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?

    - If a diamond with 1 carat heavier than with same cut and clarity, the retail price would be **8413 more**.

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?

    - One easy way to find the retail price that model predicted would be to substitute these values in the equation.

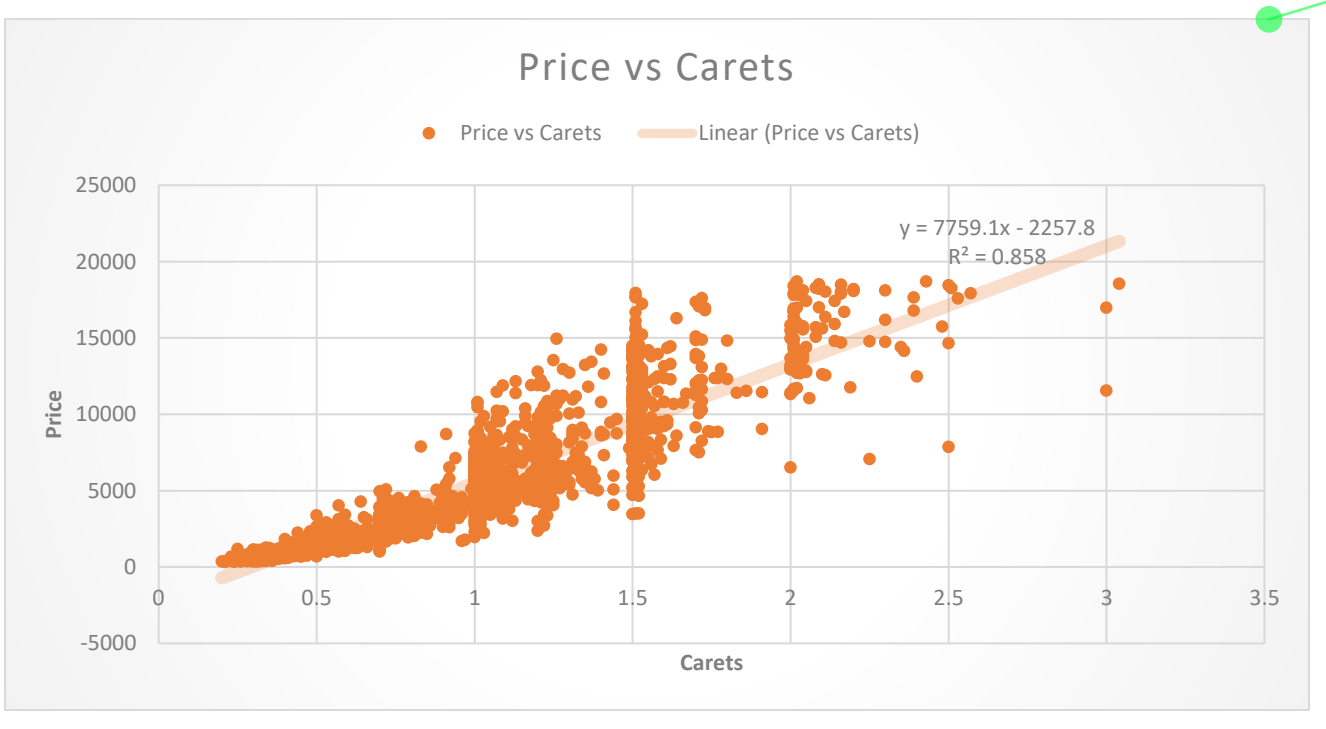
    - Price = - 5269 + 8413 x Carat + 158.1 x Cut + 454 x Clarity
      = - 5269 + 8413 x (1.5) + 158.1 x (3) + 454 x (5)
      Price = 10094.8

Andy: Your interpretation is correct! The idea here is to understand the interpretation of the coefficient. If we change a unit of any numerical variable the impact will be exactly the coefficient of the variable

Andy: The idea here is indeed to verify the application of the equation, well done!

