<u>Project: Diamond Prices</u>

# Step 1: Understanding the Model

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 diamond with one additional carat would result in an additional **$8,413** in price. The formula created by the regression determined that the coefficient for 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?

   ○ The formula is Price = -5,269 + 8,413 x Carat + 4545 x Cut + 454 x Clarity.
     • so now we will plug in the values for the different variables.
     • Price = -5,269 + 8,413 x 1.5 + 158.1 x 3 + 454 x 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.

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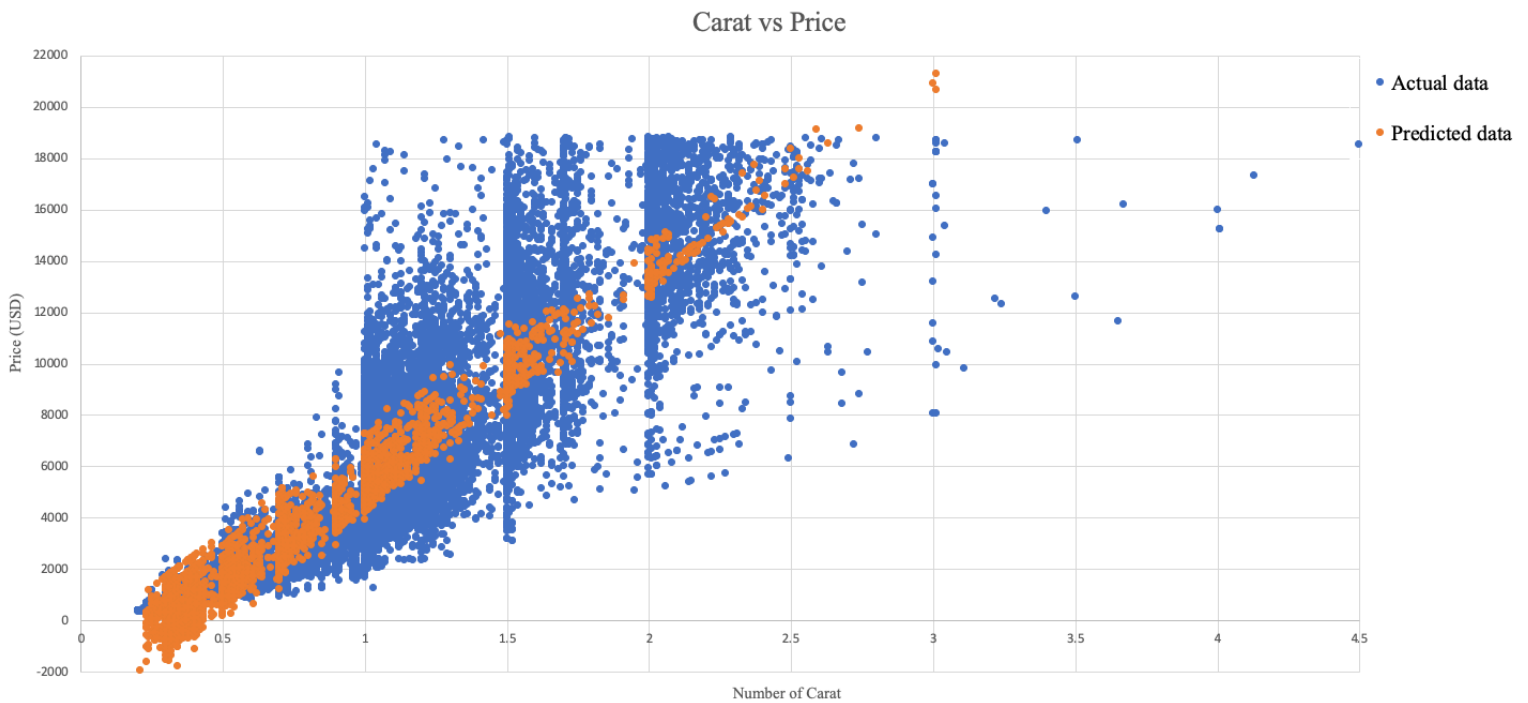
3. What strikes you about this comparison? After seeing this plot, do you feel confident in the model's ability to predict prices?

This comparison strikes me that the graph depicts big data in one picture with illustration of particular patterns which are almost impossible to see in the original table. Despite the fact that we didn't include other characteristics which we have, we can see the dense vertical areas separating the diamonds' prices with different numbers of carat. One more advantage of this plot is to consume the information given in a short period of time.

Analyzing the plot I was a little bit confused by negative numbers after my calculations which are supposed to be prices, but the average price in both data sets seem similar (average actual price = 3,939.1035, average predicted price = 3,911.1743). This thoughts lead me to the assumption that the formula works well on big data and cannot be reliable to estimate the particular price of an object or a small group of objects.

# 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 **$11,733,522.9**. I arrived at this number by using a formula from the regression model provided that was based on previous diamonds sales (Price = -5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity) and applied it to the diamonds that the jewelry company up for bid. I calculated the average predicted price by using an excel standard function for the set of predicted prices and got 3,911.1743. My next step was multiplying the average predicted price (3,911.1743) by the number of diamonds that the jewelry company up for bid (3,000) to get the final predicted bid of **$11,733,522.9.**