Tensorgo Assignment

CSV with LLM Analyzer



Submitted by :- Aditya Kumar [ 20MIP10008]

VIT Bhopal University

About project

This project is an **advanced AI CSV Analyzer** which leverage the power of LLM to understand **Natural Language** given by user and use that prompt to do several analysis on the given dataset. Users can ask any question to the LLM in the context of give dataset and LLM will extract important features related to the given prompt from dataset , hence by matching the desirability of prompt and generated answer ; LLM will generated most accurate answer.

Goals of this project

* To provide a User interface for CSV analysis using LLM
* To Plot useful and accurate charts with proper labels as per given prompt.
* To generate accurate answers asked in the context of given dataset.
* To act as companion of user in complex data analysis process.
* To summarize features of dataset.
* To find important patterns and relationships between features.

Libraries used

* **Streamlit**
* **Pandas**
* **PandasAI**
* **Torch**
* **PandasAI – SmartDataFrame**
* **PandasAI – QueryAgent**
* **os**

Components of project

* **Streamlit WebApp (FrontEnd)**
* **Bamboo LLM as primary LLM (Backend)**

Working of project

This project has primarily 2 major component Streamlit webapp from which user will be interacting and BambooLLM will be taking prompts from Webapp through an API Key , API key will help them to interact with each other remotely.

The user will upload it’s desired dataset on which he has to perform some visualization and statistical analysis. Now user will give prompts as in Natural Language keep in mind the goals of analysis of dataset, this prompt and dataset features will be transferred to BambooLLM will extract all import features from dataset by understanding the context & goal of prompt. Once LLM has understood the task it will generate a backend code using pandas and matplotlib (other visualization libraries are also supported) then this code will be running over original dataframe and eventually we will get our desired result, it could be generated text or Plots, it depends on clarity in prompt and dataset.

