Project: Diamond Prices

Complete each section. When you are ready, save your file as a PDF document and submit it in your classroom.

Step 1: Understanding the Model

Price = -5,269 + 8,413 x **Carat** + 158.1 x **Cut** + 454 x **Clarity**

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?

   We will answer here based on this equation "Price = -5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity", if we assume that the value of the Cut is the same and the Carat equals one, then the price will increase to (8,413*1 = 8413) Provided that the rest of the variables are fixed which are clarity and cut.

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 first step : is to substitute each variable for its value into this equation:
   Price **= -5,269 + 8,413 x** Carat **+ 158.1 x** Cut **+ 454 x** Clarity
   The second step: compensation
   Price = -5,269 +(8,413 x **1.5**)+ (158.1 x **3**) + (454 x **5)**
   = -5,269+12619.5+474.3+2270
   =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.

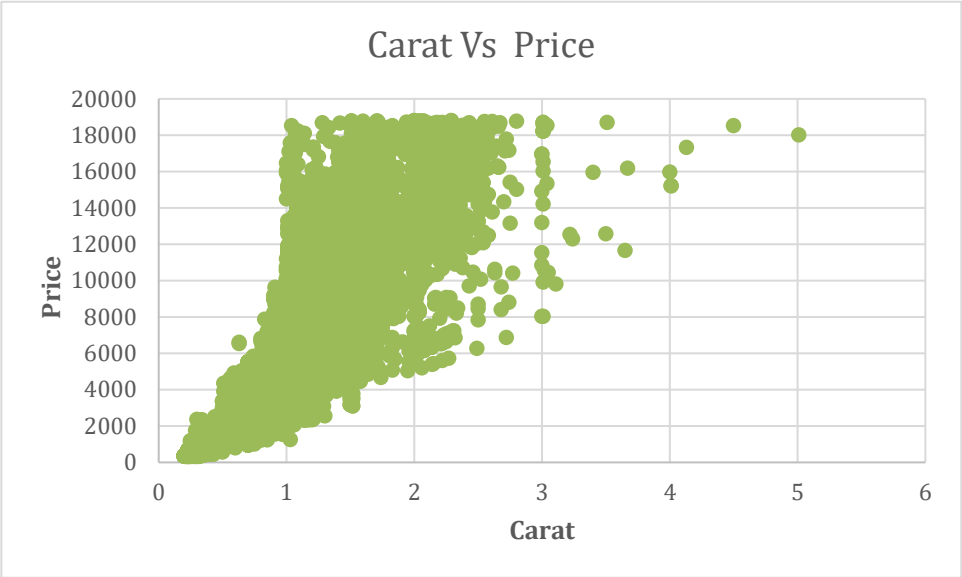


Figure 1

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.

