Project: Diamond Prices

Step 1: Understanding the Model

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?

   Price = -5,269 + (8,413 x Carat) + (158.1 x Cut) + (454 x Clarity)
   According to the given equation, we can figure the price will rise by 8,413 if these diamonds have the same cut and 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?

   Price = -5,269 + (8,413 x Carat) + (158.1 x Cut) + (454 x Clarity)
       = -5,269 + (8,413 x 1.5) + (158.1 x 3) + (454 x 5)
       = -5,269 + 12,619.5 + 474.3 + 2,270
       = 10,094.8
   You should pay 10,094.8 for this model.

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.
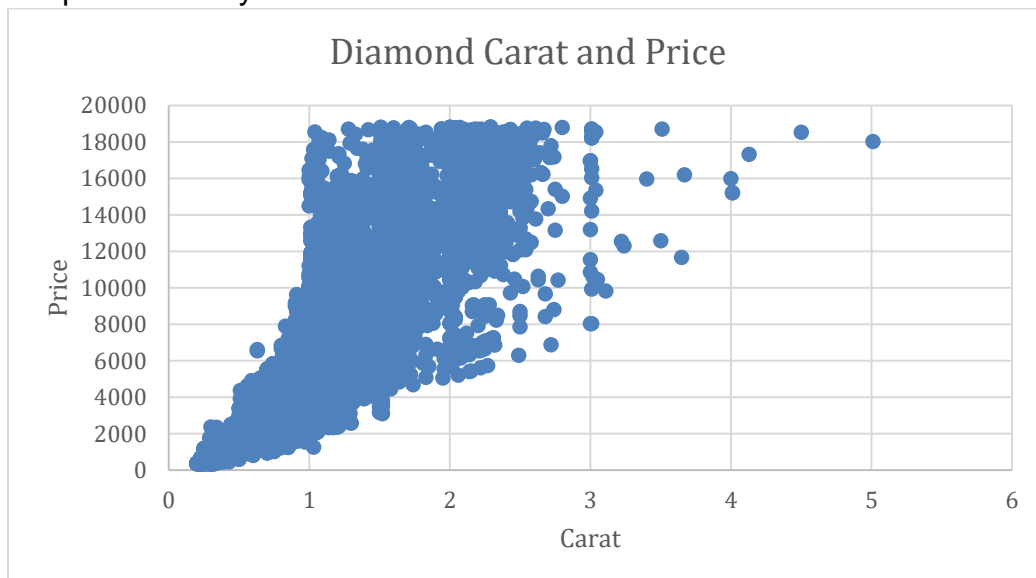
2. Plot 2 - Plot the data for the diamonds for which you are predicting prices with carat on the x-axis and predicted price on the y-axis.
   - o **Note**: You can also plot both sets of data on the same chart in different colors.


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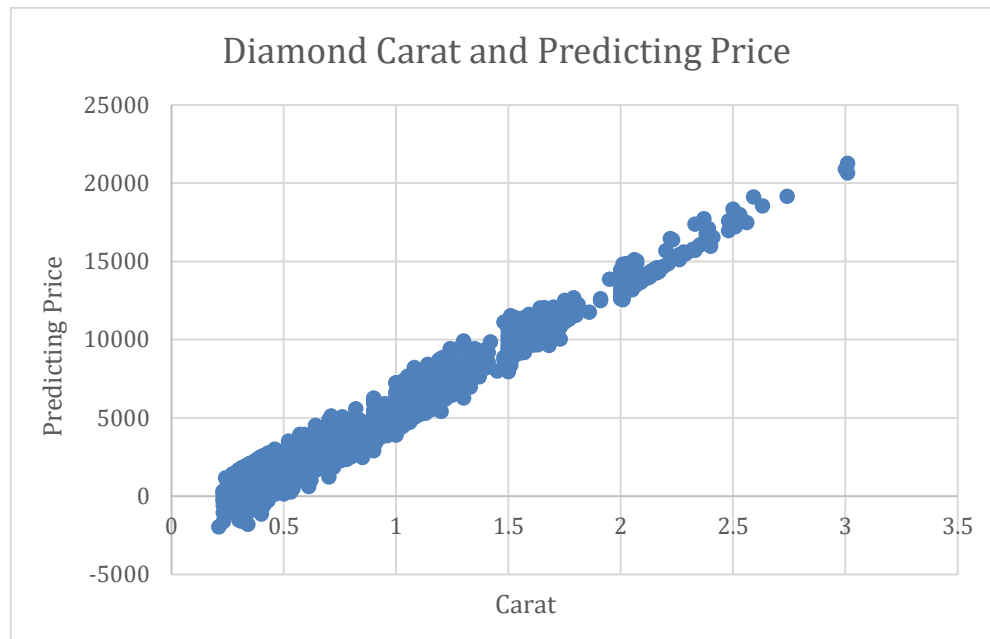
3. What strikes you about this comparison? After seeing this plot, do you feel confident in the model's ability to predict prices?

