

Udacity BAND Project 1: Diamond Prices

Jason Grenig

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?

At retail, you can expect to pay \$8,413 more if a diamond is 1 carat heavier, all else held constant. This can be found by manipulating the regression equation as follows:

Diamond A = $-5269 + 8413 * \text{Carat} + 158.1 * \text{Cut} + 454 * \text{Clarity}$

Diamond B = $-5269 + 8413*(1+\text{Carat}) + 158.1 * \text{Cut} + 454 * \text{Clarity}$

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?

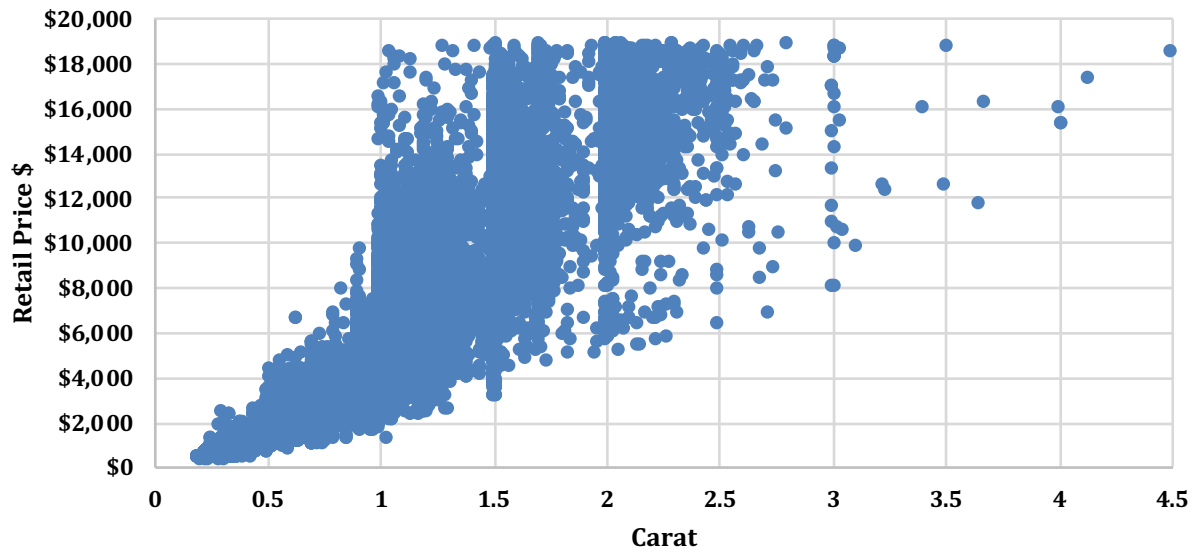
Based on the regression model, the predicted retail price for a diamond with the aforementioned attributes would be \$10,094.80.

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.

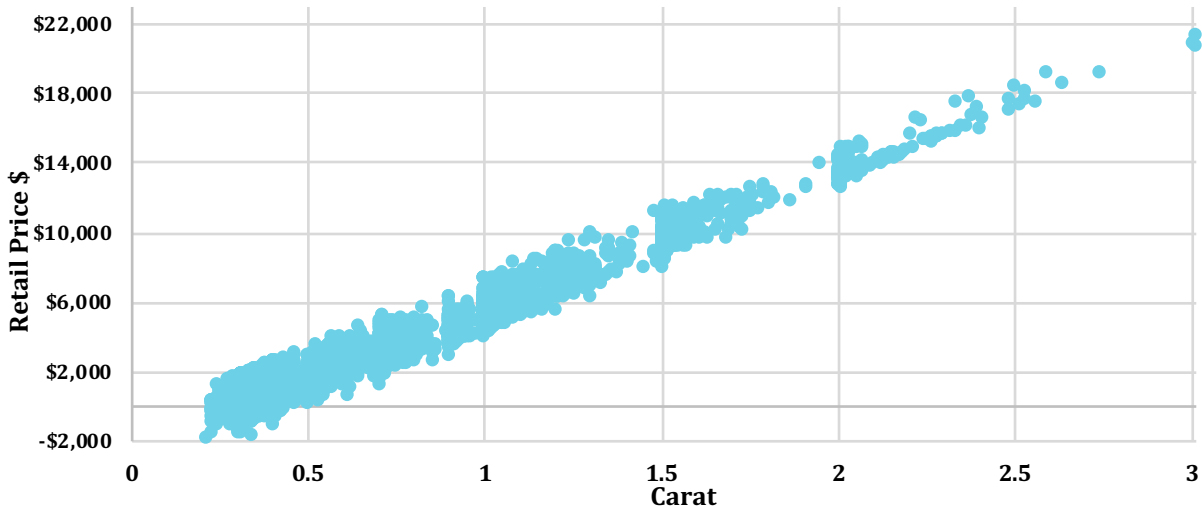
Carat vs Retail Price Comparison (Training Data)



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.

Carat vs Retail Price Comparison (Test Data)



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

In the Training Data, the relationship between retail price and carat is exponential, whereas this relationship is linear in the Test Data. For diamonds roughly 1ct or less, the regression model appears to be quite accurate between both data sets. However by comparison, prices in the Training Data model reach parity around the 1ct tick mark, and diamonds that are over 1ct in weight become exponentially more expensive than those predicted by the Test Data model. I am confident the model can accurately predict prices up to 1ct. A further question I have is whether a linear regression model should be used to predict variables that are exponentially correlated. However that may be beyond the scope here.

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.

Based on the Test Data, and without reconfiguring the given equation, the model predicts that the company should bid \$3,520,056.83 for the set of 3,000 diamonds. I arrived at this number by using the given equation to calculate each diamond's retail price, then multiplying that number by 0.3 to consider the 70% discount that the company pays.