Data dictionary

This notebook uses one of seaborn's built-in datasets called "diamonds," which contains information about diamonds. From this dataset, two files are derived in the notebook:

**diamonds.csv**

**diamonds2.csv**

The derived **diamonds.csv** dataset contains:

**39,840 rows** – each row is a unique diamond

**3 columns:**

|  |  |  |
| --- | --- | --- |
| **Column name** | **Type** | **Description** |
| color | str | Diamond color grade |
| price | int | Diamond price (US dollars) |
| log\_price | float | Logarithm of diamond price |

The derived **diamonds2.csv** dataset contains:

**34,935 rows** – each row is a unique diamond

**4 columns:**

|  |  |  |
| --- | --- | --- |
| **Column name** | **Type** | **Description** |
| color | str | Diamond color grade |
| cut | str | Diamond cut |
| price | int | Diamond price (US dollars) |
| log\_price | float | Logarithm of diamond price |

More information about this dataset can be found at [ggplot2](https://ggplot2.tidyverse.org/reference/diamonds.html) and at the [GitHub page for seaborn's datasets](https://github.com/mwaskom/seaborn-data).

Remember, you can access and download the data for any Jupyter notebook activity from within the notebook itself by navigating to the **Lab Files** dropdown menu at the top of the page, clicking into the **/home/jovyan/work** folder, selecting the relevant data file, and clicking **Download**.