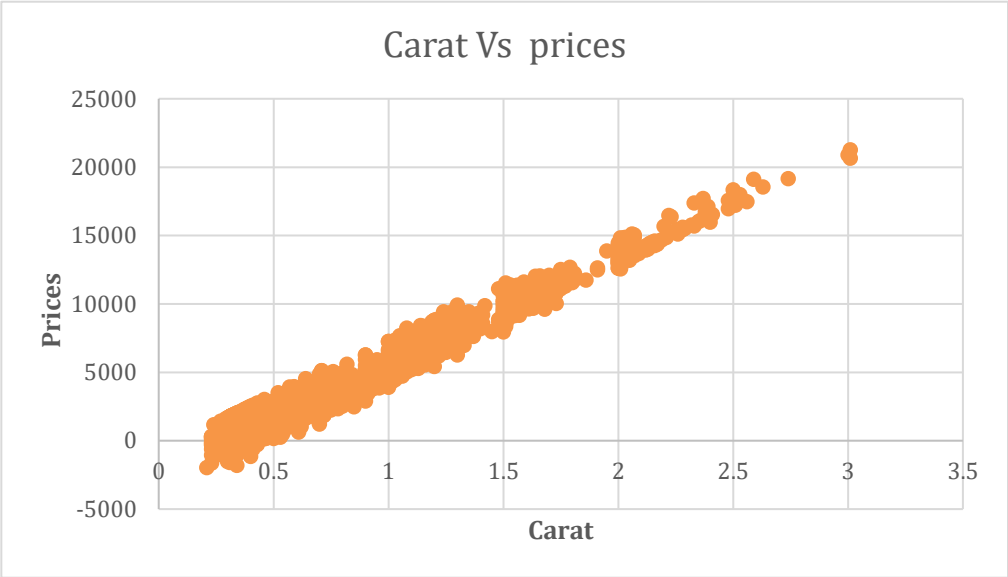


Figure 2

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

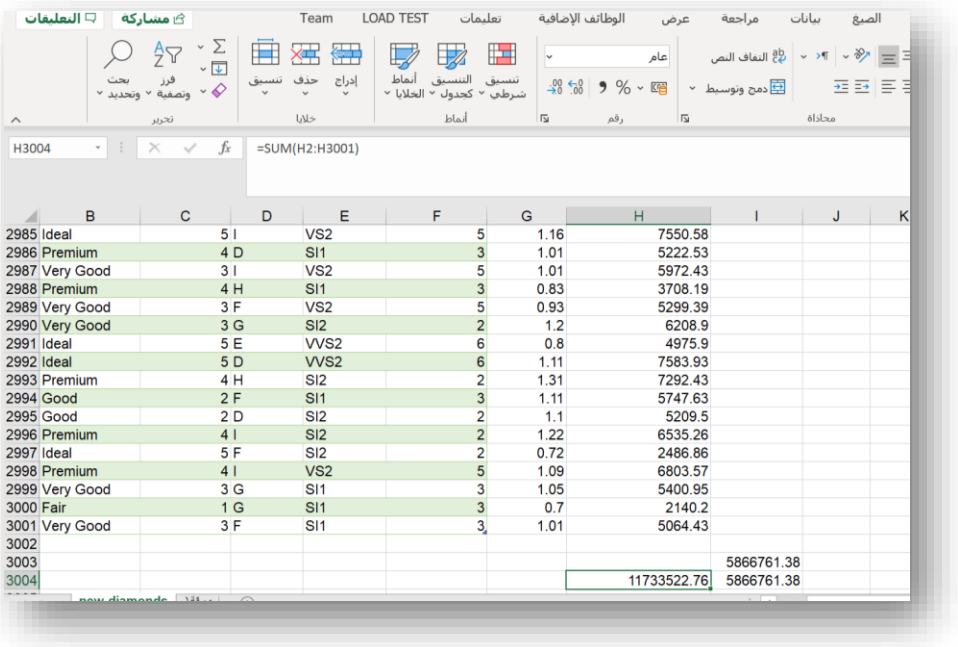
When looking at the figure 1, we notice that the correlation line appeared in the first period, after that the data began to disperse and the line appeared very weak. As for me, the figure 2 appeared, we have a very strong, positive and clear correlation line, where we can predict future values based on the chart, but we have another problem, which is the emergence of negative values in the prices variable, so this gives more than probability to solve this problem. The first probability that the linear regression does not fit that business problem where we can choose other models such as (Forest, decision tree, or boosted model) The second probability is the effected of another variable on the price, like Clarity, Color and Shape and Fluorescence. (1)
http://www.diamondc.com.hk/us/factors-affect-diamond-price

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.

Now, I changed the type of approach due to the negative results in prices . so we used the decision tree approach in forecasting and I collected all the **predicted prices that = 11733522.76** then **took 50 percent** of them and the **result is =5866761.38**
You can see sum function in figure 3



	B	C	D	E	F	G	H	I	J	K
2985	Ideal	5	I	VS2	5	1.16	7550.58			
2986	Premium	4	D	SI1	3	1.01	5222.53			
2987	Very Good	3	I	VS2	5	1.01	5972.43			
2988	Premium	4	H	SI1	3	0.83	3708.19			
2989	Very Good	3	F	VS2	5	0.93	5299.39			
2990	Very Good	3	G	SI2	2	1.2	6208.9			
2991	Ideal	5	E	VVS2	6	0.8	4975.9			
2992	Ideal	5	D	VVS2	6	1.11	7583.93			
2993	Premium	4	H	SI2	2	1.31	7292.43			
2994	Good	2	F	SI1	3	1.11	5747.63			
2995	Good	2	D	SI2	2	1.1	5209.5			
2996	Premium	4	I	SI2	2	1.22	6535.26			
2997	Ideal	5	F	SI2	2	0.72	2486.86			
2998	Premium	4	I	VS2	5	1.09	6803.57			
2999	Very Good	3	G	SI1	3	1.05	5400.95			
3000	Fair	1	G	SI1	3	0.7	2140.2			
3001	Very Good	3	F	SI1	3	1.01	5064.43			
3002										
3003								5866761.38		
3004							11733522.76	5866761.38		

Figure 3