

## Project: Diamond Prices

### Step 1: Understanding the Model

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

- 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?

the coefficient of a carat is " 8,413 " that means each carat is heavier than the other with the same cut . will have a raise in Prince and the increase will be \$ 8,413

- 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?

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

$$\text{Price} = -5,269 + 8,413 \times 1.5 + 158.1 \times 3 + 454 \times 5$$

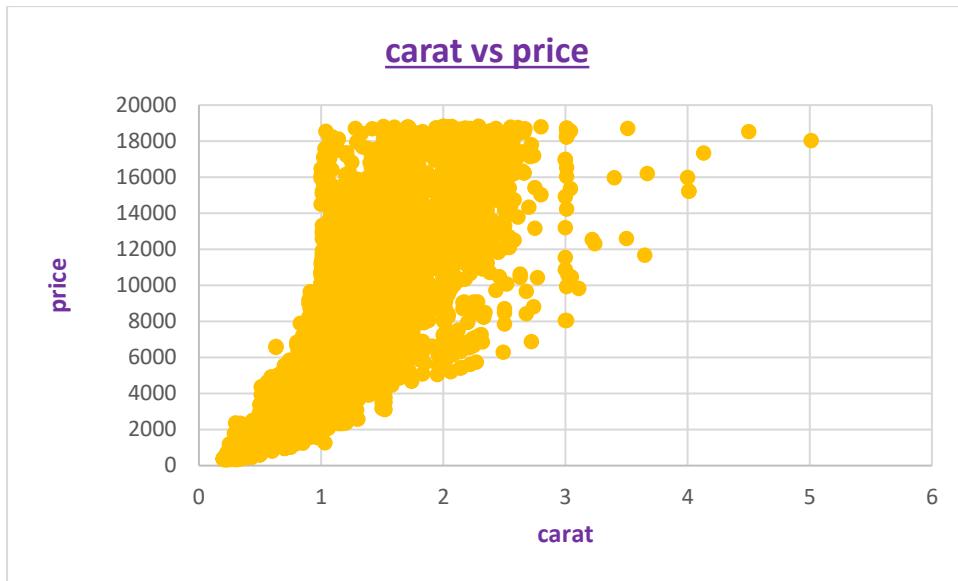
we have to pay \$ 10,094.8 for 1.5 carats of diamonds.

### 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.

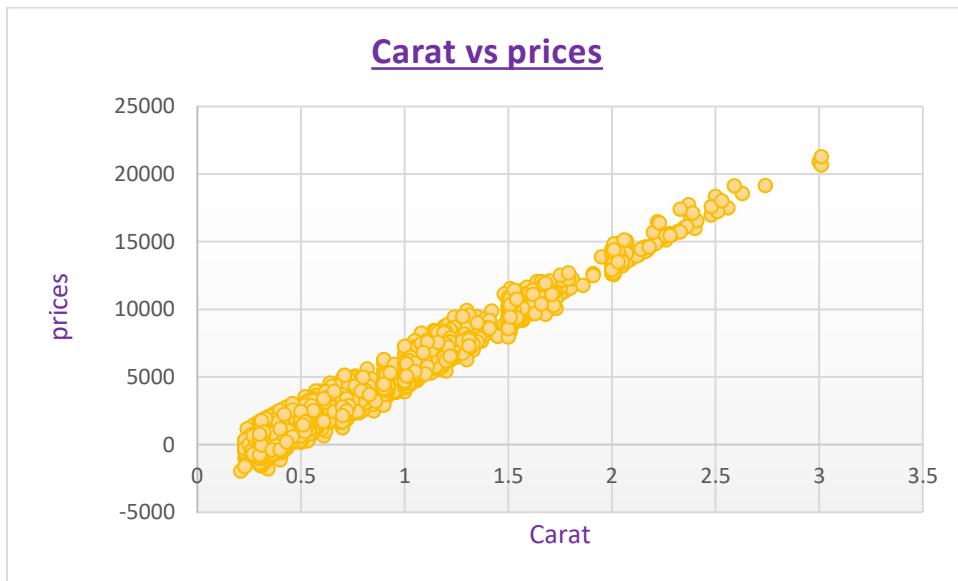
- Plot 1 - Plot the data for the diamonds in the database, with carat on the x-axis and price on the y-axis.

### diamonds in the database



- 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.

### predicting prices



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

**No, because we have prices and values less than zero, and this is an illogical loss.**

The **old price diamonds chart** show a good start from 0 to 1.5 that is look like linear but after that, from 1.5 to 5, a large dispersion appeared.

The **new diamonds chart** is that it shows a positive and strong straight line that may look good, but there are very many values less than zero and this is not good at all although it seems strong but it is not useful in this case so we should change the model "linear regression" to predict prices .

## Step 3: Make a Recommendation

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

Numeric		Classification	
Continuous	Time Based	Binary	Non Binary
Linear Regression	ARIMA	Logistic Regression	Forest Model
Decision Tree	ETS	Decision Tree	Boosted Model
Forest Model			
Boosted Model			

As we have said previously although it seems strong but it is not useful in this case so we should change the model " **linear regression** " to predict prices , So we trying to do " **logistic regression** " approach so we will sum all predict prices and take the percentage ... **Sum predict prices \* 0.8**

**So recommend price = 9,386,818\$**

	Column1	cut	cut_ord	color	clarity	clarity_	carat		predicting pric	Column2				
2975	2974	Fair	1	G	VS1	4	0.46		575.08					
2976	2975	Good	2	I	VS2	5	0.79		3963.47					
2977	2976	Ideal	5	F	VVS2	6	0.32		937.66					
2978	2977	Ideal	5	F	VS2	5	0.4		1156.7					
2979	2978	Fair	1	E	VS2	5	0.51		1449.73					
2980	2979	Premium	4	J	VS1	4	1.05		6013.05					
2981	2980	Very Good	3	H	SI1	3	0.43		184.89					
2982	2981	Ideal	5	G	VVS2	6	0.3		769.4					
2983	2982	Premium	4	E	SI1	3	1.51		9429.03					
2984	2983	Ideal	5	J	VS2	5	1.54		10747.52					
2985	2984	Ideal	5	I	VS2	5	1.16		7550.58					
2986	2985	Premium	4	D	SI1	3	1.01		5222.53					
2987	2986	Very Good	3	I	VS2	5	1.01		5972.43					
2988	2987	Premium	4	H	SI1	3	0.83		3708.19					
2989	2988	Very Good	3	F	VS2	5	0.93		5299.39					
2990	2989	Very Good	3	G	SI2	2	1.2		6208.9					
2991	2990	Ideal	5	E	VVS2	6	0.8		4975.9					
2992	2991	Ideal	5	D	VVS2	6	1.11		7583.93					
2993	2992	Premium	4	H	SI2	2	1.31		7292.43					
2994	2993	Good	2	F	SI1	3	1.11		5747.63					
2995	2994	Good	2	D	SI2	2	1.1		5209.5					
2996	2995	Premium	4	I	SI2	2	1.22		6535.26					
2997	2996	Ideal	5	F	SI2	2	0.72		2486.86					
2998	2997	Premium	4	I	VS2	5	1.09		6803.57					
2999	2998	Very Good	3	G	SI1	3	1.05		5400.95					
3000	2999	Fair	1	G	SI1	3	0.7		2140.2					
3001	3000	Very Good	3	F	SI1	3	1.01		5064.43					
3002									11733522.76					
3003										9386818.208				