

- **Question/Problem statement:**

- How to create a regression model to predict the price of diamonds?
- What is the best seller of diamonds?

- **Data Description:**

This dataset contains of 10 variables and 53940 rows. We will perform regression model to predict the price of diamonds.

- **Data type:**

- **Price:** in US dollars
- **Carat:** weight of the diamond
- **Cut:** quality of the cut (Fair, Good, Very Good, Premium, Ideal)
- **Color:** quality of the cut (Fair, Good, Very Good, Premium, Ideal)
- **Clarity:** a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))
- **x:** length in mm
- **y:** width in mm
- **z:** depth in mm
- **Depth:** total depth percentage
- **Table:** width of top of diamond relative to widest point

- **Data size:**

- **No. of rows:** 53940
- **No. of columns:** 10

- **Tools:**

- Programs: Python Jupyter Notebook
- Libraries: panda, Seaborn, Matplotlib, scikit-learn

- **MVP Goal:**

The goal of this project is to perform regression model to predict the price of diamonds.