

Unified Data Analysis Dashboard

Saptharishree M

Mtech integrated CSE with Business Analytics (2025)

VIT, Chennai

saptharishreemuthu@gmail.com

91+7708709518

LinkedIn: [saptharishree/](#)

1. OBJECTIVE:

To develop an application that can perform dynamic statistical analysis of CSV files and create dashboard , Use LLMs & the Prompt to answer questions based on the uploaded CSV, to LLM model and generate plots based on the results.

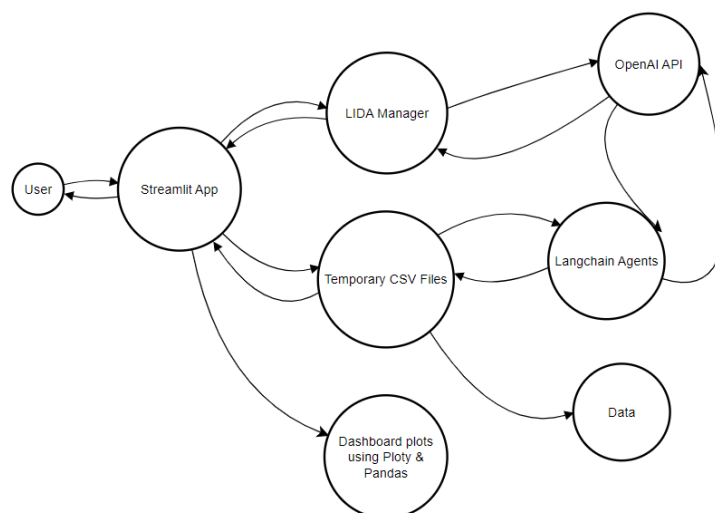
2. INTRODUCTION :

Streamlit-based Unified Data Analytics (UDA) is a web application with an easy-to-use interface that facilitates quick and thorough data analysis. Pandas is used by UDA to handle CSV files encoded with UTF-8. When the "Summarize" option is used, UDA uses ChatGPT Turbo and LIDA to generate goals and contextual summaries. The Seaborn library is used to generate visualizations, which are then supplied in base64 format and transformed into pictures. LIDA and ChatGPT Turbo handle user-defined queries for graph generation, generating pertinent visualizations.

Users can ask precise questions about the CSV data by utilizing the "Ask CSV" feature, which uses Langchain a framework to connect language models with external data sources is used to communicate with ChatGPT 3.5 Turbo as an agent. The "Dashboard" option performs internal data analysis—detecting categorical columns, numeric columns, and time stamps—using Plotly for interactive visualizations and Pandas for data segmentation and overview.

3. SYSTEM DESIGN

Architecture of UDA :



The architecture of Unified Data Analytics (UDA) consists of a Streamlit-based web application that facilitates user interaction and data processing. It integrates with various components such as LIDA Manager for summarization and goal-setting, and the OpenAI API

for advanced language model capabilities. The system manages CSV files, performs data analysis and visualization using libraries like Plotly and Seaborn, and allows users to query and generate insights through a comprehensive dashboard and interactive visualizations.

