

Project: Diamond Prices

Complete each section. When you are ready, save your file as a PDF document and submit it here: <https://classroom.udacity.com/nanodegrees/nd008/parts/235a5408-0604-4871-8433-a6d670e37bbf/project#>

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?

Since Price = $-5,269 + 8,413 \times \text{Carat} + 158.1 \times \text{Cut} + 454 \times \text{Clarity}$

Then, Lighter one price= $-5,269 + 8,413 \times \text{Carat} + 158.1 \times \text{Cut} + 454 \times \text{Clarity}$

Heavier one price= $-5,269 + 8,413 \times (\text{Carat}+1) + 158.1 \times \text{Cut} + 454 \times \text{Clarity}$

So, we use heavier one price minus the lighter one price= $-5,269 + 8,413 \times (\text{Carat}+1) + 158.1 \times \text{Cut} + 454 \times \text{Clarity} - (-5,269 + 8,413 \times \text{Carat} + 158.1 \times \text{Cut} + 454 \times \text{Clarity}) = 8,413 \times (\text{Carat}+1) - 8,413 \times \text{Carat} = 8,413 \times 1 = 8,413$

Therefore, I expect to pay 8,413 more.

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?

Since Price = $-5,269 + 8,413 \times \text{Carat} + 158.1 \times \text{Cut} + 454 \times \text{Clarity}$

Carat=1.5, cut=3, clarity=5

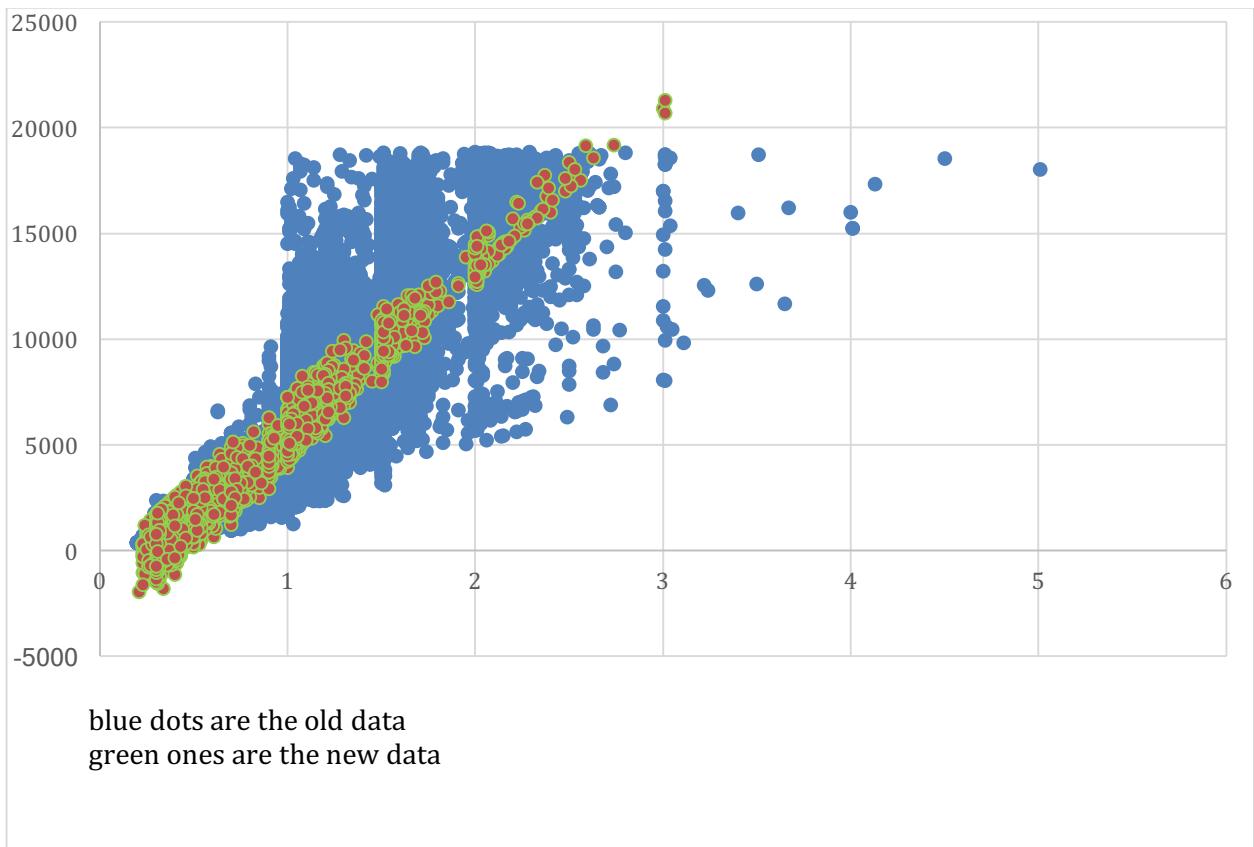
Then, the price= $-5,269 + 8,413 \times 1.5 + 158.1 \times 3 + 454 \times 5 = 10094.8$

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.

