TASK:

This assignment requires a certain amount of research and initiative, as not all tasks and subtasks are fully defined. You are expected to employ critical and creative thinking skills to complete this assignment in a comprehensive manner.

Tasks:

1. Refer to the assigned readings for examples on how to approach EDA in a professional manner.
2. Select a dataset from a reputable source (e.g., from Kaggle, UCI Machine Learning Repository) from the list of "Repositories for Finding Suitable Datasets," located in Class Resources, that is suitable for exploratory data analysis. The dataset should have at least 10 variables and 1,000 rows.
3. Conduct exploratory data analysis on the selected dataset using Python libraries (e.g., Pandas, Matplotlib, Seaborn). The analysis should include data cleaning, summary statistics, and visualizations to understand the data distribution and patterns.
4. Identify potential research questions or hypotheses that could be addressed through this analysis, and conduct correlation analysis to explore the relationships between variables. Use appropriate statistical techniques to measure the strength and direction of the relationship.
5. Fit a regression model to investigate the relationship between two or more variables in the dataset. Use appropriate techniques to validate the assumptions of the regression model, such as normality, linearity, and homoscedasticity.
6. Document the analysis process and findings in a report, including data visualizations, summary statistics, correlation analysis, and regression modeling. The report should provide clear and concise explanations of the analysis methods and results.