These two plots help us notice some information about our data. From the first plot, we can notice that the correlation is strong from 0 to 1 carat and become weaker through 1 to 3 carat and the prices are not linear anymore which kind of realistic data since the diamond price depends on other factors of diamonds. But the other plot which displays the predicate prices shows an almost perfect linear relation with a strong positive correlation. Moreover, it shows a negative value for diamond prices.

After seeing this model, I am not confident in its ability to predicate prices. Especially, since this model predicate negative values for the price. Because it does not make sense to have a negative price for a diamond. You are not supposed to pay people for them to buy a diamond from you.

Since the linear regression model does not give accurate predicted prices. I think using another model to predicate the price is better. From the Analyst Methodology Map, we can choose another model to predict diamond price. The other models as shown in the figure below are the Decision tree, Forest Model and Boosted Model.
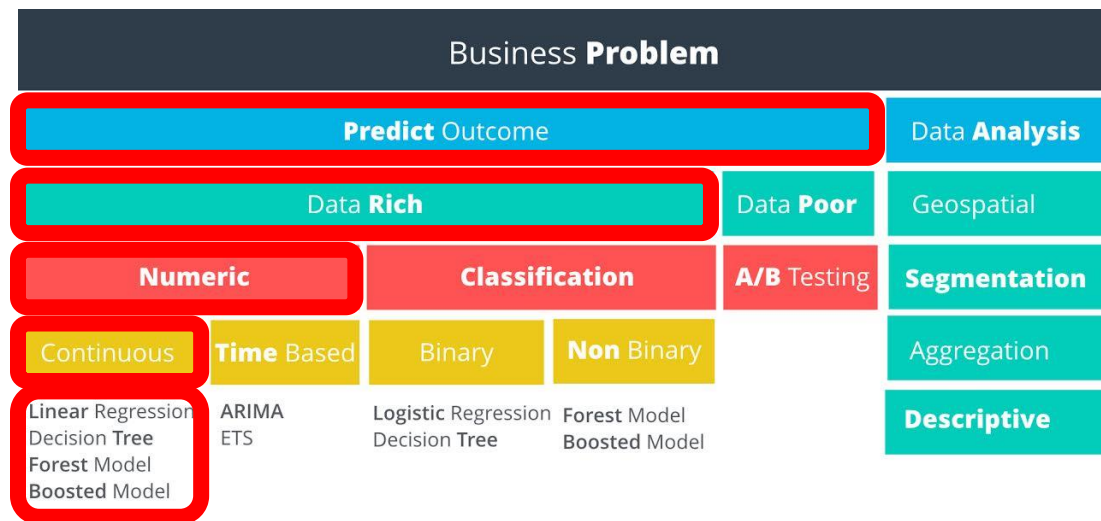
Step 3: Make a Recommendation

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

I recommend the jewelry company to bid a price of 70% of the total predicted prices of the whole set. The total of predicated prices is 11,733,522.76$. So, the bid price should be 70% of this price which equals 8,213,465.932$.

Bid price = (total price * 70) / 100 = (11,733,522.76 * 70) / 100 = 8,213,465.932