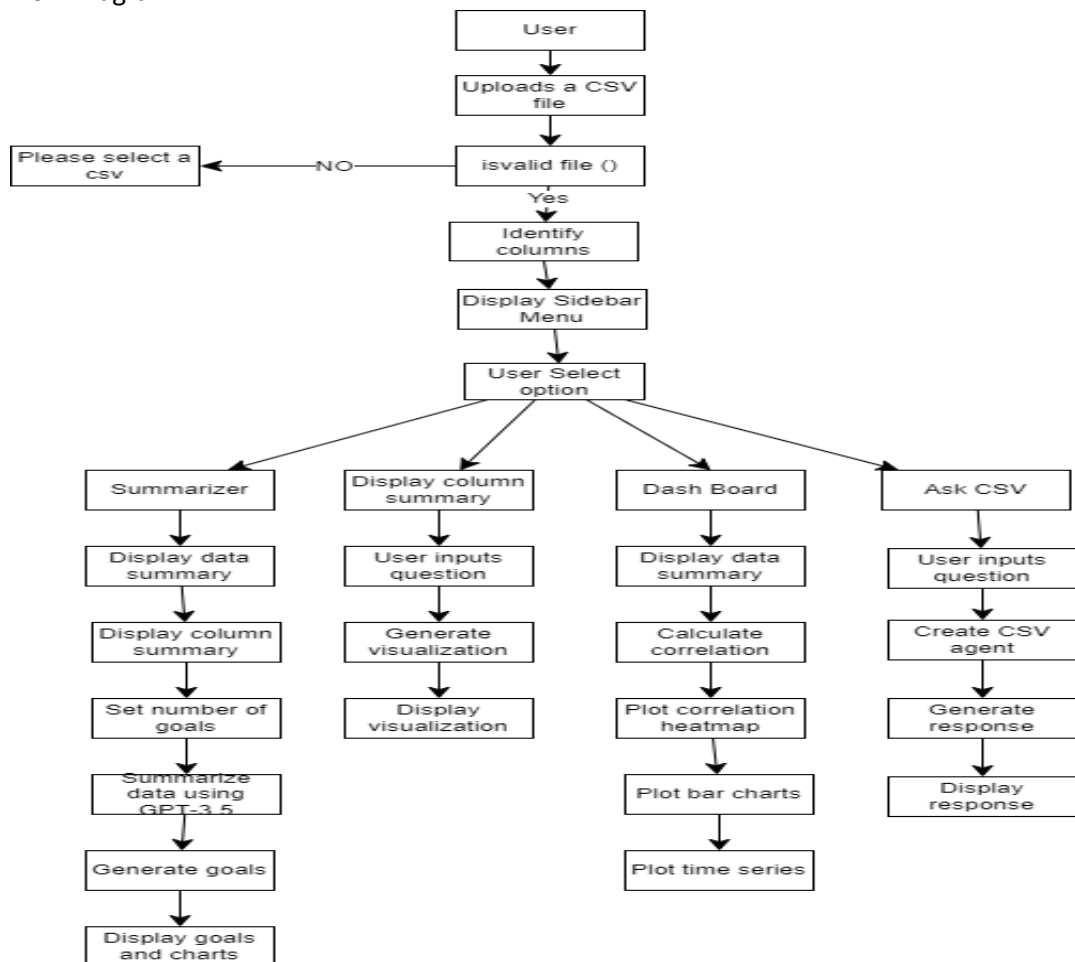
4. TECHNOLOGIES USED

- **Streamlit**: Used for building web interface.
- **Pandas**: Utilized for data manipulation and CSV file handling.
- **Plotly**: Provides interactive visualizations used in dashboards.
- **Seaborn**: Visualization library, here it is used with Plot Generation and LIDA .
- **GPT 3.5 Turbo** : LLM Backend of LIDA and AI Agent used in Ask CSV
- **LIDA Manager**: Assists in summarizing data and generating visualizations using GPT-3.5 Turbo.
- **Langchain**: Connecting CSV with LLM and Creating OpenAI Agent in Ask CSV .

5. IMPLEMENTATION

The Unified Data Analytics (UDA) application uses several key technologies. Streamlit creates a simple and interactive web interface. Pandas processes CSV files for analysis and visualization. Plotly and Seaborn generate various charts and plots to show data insights. LIDA Manager and GPT-3.5 Turbo are used for summarizing data and setting goals, while Langchain helps with interactive querying

Flow Diagram



The flow of UDA starts with uploading a CSV file, validating it, and proceeds through various features such as summarization, question-based graph generation, data dashboard, and querying with CSV agents. Each option leads to specific data analysis and visualization tasks .

6. FEATURES

1. Summarize

The Summarize menu provides an overview of the data with basic statistics and a column summary, and allows users to set goals for data summarization and visualization. It uses LIDA Manager and Seaborn to generate charts based on these goals.

2. Question based Graph

The Question Based Graph option displays a column summary and lets users input queries about the data, creating visualizations based on these questions with LIDA and GPT-3.5 Turbo.

