**DSI -06 Homework 2:**

Introduction with Statistical Learning with Applications in Python (ISLP) pg. 129-130

9. This question involves the use of multiple linear regression on the Auto data set.

1. Produce a scatterplot matrix which includes all of the variables in the data set.
2. Compute the matrix of correlations between the variables using the DataFrame.corr() method.
3. Use the sm.OLS() function to perform a multiple linear regression with mpg as the response and all other variables except name as the predictors. Use the summarize() function to print the results. Comment on the output. For instance:
4. Is there a relationship between the predictors and the response? Use the anova\_lm() function from statsmodels to answer this question.
5. Which predictors appear to have a statistically significant relationship to the response?
6. What does the coefficient for the year variable suggest?
7. Produce some of diagnostic plots of the linear regression fit as described in the lab. Comment on any problems you see with the fit. Do the residual plots suggest any unusually large outliers? Does the leverage plot identify any observations with unusually high leverage?
8. Fit some models with interactions as described in the lab. Do any interactions appear to be statistically significant?
9. Try a few different transformations of the variables, such as . Comment on your findings.

Additional Practice Questions:

Explain the difference between MLR and OLS to a non-technical audience.

How would you describe this exercise in an interview to both a technical and non-technical interviewer? What are the key insights you would want to show?

Can you think of a business context where this exercise would have applications?