Colab AI Prompts for Data Analysis & Visualization

This guide provides useful prompts and tips for using Google Colab's AI assistant to interact with datasets. The goal is to help participants perform exploratory data analysis (EDA) and generate meaningful visualizations.

📊 Suggested Dataset: Student Performance

Dataset: `student_performance.csv`
Columns include:

- **`gender`**: Gender of the student (categorical: Male or Female).
- **`study_hours`**: Average number of hours the student studies per week (numerical).
- **`attendance_rate`**: Percentage of classes attended by the student (numerical, 0 to 100).
- **`quiz1`**: Score obtained in Quiz 1 (numerical, out of 10).
- **`quiz2`**: Score obtained in Quiz 2 (numerical, out of 10).
- **`assignment`**: Score for assignments submitted (numerical, out of 20).
- **`project`**: Score for the final project (numerical, out of 30).
- **`final_exam`**: Score in the final exam, influenced by `study_hours` and `attendance_rate` (numerical, out of 30).
- **`final_score`**: Total score combining quiz1, quiz2, assignment, project, and `final_exam` (numerical, out of 100).
- **`grade_category`**: Final letter grade based on `final_score` (categorical: A, B, C, D, F).

This dataset mimics academic data and is ideal for demonstrating basic analysis tasks.

🧠 Prompts for Using Colab AI Assistant

1. Load the Data

• Upload the file using the file upload widget in Colab.

• Prompt: 'Load the uploaded CSV file into a pandas DataFrame.'

2. Get Summary Statistics

• Prompt: 'Show the first five rows of the dataset.'

- Prompt: 'Display summary statistics for the numerical columns.'
- Prompt: 'Check for missing values in the dataset.'

3. Visualize Categorical Data

- “Plot a bar chart showing the number of students in each grade_category.”
- “Create a pie chart to visualize the distribution of students by gender.”
- “Group students by grade_category and show their count as a horizontal bar chart.”

4. Visualize Numerical Data

- “Create a histogram of the final_score to analyze overall grade distribution.”
- “Plot a box plot of study_hours across different grade_category values.”
- “Generate a scatter plot of attendance_rate vs final_score to examine correlation.”
- “Visualize how final_exam scores vary with study_hours using a scatter or line plot.”
- “Compare average project scores across grade_category using a bar chart.”