Predicting Diamond Prices

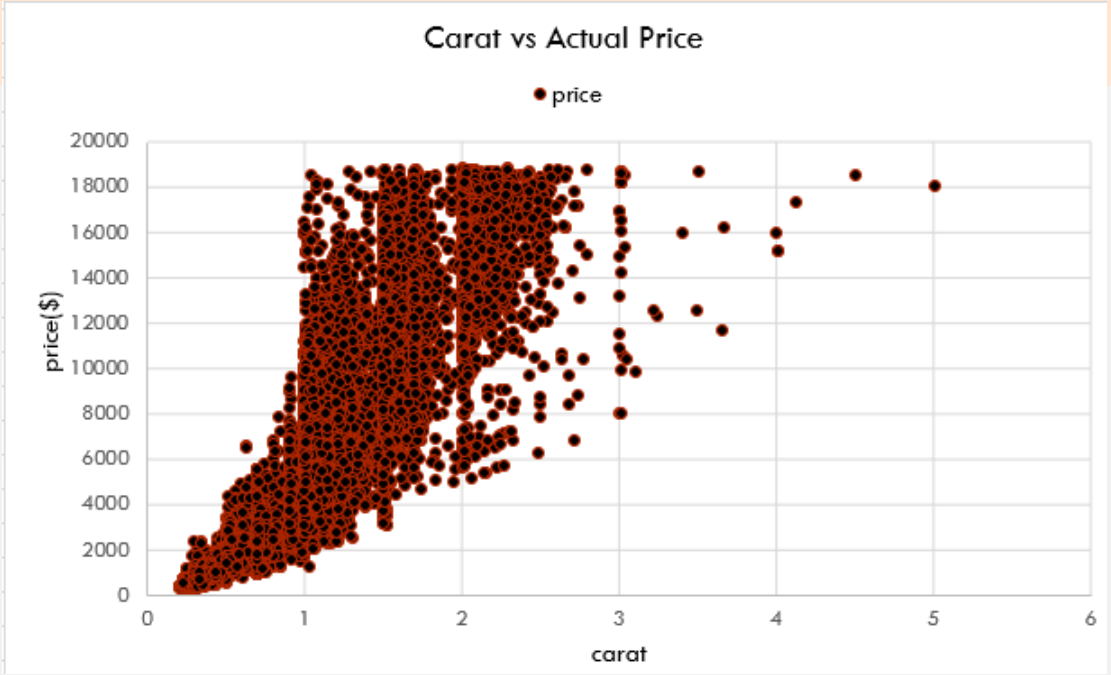
Detail Analysis using Excel

by Hafizah Ab Rahim, MARCH 2021

TRAINING DATA:
- 50,000 rows of data information
- 6 predictors
- target variable is price
- contains the data used to build regression model

Equation provided:
Price = -5269 + 8413*Carat + 158.1*Cut + 454*Clarity

Variables:
Carat - the weight of the diamond, and is a numerical variable.
Cut - the quality of the cut of the diamond, and falls into 5 categories: fair, good, very good, ideal, and premium. Each of these categories are represented by a number, 1-5, in the *Cut_Ord* variable.
Clarity - the internal purity of the diamond, and falls into 8 categories: I1, SI2, SI1, VS1, VS2, VVS2, VVS1, and IF. Each of these categories are represented by a number, 1-8, in the *Clarity_Ord* variable.



	carat	cut	cut_ord	color	clarity	clarity_ord	price
1	0.51	Premium	4	F	VS1	4	1749
2	2.25	Fair	1	G	I1	1	7069
3	0.7	Very Good	3	E	VS2	5	2757
4	0.47	Good	2	F	VS1	4	1243
5	0.3	Ideal	5	G	VVS1	7	789
6	0.33	Ideal	5	D	SI1	3	728
7	2.01	Very Good	3	G	SI1	3	18398
8	0.51	Ideal	5	F	VVS2	6	2203
9	1.7	Premium	4	D	SI1	3	15100
10	0.53	Premium	4	D	VS2	5	1857
11	0.39	Premium	4	H	SI1	3	834
12	1.5	Very Good	3	H	SI1	3	7708
13	1	Premium	4	E	VS2	5	6272
14	1.29	Ideal	5	J	VS1	4	5676
15	2.01	Good	2	D	SI2	2	16776
16	1.13	Ideal	5	G	VS1	4	7404
17	0.7	Ideal	5	I	SI2	2	1702
18	0.38	Very Good	3	I	VS1	4	606
19	1.17	Ideal	5	H	SI2	2	5423
20	1.51	Premium	4	F	SI1	3	8033
21	0.4	Ideal	5	D	VVS1	7	1279
22	0.41	Very Good	3	F	VS2	5	863
23	0.51	Ideal	5	G	VVS1	7	1893

