<u>Project: Diamond Prices</u>

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**

    Expect to pay an increase $8,413 over the price with the same cut, because the fixed value per carat is 8,413 so if one carat increased from the another, it will the price increase by this amount.

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 * **1.5)** + (158.1 x **3)** + (454 x **5) = 10,094.8**

    To get 1.5 carat diamond we will pay $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.

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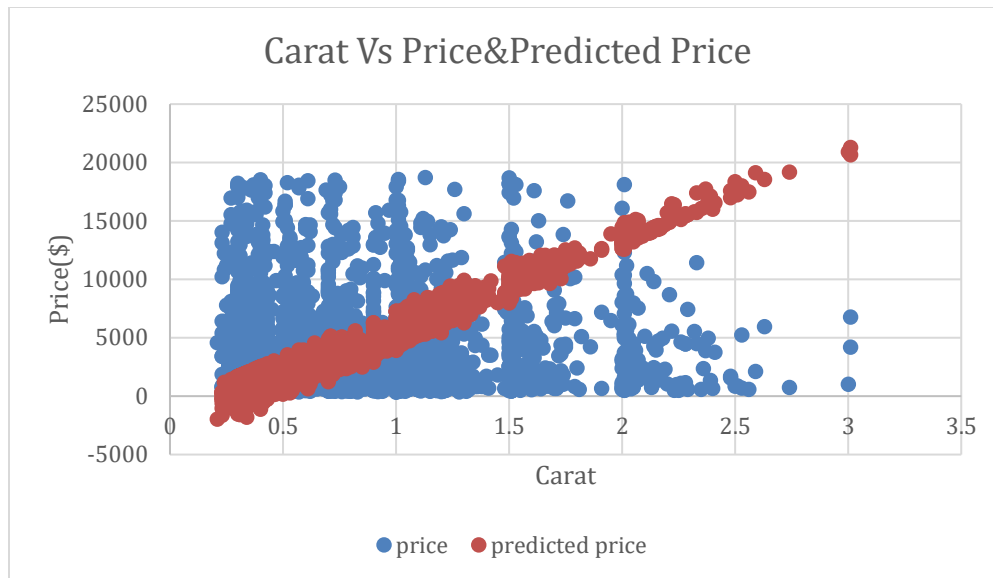
3. What strikes you about this comparison? After seeing this plot, do you feel confident in the model's ability to predict prices?

In the price of old diamonds, we notice in the chart that the carat and price correlation is not strong, explains it the diverse range and regression is not linear. that affected by factors such as cut, color and clarity.

As for the expected price, there is a linear regression and a strong relationship between the prices and the carat, but we notice the presence of negative prices, which makes linear regression not the correct choice to predict the new prices.

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.

I recommend a bid of $8,800,142.

By using the Methodology Map to choose the correct methodology to solve the business problem, the outcome to predict is the amount of prices, which is continuous, numeric variable.
Since the linear regression method did not solve the problem, I used the random Decision Forest (75% of sum the predicting prices).