Project: Diamond Prices

Step 1: Understanding the Model

Answer the following questions:

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**

- **The one additional Carat would result in an additional $8,413 in price. The formula created by the regression determined that the coefficient for a Carat is 8,413, So for every increase in the number of Carat the price will increase by the amount 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 formula is 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)

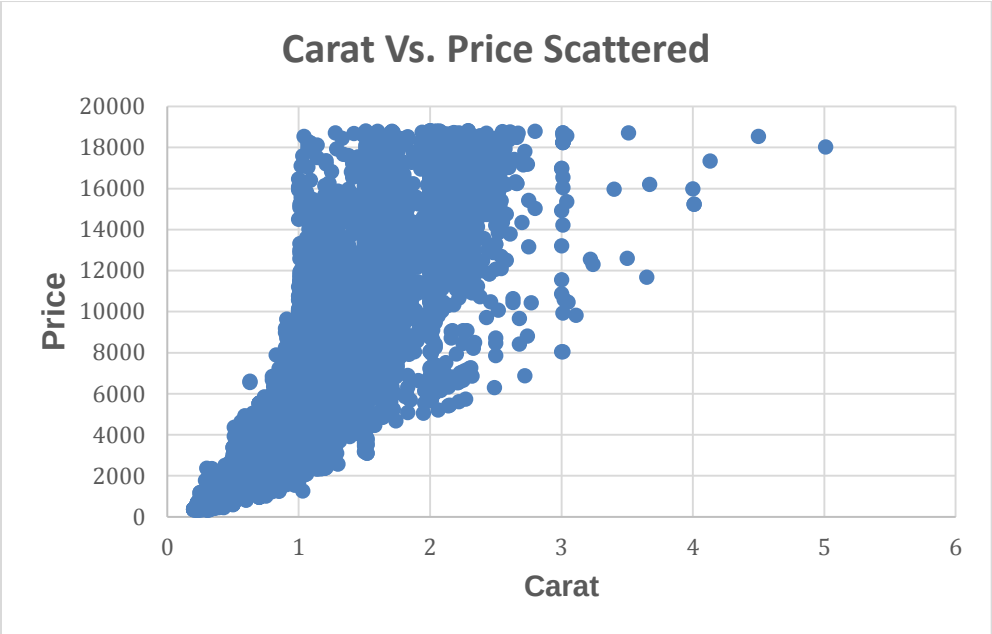
  = 10,094.8

  The formula predict that I will pay the amount of $ 10,094.8

Step 2: Visualize the Data

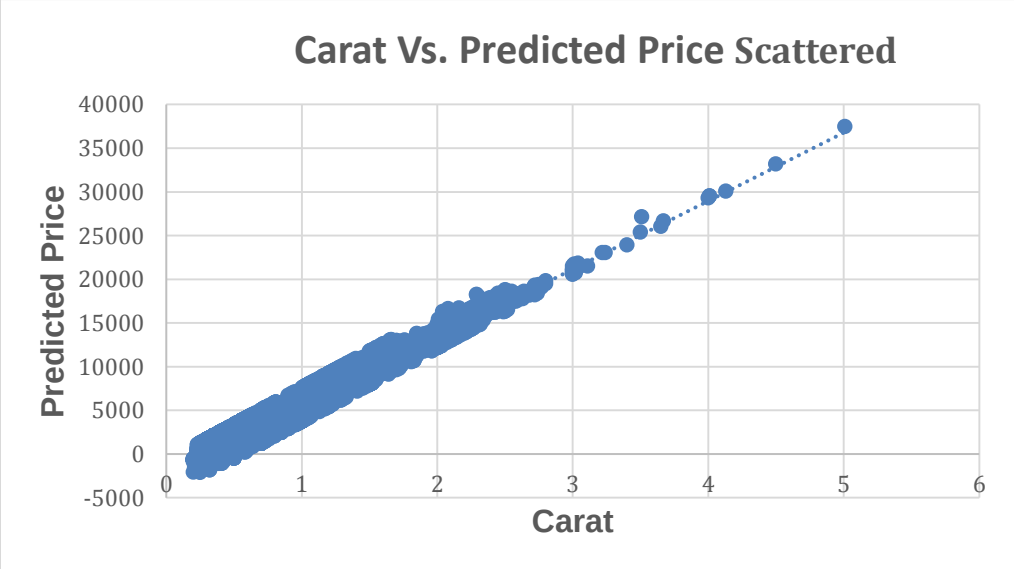
Make sure to plot and include the visualizations in this report. For example, you can create graphs in Excel and copy and paste the graphs into this Word document.

