## Project: Diamond Prices

Step 1: Understanding the Model

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?

Ans:- As we studied, the regression analysis model is a statistical method that allows businesses to analyse and discover the relationship b/w different variables in their business, so we will first find the association between diamond Carat and Cut variables to know the dependencies between them before regression. Here the Correlation between Diamond Cut order and Carat has not a very strong relation with just having 18% of association which is -0.1357 but with a Slope of -0.0577 which is rate of change along the regression line, showing week relation with each other, as shown below with scatter graph
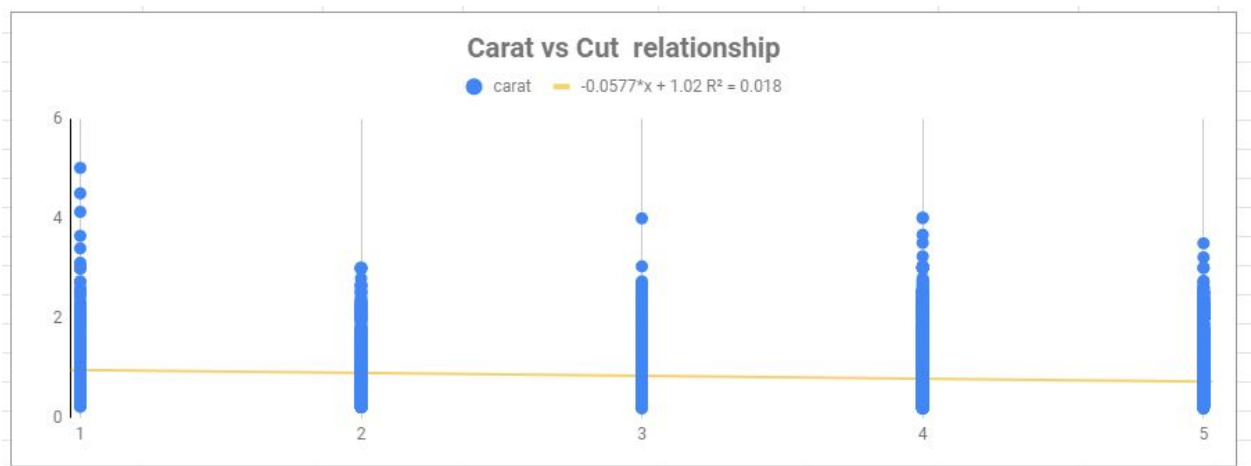
Now we apply the mathematical equation to predict the value of Price as a dependent and Carat as an independent variable to find and check the degree of association between Price and Carat variables we have seen that there exists a strong positive linear association between price of diamond and carat of a diamonds attribute and the Price showing increasing at a very straight line of slope with R-Square of 96% of a variability in Price regression on Carat as shown below.

r=Correlation	0.9803065341
a=Y-Intercept	-2281.958355
b=Slope	7785.794791
Y=Regression	
r=Square	0.9610009008
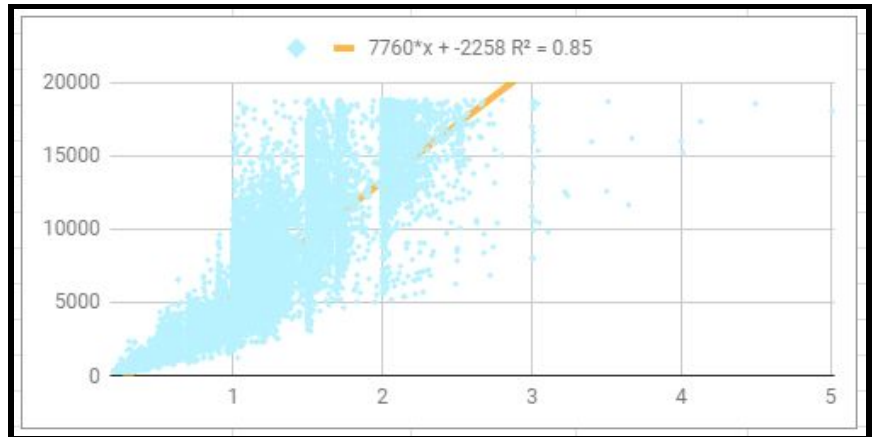
In a figure (r=0.9), the data points are practically in a straight line and this kind of association or correlation is near perfect. So according to the model; to purchase 1 carat diamond heavier than another with the same cut, the company should expect to pay 90% above the price. Here we can see the linear relationship between Price and Carat which means that as the Carat increases the PRice would also increase. This should also mean that if a diamond has high Carat, would also have a higher probability of getting a high Price . Hence we have clearly understood, how does the value of "Y" change with effect to change in "X" attribute.

