Q1: To answer the assignment, use hw_4.csv.

The data set consists of the sales of a product in 200 different markets, along with advertising budgets (in $1,000) for the product in each of those markets for three different media: TV, radio, and newspaper.

The goal of the assignment is to develop an accurate model that can be used to predict sales on the basis of the three media budgets.

   a) Do exploratory analysis to understand your data.
   b) Split data into test and train. Build a model to predict sales based on TV budget. Assess linear regression assumptions of the model. Based on your model, how much increase in sales of the product do you for expect every $1,000 increase in the TV advertising budget?
   c) Use all features to build a multiple linear regression model. Assess the assumptions of your model
   .
   d) Add interaction effect to the model in part c. Does it improve your model? How do you interpret interaction effect in your model?
   e) Use the best model you have developed to predict sales on test data set. Report MAE and MAPE. Plot actual (y) vs. prediction (y_hat).

Q2: The Gunderson Plant manufactures the industrial product line of FGT Industries. Plant management wants to be able to get a good, yet quick, estimate of the manufacturing overhead costs that can be expected each month. The easiest and simplest method to accomplish this task is to develop a flexible budget formula for the manufacturing overhead costs. The plant's accounting staff has suggested that simple linear regression be used to determine the behavior pattern of the overhead costs. The regression data can provide the basis for the flexible budget formula. Sufficient evidence is available to conclude that manufacturing overhead costs vary with direct labor hours. The actual direct labor hours and the corresponding manufacturing overhead costs for each month of the last three years have been used in the linear regression analysis.

The three-year period contained various occurrences not uncommon to many businesses. During the first year, production was severely curtailed during two months due to wildcat strikes. In the second year, production was reduced in one month because of material shortages, and increased significantly (scheduled overtime) during two months to meet the units required for a one-time sales order. At the end of the second year, employee benefits were raised significantly as the result of a labor agreement. Production during the third year was not affected by any special circumstances.

Various members of Gunderson's accounting staff raised some issues regarding the historical data collected for the regression analysis. These issues were as follows.

   • Some members of the accounting staff believed that the use of data from all 36 months would provide a more accurate portrayal of the cost behavior. While they recognized that any of the monthly data could include efficiencies and inefficiencies, they believed these efficiencies and inefficiencies would tend to balance out over a longer period of time.
   • Other members of the accounting staff suggested that only those months that were considered normal should be used so that the regression would not be distorted. Still other members felt that only the most recent 12 months should be used because they were the most current.

- Some members questioned whether historical data should be used at all to form the basis for a flexible budget formula.

The accounting department ran two regression analyses of the data—one using the data from all 36 months and the other using only the data from the last 12 months. The information derived from the two linear regressions is shown below (t-values shown in parentheses). The 36-month regression is:

$$OH_t = 123,810 + 1.60\,DLH_t, \quad R^2 = 0.32$$
$$\scriptstyle(1.64)$$

The 12-month regression is:

$$OH_t = 109,020 + 3.00\,DLH_t, \quad R^2 = 0.48$$
$$\scriptstyle(3.01)$$

a) Which of the two results (12 months versus 36 months) would you use as a basis for the flexible budget formula?

b) How would the four specific issues raised by the members of Gunderson's accounting staff influence your willingness to use the results of the statistical analyses as the basis for the flexible budget formula? Explain your answer.