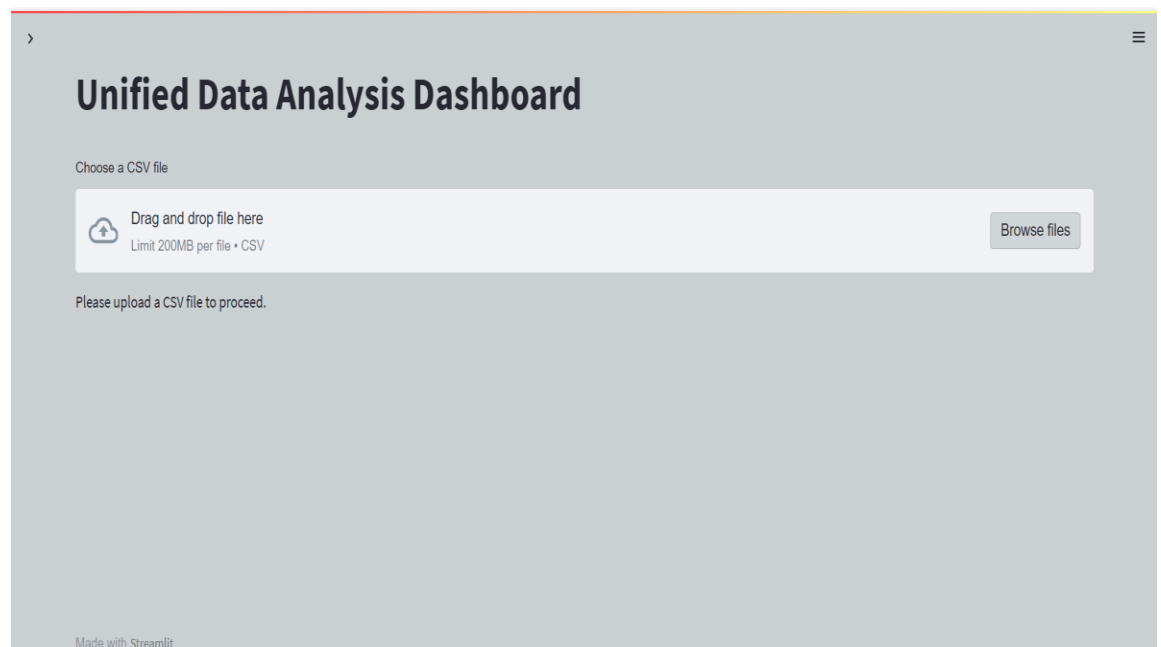
3. Data Dashboard

The Data Dashboard shows statistical summaries, plots correlation heatmaps for numeric data, creates bar charts for categorical data, generates time series plots, and decomposes time series data into its components.

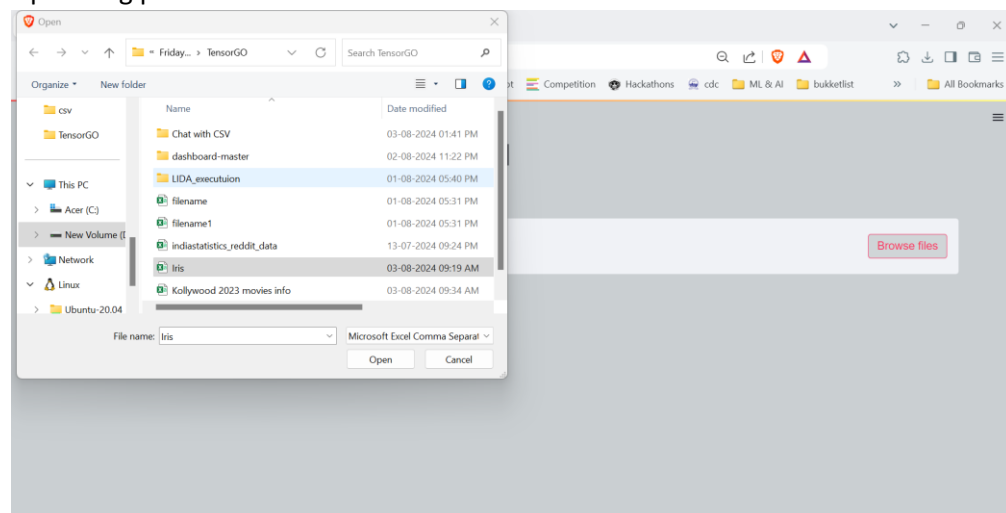
4. Ask CSV

The Ask CSV feature lets users input questions about the CSV file, using GPT-3.5 Turbo to generate and display responses.

