

# Predicting Diamond Prices (Excel Linear Regression)

## Project Overview

A jewelry company wants to put in a bid to purchase a large set of diamonds, but is unsure how much it should bid. In this project, I will use the results from a predictive model to make a recommendation on how much the jewelry company should bid for the diamonds. An Linear regression model has already been built :

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

There are two datasets.

**diamonds.csv** contains the data used to build the regression model.

**new\_diamonds.csv** contains the data for the diamonds the company would like to purchase.

## Step 1: Understanding the Model

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

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

Assume the cut is 1, clarity is 1

When Price with carat 1=  $-5,269 + 8,413 \times 1 + 158.1 \times 1 + 454 \times 1 = 3756.1$

When Price with carat 2=  $-5,269 + 8,413 \times 2 + 158.1 \times 1 + 454 \times 1 = 12169.1$

So we need to pay  $12169.1 - 3756.1 = 8413$  more

Because when others being equal, 1 carat heavier means \$8413 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?

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

## Step 2: Visualize the Data

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

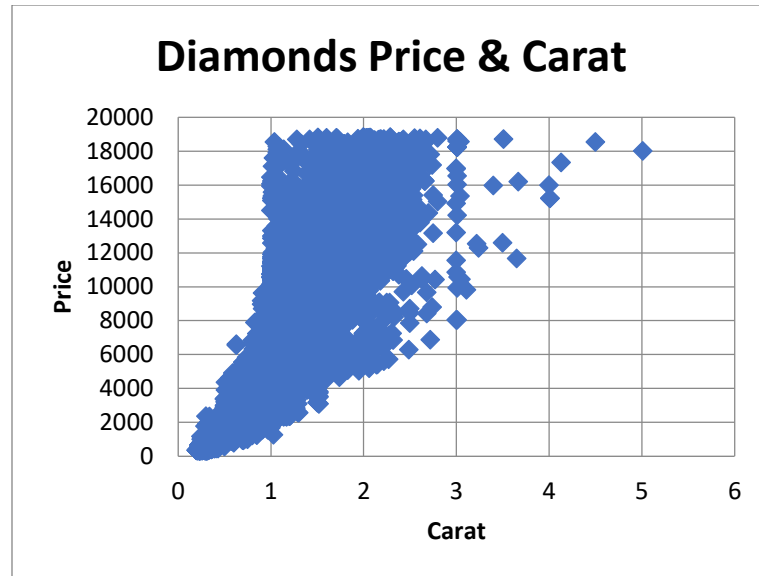


Figure 1.

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.

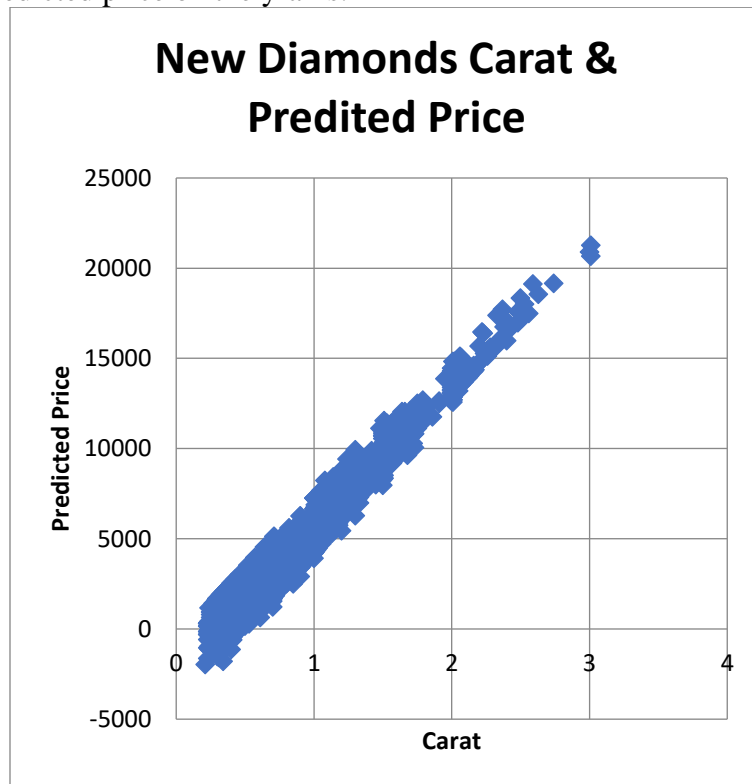


Figure 2.

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

Based on the above figure, the model is somewhat useful to set a bid price, but confidence in the predicted prices is not too high. This is because 1) the model predicted

negative prices, this would not happen in reality. 2) the observed predicted prices are in much narrow range than the actual prices, this means the factors which are not included could help to improve the accuracy of the model.

## Step 3: Make a Recommendation

Predicted the price of each new diamond using linear regression

Calculated the bid price, which is 70% of the predicted price.

Sum the bid prices, and result in a recommendation bid price of 8213465.93