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.

The formula is price = -5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity



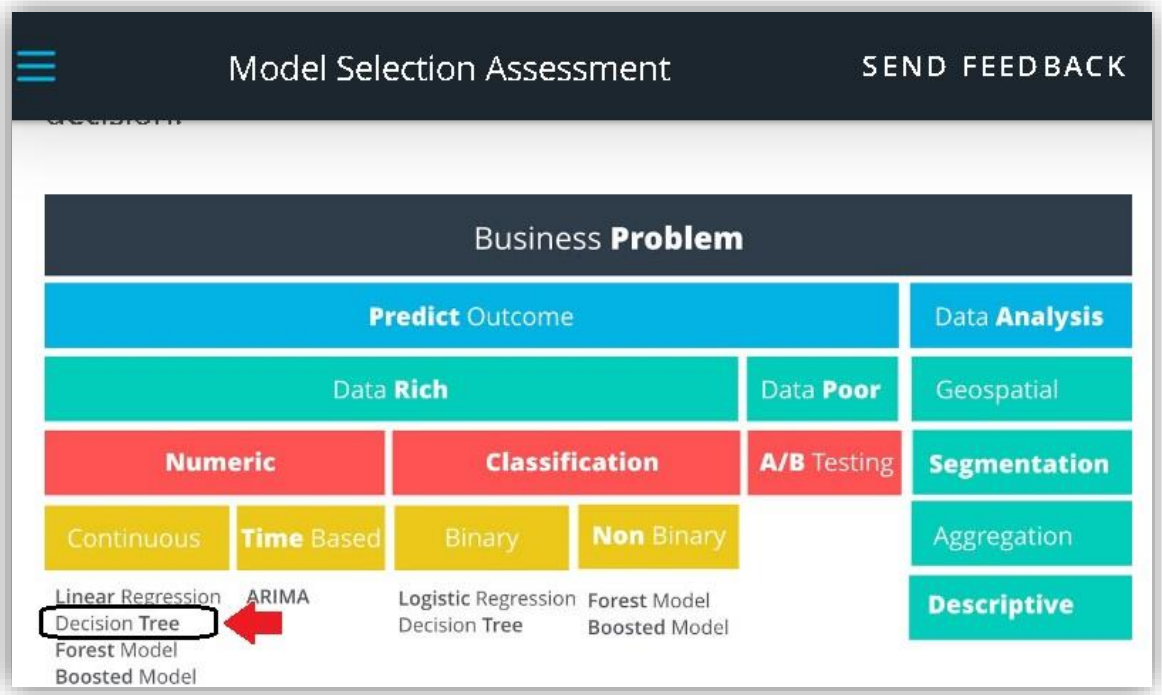
3. What strikes you about this comparison? After seeing this plot, do you feel confident in the model's ability to predict prices?
   - **<u>On Plot#1 (Carat Vs. Price Scattered) - Old available data:</u>**

     We can see that Linear Regression Model working perfect for variables 0 & 1 in X-axis but starting from variable 2 and further there were a very high diversity caused by other variables such as Color & Clarity and we can notice that the linear is not fine and the correlation is bad.

   - **<u>On Plot#2 (Carat Vs. Predicted Price Scattered) - New data:</u>**

     We can see a very strong Positive correlation between Carat & predicted price in the plot also noticed that many Negative variables in the predicted price (y-axis) which is not reasonable & acceptable.

In conclusion, I found that since there are a big diversify in plot#1 and many negative variables in predicted price in plot#2 that Linear Regression Model is not the right Model to solve this problem therefor I choose to solve it Using Decision Tree Model mentioned in the Methodology Map as shown below.



