

Approach for Data Analysis and LLM Interaction

The following approach outlines how to analyse data from a CSV file, perform statistical analysis, visualize the data, and interact with a language model using the langchain library and OpenAI.

1. Import Necessary Libraries

- Use libraries such as Pandas for data manipulation, Matplotlib for visualization, and OpenAI for language model interaction.

2. Read and Explore CSV Data

- Load the CSV file into a Pandas DataFrame to examine the structure and summarize the data.

3. Define Functions for Statistical Calculations

- Create a function to compute basic statistics like mean, median, mode, standard deviation, and correlation.

4. Define Functions for Data Visualization

- Implement functions to generate histograms, scatter plots, and line plots to visualize different aspects of the data.

5. Perform Statistical Analysis and Create Visualizations

- Compute statistics for the DataFrame and convert the results into a format suitable for the language model.
- Generate visualizations for specified columns to understand the data distribution and relationships.

6. Set Up OpenAI API Key

- Securely obtain and set the OpenAI API key for accessing the language model.

7. Configure and Utilize the Language Model

- Specify the language model to use and prepare a prompt containing the calculated statistics.
- Instantiate a language model object and invoke it with the prepared prompt to generate a response.

8. Interactive Questioning with the Language Model

- Allow for user interaction by enabling the user to input questions, which are then processed by the language model to provide contextual answers.