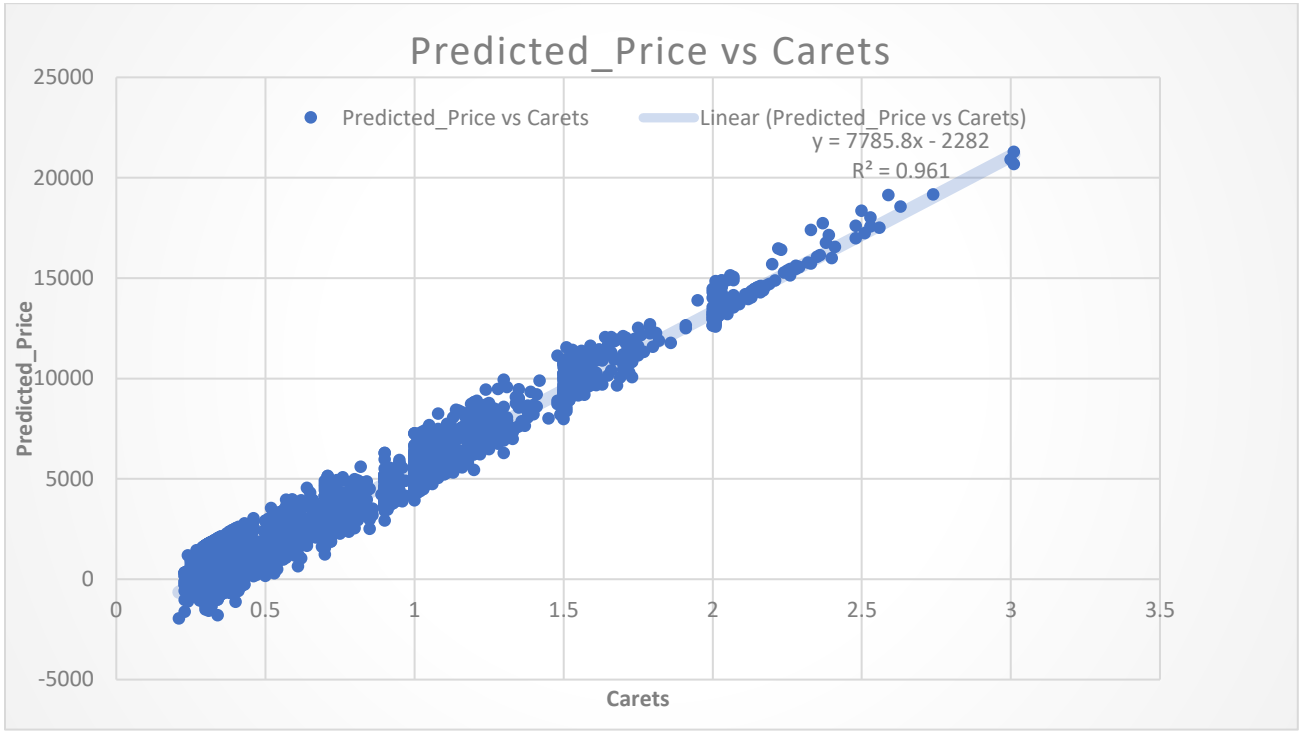
Step 2 - Visualize the Data: Create two scatter plots. If you're not sure what a scatter plot is, see here.

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



Andy: Great work including all the relevant labels in both chart! This helps with the chart interpretability and improves the overall quality of our report.

- 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?
    - This linear regression model has the strong correlation between caret and its price which tells us that we have the caret should be considered as one of the features. If we look at some of the predicted data for carets less than 0.5, we can see negative values, so we still need to include more features to this model.

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

- I recommend a bid of $8213465.93 and I came to that number based on the historical data provided and applied it the diamonds available for auctions. Given the company purchases diamonds at 70% of that price, so

    $11733522.8*70% = $8213465.93

Andy: The idea here is to see that the predicted prices are in a much narrower range than actual prices (i.e. they are less spread out). This reinforces the fact that there are likely other factors omitted from the model that would help improve our accuracy.

We can also see that, although the model can do a good job on average, for any particular diamond the forecast may be very wrong, including negative predictions.

With this in mind, while I would not be confident using the model to come up with the price for a single diamond, it can still be useful to recommend the bid price for the whole set of diamonds.

As we are talking about a large number of diamonds, this would allow for the errors to average out which results in a good level of confidence in the final price predicted.

Andy: The expected offer is indeed $8,213,466 excellent work!

Obs:

While not a requirement one way we could deal with the negative predictions is to replace the negative predictions with a value of zero.

On the other hand, we could argue that some of the prices predicted by the model are overestimated so the negative prices act as a counter balance in the final bid price and, therefore, we should keep the negative prices for the final bid.