	carat	cut	cut_ord	color	clarity	clarity_ord	predicted_price
1	1.22	Premium	4	G	SI1	3	6989.26
2	1.01	Good	2	G	VS2	5	5814.33
3	0.71	Very Good	3	I	VS2	5	3448.53
4	1.01	Ideal	5	D	SI2	2	4926.63
5	0.27	Ideal	5	H	VVS2	6	517.01
6	0.52	Premium	4	G	VS1	4	1554.16
7	1.01	Premium	4	F	SI1	3	5222.53
8	0.59	Ideal	5	D	SI1	3	1847.17
9	1.01	Good	2	E	SI1	3	4906.33
10	2.03	Ideal	5	F	SI2	2	13507.89
11	1.35	Premium	4	H	VS2	5	8990.95
12	0.74	Ideal	5	G	SI1	3	3109.12
13	0.9	Premium	4	D	SI1	3	4297.1
14	0.3	Good	2	G	VS2	5	-158.9
15	1.01	Good	2	F	VS2	5	5814.33
16	1.02	Good	2	H	SI2	2	4536.46
17	2.05	Premium	4	G	SI1	3	13972.05
18	0.54	Ideal	5	I	SI1	3	1426.52
19	0.72	Ideal	5	G	VS2	5	3848.86
20	2	Premium	4	J	SI2	2	13097.4
21	1.57	Premium	4	G	SI2	2	9479.81
22	0.89	Premium	4	G	SI1	3	4212.97

According to the model, if a diamond is 1 carat heavier than another with the same cut and clarity, how much more should I expect to pay? Why?

You are expected to pay an extra of $8413 for every increment of 1 carat with the assumption that the cut and the clarity of the diamond are the same. This extra value comes from the coefficient of Carat in the model.

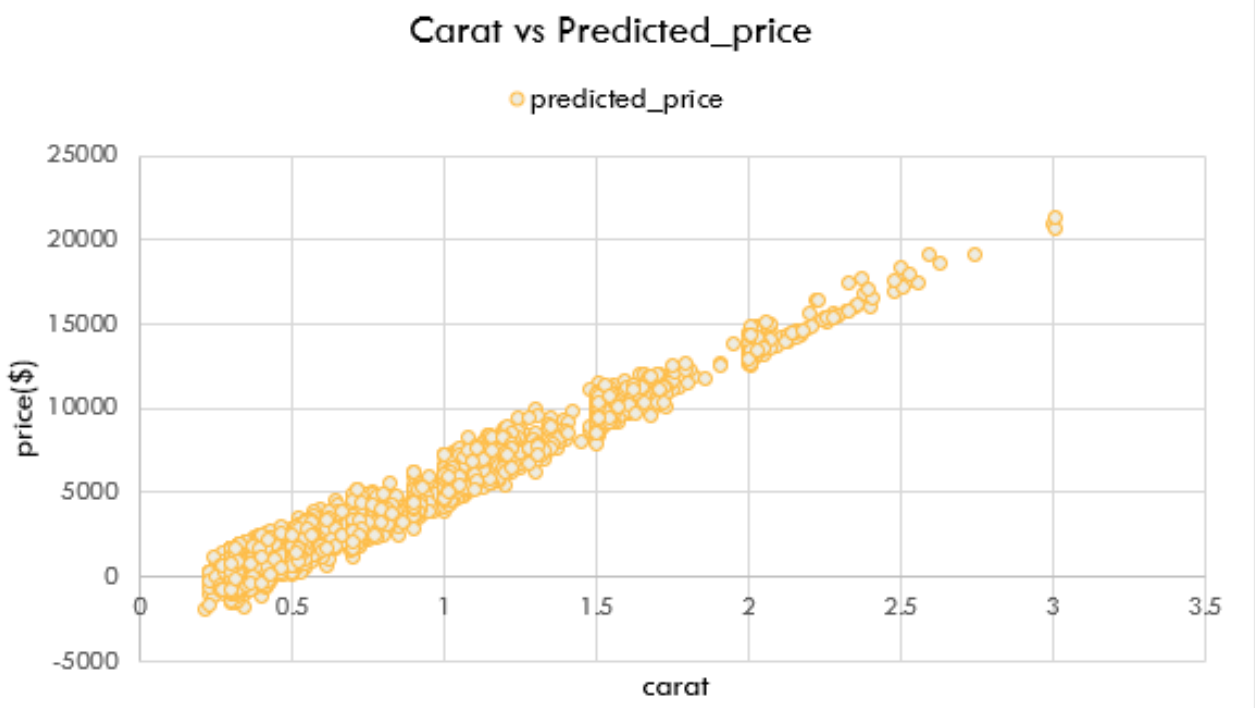
Price of diamond with carat = 1:
= -5269 + 8413*Carat + 158.1*Cut + 454*Clarity
= -5269 + 8413*1 + 158.1*4 + 454*3
= **$ 5138.4**

