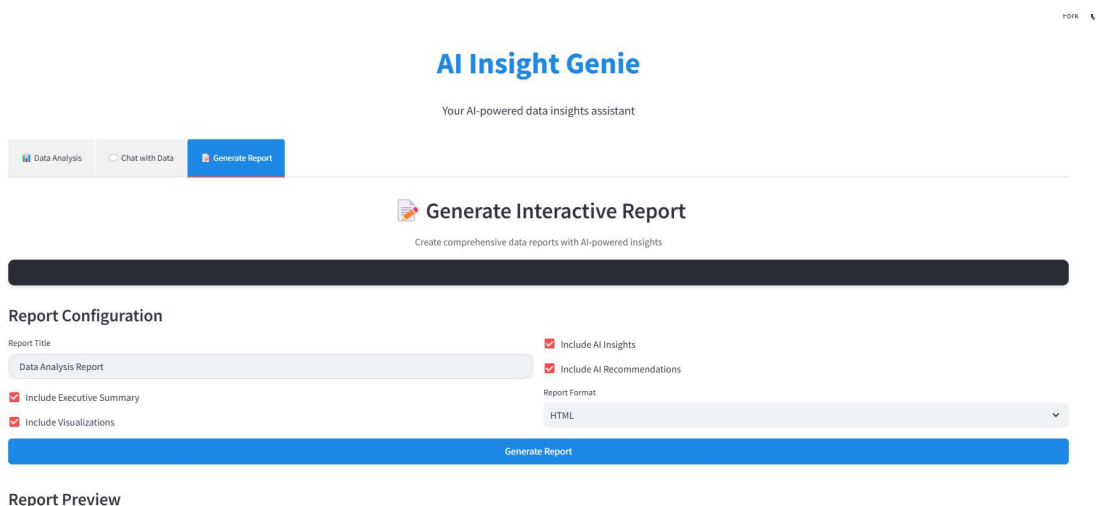
# PROJECT REPORT

Chat tab requires some enhancements but it uses Gemini model “gemini-1.5-flash-latest” for answering the queries.



Generate Reports tab

Generates report in html format but in future upgradation, pdf format is also available.



SUBMITTED TO: DATA INSIGHTS