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?
    - Considering the equation of the linear regression model (-5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity), for every 1 carat weight heavier than another, an additional $8,413 is expected to be paid. This is because the coefficient of the carat predictor variable is 8,413 and it implies that every unit increment in the carat weight attracts and additional price of the value of the coefficient.

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?
    - The equation of the linear regression model is; Price = -5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity.
      Price = -5,269 + 8,413 x 1.5 + 158.1 x 3 + 454 x 5
      Price = -5,269 + 12,619.5 + 474.3 + 2,270
      Price = 10,094.8
    - From the above calculation, the model would predict a price of $10,094.8

Step 2: Visualize the Data

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.
    - **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?

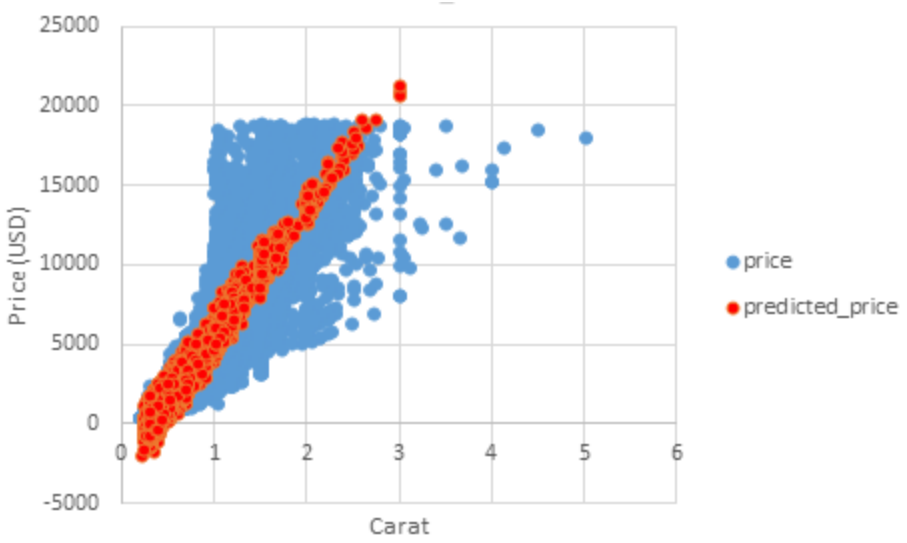

Figure 1: (Carat vs Price) Plot

- From my observation of the plot, the predicted prices seem to display a high positive correlation compared to the existing prices. Although there are several outliers (such as the 5 carat diamond with a price of about $18,000), the plot generally indicates that the higher the weight of a diamond, the higher the price. I believe that if other predictor variables are put into consideration, the strength of the correlation could be affected.
- Also from the plot, I am pretty optimistic that the model will do well in predicting prices even though it contains some outliers.

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 would recommend that the jewelry company bid a price of $8,213,465.932 for the set of new diamonds.
    - I was able to use the provided equation generated from the linear regression model of the 50,000 diamond dataset to predict the prices for the 3,000 new diamond dataset and obtained the total predicted price of $11,733,522.76. Since the jewelry company generally purchases diamonds from distributors at 70% of the price, I multiplied the total predicted price of $11,733,522.76 by 0.7 to obtain the desired bid price of $8,213,465.932.