Sr. Analyst BI and Strategic Data Insights Skill Assessment:

Summary:

This is a quick test that will determine your abilities in a programming language (choose any high level language you prefer, R, Python, Julia, etc, VBA is not allowed), and your ability to learn programming/mathematical methods quickly. Below is attached a dataset link on Kaggle, and instructions on what kind of model I would like to see built and its respective output. If you get stuck on a section do as much as you can, and in the code comment out what you're trying to do and how you would complete the program had you not gotten stuck. Use every tool at your disposal here, Google, textbooks, even the code on Kaggle already written. A work environment has no limits as to where you can gather information and this is meant to simulate that, just know that if you copy someone else's code you should be able to explain it. Reach out to me at nicholasbhandari@pvh.com if you have any questions and good luck!

Data:

https://www.kaggle.com/datasets/uciml/autompg-dataset

The data here is a mpg dataset, I'd like you to only use the numerical variables and not worry about the character variable at the end (car type).

Steps:

Follow each of the steps listed and paste the code relevant to each section beneath the section description.

Step 1: *Write code to pull the csv of your data into the program.*

Step 2: Split your data into training and testing sets (use a 70/30 split).

Step 3: Using your training data run a linear regression using mpg as your predicted variable and cylinders, displacement, horsepower, weight, and acceleration as your predictor variables.

Step 4: Using your testing data, calculate the mean squared error (MSE) of your model. As a bonus perform 10 fold cross validation if you can.

Step 5 (bonus): Transform mpg into a binary variable (splitting at the median to define "high" mpg and "low" mpg as 1, 0). Perform steps 1-4 again but instead of MSE use prediction accuracy as your performance metric and logistic regression as your model choice.