Multiple Regression Assessment

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 regression.
2. Define the objectives or goals of the data analysis. Ensure that your objectives or goals are reasonable within the scope of the data dictionary and are represented in the available data.

Part II: Method Justification

B. Describe multiple regression methods by doing the following:

1. Summarize the assumptions of a multiple regression model.
2. Describe the benefits of using the tool(s) you have chosen (i.e., Python, R, or both) in support of various phases of the analysis.
3. Explain why multiple regression is an appropriate technique to analyze the research question summarized in Part I.

Part III: Data Preparation

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

1. Describe your data preparation goals and the data manipulations that will be used to achieve the goals.
2. Discuss the summary statistics, including the target variable and all predictor variables that you will need to gather from the data set to answer the research question.
3. Explain the steps used to prepare the data for the analysis, including the annotated code.
4. Generate univariate and bivariate visualizations of the distributions of variables in the cleaned data set. Include the target variable in your bivariate visualizations.
5. Provide a copy of the prepared data set.

Part IV: Model Comparison and Analysis

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

1. Construct an initial multiple regression model from all predictors that were identified in Part C2.

2. Justify a statistically based variable selection procedure and a model evaluation metric to reduce the initial model in a way that aligns with the research question.

3. Provide a reduced multiple regression model that includes both categorical and continuous variables.

> Note: The output should include a screenshot of each model.

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

1. Explain your data analysis process by comparing the initial and reduced multiple regression models, including the following elements:

   • the logic of the variable selection technique

   • the model evaluation metric

   • a residual plot

2. Provide the output and any calculations of the analysis you performed, including the model's residual error. (**Note**: The output should include the predictions from the refined model you used to perform the analysis.)

3. Provide the code used to support the implementation of the multiple regression models.

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 coefficients of the statistically significant variables of the model

   • the statistical and practical significance of the model

   • the limitations of the data analysis

2. Recommend a course of action based on your results.