

1. Executive Summary:

This project aims to develop an AI-powered web application that transforms raw data into actionable insights through natural language interaction. Built with Flask and powered by Groq-hosted Large Language Models (LLMs), it allows users to upload CSV or Excel files and ask analytical questions in plain English. The system automatically generates and executes the relevant Python (Pandas) code, displaying both the code and its results for full transparency. Designed for professionals and students alike, Analyza simplifies complex data exploration, making self-service analytics intuitive, conversational, and accessible, bridging the gap between data and decision-making.

2. Problem Statement:

In today's data-driven world, many users struggle to extract meaningful insights from datasets due to limited coding knowledge and complex analytical tools. Traditional data analysis methods require programming expertise, making it difficult for non-technical users to interpret and explore data effectively. To address this gap, there is a need for an AI-powered chatbot that enables users to analyse data through simple natural language conversations. Such a system should automatically generate and execute analytical code, visualise results, and deliver insights, making data analysis interactive, accessible, and efficient for everyone.

Background:

In today's data-driven corporate environment, organisations generate massive volumes of data across various departments, including finance, marketing, operations, and HR. However, the process of extracting actionable insights often requires technical expertise in programming and data analysis, creating a dependency on specialized data teams. This slows down decision-making and limits data accessibility for non-technical professionals. To bridge this gap, businesses are increasingly adopting AI-driven tools that simplify data analysis through natural language interactions, enabling faster and smarter business insights.

Objective:

The objective of this project is to develop Analyza, an AI-powered data analyst assistant that allows corporate users to perform complex data analysis through simple conversational queries. By leveraging Large Language Models (LLMs) and Flask-based integration, Analyza automates the process of generating and executing analytical Python (Pandas) code, thereby transforming natural language questions into data insights enhancing productivity, accuracy, and decision-making efficiency.

Scope:

This project involves the development of an interactive web-based chatbot interface that enables users to upload datasets (in CSV/Excel format), generate visual and statistical analyses, and interpret results in real-time. It covers features such as natural language query processing, automated code generation, and result visualization. The tool is designed for corporate data teams, business analysts, and decision-makers, empowering them to derive insights quickly without requiring coding expertise, ultimately promoting self-service analytics within organizations.

3. Data Sources:

Primary Data:

User-uploaded datasets in CSV or Excel format containing organizational or business data for analysis.

Secondary Data:

Public datasets (for testing and validation) sourced from platforms like Kaggle , used to evaluate the accuracy and versatility of Analyza's analytical capabilities.

4. Methodology:

Data Integration:

The system allows users to upload datasets, which are automatically read, validated, and prepared for analysis.

AI-Powered Query Interpretation:

Natural language queries entered by the user are processed through a Groq-hosted LLM, which translates them into executable Python (Pandas) code.

Execution and Output Generation:

The generated code is executed securely within the Flask backend, and both the code and results are displayed to the user for transparency.

User Interface Design:

An intuitive and minimalistic interface was developed using HTML, CSS, and Tailwind CSS, ensuring ease of use and smooth interaction for all user types.

5. Expected Outcomes:

- A functional AI-based chatbot capable of performing data analysis using natural language queries.
- Reduced dependency on technical analysts for routine data exploration.
- Faster decision-making through instant insights and visual results.
- Promotion of self-service analytics across corporate departments.

6. Tools and Technologies:

- **Flask** – Web framework for backend development
- **Python (Pandas, NumPy)** – Data analysis and computation
- **Groq-hosted LLM** – Natural language processing and code generation
- **HTML, CSS, Tailwind CSS** – Frontend design and user interface
- **GitHub** – Version control and project hosting
- **CSV/Excel datasets** – User-uploaded data for analysis

7. Risks and Challenges:

Handling user-uploaded datasets poses risks of data leakage or misuse. Ensuring secure storage, encryption, and compliance with data protection standards is essential.

The AI model may misinterpret natural language queries or generate incorrect analytical code, leading to inaccurate results and potentially misinformed decisions.

Processing large datasets and executing real-time analyses can create performance bottlenecks, requiring robust backend optimisation and scalable infrastructure.

Non-technical users may rely too heavily on AI-generated insights without validating results, which can introduce analytical errors or bias in business decisions.

8. Conclusion:

The Analyza – AI Data Analyst chatbot project provides organisations with an innovative and intelligent solution to simplify data analysis through natural language interaction. By integrating AI-powered automation with conversational analytics, it eliminates the need for manual coding and makes data-driven insights accessible to all business users. This tool enables faster decision-making, enhances productivity, and promotes a culture of self-service analytics within corporate environments. Ultimately, Analyza empowers teams to transform raw data into actionable insights efficiently, bridging the gap between technical complexity and strategic decision-making.

Name: [DEEP UTPAL DEY](#)

BATCH: [PGA-29\(THANE\)](#)