

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?
 - o The one carat heavier would result in an additional \$8413 in price. The formula created by the regression determined that the coefficient for a carat is 8,413.00, so for every increase in the weight of diamond 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?
 - o The formula is $\text{price} = -5,269 + 8,413 * \text{Carat} + 158.1 * \text{Cut} + 454 * \text{Clarity}$
 - 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

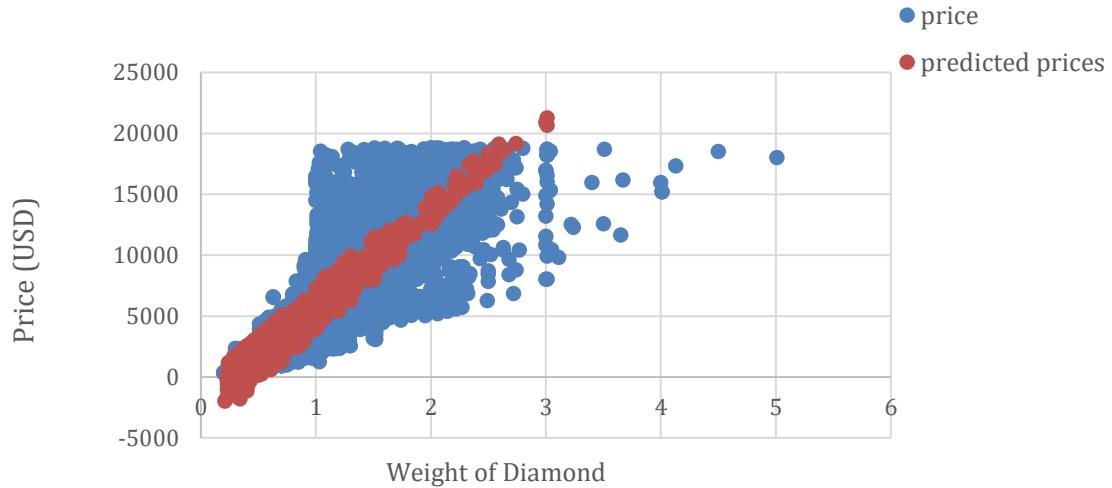
Step 2: Visualize the Data

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.
 - o **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?

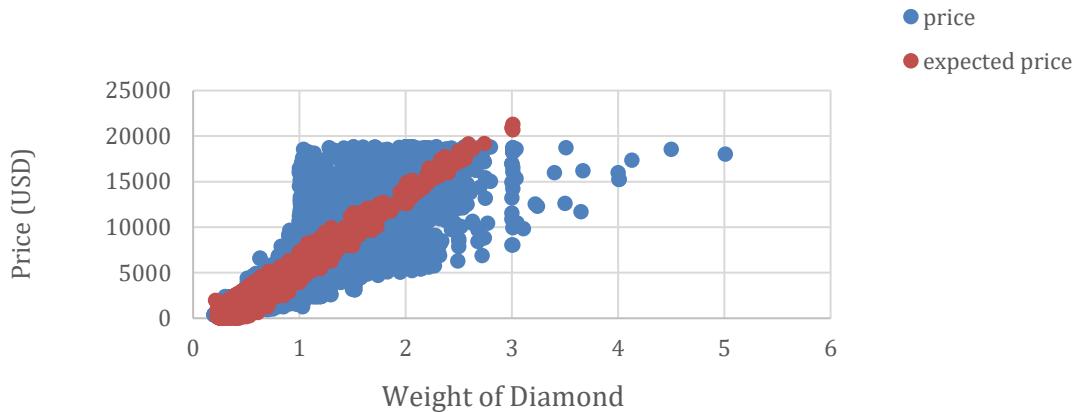
Carat vs Price



The predicted prices are more compact than the actual data is. This is because we are not accounting for everything that effects prices. There are many more things than carat that effect 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 the country we are training the model on.

I noticed the linear regression model produces a negative price values when Carat value is smaller than or equals 0.4 as shown on the graph above. I then tried to fix this by removing negative signs from price values and got a graph as below:

Carat vs Price



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.
 - a. I recommend a bid of \$8,220,532.29. I arrived at this number by using a formula from the regression model provided that was based on previous diamond sales and applied it to the diamonds that were up for bid. I then factored in the margin the investors were looking for which was 30% so I multiply the predicted amount 11743617.6 by .70 to get the final predicted bid of \$8,220,532.29