

Project: Diamond Prices

Complete each section. When you are ready, save your file as a PDF document and submit it here: <https://classroom.udacity.com/nanodegrees/nd008/parts/235a5408-0604-4871-8433-a6d670e37bbf/project#>

Step 1: Understanding the Model

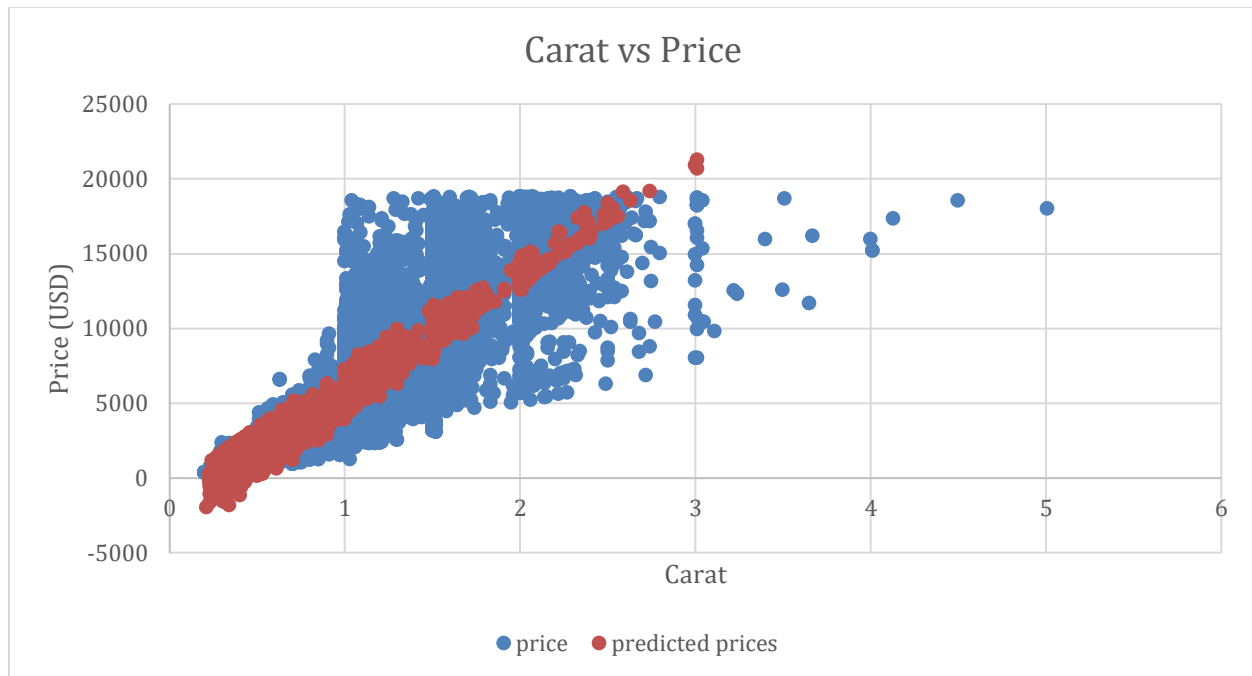
Answer the following questions:

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 one additional carat would result in an additional \$8413 in price. The formula created by the regression determined that the coefficient for a carat is \$8413, so for every increase in the number of carats 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 $\text{price} = -5269 + 8413 * \text{carat} + 158.1 * \text{cut} + 454 * \text{clarity}$
 - so now we will plug in the values for the different variables.
 - $\text{Price} = -5269 + 8413 * 1.5 + 158.1 * 3 + 454 * 5$
 - $\text{Price} = 10094.80$

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 prices are more linear and compact than the actual data is. This is because we are not accounting for everything that affects prices. There are many more things than just carats that affect it. We had cut and clarity factored into our formula but not even that will account for all the variation. For instance, this formula might look very different depending on where the data is coming from (ex. Location of sellers, jewelry store and reputation, online or in-store).

After looking at this plot the model appears on average to predict the prices ok, but it can be very off for certain diamonds. There appears to be a couple diamonds of 3 carats but sold for only \$8000 approximately. Also, there seems to be a few diamonds less than 1 carat sold for almost \$19000. The formula gives negative prices for some diamonds less than a half carat; obviously this will never happen. While the formula may not be completely accurate for an individual diamond, it should do a decent job at predicting the price for a set of diamonds since it on average looks representative.

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 \$8,213,465.93. I arrived at this number by using a formula from the regression model provided that was based on the previous diamond sales and applied it to the diamonds that were up for bid. I then factored in the percentage of the full price that the company generally purchases diamonds from distributors for which was 70%, so I multiplied the amount 11,733,522.76 by .80 to get the final predicted bid of \$8,213,465.93.