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?

Ans:- by putting the given variables values (Carat=1.5, Cut=3 and Clarity=5) into a Price formula; which is

```
Price = -5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity
      =-5269+(8413*1.5)+(158.1*3)+(454*5)
      = 10094.8
```

Now, for calculating how much price the model predicts for this diamond, we have Y-intercept a= -2281.958355 and a regression Slope line at 7786, by putting these values using a linear regression formula for finding the strength of the association. so we have;

Y-Intercept a=	-2281.958355
Slope b=	7785.794791

```
Y = a + bX
Y = -2281.96 + (7786*1.5)
```
we have,   `= 9396.73383`

So, the best fit price to pay under the observation the Predict Price for 1.5 Carat diamond with a Very Good cut (3 in the model) and a VS2 clarity rating (5 in the model) to pay for it is **=9396.73**.

| Price = -5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity			
Price=	10094.8	Predict price to Bid =	9396.733831
Price for carat Diamond with Very Good cut (3) and clarity rating VS2 (5)			

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.
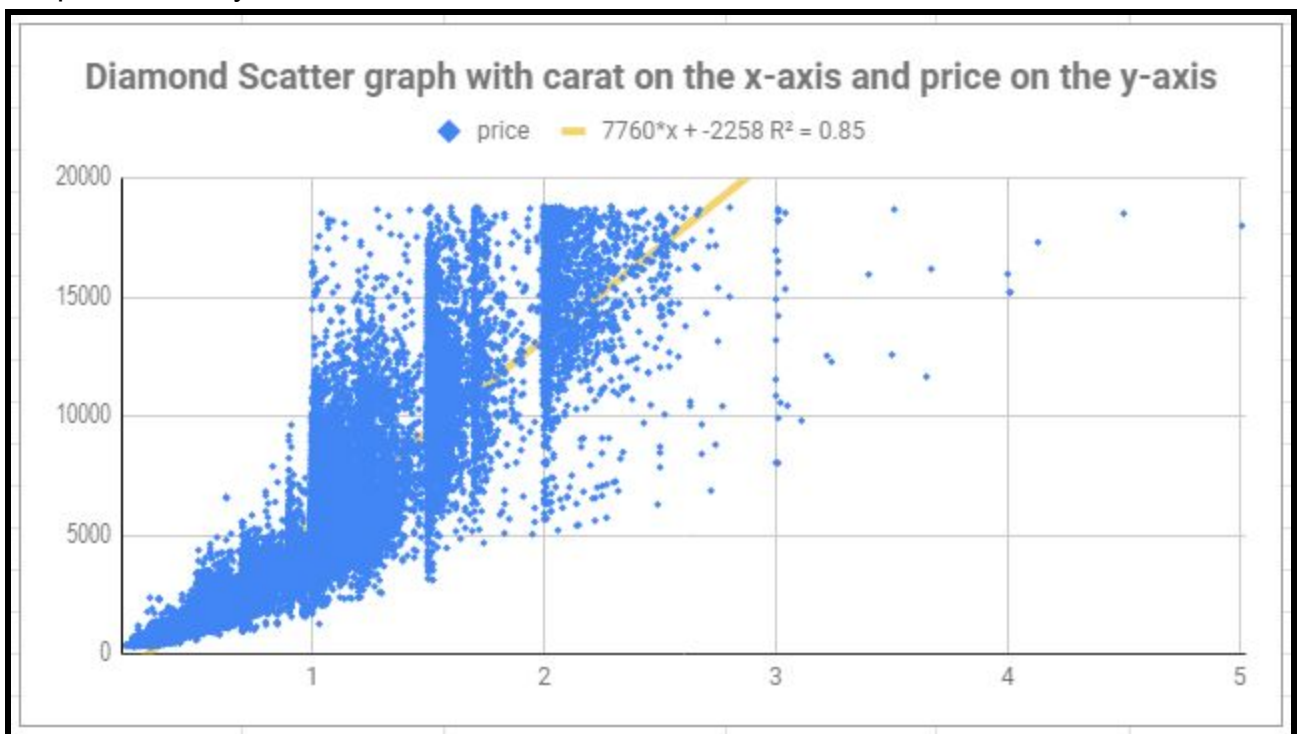
2. Plot 2 - Plot the data for the diamonds for which you are predicting prices with carat on the x-axis and predicted price on the y-axis.