7. DEMONSTRATION



- 1.
2. Uploading prominent iris data



3. Once you Upload , click on the side bar and input Goals .

Choose an Option

Summarize

Number of Goals

035

Choose a CSV file

Drag and drop file here

Limit 200MB per file • CSV

Browse files

Iris.csv

5.0KB

Summarization of your Data

Data Summary:

	Id	SepalLengthCm	SepalWidthCm
count	150	150.0000	150.00
mean	1970-01-01 00:00:00.0000000075	5.8433	3.05
min	1970-01-01 00:00:00.0000000001	4.3000	2.00
25%	1970-01-01 00:00:00.0000000038	5.1000	2.80
50%	1970-01-01 00:00:00.0000000075	5.8000	3.00
75%	1970-01-01 00:00:00.0000000112	6.4000	3.30

Uploaded Data Contains

	Column Name	Unique Values	Column Type
0	Id	150	Categorical
1	SepalLengthCm	35	Numeric
2	SepalWidthCm	23	Numeric
3	PetalLengthCm	43	Numeric
4	PetalWidthCm	22	Numeric

- 4.

Choose an Option

Summarize

Number of Goals

035

Goal 1: Goal(question="How do the Sepal Lengths vary across different species?", visualization="Bar chart showing average Sepal Length for each Species", rationale="By comparing the Sepal Lengths of different species, we can identify any potential patterns or differences that may exist between them.", index=0)

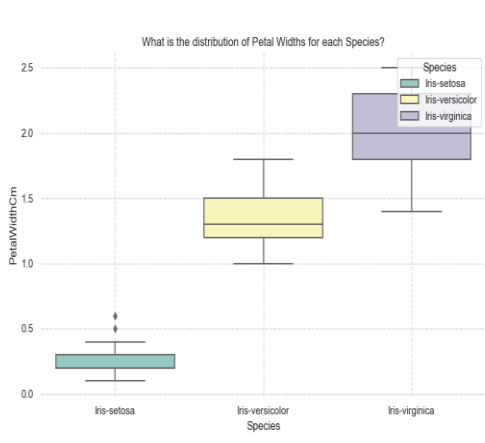
How do the Sepal Lengths vary across different species?

Mean Sepal Length: 5.84

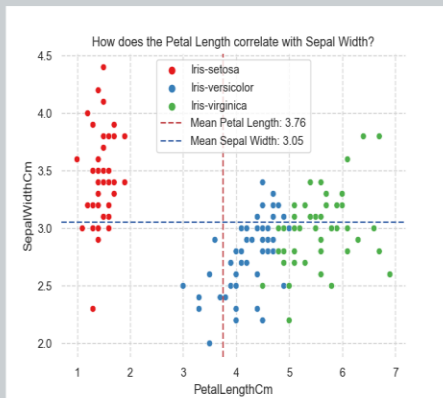


Goal 2: Goal(question="What is the distribution of Petal Widths for each Species?", visualization="Box plot of Petal Width grouped by Species", rationale="Examining the distribution of Petal Widths for each species can provide insights into the variability and spread of this feature within each category.", index=1)

What is the distribution of Petal Widths for each Species?



Goal 3: Goal(question="How does the Petal Length correlate with Sepal Width?"; visualization="Scatter plot of Petal Length against Sepal Width"; rationale="Understanding the relationship between Petal Length and Sepal Width can help in determining if there is any correlation or dependency between these two features."; index=2)



All these 3 Goals and Graphs were generated.

