

## 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?
  - *If the weight of the diamond (carat) is heavier by 1, I would expected to pay 8,413\$ more for it. This is in accordance with regression and coefficient value of 8,413\$ which means that if we decide to bought 1 carat heavier diamond than another the price will go up for that amount.*
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?
  - *According to the regression formula I would have to pay 10,094.8\$ for such a diamond. The formula in that case would look like this:*  
$$\text{Price} = -5269 + 8413 \times 1.5 + 158.1 \times 3 + 454 \times 5$$

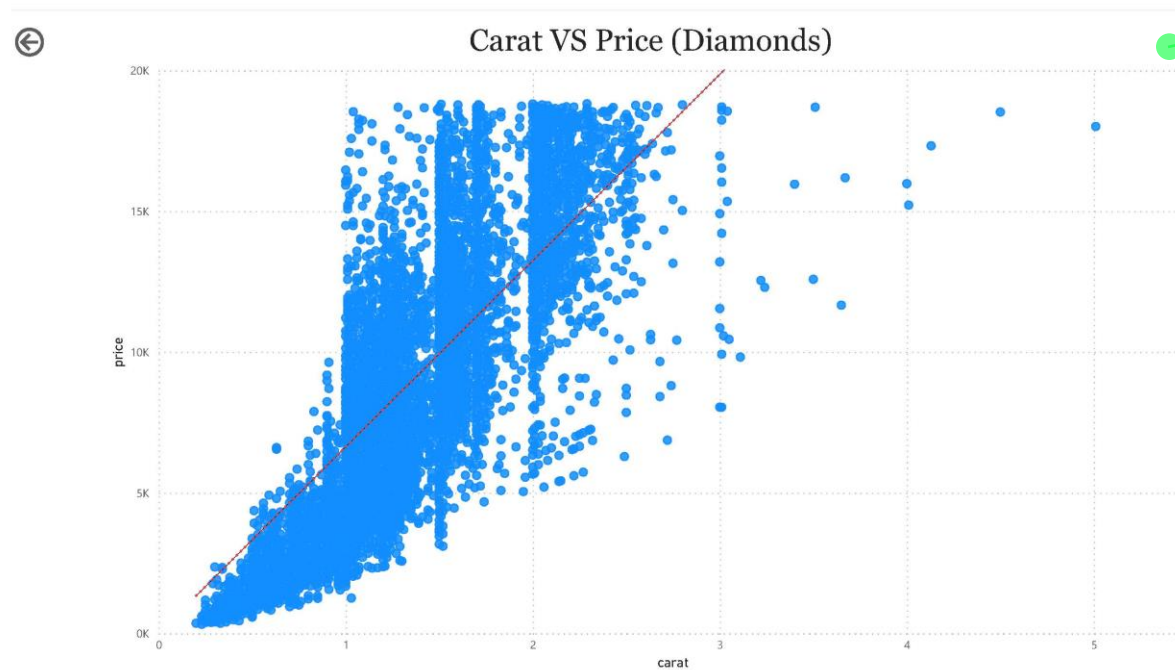
Andy: Your interpretation is correct! The idea here is to understand the interpretation of the coefficient. If we change a unit of any numerical variable the impact will be exactly the coefficient of the variable

Andy: The idea here is indeed to verify the application of the equation, well done!