Architecture of project

**NLP + Feature Extraction**

**Query Agent**

Img/Text ouptut

Backend LLM Processing Engine

Bamboo LLM

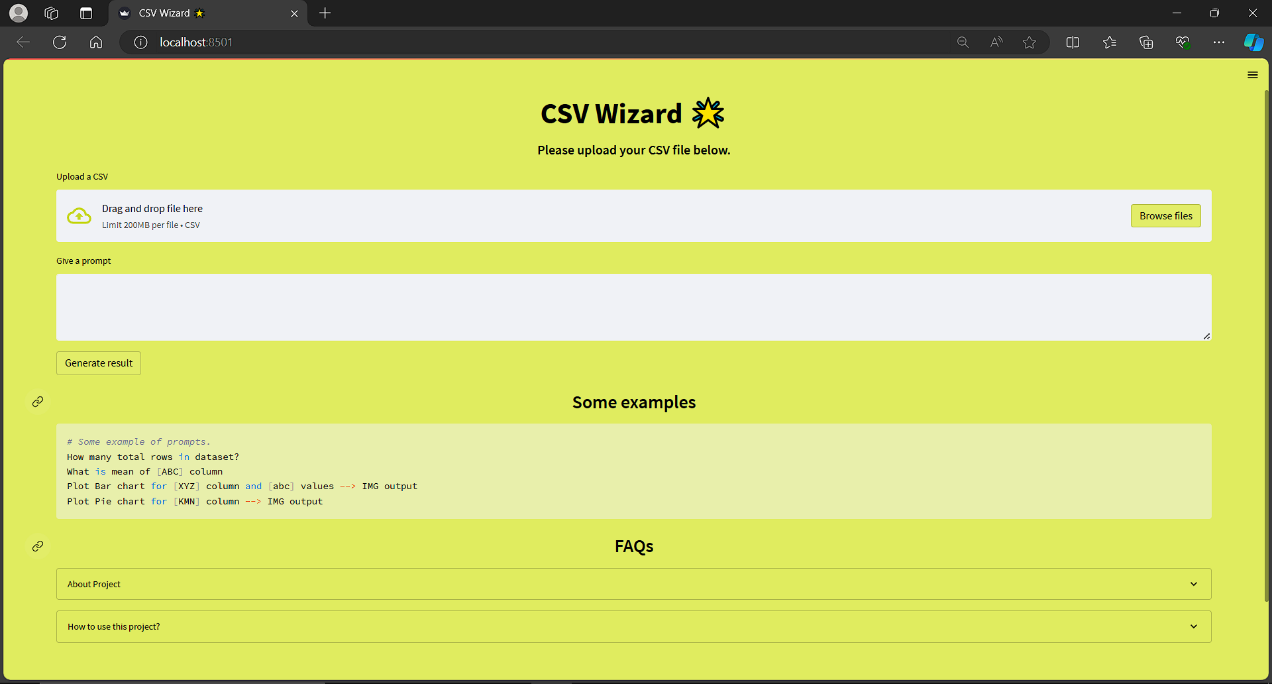
**WebAPP(Streamlit)**

**User**

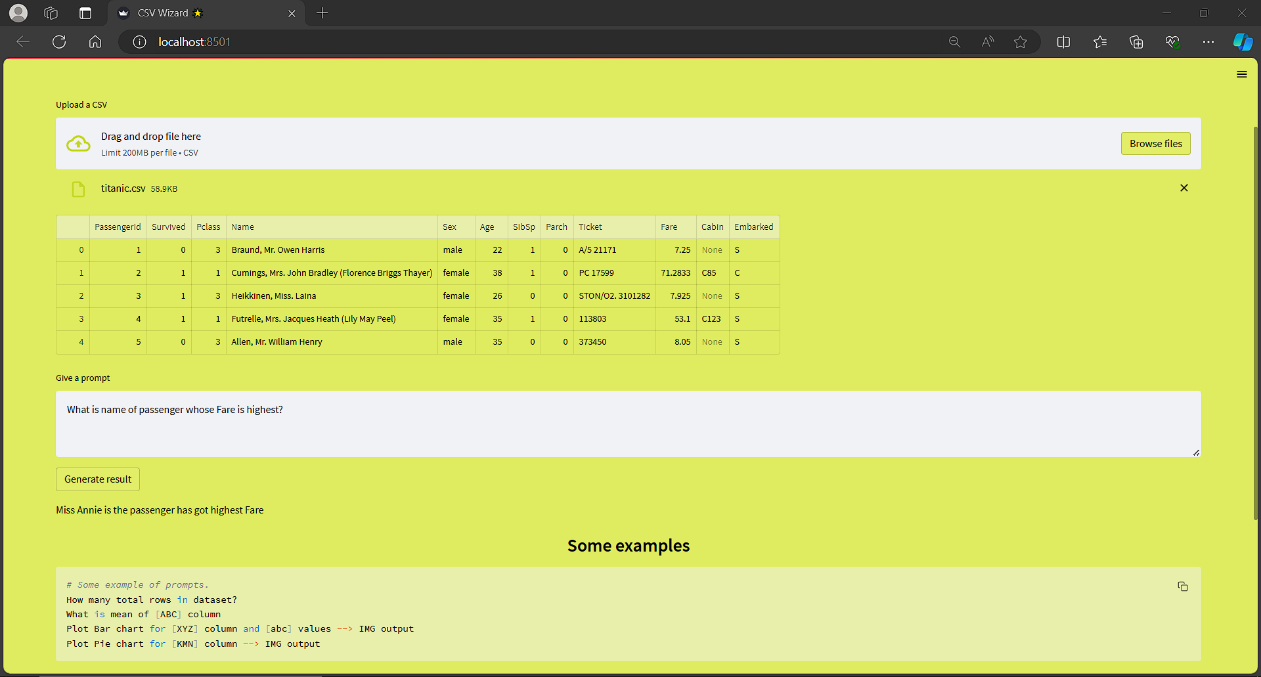
Prompt

Outputs

