Project: Diamond Prices

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?

   The expected to pay for a diamond is 1 carat heavier than another but they have the same cut, the price will increase by $ 8,413. The coefficient for the carat in the formula that we created by the regression model is equal to 8,413, so for every increase in the carat weight, 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?

   The linear regression model equation is:

   Price = -5,269 + 8,413***Carat*** + 158.1***Cut*** + 454* **Clarity**

   so now we will plug in the values for the different variables.

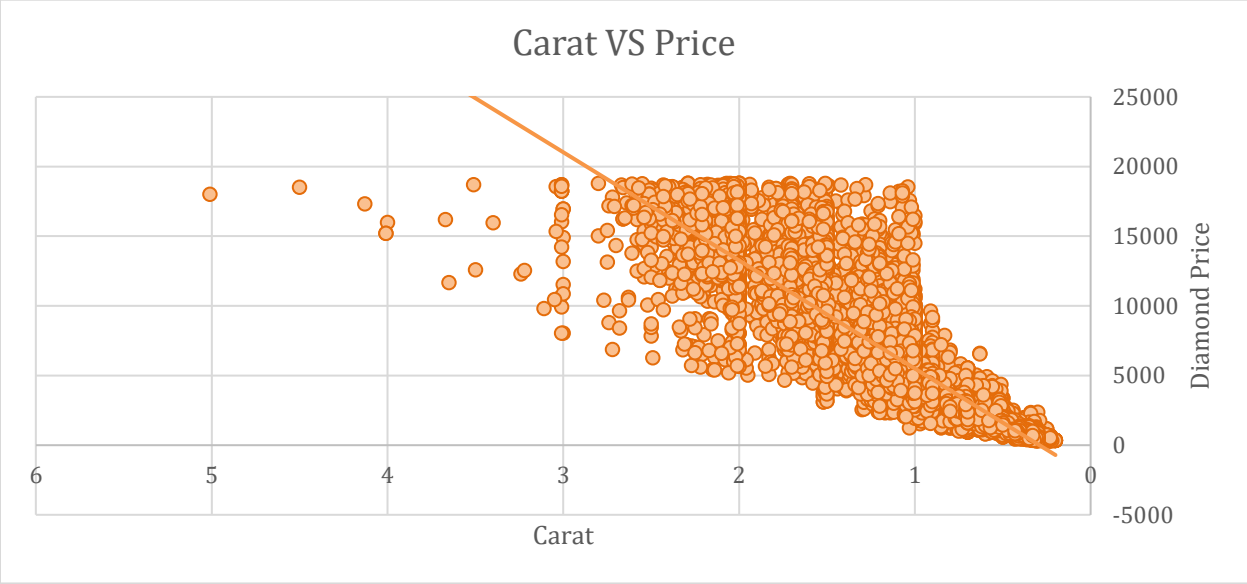
   Price = -5,269 + 8,413***1.5*** + 158.1***3*** + 454* **5**

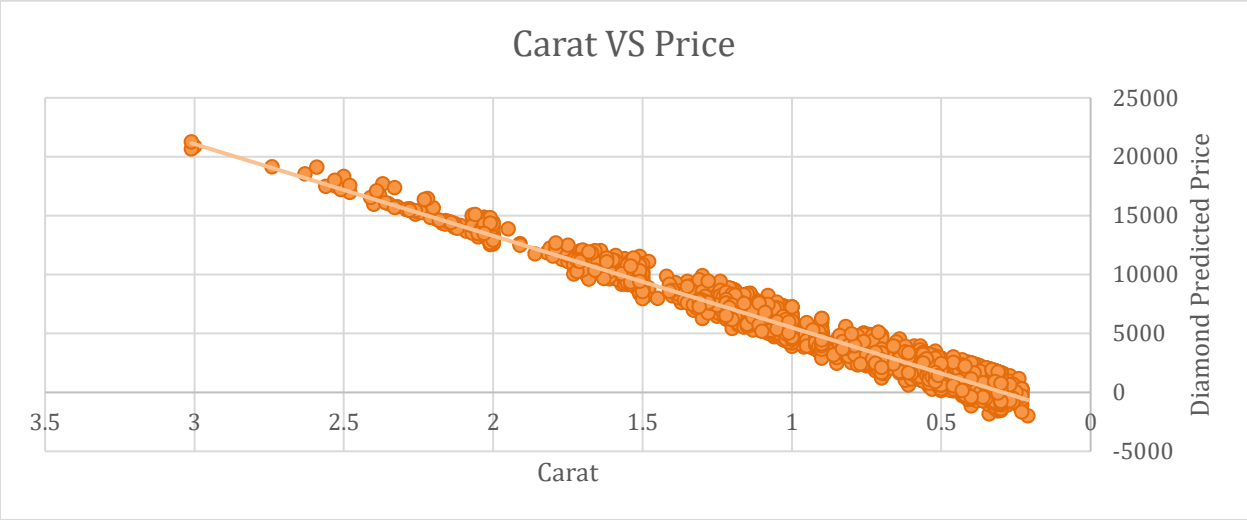
   Price = $10,094.8

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?





I calculated the correlation between the Predicted Price and Carat to validate and measure the quality of the result of the model. In this model, the value of r is (0.980), which is means there is a strong correlation between Predicted Price and Carat. Then I calculated the R-squared value (0.961), to ensure that the data fits the line and it shows that it is close to 1 which means nearly all variance in the Predicted Price is explained by the model. I also noticed that there is a positive relationship between Predicted Price and Carat. The higher the carat, the higher the price. Otherwise, I noticed that some predicted prices are negative numbers, which means that the linear regression model may not fit some prices. In addition to that, the carat alone may not affect the price of the diamond, as there are other factors such as the quality of the cut of the diamond and the internal purity of the diamond that is taken into account.

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.

   I recommend a bid of $8,213,465.932. I arrived at this number by using a formula from the regression model provided to estimate the predicted price for the 3,000 diamonds in New-Diamonds. Then I did a regression for Carat and Predicted Price data in New-Diamond. After that, I summed all Predicted Price which is equal to $11,733,522.76 then I multiplied it by 70% to get the final predicted bid of $8,213,465.932