<u>Project: Diamond Prices</u>

# Step 1: Understanding the Model

- ***Answer***

From the equation in our section (project details)

**Price** = -5,269 + 8,413 x **Carat** + 158.1 x **Cut** + 454 x **Clarity**

========================================================================

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?

    - *Answer*

We clearly notice here that the parameter per carat is 8413, which means that each carat is heavier than the other with the same cut, and this also means that it will increase in price and the increase will be 8413 dollars.

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?

    - *Answer*

From the equation we substitute the variable value above

**Price** = -5,269 + 8,413 x **Carat** + 158.1 x **Cut** + 454 x **Clarity**

 = -5,269 +( 8,413 1*5) + x (158.1 x**3)** + (454 x *5)

SO, we should pay $10,094.8 to got 1.5 carat diamond

So, we should pay $10,094.8 to got 1.5 carat diamond

# 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.
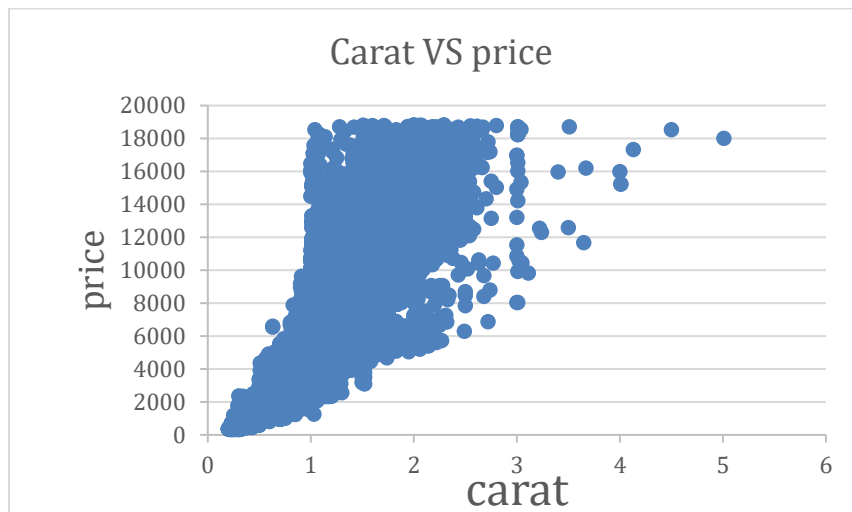


2. Plot 2 - Plot the data for the diamonds for which you are predicting prices with carat on the x-axis and predicted price on the y-axis.

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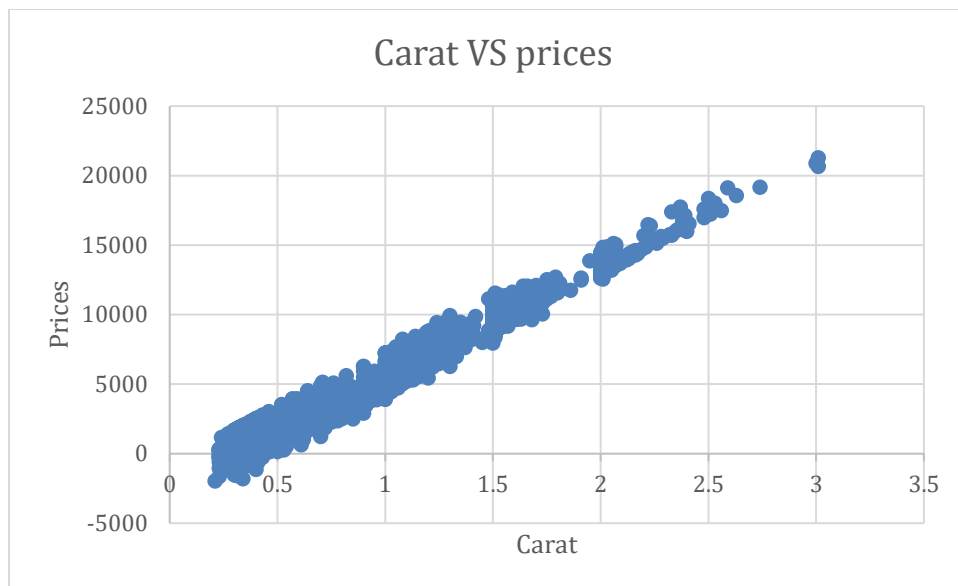
3. What strikes you about this comparison? After seeing this plot, do you feel confident in the model's ability to predict prices?

