Python Data Analysis Assignment

Objective:
The purpose of this assignment is to evaluate your ability to manipulate and visualize data using Python libraries such as NumPy, pandas, and Seaborn. You will be required to perform both univariate and bivariate analyses on a dataset to extract meaningful insights.

Dataset:
For this assignment, you can use any publicly available dataset. As an example, you might consider the "Titanic" dataset or any other dataset with numeric and categorical variables. Ensure you have at least 1000 rows and 10 columns in the dataset.

Tasks:

1. Data Cleaning and Preliminary Analysis:

- Load the dataset into a pandas DataFrame.
- Handle any missing values.
- Provide a brief description of each column (datatype, number of unique values, mean, and standard deviation for numerical columns).

2. NumPy Tasks:

- Convert relevant columns into NumPy arrays.
- Compute the mean, median, and standard deviation for at least three numerical columns using NumPy.
- Find the correlation coefficient between two numerical columns using NumPy.

3. Univariate Analysis:

- Use Seaborn to plot the distribution of three numerical columns of your choice.
- Plot the count distribution of two categorical columns.

4. Bivariate Analysis:

- Plot a scatter plot between two numerical columns.
- Plot a boxplot showcasing the distribution of a numerical column against a categorical column.
- Compute and visualize the correlation matrix of the numerical columns.

5. Advanced Questions:

- For a chosen numerical column, fit a simple linear regression model to predict its values based on another numerical column. Comment on the fit of the model.
- Group the data by a categorical column and obtain the mean and standard deviation of all numerical columns for each group.

6. Bonus:

- Visualize a pair plot of the dataset (using Seaborn) to showcase relationships between all numerical columns.

- Generate a heat map of missing values in the dataset.

Submission:

- Submit a Jupyter notebook containing all your code, visualizations, and explanations. Notebook should be in Html format.
- Please share the dataset as well that you have used for the assignment.
- Ensure your code is well-commented, organized, and easily readable.
- Your explanations and insights derived from the analyses are crucial. Ensure you interpret each visualization and the results of your analyses.

Next Steps:

Once you submit the assignment and its found to be satisfactory, you will be invited for a 1-2-1 interview, where one section would be to discuss your assignment. Thanks!