Diamond Price Prediction

**Assignment:** Engineer at least one feature and build a model that will predict price for each record, given an input file with the same layout as diamonds.txt minus price. The model does not have to use any fancy machine learning. Your feature engineering and model-building may be done in any programming language, although we prefer either Python or R. Please comment the code and make it easily readable.

**Data:** Tab delimited text file with header and 40,000 records of “round” diamonds

**Price:** in US dollars

**Cut:** quality of the cut (Fair, Good, Very Good, Premium, Ideal)

**Color:** diamond color, from J (worst) to D (best)

**Clarity:** a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))

**Depth:** total depth percentage = z / mean(x, y) = 2 \* z / (x + y)

**Table:** width of top of diamond relative to widest point

**X:** length in mm

**Y:** width in mm

**Z:** depth in mm