**In notice that:**

================

The old price diamonds

- The graph shows from (0 to 1) that the **caret and prices** are almost linear with a clear and positive flip, compared to the other graph, we find that it starts from (1 to 3).  We can see that the caret and prices are not linear and the correlation is not strong.


- And when we **want to talk about the range**, we see that it indicates a diversity in the distributed prices, which indicates that there are some other independent factors that clearly affect the prices of diamonds (example: diamond color, clarity, transparency, etc.)


 **The New - Diamonds**:

================

- We will clearly notice that **the relationship between sponsorship and pricing is** positive, **illustrated in the original new diamond price chart**.  Likewise, we can clearly see some examples where the price becomes negative, **indicating that linear regression is not a completely** correct approach to solving this problem (forecasting diamond prices).
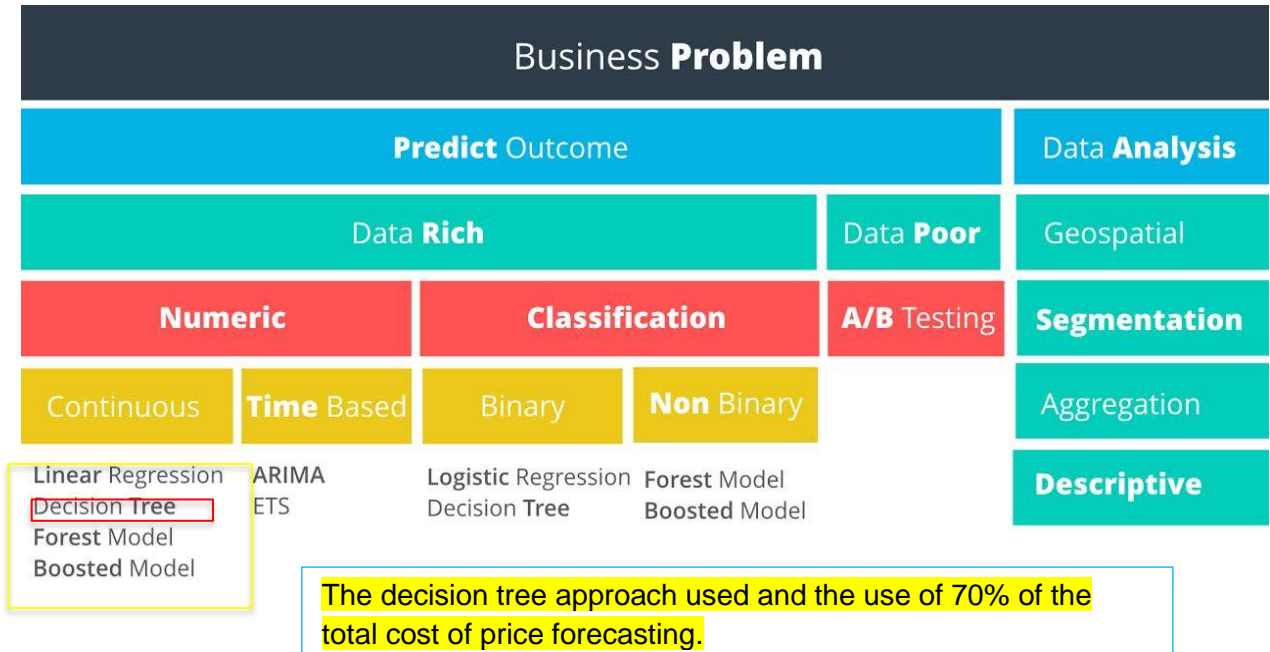
# Step 3: Make a Recommendation

