<u>Project: Diamond Prices</u>

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?

   The regression formula is
           Price = -5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity
   So if we hold cut and clarity constant, then 1 carat heavier adds 8413 into the price since carat's coefficient is 8413.

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?
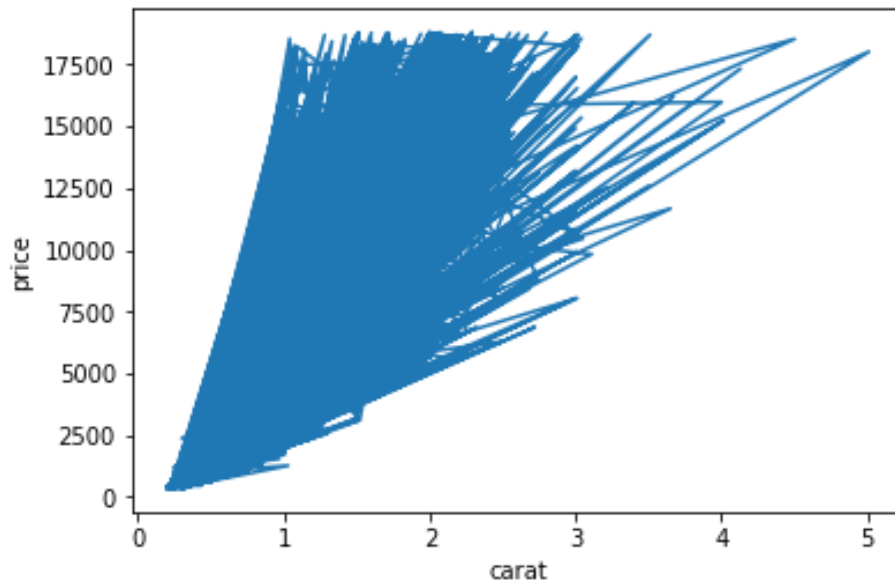
   According to the formula:

   Price = -5269 + 8413*1.5 + 158.1*3 + 454*5 = 10094.8

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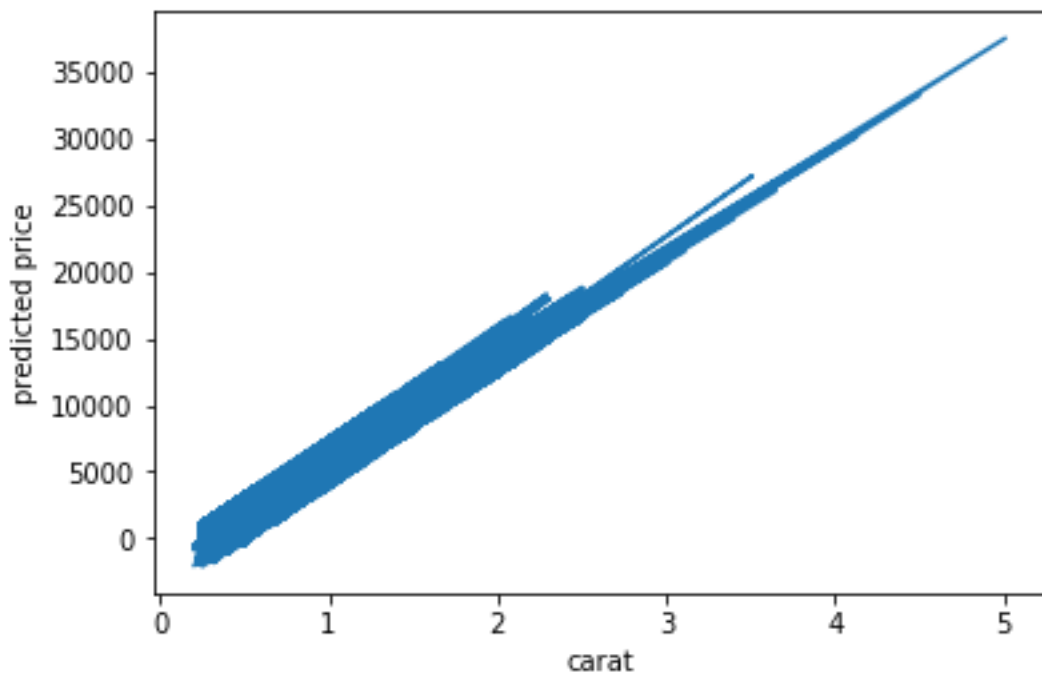
3. What strikes you about this comparison? After seeing this plot, do you feel confident in the model's ability to predict prices?

The predicted price has a very obvious linear dependency on carat, while the true price does not. Thus, the model is not capable enough to predict the 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.

   We will still use the linear regression formula. So we apply the formula to each diamond in the dataset to get a predicted price, then we add up these 3000 predicted prices and multiply the sum by 0.7. The finally total price is

   $ 137,911,875