Part I: Research Question

A. Describe the purpose of this data analysis by doing the following:

1. Summarize **one** research question that is relevant to a real-world organizational situation captured in the data set you have selected and that you will answer using multiple linear regression in the initial model.
2. Define the goals of the data analysis.

Note: Ensure that your goals are within the scope of your research question and are represented in the available data.

Part II: Method Justification

B. Describe multiple linear regression methods by doing the following:

1. Summarize **four** assumptions of a multiple linear regression model.
2. Describe **two** benefits of using Python or R in support of various phases of the analysis.
3. Explain why multiple linear regression is an appropriate technique to use for analyzing the research question summarized in part I.

Part III: Data Preparation

C. Summarize the data preparation process for multiple linear regression analysis by doing the following:

1. Describe your data cleaning goals and the steps used to clean the data to achieve the goals that align with your research question including your annotated code.
2. Describe the dependent variable and *all* independent variables using summary statistics that are required to answer the research question, including a screenshot of the summary statistics output for each of these variables.
3. Generate univariate and bivariate visualizations of the distributions of the dependent and independent variables, including the dependent variable in your bivariate visualizations.
4. Describe your data transformation goals that align with your research question and the steps used to transform the data to achieve the goals, including the annotated code.
5. Provide the prepared data set as a CSV file.

Part IV: Model Comparison and Analysis

D. Compare an initial and a reduced linear regression model by doing the following:

1. Construct an initial multiple linear regression model from *all* independent variables that were identified in part C2.
2. Justify a statistically based feature selection procedure or a model evaluation metric to reduce the initial model in a way that aligns with the research question.
3. Provide a reduced linear regression model that follows the feature selection or model evaluation process in part D2, including a screenshot of the output for each model.

E. Analyze the data set using your reduced linear regression model by doing the following:

1. Explain your data analysis process by comparing the initial multiple linear regression model and reduced linear regression model, including the following element:
   - a model evaluation metric
2. Provide the output and *all* calculations of the analysis you performed, including the following elements for your reduced linear regression model:
   - a residual plot
   - the model's residual standard error
3. Provide an executable error-free copy of the code used to support the implementation of the linear regression models using a Python or R file.

Part V: Data Summary and Implications

F. Summarize your findings and assumptions by doing the following:

1. Discuss the results of your data analysis, including the following elements:
   - a regression equation for the reduced model
   - an interpretation of the coefficients of the reduced model
   - the statistical and practical significance of the reduced model
   - the limitations of the data analysis
2. Recommend a course of action based on your results.