Predicting Diamond Prices

Step 1 - Understanding the Model

- 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 is one carat heavier than another with the same cut and clarity, the retail price will be $8,413 higher because according to the linear regression equation, the coefficient of the carat variable is 8,413, this means for a one unit change in carat size, they'll be an $8,413 change in price and since the relationship between carat size and price is a positive one (shown by the + sign attached to the carat coefficient), a one-unit increase in carat size indicates an $8,413 increase in price.

- 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?

  - The formula is
    $$Price = -5,269 + 8,413 * Carat + 158.1 * Cut + 454 * Clarity$$
   Given the regression model above, we would substitute the values for the different variables into the equation.
   - $Price = -5,269 + 8413 * (1.5) + 158.1 * (3) + 454 * (5)$
    - $Price = 10,094.8$

Step 2: Visualize the Data

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

- 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 colours.

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

1

Scatterplot of carat versus price

Scatterplot of Carat versus Predicted Price
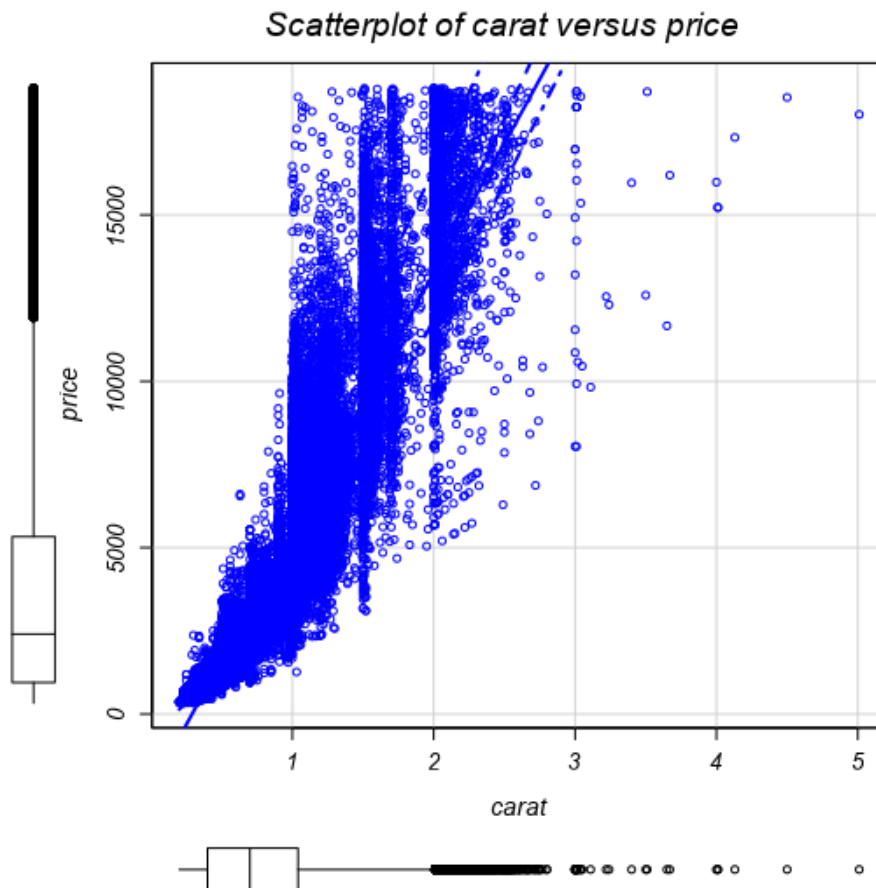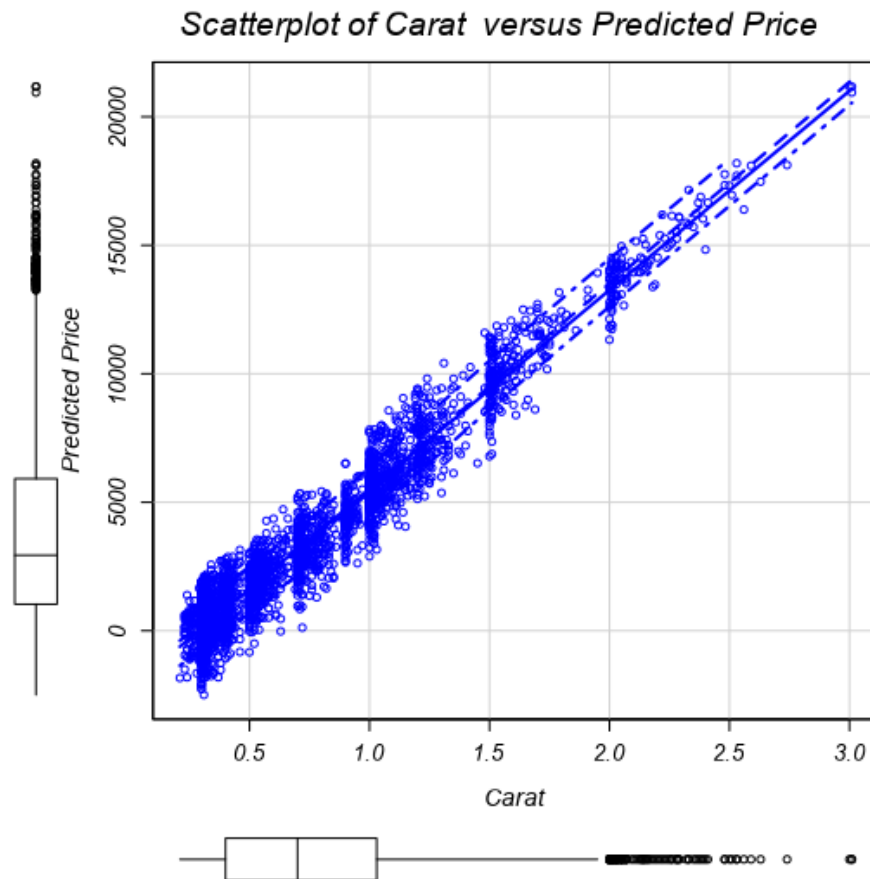
The scatter plot of the predicted prices show a highly positive linear relationship between the Carat size and the price of the diamonds however the scatter plot of the known prices, that is the first data set do not show a very strict linear relationship, it's positively linear at smaller carat sizes but they are diamonds with very high prices even if they have small carat sizes, this difference in the two plots show that they are other attributes of a diamond that can affect its price such as the cut and clarity which I didn't factor into the scatterplots above. Overall I think the model can predict prices well because the two scatter plots still share similarities that show a positive linear relationship between carat size and price.

Step 3: Make a Recommendation

- What bid do you recommend for the jewelry company? Please explain how you arrived at that number.

    - The Jewelry Company should make a bid of $8,235,352.05. Below is how I arrived at my recommended price;

        1. I made use of an old dataset of diamonds which contains the prices of different diamonds alongside their attributes, such as carat size, cut and color.

        2. I trained a linear regression model to predict the price of a diamond based on its attributes.

        3. I used the regression model above to predict the price of each diamond in the set the jewelry company wanted to purchase.

        4. I summed up the prices of the diamonds to get the total retail price of the set.

        5. I multiplied this sum by 70% to get the price the company should pay to the jewelry distributor, since they generally pay 70% of the retail price.