5. Option 2 Question based Graph : Enter a plot to be generated

Choose an Option
Question based Graph

Choose a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Iris.csv 5.0KB

Query your Data to Generate Graph

Column Name	Unique Values	Column Type
0 Id	150	Categorical
1 SepalLengthCm	35	Numeric
2 SepalWidthCm	23	Numeric
3 PetalLengthCm	43	Numeric
4 PetalWidthCm	22	Numeric
5 Species	3	Categorical

Query your Data to Generate Graph
Comparison on various species based on Petal length and width
Generate Graph

Choose an Option
Question based Graph

Generate Graph
Your Query: Comparison on various species based on Petal length and width

Comparison on various species based on Petal length and width

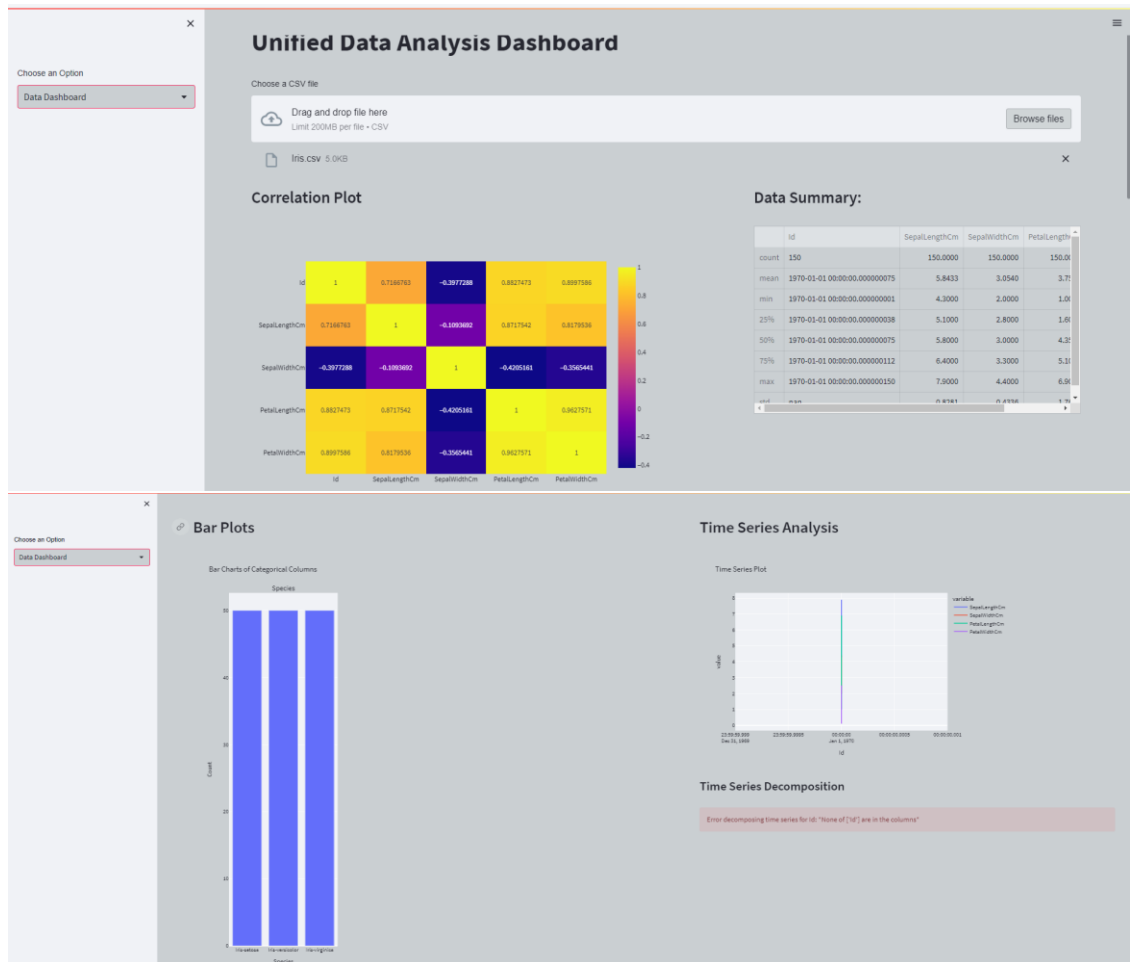
Comparison on various species based on Petal length and width

