

PROJECT 1: PREDICTING DIAMOND PRICES

By Ivy Huong Nguyen

Step 1 - Understanding the Model:

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 would the retail price of the heavier diamond be?

Why?

Given model: $\text{price} = -5269 + 8413 \cdot \text{carat} + 158.1 \cdot \text{cut} + 454 \cdot \text{clarity}$

Known: Price 2 has $\text{price}_2 = \text{price}_1 + 1$ and $\text{cut} = \text{cut}_1 = \text{cut}_2$, $\text{clarity} = \text{clarity}_1 = \text{clarity}_2$

Let call price_1 and price_2 as the price of the regular diamond and the diamond with 1 carat heavier than the regular one, respectively,

Then we have:

$$\text{price}_1 = -5269 + 8413 \cdot \text{carat}_1 + 158.1 \cdot \text{cut}_1 + 454 \cdot \text{clarity}_1$$

$$\rightarrow \text{price}_1 = -5269 + 8413 \cdot \text{carat}_1 + 158.1 \cdot \text{cut} + 454 \cdot \text{clarity}$$

$$\text{price}_2 = -5269 + 8413 \cdot \text{carat}_2 + 158.1 \cdot \text{cut}_2 + 454 \cdot \text{clarity}_2$$

$$\rightarrow \text{price}_2 = -5269 + 8413 \cdot (\text{carat}_1 + 1) + 158.1 \cdot \text{cut} + 454 \cdot \text{clarity}$$

$$\rightarrow \text{price}_2 = -5269 + 8413 \cdot \text{carat}_1 + 8413 + 158.1 \cdot \text{cut} + 454 \cdot \text{clarity}$$

$$\rightarrow \text{price}_2 = (-5269 + 8413 \cdot \text{carat}_1 + 158.1 \cdot \text{cut} + 454 \cdot \text{clarity}) + 8413$$

$$\rightarrow \text{price}_2 = \text{price}_1 + 8413$$

Therefore, if a diamond is 1 carat heavier than another with the same cut and clarity then the retail price of the heavier diamond would be \$8413 in addition.

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), what retail price would the model predict for the diamond?

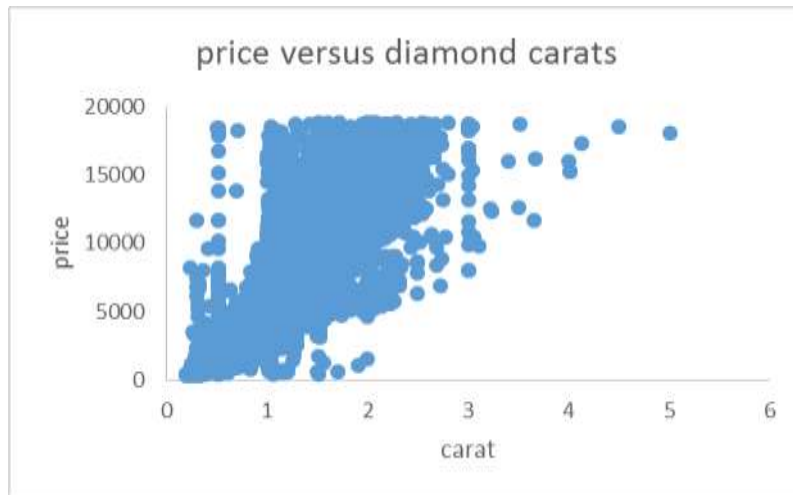
Given: $\text{carat} = 1.5$, $\text{cut} = 3$, $\text{clarity} = 5$

Therefore, based on the given model, the retail price should be \$10,094.8.

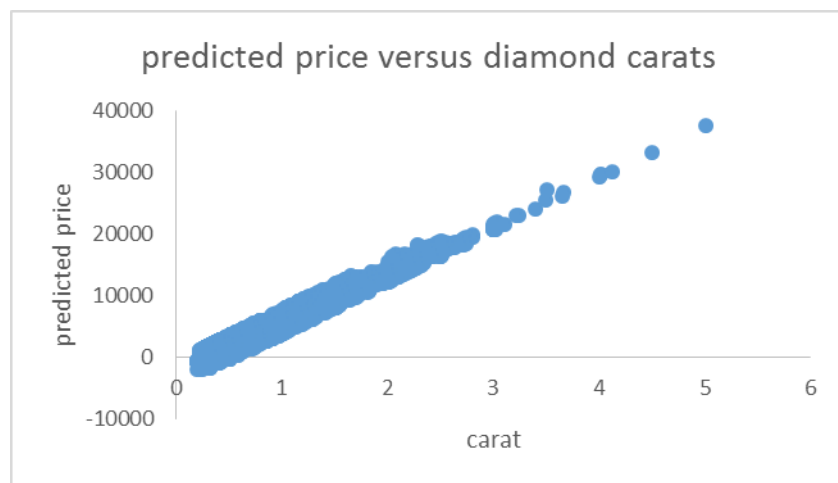
$$\text{Price} = -5269 + 8413 \cdot 1.5 + 158.1 \cdot 3 + 454 \cdot 5 = 10,094.8$$

Step 2 - Visualize the Data: Create two scatter plots.

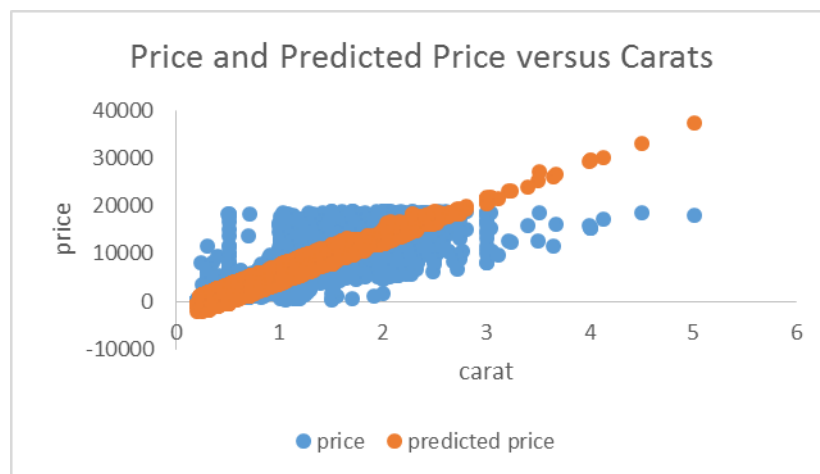
- Plot 1 - Plot the data for the diamonds in the database, with carat on the x-axis and price on the y-axis.



- 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.



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

By plotting the two plots together, I notice a huge bias in using the model to predict the price. The price tends to be either lower or higher than the true values especially near the ends of the graph. With this behavior, I don't feel confident in the model's ability to predict prices for diamonds in the very light or very heavy weight.

Step 3 - The Recommendation: What bid do you recommend for the jewelry company? Please explain how you arrived at that number.

For this auction, I would recommend the jewelry company to bid at the price of \$11,733,523 for the total of 3000 diamonds. I came up with this number by using the given linear regression model to calculate the predicted price for each diamond in the given new_diamond.csv dataset. Then, I summed the prices of the 3000 diamonds together to get the recommended value.