### Step 2: Visualize the Data

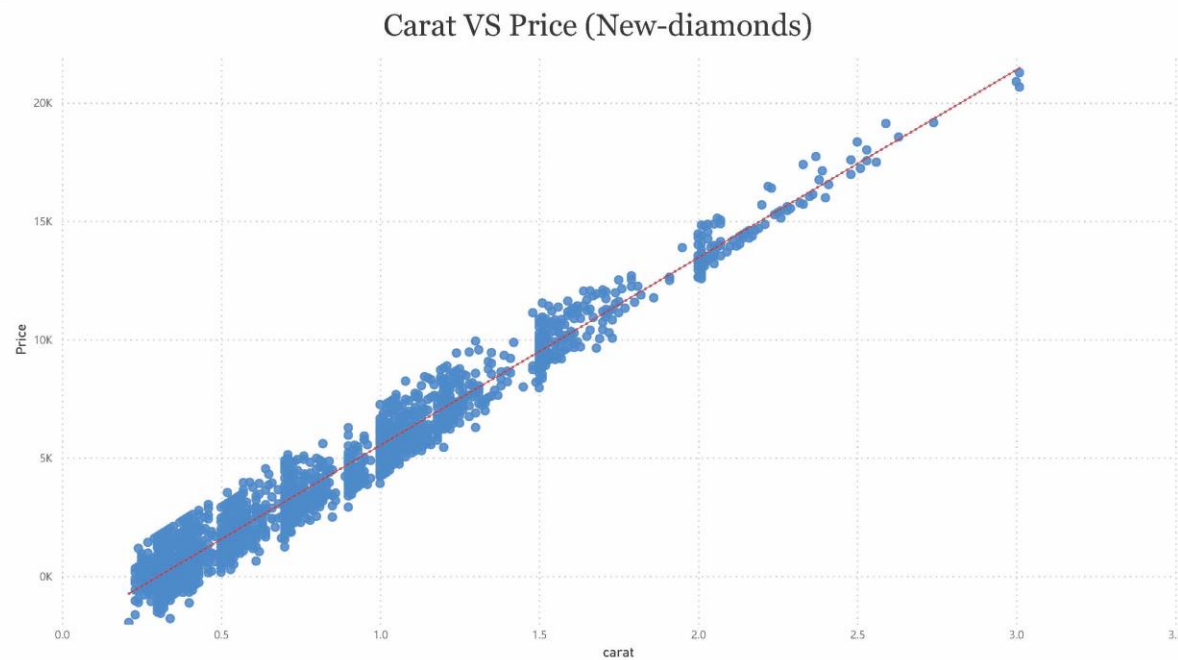
Make sure to plot and include the visualizations in this report. For example, you can create graphs in Excel and copy and paste the graphs into this Word document.

1. Plot 1 - Plot the data for the diamonds in the database, with carat on the x-axis and price on the y-axis.



Andy: Great work including all the relevant labels in both charts! This helps with the chart interpretability and improves the overall quality of our report.

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?

- *We can see that two scatter charts showing that there is a relationship in Price and Carat. First graph related to Diamonds table showing good relationship but not the best like the second graph (New-diamonds). The reason for this could be the less population data (50,000 versus 3,000). Anyway we can conclude that there is a good relationship for building predicted prices.*

## 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.

- After I analyzed the relationship and done calculations, I would recommend 8,213,465.93\$ for a bid. At first I have conducted regression formula to find a predicted price for each diamond. After it I summed up prices for all 3,000 diamonds and multiply it by 0.70, because the margin is 30%. After that I came up with a value of 8,213,465.93\$.

Andy: The idea here is to see that the forecast prices are in a much narrower range than the actual set of prices. This reinforces the fact that there are likely other factors omitted from the model that would help improve our accuracy.

We can also see that, although the model can do a good job on average, for any particular diamond the forecast may be very wrong, including negative predictions.

With this in mind, while I would not be confident using the model to come up with the price for a single diamond, it can still be useful to recommend the bid price for the whole set of diamonds.

As we are talking about a large number of diamonds, this would allow for the errors to average out which results in a good level of confidence in the final price predicted.

Andy: The expected offer is indeed \$8,213,466 excellent work!

Obs:

While not a requirement one way we could deal with the negative predictions is to replace the negative predictions with a value of zero.

On the other hand, we could argue that some of the prices predicted by the model are overestimated so the negative prices act as a counter balance in the final bid price and, therefore, we should keep the negative prices for the final bid.