

Project: Diamond Prices

Overview

This project aims to predict and recommend the best price that a jewelry company should bid for in its purchase of diamonds.

The software applications used for the entire analytical process and model building are Excel and Alteryx.

The following steps were taken to get the desired results.

Step 1: Understanding the Model

1. **Question:** 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?

Answer: You should expect to pay \$8,413 more. This is so because the coefficient for a carat in the regression equation is 8,413. Thus, for every increase in the weight of a carat, the price will increase by the amount of the coefficient.

2. **Question:** 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?

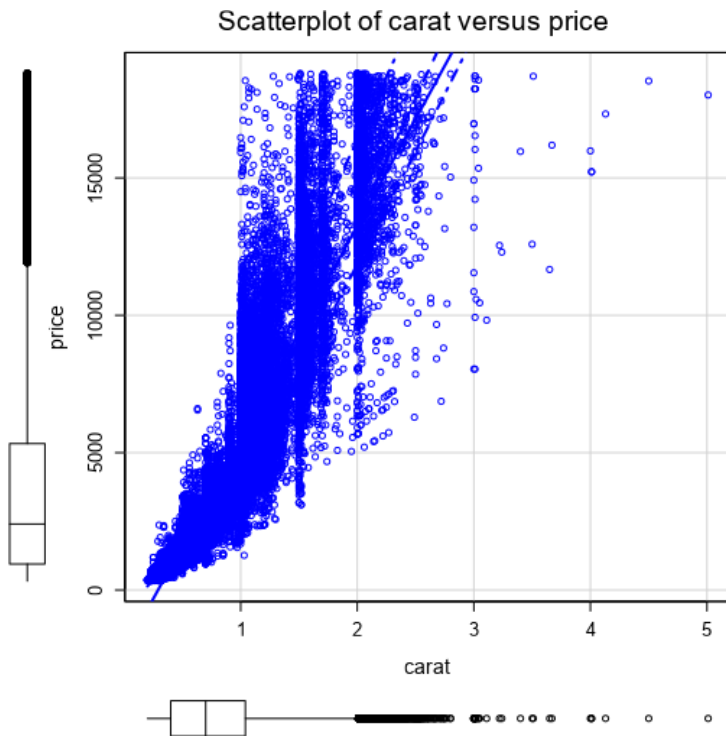
Answer: The model formula is **Price** = -5,269 + 8,413 x **Carat** + 158.1 x **Cut** + 454 x **Clarity**

Now, plugging in carat = 1.5, cut = 3 and clarity = 5 into the above model, we have

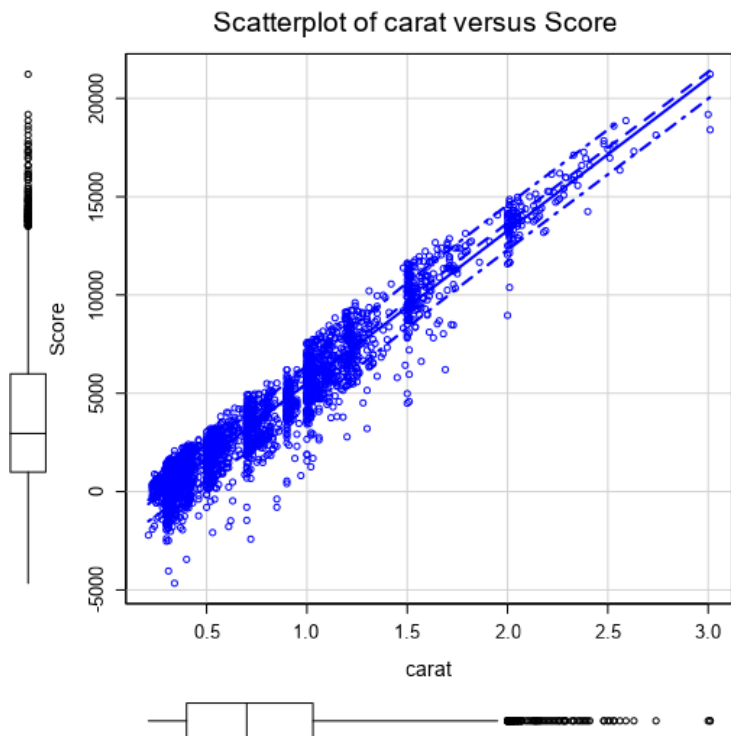
$$\begin{aligned}\text{Price} &= -5,269 + 8,413 \times 1.5 + 158.1 \times 3 + 454 \times 5 \\ &= \$10,094.80\end{aligned}$$

Step 2: Visualize the Data

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



2. Plot 2 - Plotting the data for the diamonds for which you are predicting prices with carat on the x-axis and predicted price on the y-axis.



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

Answer: There seems to be a better correlation between Score (predicted prices) and carat in the test dataset than between price and carat in the training dataset. This is so because not all predictor variables that appear in our model were used in the scatterplot. Even so, the list of predictor variables used in the model is not exhaustive as there could be some other latent factors to consider, depending on where the predictive analysis is being carried out.

Having stated the above, I must state that I am fairly confident in the model's ability to predict the diamond prices. On the average, the differences between the scatterplots are negligible as they both display an obvious similar trend.

Step 3: Make a Recommendation

1. **Question:** What price do you recommend the jewelry company to bid? Please explain how you arrived at that number.

Answer: I recommend the jewelry company to bid for \$8,211,163.14591454.

I arrived at this figure by using the summarize tool in the Alteryx analytical software to sum up all the scores (predicted prices) for all the diamonds (which gave \$11,730,233.065592) and then using the formula tool to multiply this sum by 70% (or 0.7). This gave the recommended bid price of \$8,211,163.14591454.

NB: Bid_price value re-checked and found to be correct from given data.