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?
    - If a diamond is 1 carat heavier than the other with same cut, one can expect to pay an extra amount of 8,413. It is evident from the formula created by the regression obtained from the model, that the co-efficient of the variable Carat is 8,413, indicating the increase in price for increase in each carat

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?
    - The regression formula is given by

$$\boldsymbol{Price} = -5{,}269 + 8{,}413\, x\, \boldsymbol{Carat} + 158.1\, x\, \boldsymbol{Cut} + 454\, x\, \boldsymbol{Clarity}$$

    - Plugging in the given values for $Carat = 1.5; Cut = 3; Clarity = 5$

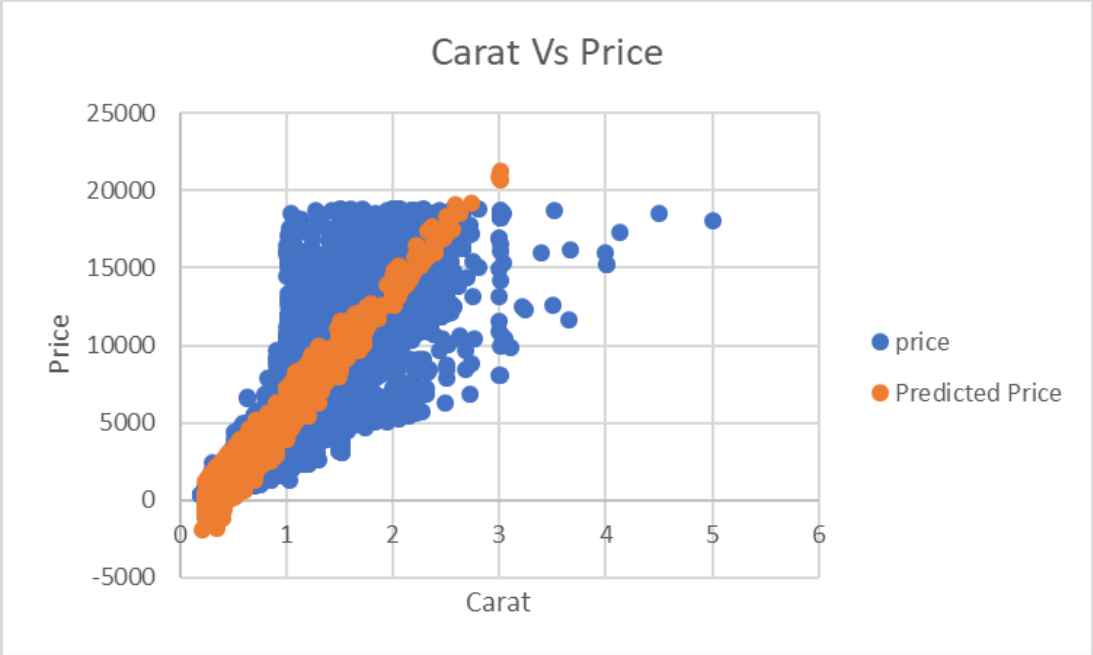
$$\boldsymbol{Price} = -5{,}269 + 8{,}413\, x\, \mathbf{1.5} + 158.1\, x\, \mathbf{3} + 454\, x\, \mathbf{5} = \$\,\mathbf{10{,}094.8}$$

This is the price the model would predict for given attribute values

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?



The predicted price has more linear relationship with Carat than the actual Price. This is because we used a linear regression model, and also the carat attribute contributes more to the pricing as we see from the regression formula. There might be other factors influencing the pricing of diamonds such as the current market condition and inflation in the money value, $ in US.

We also see some the predicted price going below the zero line or negative pricing, which is logically wrong and this is a drawback of using the prediction model. That is the model accuracy reduces for Carat values less than 0.5, while it improves as the carat value increases.

We see that the actual price varies broadly in the 1-3 Carat range, but our model predicts the price on average (It is splitting through a half line), which is desirable.

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.
   Recommended Price to bid = $ 8,213,465.932
   - The total amount of the whole lot as predicted by the model is $ 11733522.76, arrived by using SUM formula in excel. Considering that the company buys diamonds at 70% of the total retail cost. The recommended buy price is $0.7 \times 11733522.76 = \$\ 8213465.932$