***AI Powered Data Analytics Chatbot***

***Project Description:***

This project is aimed at developing an AI-powered Data Analytics Chatbot which helps users, students to learn and apply Data Analytics and Data Science Techniques on various kinds of data to fetch the optimal and useful output in a matter of seconds by providing data relative prompts.

This bot is built using several powerful libraries such as Numpy, Pandas, Seaborn, Matplotlib, Scikit-Learn, Scipy, Plotly, Streamlit, Regex, and many more. It uses an LLM integration model (Gemini 1.5 Flash) using the model Google Generative AI module.

This bot will assist students in generating insights analytics performing statistical analysis visualizations and major understanding core ML functions.

***Tools and Technologies Used:***

To perform this project Jupyter Notebook cannot be used because of its limited integrations. Instead, tools like Visual Studio Code as well as PyCharm can be used to perform this project since it comes with various integrations

In this project, we will be moving forward with the **Visual Studio Code** as the IDE.

***Objective:***

To develop an **AI-integrated, interactive chatbot** that helps users perform Exploratory Data Analysis (EDA), generate visualizations, and build ML models through natural language queries.

***Modules Used for this Project:***

* Python Fundamentals
* Numpy
* Pandas
* Matplotlib
* Seaborn
* Scipy
* Statistics
* Scikit-Learn
* Plotly
* Google Generative AI
* Regex
* Base64
* IO
* PIL

***Key Features of This Project:***

* Data upload and handling – The users can upload datasets or files, and the bot will display the first few rows as a preview and generate insights based on the query provided
* EDA – Users can request statistics, visualizations, and several other Data Analysis tools or functions
* Statistical Analysis – Users can fetch descriptive statistics as well as testing statistics
* ML Assistance – The user can build or utilize basic ML models such as supervised models or unsupervised models. Also, it provides and utilizes model evaluation techniques such as accuracy classification report and many more
* Gemini Integration – If the user has some queries regarding the data, the Gemini API can be utilized as Gemini 1.5 Flash can be used to generate human like responses
* Query parsing with regex – This bot can recognize patterns structures provided by the user such as several data queries as well as statistical queries
* Visualizations - Plotly for interactive charts, Matplotlib/Seaborn for static charts (now viewable using PIL + Base64 workaround), Supports violin plot, histogram, boxplot, scatterplot, etc.

***Future Enhancements for this project:***

* Memory persistence via local JSON or database (e.g., SQLite).
* User authentication and role-based access.
* Enhanced query understanding via LangChain or semantic parsers.
* Autosuggestion of visualizations from data patterns.
* Export chat and analysis as PDF or Excel reports.
* Prompt Memory Saving as well as clear button.
* Voice Commands for Prompts.
* Advanced ML, DL Modules Integration

***Project File Structure:***

* Data Analytics Chatbot (Folder)
  + streamlit (Folder)
    - secrets.toml
  + assets (Folder)
  + static\_plots (Folder)
  + venv (Folder)
  + app.py
  + chatbot\_logic.py
  + gemini\_handler.py
  + utils.py
  + requirements.txt

***References***

* Python Documentation
* Streamlit Documentation
* Pandas Documentation
* Plotly Documentation
* Google AI Studio for API Key
* Gemini LLM (gemini.com)