Predicting Price with Carat on X-axis and Price on Y-axis

Price — $7786*x + -2282\ R^2 = 0.961$
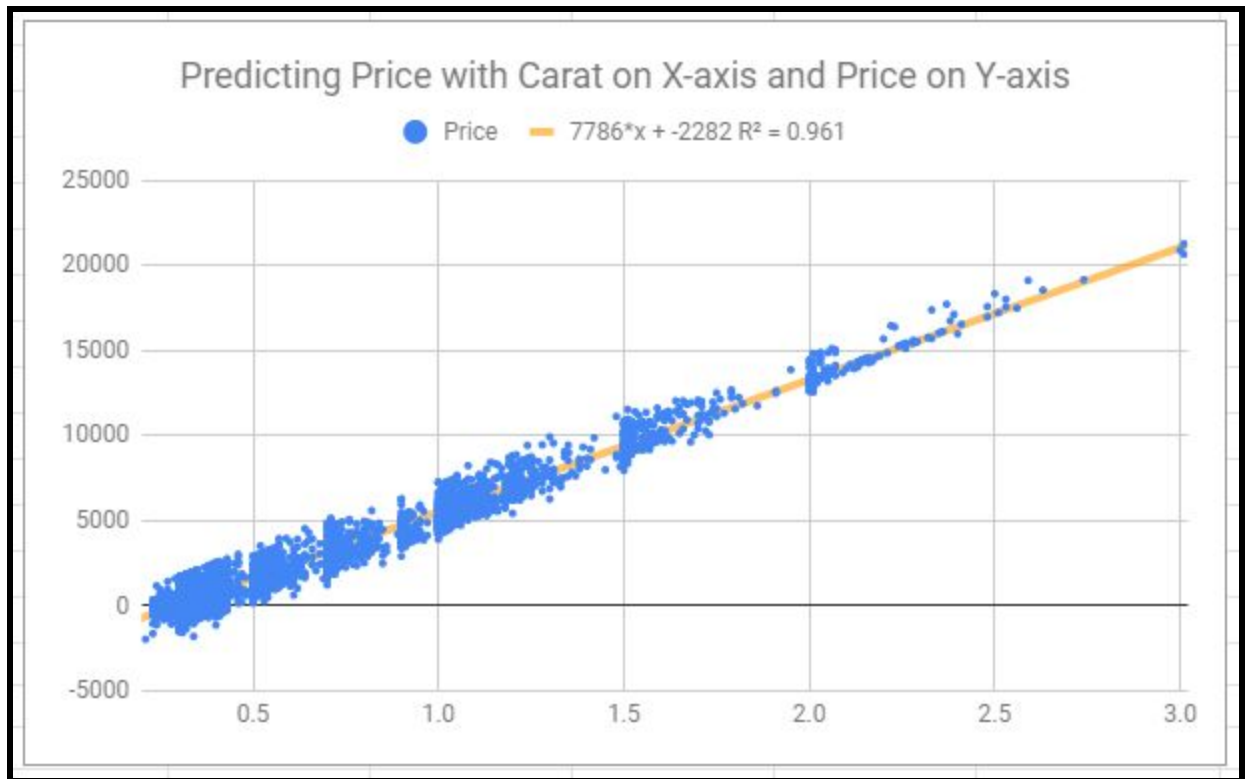
3. What strikes you about this comparison? After seeing this plot, do you feel confident in the model's ability to predict prices?

Ans:- The Correlation describes the strength of the association between the dependent variables on Y-axis and independent variables on X-axis; and analysis of the diamond price using a linear regression model for finding the most relative value to predict the dependent variable which is *'Price'* in our case. After a regression we have found that there is much strong relativity and high degree of association where Correlation 'r' is 0.98 which is very near to the perfect association, then we find the **Slope** of the regression line denoted by 'b' which is 7785.80 and the **Intercept** of the regression line denoted by 'a' which is -2281.96 which means that the specified model gave us the ability for taking confidential decisions to predict and forecast the prices of diamonds.

Step 3: Make a Recommendation

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

We have learned that regression analysis is a statistical method of mathematical equations that allow businesses to analyse and discover the relationship b/w different attributes, and forecast the future outcomes . After a detailed study and analysis of diamond

forecast price based on its attributes, the jewelry company's analytics team has a clear picture of a data set; for which the total price of entire set of 3000 diamonds became **11.73 (M)** as shown in new-diamond data set but the company's analytics team would like to bid for amounting **8.2 (M)** for the entire set of 3000 of diamonds to purchase from distributor at 70% of that price in order to raise their inventory, as detail shown below.

Total Price of 3000 set of Diamonds   11733522.76

Predict Price of 3000 set of Diamonds   11733522.76 (**11.73M**)

Diamond purchase price from distributor @ 70%   8213465.932 (**8.2M**)