**Information on the Diamond Dataset**

In this lesson, you'll be working with a dataset regarding the prices and attributes of approximately 54,000 round-cut diamonds. You'll go through the steps of an explanatory data visualization, systematically starting from univariate visualizations, moving through bivariate visualizations, and finally multivariate visualizations. Finally, you'll work on polishing up selected plots from the analysis so that their main points can be clearly conveyed to others.

You can find a copy of the dataset in the Resources tab of the classroom; it will automatically be available to you in the workspaces of this lesson. The dataset consists of almost 54,000 rows and 10 columns:

* **price**: Price in dollars. Data was collected in 2008.
* **carat**: Diamond weight. 1 carat is equal to 0.2 grams.
* **cut**: Quality of diamond cut, affects its shine. Grades go from (low) Fair, Good, Very Good, Premium, Ideal (best).
* **color**: Measure of diamond coloration. Increasing grades go from (some color) J, I, H, G, F, E, D (colorless).
* **clarity**: Measure of diamond inclusions. Increasing grades go from (inclusions) I, SI2, SI1, VS2, VS1, VVS2, VVS1, IF (internally flawless).
* **x**, **y**, **z**: Diamond length, width, and depth, respectively, in mm.
* **table**: Ratio of width of top face of diamond to its overall width, as a percentage.
* **depth**: Proportional depth of diamond, as a percentage. This is computed as 2 \* z / (x + y), or the ratio of the depth to the average of length and width.

For the case study, we will concentrate only the variables in the top five bullet points: price and the four 'C's of diamond grade. Our focus will be on answering the question about the degree of importance that each of these quality measures has on the pricing of a diamond. You can see an example report covering all of the variables in the project information lesson