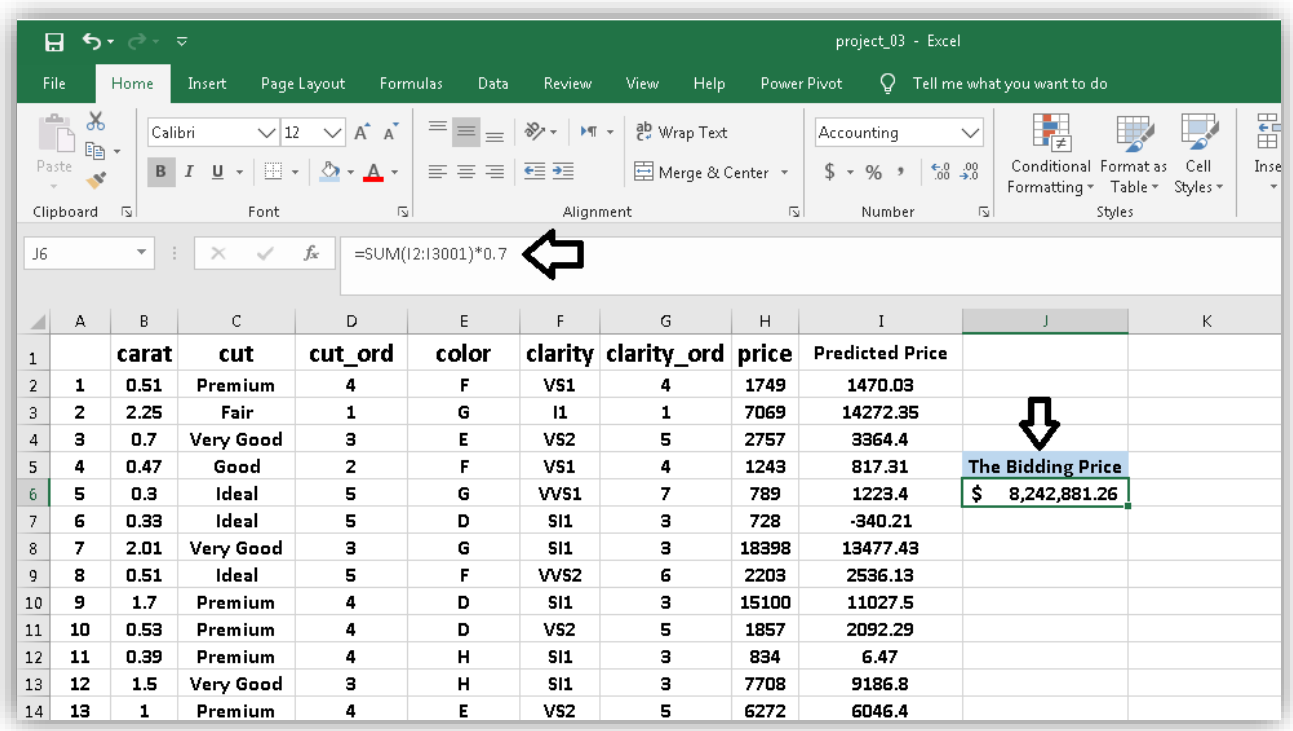
Step 3: Make a Recommendation

Answer the following questions:

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

- **By applying Decision Tree Model, I calculate the prices of 3000 Diamonds from predicted prices shown below and multiply it with % 70 to get the Bid price of:**

  $ 8,242,881.26



	A	B	C	D	E	F	G	H	I	J	K
1		carat	cut	cut_ord	color	clarity	clarity_ord	price	Predicted Price		
2	1	0.51	Premium	4	F	VS1	4	1749	1470.03		
3	2	2.25	Fair	1	G	I1	1	7069	14272.35		
4	3	0.7	Very Good	3	E	VS2	5	2757	3364.4		
5	4	0.47	Good	2	F	VS1	4	1243	817.31	The Bidding Price	
6	5	0.3	Ideal	5	G	VVS1	7	789	1223.4	$ 8,242,881.26	
7	6	0.33	Ideal	5	D	SI1	3	728	-340.21		
8	7	2.01	Very Good	3	G	SI1	3	18398	13477.43		
9	8	0.51	Ideal	5	F	VVS2	6	2203	2536.13		
10	9	1.7	Premium	4	D	SI1	3	15100	11027.5		
11	10	0.53	Premium	4	D	VS2	5	1857	2092.29		
12	11	0.39	Premium	4	H	SI1	3	834	6.47		
13	12	1.5	Very Good	3	H	SI1	3	7708	9186.8		
14	13	1	Premium	4	E	VS2	5	6272	6046.4		