Project: Diamond Prices

Complete each section. When you are ready, save your file as a PDF document and submit it in your classroom.

Step 1: Understanding the Model

Answer the following questions:

1. According to the model, if a diamond is 1 carat heavier than another with the same cut, how much more should I expect to pay? Why?

    - ✓ Considering the linear regression model equation, (**Price** = -5,269 + 8,413 x **Carat** + 158.1 x **Cut** + 454 x **Clarity**), the one additional carat would result in an additional $8,413 in price. The formula created by the regression determined that the coefficient for a carat is 8,413, so for every increase in the number of carat the price will increase by the amount of the coefficient.

2. If you were interested in a 1.5 carat diamond with a **Very Good** cut (represented by a 3 in the model) and a **VS2** clarity rating (represented by a 5 in the model), how much would the model predict you should pay for it?

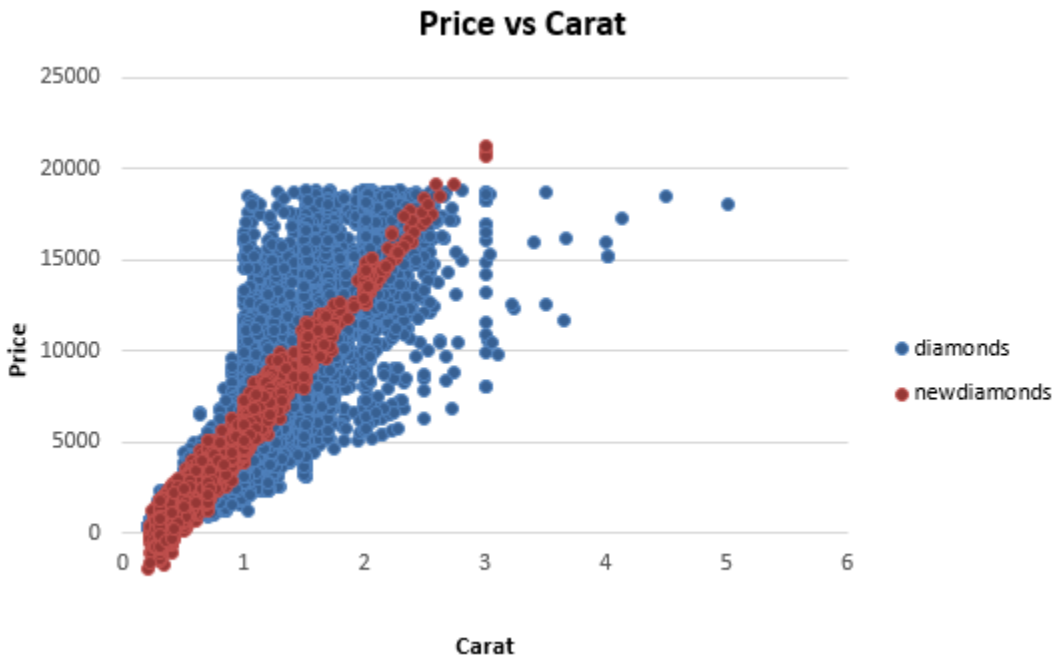
    - ✓ Considering the linear regression model equation, (**Price** = -5,269 + 8,413 x **Carat** + 158.1 x **Cut** + 454 x **Clarity**), by plugging in the values for different variables we get:
      Price = -5,269 + 8,413 x **1.5** + 158.1 x **3** + 454 x **5**
      Price = $10094.8

Step 2: Visualize the Data

Make sure to plot and include the visualizations in this report. For example, you can create graphs in Excel and copy and paste the graphs into this Word document.

1. Plot 1 - Plot the data for the diamonds in the database, with carat on the x-axis and price on the y-axis.
2. Plot 2 - Plot the data for the diamonds for which you are predicting prices with carat on the x-axis and predicted price on the y-axis.

- **Note**: You can also plot both sets of data on the same chart in different colors.



3. What strikes you about this comparison? After seeing this plot, do you feel confident in the model's ability to predict prices?

Some observations I can make about the model:

- ✓ For low carat values it predicts negative prices (which does not make sense)
- ✓ The new diamond datasets has a strong positive correlation with carat
- ✓ The price dispersion is much higher with the real data
- ✓ Considering it predicts negative prices, if it was for me, I would probably test other models.

Step 3: Make a Recommendation

Answer the following questions:

1. What price do you recommend the jewelry company to bid? Please explain how you arrived at that number.

- ✓ If I would take this model in consideration, I would recommend a bidding of **$8,213,465.93** Summing up all the prices of 3000 diamonds in the new_diamonds.csv, we get 11,733,522.76 Since this is the consumer price,

taking 70% of the sum and that is the amount a company should bid for diamonds. (8,213,465.93)

Bid sum = 0.7 * 11,733,522.76 = **8,213,465.93**