<p align="center"><u>Project: Diamond Prices</u></p>

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?

   - According to the linear regression model developed, if one diamond is 1 carat heavier than another with the same cut (and clarity), then it would result in an additional $8,413 in price. The formula created by the regression determined that the coefficient for a carat is 8413, so for every increase in 1 in the number of carats, the price will increase by the amount of the coefficient.
   - Regression formula is as given below:
     Price = -5269 + 8413 x **Carat** + 158.1 x **Cut** + 454 x **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?

   - Regression formula is as given below:
     Price = -5269 + 8413 x **Carat** + 158.1 x **Cut** + 454 x **Clarity**
   - So, now we plug in the values for the different variables as given in the question (Carat = 1.5, Cut = 3, Clarity = 5)
   - **Price** = 10094.80

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.
   - **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?

Carats vs. Price
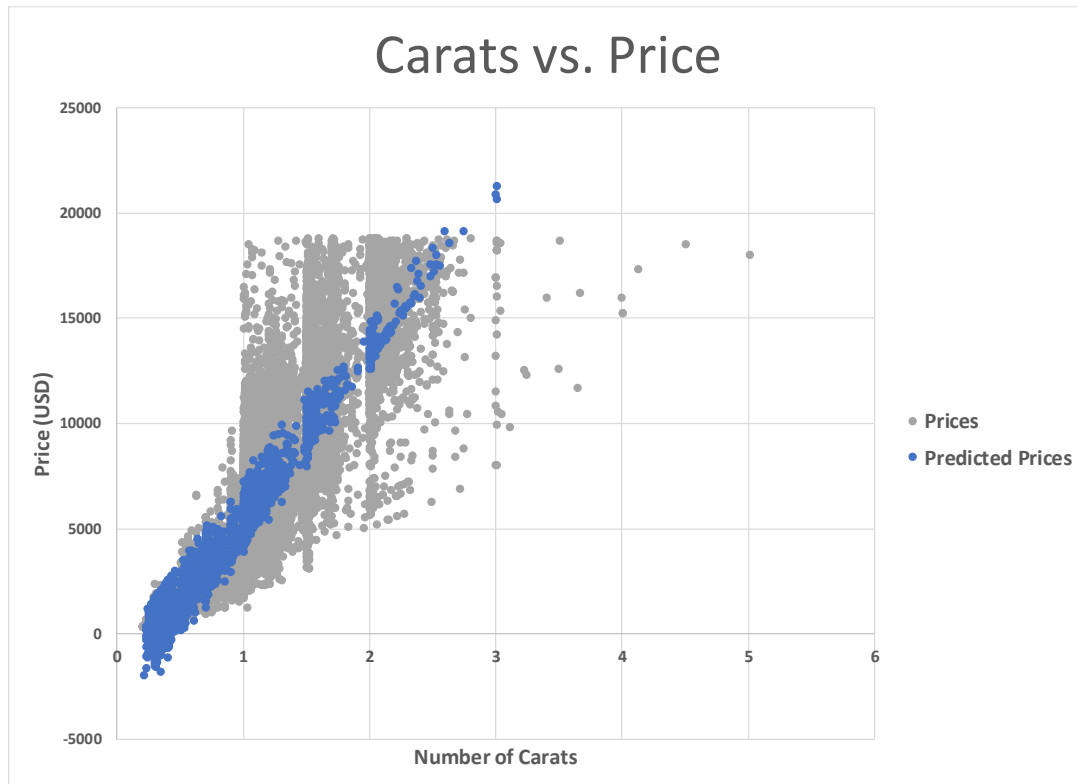
The predicted prices are much more compact than the actual data and conforms much more strongly to a linear model than the actual prices. The actual prices data is very spread out from the model and shows a large variance. This is possibly because we are not accounting for all factors that affect the diamond prices. There are many more factors other than the carats that affect the prices. Though we are accounting for clarity and cut, there may be other factors that affect prices such as the shape, the fluorescence, color etc. There may also be other factors such as the mine from which it was sourced, the specific bourse where it was first sold.

It must also be noted that the model predicts negative prices for several diamonds. This is the portion of the blue dots in the plot that are below the 0 line of the prices. For actual diamonds, the prices can never be negative. So this part of the model is erroneous. Therefore, it might be prudent to ex

Given these factors, the model on averages seems to predict the prices ok. But it is off for a significant portion of the diamonds dataset. There are many outliers. While it may not predict individual diamonds accurately, it can be used to predict the price for a large and varied set of diamonds, since on average the model works. Still, it is recommended that other factors be included in the model to make the predictions better.

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 $ 8,213,465.93. I arrived at this figure by using the formula from the regression model provided that was determined based on previous diamond prices and applied it to the set of diamonds that were up for bid. This determined the retail price that the customers will pay for the diamonds. I then factored in the margin the investors were looking for which was 30%, since the company generally purchases diamonds from distributors at 70% of the retail price. So, I multiplied the predicted amount (retail price) of $ 11,733,522.76 by 0.70 to get the final predicted bid of $ 8,213,465.93