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

- Answer
  ============

We **conclude from the previous steps**, and we clearly notice some examples where prices have a negative value, **so of course we will understand** that linear regression is not the correct way to solve this problem **(predicting diamond prices**) as well as from the lessons that we used the decision tree approach and the expected prices using the random forest model (70% Of the total price forecast) as shown in the pictures below $5.213465.



| Business **Problem** | | | | |
|---|---|---|---|---|
| **Predict** Outcome | | | | Data **Analysis** |
| Data **Rich** | | | Data **Poor** | Geospatial |
| **Numeric** | | **Classification** | **A/B** Testing | **Segmentation** |
| Continuous | **Time** Based | Binary | **Non** Binary | Aggregation |
| Linear Regression<br>Decision **Tree**<br>Forest Model<br>Boosted Model | ARIMA<br>ETS | Logistic Regression<br>Decision **Tree** | Forest Model<br>Boosted Model | **Descriptive** |

The decision tree approach used and the use of 70% of the total cost of price forecasting.

Cell reference: I3002 — Formula: `=SUM(I2:I3001)`

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 83 | 2982 | 1.51 | Premium | 4 | E | SI1 | 3 | 1.51 | 9429.03 | | | | |
| 84 | 2983 | 1.54 | Ideal | 5 | J | VS2 | 5 | 1.54 | 10747.52 | | | | |
| 85 | 2984 | 1.16 | Ideal | 5 | I | VS2 | 5 | 1.16 | 7550.58 | | | | |
| 86 | 2985 | 1.01 | Premium | 4 | D | SI1 | 3 | 1.01 | 5222.53 | | | | |
| 87 | 2986 | 1.01 | Very Good | 3 | I | VS2 | 5 | 1.01 | 5972.43 | | | | |
| 88 | 2987 | 0.83 | Premium | 4 | H | SI1 | 3 | 0.83 | 3708.19 | | | | |
| 89 | 2988 | 0.93 | Very Good | 3 | F | VS2 | 5 | 0.93 | 5299.39 | | | | |
| 90 | 2989 | 1.2 | Very Good | 3 | G | SI2 | 2 | 1.2 | 6208.9 | | | | |
| 91 | 2990 | 0.8 | Ideal | 5 | E | VVS2 | 6 | 0.8 | 4975.9 | | | | |
| 92 | 2991 | 1.11 | Ideal | 5 | D | VVS2 | 6 | 1.11 | 7583.93 | | | | |
| 93 | 2992 | 1.31 | Premium | 4 | H | SI2 | 2 | 1.31 | 7292.43 | | | | |
| 94 | 2993 | 1.11 | Good | 2 | F | SI1 | 3 | 1.11 | 5747.63 | | | | |
| 95 | 2994 | 1.1 | Good | 2 | D | SI2 | 2 | 1.1 | 5209.5 | | | | |
| 96 | 2995 | 1.22 | Premium | 4 | I | SI2 | 2 | 1.22 | 6535.26 | | | | |
| 97 | 2996 | 0.72 | Ideal | 5 | F | SI2 | 2 | 0.72 | 2486.86 | | | | |
| 98 | 2997 | 1.09 | Premium | 4 | I | VS2 | 5 | 1.09 | 6803.57 | | | | |
| 99 | 2998 | 1.05 | Very Good | 3 | G | SI1 | 3 | 1.05 | 5400.95 | | | | |
| 00 | 2999 | 0.7 | Fair | 1 | G | SI1 | 3 | 0.7 | 2140.2 | | | | |
| 01 | 3000 | 1.01 | Very Good | 3 | F | SI1 | 3 | 1.01 | 5064.43 | 8213465.932 | | | |
| 02 | | | | | | | | | 11733522.76 | | | | |
| 03 | | | | | | | | | | | | | |
| 04 | | | | | | | | | | | | | |

70% of the statistics if the price is previously