Price of diamond with carat = 2:
= -5269 + 8413*Carat + 158.1*Cut + 454*Clarity
= -5269 + 8413*2 + 158.1*4 + 454*3
= **$ 13551.4**

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 model predicts that I should pay $ 10094.80 for a diamond with the above specification.

Calculating price when carat = 1.5, cut = 3, clarity = 5
= -5269 + 8413*Carat + 158.1*Cut + 454*Clarity
= -5269 + 8413*1.5 + 158.1*3 + 454*5
= $ 10094.8



by Hafizah Ab Rahim, MARCH 2021

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?

--Answer--
You are expected to pay an additional amount between the range of $ 8413 and $ 11591 inclusive. Based on the model provided, the predictor variables that affect the model are carat, cut and clarity. It is important to take into consideration the possible changes in the value of clarity which affect the final additional amount as the carat increased by 1 carat.

--Reason and explanation--
The minimum and maximum values of clarity are 1 and 8 respectively.

Calculating base amount
Price of diamond with Carat = 1, Cut = 1 Clarity = 1 :
= -5269 + 8413*Carat + 158.1*Cut + 454*Clarity
= -5269 + 8413*1 + 158.1*1 + 454*1
= $ 3756.1

Carat is increased to 2 and other values stay the same
Price of diamond with Carat = 2, Cut = 1 Clarity = 1 :
= -5269 + 8413*Carat + 158.1*Cut + 454*Clarity
= -5269 + 8413*2 + 158.1*1 + 454*1
= $ 12169.1

So here, we can find the minimum amount of additional price when the diamond is 1 carat heavier and clarity is at minimum.