**Overall UI**

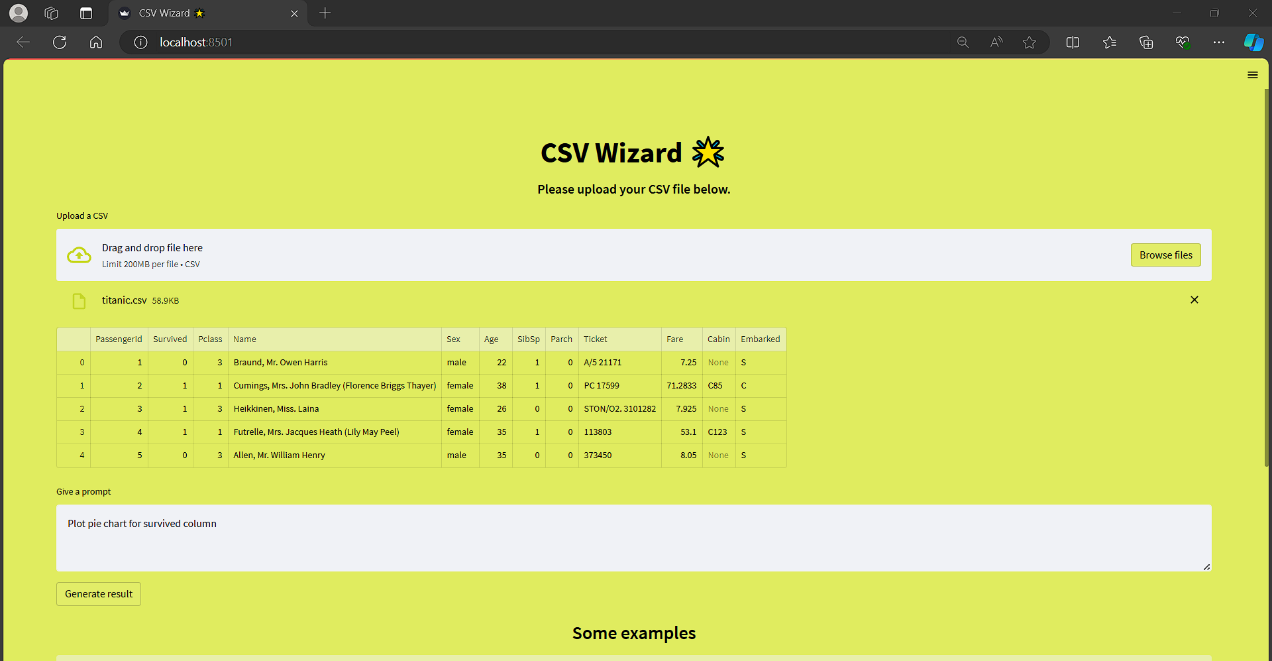


**Example : Prompt = What is the passenger name with highest Fare value?**

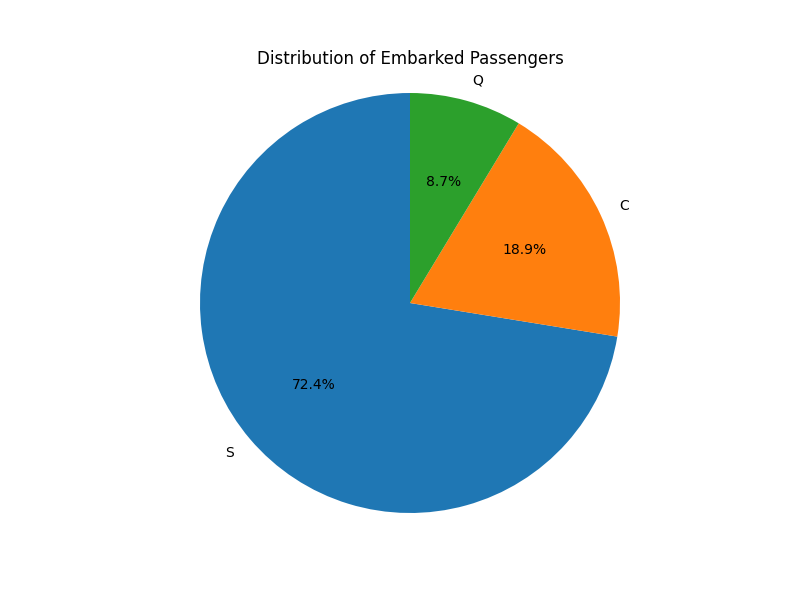
****

**Output : Miss Annie**

**Example 2: Prompt = Plot Pie chart for Survived column**

****

**Output Generated Chart:**

****

**THANK YOU!!**