- 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.



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

I have two questions

- Why there are some negative numbers in the new data set. Negative price does not exist, there must be something not so good with the equation.
- The new data trend is too concentrated to be true. The old data seems more natural. That means, the equation surely needs to be fine-tuned.

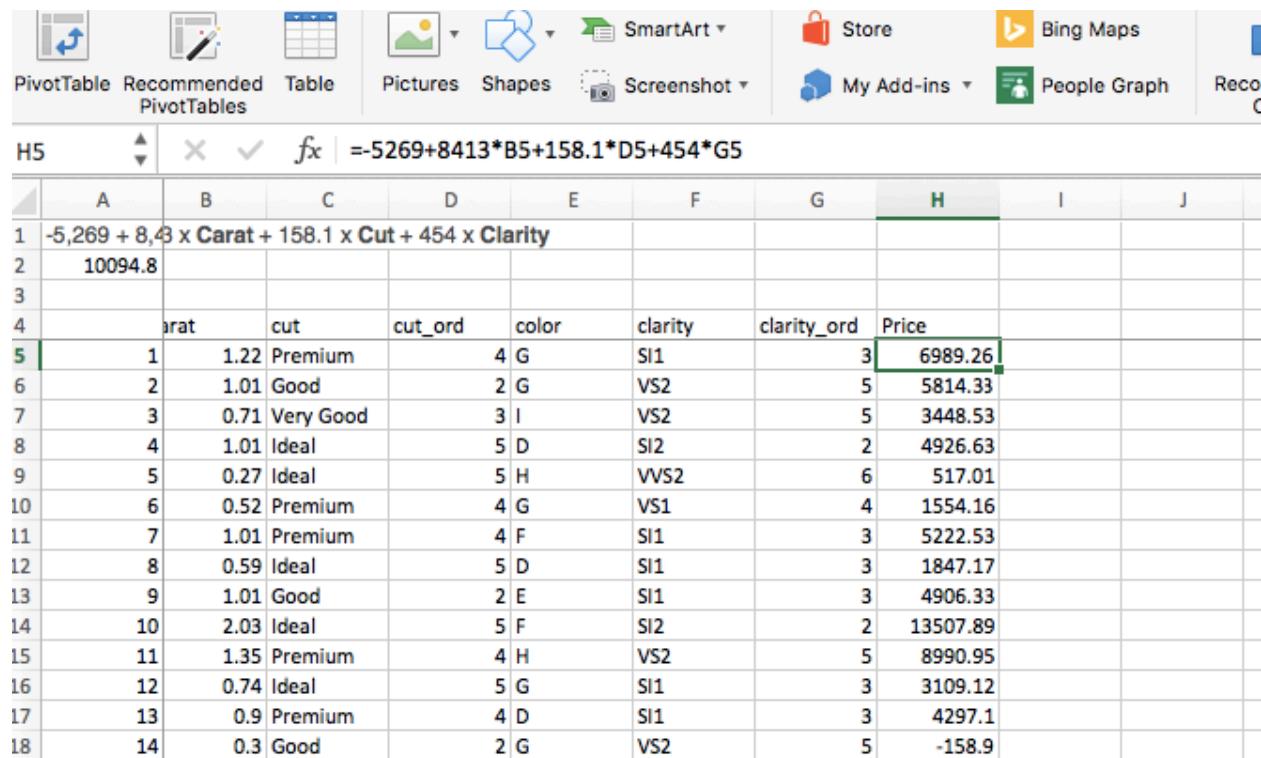
I am not so confident about the model's ability to predict prices. It may result with an approximate number, but could be better. Good news is the two are in same trend line. So the guess would not be too far.

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 put the equation in the new data excel and calculated all the individual prices of the target variables.



The screenshot shows a Microsoft Excel spreadsheet. The formula bar at the top displays the formula $=-5269+8413*B5+158.1*D5+454*G5$. The main area contains a data table with columns labeled A through J. Row 1 contains the header: "1 -5,269 + 8,43 x Carat + 158.1 x Cut + 454 x Clarity". Row 2 contains the value "2 10094.8". Rows 4 through 18 contain data points for various diamonds, with the last row (H18) showing a negative value of "-158.9". The columns represent Carat, cut, cut_ord, color, clarity, clarity_ord, and Price.

