

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?

Given the linear regression formula is

$$\text{Price} = -5,269 + 8,413 \times \text{Carat} + 158.1 \times \text{Cut} + 454 \times \text{Clarity}$$

The coefficient for carat is 8413, so with Carat and Clarity being equal, 1 Carat heavier diamond will result in \$8,413 higher in predicted price.

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?

Carat = 1.5

Cut = 3

Clarity = 5

$$\text{Price} = -5,269 + 8,413 \times \text{Carat} + 158.1 \times \text{Cut} + 454 \times \text{Clarity}$$

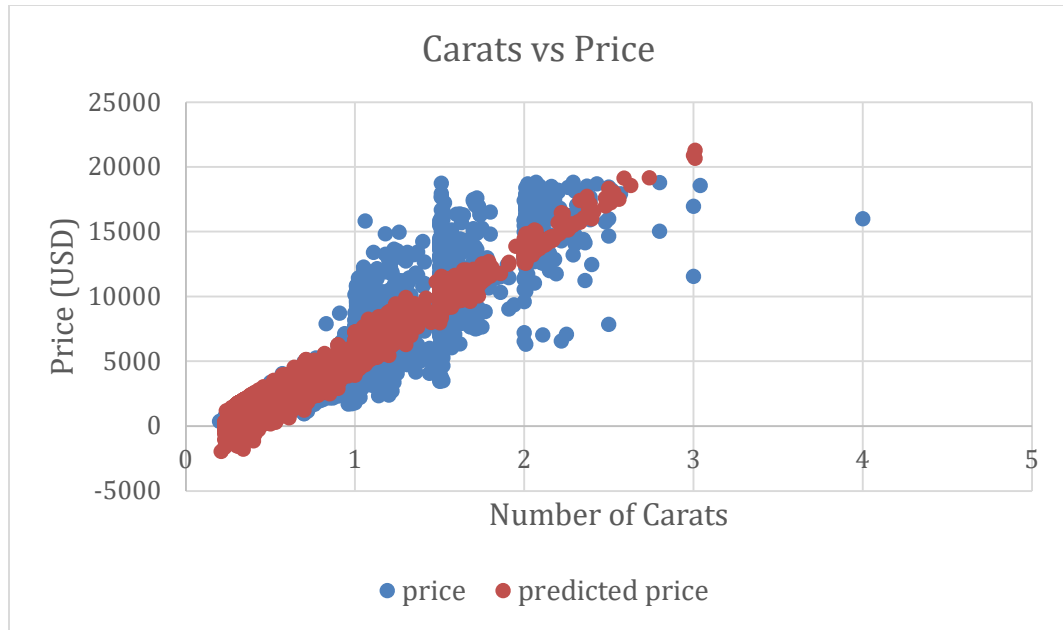
$$\text{Price} = -5,269 + 8,413 \times 1.5 + 158.1 \times 3 + 454 \times 5$$

$$\text{Price} = \$10,094.80$$

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.



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

Looking at the plot, the predicted prices are in a much narrower range than the actual prices with respect to the number of Carats. This may be due to the predicted price having more dependencies other than the number of Carats, Cut and Clarity. Also, the linear regression formulation may be too simplistic and thus underperforming. A more complex formulation and/or more dependencies may improve predictive accuracy.

While the linear model is able to generalize the price decently on average with overpredictions cancelling out underpredictions. Predictions perform well for less than 1 Carat, but beyond 1 Carat, the performance of the model starts to suffer, fairing more worse with more Carats. There is a clear outlier at 4 Carats and even some negative prices for less than 0.5 Carts.

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.

Summing up all the predicted prices for the 3000 diamonds gives \$11,733,522.76. Given the high number of diamonds, the overpredictions and underpredictions should cancel out to result in a relatively accurate total price.