

STEP 1 - Understanding the Model:

1. According to the linear formula provided, if a diamond is 1 carat heavier than another with the same cut and clarity, how much more would the retail price of the heavier diamond would be? Why?
2. If you were interested in a 1.5 carat diamond with a very good cut (represented by 3 in a model) and a VS2 clarity rating (represented by a 5 in the model), what retail price would the model predict for a diamond.

Answer:

If a diamond is 1 carat heavier than another with the same cut and clarity, then the retail price of the heavier diamond would be increase in 8413.0 bucks. Because, based on the given data and the given linear formula, the coefficient of carat is 8413.0 .

For example,

Let's say, the price of a diamond of 1.5 carat with very good cut (3) and a VS2 clarity rating (5) is - **\$10094.8**

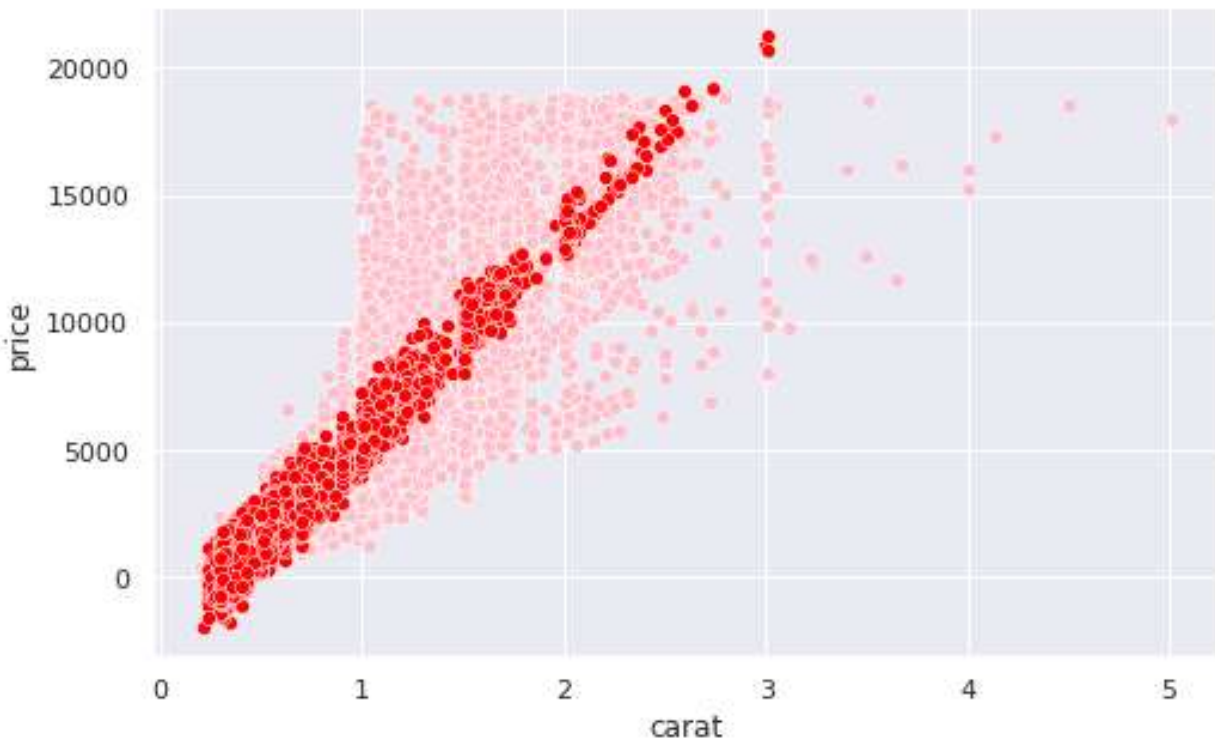
While, the price of a diamond with the same cut and clarity but with additional one carat (1.5 +1) is - **\$18507.8 bucks**

Price Difference = **18507.8 - 10094.8 = 8413.0 //**

Hence, the retail price difference between two diamonds with same cut and clarity except the increase in one carat is - **\$8413.0.**

STEP 2 - Visualize the Data:

- 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?



- For using linear regression, the independent and dependent variable must be linearly correlated – either positively or negatively.
- For training dataset, it is clear that the x and y axis are positively correlated from 0 to 1 very well, while it got distorted after 1 carat. Furthermore, the predicted values and the carat values from the testing dataset are very well linearly correlated and directly proportional. Judging from the above plot, linear regression is a good choice to build this model while it is not the best choice.

STEP 3 - The Recommendation:

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

RECOMMENDATION

From my analysis, the predicted sum of all 3000 diamonds together is \$11733522.76. However, the above value of \$11733522.76 is a customer selling price which is 30 percent higher than the company's buying price. Hence the company should bid 70 % of the above mentioned price, which is **\$8213465.932**.