**Mini Assignment: Linear Regression Analysis**

**Objective:**

Your task is to identify a suitable dataset from either Kaggle or the UCI Machine Learning Repository that can be analyzed using linear regression. You will then conduct a preliminary analysis to understand the dataset better, focusing on the dependent and independent variables, and prepare your data for linear regression analysis.

**Instructions:**

**1. Dataset Selection:**

* **Source:** Choose a dataset from either Kaggle or the UCI Machine Learning Repository.
* **Criteria:** The dataset should have a mix of numerical and categorical variables and should be suitable for a linear regression analysis. Ensure the dataset has a clear potential dependent variable (the outcome you are trying to predict or understand) and multiple independent variables.

**2. Research Question:**

* **Description:** Clearly define the research question you are investigating with this dataset. What are you trying to predict or understand through linear regression analysis?
* **Objective:** This should guide your analysis and choice of independent variables.

**3. Data Exploration and Preprocessing:**

* **Variables Identification:** Identify and list all the independent variables in the dataset. For each variable, specify whether it is numerical or categorical.
* **Dependent Variable:** Clearly state the dependent variable in your dataset.
* **Data Cleaning:** Briefly describe any data cleaning steps you performed to prepare the dataset for analysis.

**4. Variable Encoding:**

* **Need for Encoding:** Discuss which categorical variables (if any) required encoding to be used in the linear regression model.
  + **One-Hot Encoding:** Identify which categorical variables you chose for one-hot encoding and explain why.
  + **Ordinal Encoding:** Identify any variables that were more suitable for ordinal encoding and justify your choice.

**5. Preliminary Analysis:**

* **Statistical Summary:** Provide a statistical summary of your independent variables. This should include mean, median, mode, standard deviation, or any other relevant statistics.
* **Visual Analysis:** Include at least one visual (e.g., scatter plot, histogram) that you think provides insight into your dataset or research question.

**6. Report Submission:**

* Format: Your report should be submitted as a PDF document.
* Length: The report should not exceed 3 pages, including any figures or tables.
* Components: Ensure your report includes the following:
  + Introduction to the research question and dataset
  + Description of the independent and dependent variables
  + Discussion on variable encoding
  + Preliminary data analysis (statistical summary and visual analysis)
  + Conclusion with potential insights or findings from your preliminary analysis

**7 Deadline:**

**Please submit your completed report by 4/9/2024.**