Legend:

- Iris-setosa (Red dots)
- Iris-versicolor (Blue dots)
- Iris-virginica (Green dots)
- Mean Petal Length: 3.76 (Blue dashed line)
- Mean Petal Width: 1.20 (Red dashed line)

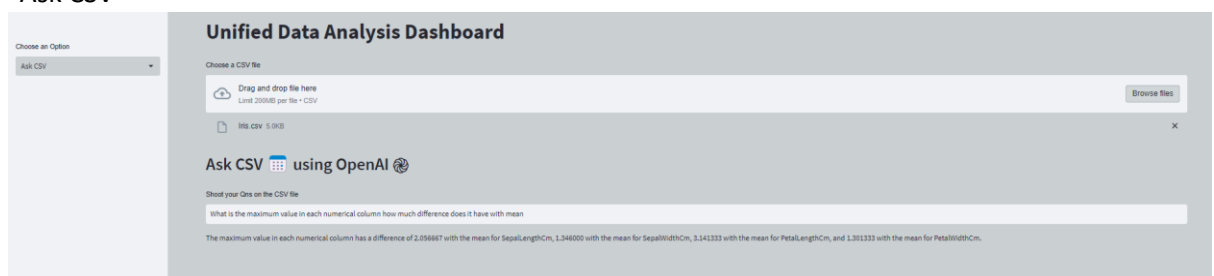
The x-axis is labeled 'Petal Length (cm)' and ranges from 1 to 7. The y-axis is labeled 'Petal Width (cm)' and ranges from 0.0 to 2.5.

6. Option 3 : Data Dashboard



Since the Data Doesn't have a Time series data ,It does not make a meaning full chart

7. "Ask CSV



REFERENCES

1. Query Your CSV using LIDA: Automatic Generation of Visualizations with LLMs. (2023). *YouTube*. Retrieved from <https://www.youtube.com/watch?v=U9K1Cu45nMQ>
2. LAMBDA - Multi-Agent Data Analysis System. (2023). *GitHub*. Retrieved from <https://github.com/Stephen-SMJ/LAMBDA>
3. Understanding LIDA: A Complete Setup to Data Visualization. (2023). *Stackademic Blog*. Retrieved from <https://blog.stackademic.com/understanding-lida-a-complete-setup-to-data-visualization-f81ffa1748e8>
4. Meta LLaMA 2 vs OpenAI GPT-4. (2023). *Medium*. Retrieved from <https://medium.com/@meetdianacheung/meta-llama-2-vs-openai-gpt-4-785589efe15e>

5. LIDA: A Comprehensive Overview. (2023). *arXiv*. Retrieved from <https://arxiv.org/abs/2303.02927>
6. LIDA GitHub Repository. (2023). *GitHub*. Retrieved from <https://github.com/microsoft/lida>
7. LangChain Explained in 13 Minutes | QuickStart Tutorial for Beginners. (2023). *YouTube*. Retrieved from <https://www.youtube.com/watch?v=aywZrzNaKjs>
8. Chat with CSV Streamlit Chatbot using Llama 2: All Open Source. (2023). *YouTube*. Retrieved from <https://www.youtube.com/watch?v=WB10mFa4T8>
9. How to Build a Dashboard Web App in Python with Streamlit. (2023). *YouTube*. Retrieved from <https://www.youtube.com/watch?v=fThcHGtOeQ>
10. LangChain Documentation. (2023). *LangChain*. Retrieved from <https://python.langchain.com/v0.2/docs/introduction/>
11. Streamlit Documentation. (2023). *Streamlit*. Retrieved from <https://docs.streamlit.io/>
12. OpenAI API Documentation. (2023). *OpenAI*. Retrieved from <https://platform.openai.com/docs/api-reference/introduction>