Project: Diamond Prices

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#](https://classroom.udacity.com/nanodegrees/nd008/parts/235a5408-0604-4871-8433-a6d670e37bbf/project)

# Step 1: Understanding the Model

*Answer the following questions:*

**Model learned:**

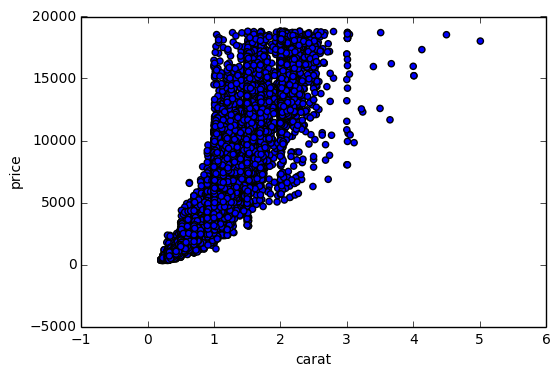
**Price = 8309.62 \* carat + 171.08 \* cut\_ord + 468.04 \* clarity\_ord + -5230.22**

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 change in the carat value corresponds to the weight to this feature in the linear equation learned. In the model I learned using a SGDRegressor, the carat coefficient was 8,309.62, and thus, an increase of 1 in the carat would correspond with an increase of 8,309.62 in the final 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? **According to my model, the final price of a diamond with such characteristics would be of 10087.65.**

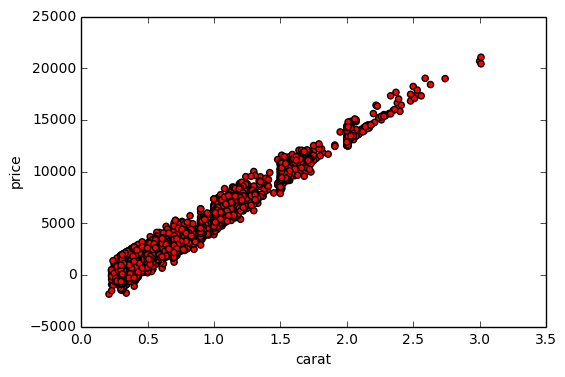
# Step 2: Visualize the Data

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.



1. 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.



1. What strikes you about this comparison? After seeing this plot, do you feel confident in the model’s ability to predict prices? **Two things bother me in these plots: the first is that in the predicted values, some diamonds are predicted to have negative values, something that doesn’t make any sense. The other one is that the distribution in the carats in the training data seems a little skewed towards the [0; 3] interval. This might lead to inefficient predictions if diamonds with higher carats are presented to the learned regression model.**

# 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. **After learning the regression model, I’ve summed up the predicted values and then deduced 30% of it. The final value is 8,268,709.99.**