Minimum additional price: $ 12169.1 - $ 3756.1 = $ 8413

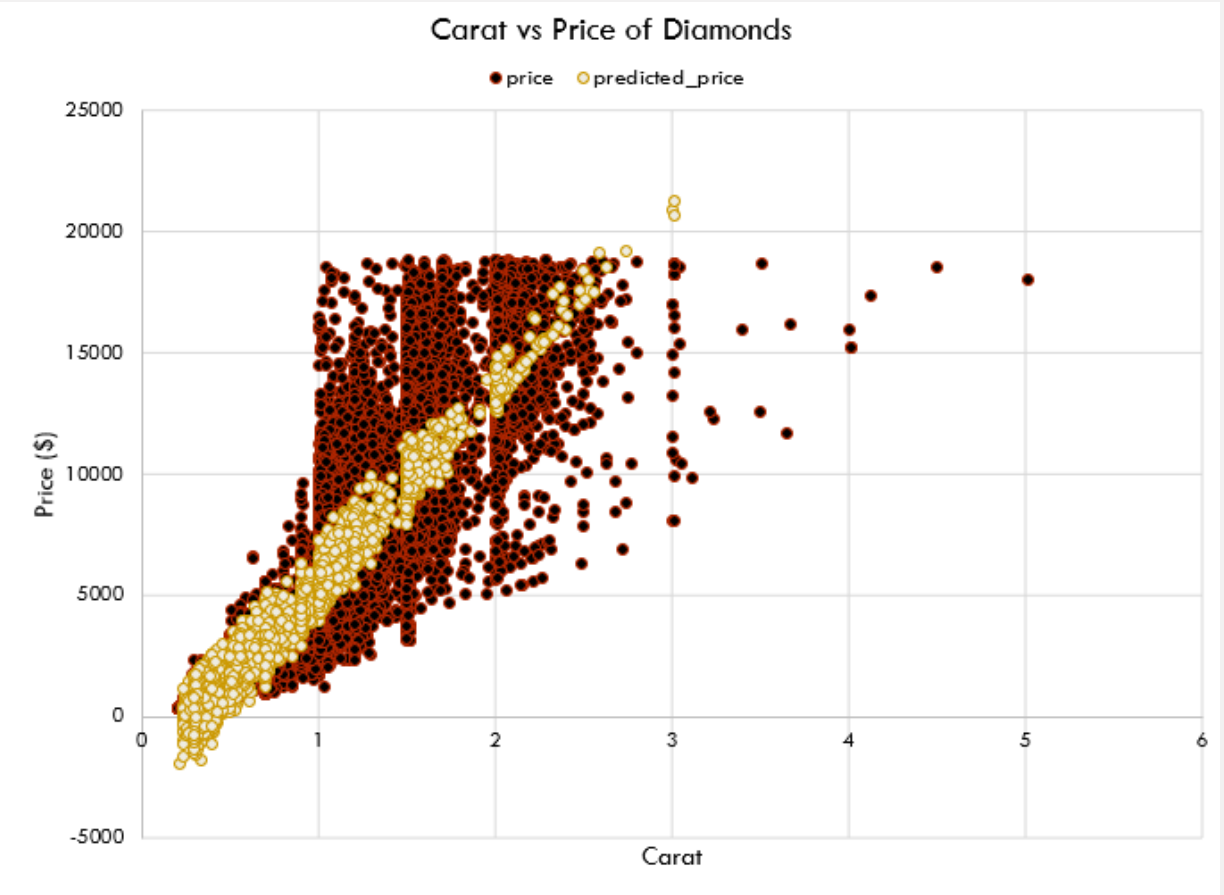
Carat stays at 2 and clarity is at maximum
Price of diamond with Carat = 2, Cut = 1 Clarity = 8 :
= -5269 + 8413*Carat + 158.1*Cut + 454*Clarity
= -5269 + 8413*2 + 158.1*1 + 454*8
= $ 15347.1

Similarly here, we can find the maximum amount of additional price when the diamond is 1 carat heavier and clarity is at maximum.

Maximum additional price : $ 15347.1 - $ 3756.1 = $ 11591

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.

The predicted prices for diamonds are closer together compared to the actual prices. I noticed a linear trend associated with both, the actual prices and predicted prices. However, the actual prices are more spread out. Next, the model predicted negative prices for 291 diamonds that have less than 1 carat.
Thus, I do not feel confident with the model's ability to predict prices and believe that a better model can be generated by including additional predictors.



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

Answer:
I would recommend the jewelry company to offer a bid of $ 8309267.40 for the whole set of diamonds.

Explanation:
The model predicted negative prices for 291 count of diamonds totaling to $ 136860. There are 2709 count of diamond with positive prices totaling to $ 11870382. Since only 10.7% of the diamonds are negatively priced, I approximated these prices to $ 0. The diamonds are usually bought at 70% of the price, so the final price is $ 8309267.40.

=SUM(D2:D3001) → $11,733,523 (All prices included)
=SUMIF(D:D,">0") → $11,870,382 (Only positive prices included)
=SUMIF(D:D,"<0") → $136,860 (Only negative prices included)
=COUNTIF(D:D,">0") → 2709 (Total diamond with positive prices)
=COUNTIF(D:D,"<0") → 291 (Total diamond with negative prices)
=$11,870,382 *0.7 → $8,309,267.40 (Final price for the whole set)

49989	1.03	7328		
49990	0.47	1058		
49991	1.01	4989		
49992	1.21	7786		
49993	0.43	848		
49994	0.7	3105		
49995	0.5	1449		
49996	0.33	692		
49997	0.71	2918		
49998	0.43	1056		
49999	1.14	6619		
50000	1.01	6787		
50001	1.77	9428		
50002	1.22		6989.26	
50003	1.01		5814.33	
50004	0.71		3448.53	
50005	1.01		4926.63	
50006	0.27		517.01	
50007	0.52		1554.16	
50008	1.01		5222.53	
50009	0.59		1847.17	
50010	1.01		4906.33	
50011	2.03		13507.89	
50012	1.35		8990.95	
50013	0.74		3109.12	
50014	0.9		4297.1	
50015	0.3		-158.9	
50016	1.01		5814.33	
50017	1.02		4536.46	
50018	2.05		13972.05	
50019	0.54		1426.52	

Make sure to add these data separately to view 2 graphs