	A	B	C	D	E	F	G	H	I	J
1	-5,269 + 8,43 x Carat + 158.1 x Cut + 454 x Clarity									
2	10094.8									
4	Carat	cut	cut_ord	color	clarity	clarity_ord	Price			
5	1	1.22	Premium	4 G	SI1	3	6989.26			
6	2	1.01	Good	2 G	VS2	5	5814.33			
7	3	0.71	Very Good	3 I	VS2	5	3448.53			
8	4	1.01	Ideal	5 D	SI2	2	4926.63			
9	5	0.27	Ideal	5 H	VVS2	6	517.01			
10	6	0.52	Premium	4 G	VS1	4	1554.16			
11	7	1.01	Premium	4 F	SI1	3	5222.53			
12	8	0.59	Ideal	5 D	SI1	3	1847.17			
13	9	1.01	Good	2 E	SI1	3	4906.33			
14	10	2.03	Ideal	5 F	SI2	2	13507.89			
15	11	1.35	Premium	4 H	VS2	5	8990.95			
16	12	0.74	Ideal	5 G	SI1	3	3109.12			
17	13	0.9	Premium	4 D	SI1	3	4297.1			
18	14	0.3	Good	2 G	VS2	5	-158.9			

Then I calculated the total of the individual prices to get a total set price for retail.

Table Recommended Table Pictures Shapes Screenshot My Add-ins People Graph

D5 $\Delta \times \checkmark f\!x =\text{SUM}(\text{H5:H3004})$

A	B	C	D	E	F	G	H	I
5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity								
	carat	cut	cut_ord	color	clarity	clarity_ord	Price	
2981	0.3	Ideal		5 G	VVS2	6	769.4	
2982	1.51	Premium		4 E	SI1	3	9429.03	
2983	1.54	Ideal		5 J	VS2	5	10747.52	
2984	1.16	Ideal		5 I	VS2	5	7550.58	
2985	1.01	Premium		4 D	SI1	3	5222.53	
2986	1.01	Very Good		3 I	VS2	5	5972.43	
2987	0.83	Premium		4 H	SI1	3	3708.19	
2988	0.93	Very Good		3 F	VS2	5	5299.39	
2989	1.2	Very Good		3 G	SI2	2	6208.9	
2990	0.8	Ideal		5 E	VVS2	6	4975.9	
2991	1.11	Ideal		5 D	VVS2	6	7583.93	
2992	1.31	Premium		4 H	SI2	2	7292.43	
2993	1.11	Good		2 F	SI1	3	5747.63	
2994	1.1	Good		2 D	SI2	2	5209.5	
2995	1.22	Premium		4 I	SI2	2	6535.26	
2996	0.72	Ideal		5 F	SI2	2	2486.86	
2997	1.09	Premium		4 I	VS2	5	6803.57	
2998	1.05	Very Good		3 G	SI1	3	5400.95	
2999	0.7	Fair		1 G	SI1	3	2140.2	
3000	1.01	Very Good		3 F	SI1	3	5064.43	
							11733522.8	

At the end, I include the 75% discount for wholesale price, and get the final bid price as 8,800,142.07.

If you prefer the US number system with decimal separate, I could adjust the cell format.

5	D	VVS2	6	7583.93
4	H	SI2	2	7292.43
2	F	SI1	3	5747.63
2	D	SI2	2	5209.5
4	I	SI2	2	6535.26
5	F	SI2	2	2486.86
4	I	VS2	5	6803.57
3	G	SI1	3	5400.95
1	G	SI1	3	2140.2
3	F	SI1	3	5064.43
				11733522.8
				8,800,142.07

PivotTable Recommended Table | Pictures Shapes Screenshot ▾ My Add-ins ▾ People Graph

A1006 fx =0.75*H3005

	A	B	C	D	E	F	G	H	I
		-5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity							
	1006								
		10094.8							
		carat	cut	cut_ord	color	clarity	clarity_ord	Price	
1	2981	0.3	Ideal	5	G	VVS2	6	769.4	
2	2982	1.51	Premium	4	E	SI1	3	9429.03	
3	2983	1.54	Ideal	5	J	VS2	5	10747.52	
4	2984	1.16	Ideal	5	I	VS2	5	7550.58	
5	2985	1.01	Premium	4	D	SI1	3	5222.53	
6	2986	1.01	Very Good	3	I	VS2	5	5972.43	
7	2987	0.83	Premium	4	H	SI1	3	3708.19	
8	2988	0.93	Very Good	3	F	VS2	5	5299.39	
9	2989	1.2	Very Good	3	G	SI2	2	6208.9	
10	2990	0.8	Ideal	5	E	VVS2	6	4975.9	
11	2991	1.11	Ideal	5	D	VVS2	6	7583.93	
12	2992	1.31	Premium	4	H	SI2	2	7292.43	
13	2993	1.11	Good	2	F	SI1	3	5747.63	
14	2994	1.1	Good	2	D	SI2	2	5209.5	
15	2995	1.22	Premium	4	I	SI2	2	6535.26	
16	2996	0.72	Ideal	5	F	SI2	2	2486.86	
17	2997	1.09	Premium	4	I	VS2	5	6803.57	
18	2998	1.05	Very Good	3	G	SI1	3	5400.95	
19	2999	0.7	Fair	1	G	SI1	3	2140.2	
20	3000	1.01	Very Good	3	F	SI1	3	5064.43	
								11733522.8	
								8800142.07	