<u>Project: Diamond Prices</u>

Complete each section. Then save your file as a PDF document and submit it here: https://classroom.udacity.com/nanodegrees/nd008/parts/235a5408-0604-4871-8433-a6d670e37bbf/project# by Ursula Perez Ramirez

Step 1: Understanding the Model

Answer the following questions:

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

   The formula created by the linear regression model is: **Price** = -5,269 + 8,413 x **Carat** + 158.1 x **Cut** + 454 x **Clarity.** The coefficient for Carat is 8,413, therefore if cut and clarity are the same in the two diamonds, and the first element is a constant, one carat heavier will increase the price in $8,413. That is, for each additional carat, holding the rest of properties equal, the price will increase by the amount of the Carat 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** + 158.1 x **Cut** + 454 x **Clarity.** So, entering the value of the corresponding coefficients in the formula we obtain the price for the specified diamond.
   The predicted price is Price = -5,269 + 8,413 x **1.5** + 158.1 x **3** + 454 x **5 = $10,094.80.**

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.
   o **Note**: You can also plot both sets of data on the same chart in different colors.
   Figure 1 shows the scatter plots (plot 1 in blue and plot 2 in orange) of the representation of number of diamond carats vs diamond price.
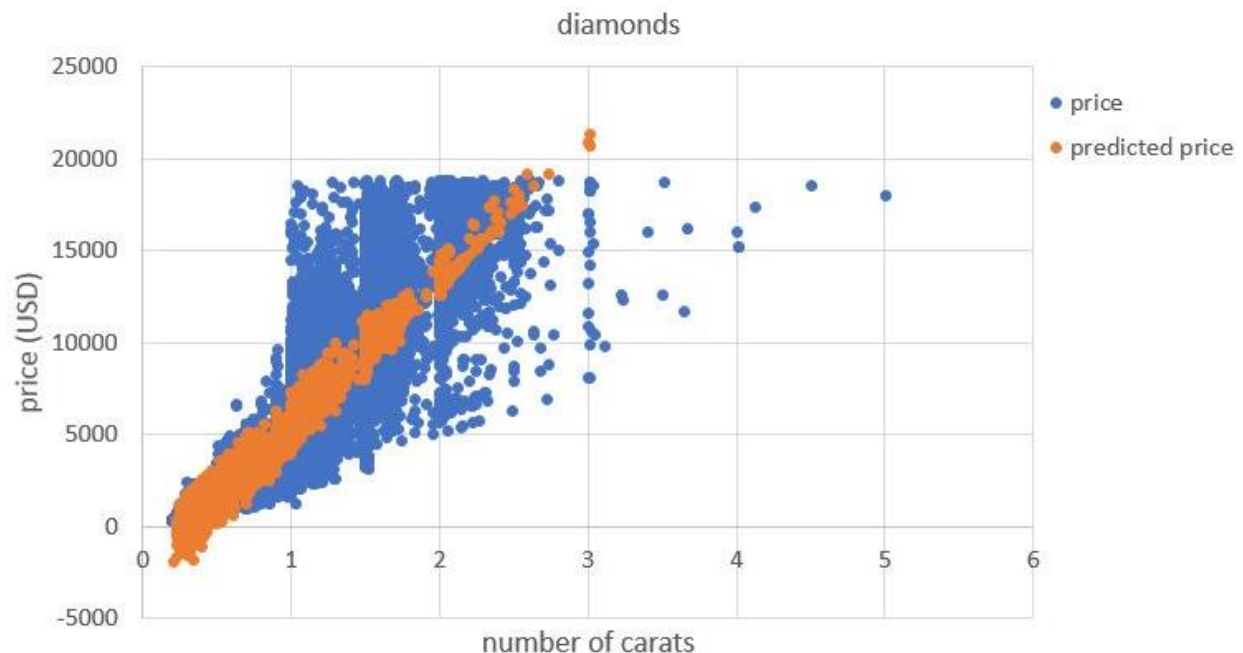
Figure 1. Scatter plot of the price for 50,000 diamonds and the predicted prices for 3,000 diamonds that the company is willing to purchase, according to their number of carats.

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

Figure 1 shows that the predicted prices (in orange) in each interval of number of carats are much more compact than the known prices of the already sold diamonds (in blue). In the formula from the linear regression model, not only carats are taking into account, but cut, clarity and a negative constant. So, we are not considering the necessary factors to account for all the variance and predict the prices. Moreover, even if we considered all these factors in the formula, in my opinion other characteristics influence the price of a diamond, for example the place where the diamond was collected, location of the jewelry shop and treatments.

The model predicts very well the prices for the diamonds whose number of carets is between 0.25 and 0.5, and then, the thick line of predicted values seems to be the diagonal of the price shape (in blue). The diamond to predict with the maximum number of carets had 3.01 carets. The model gave the best fit line but sometimes it underestimated the price of the diamonds and even gave negative prices for those diamonds with less desirable characteristics. But, although the model mispredicted the value of many diamonds, the prediction of the total price to pay for the 3,000 diamonds seems to be accurate (good average).

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. HINT: The number should be 7 digits.

   ○ The company usually purchases diamonds from distributors at 70% of that price, to get a margin of 30%, so my recommended bid price should represent that. Therefore I recommend a bid of $8,213,465.932 which is the 70% of the total predicted price for the 3,000 diamonds: $11,733,522.76. To obtain the predicted value of the 3,000 diamonds I used the formula from the linear regression model (based on the characteristics and prices of the previous diamond sales), applied it to the 3,000 diamonds and then I added up all those predicted prices. Finally, I multiplied that amount ($11,733,522.76) by 0.70 to get the final predicted bid of $8,213,465.932.

I acknowledge the explanation project to predict the price of several houses according to their number of bedrooms, by Udacity (house-example.docx).