**Experiment No:2**

**Aim**: To predict Diamond on Price Prediction Dataset using linear regression.

**Theory:**

Diamond price prediction involves estimating the market value of a diamond based on its physical and qualitative attributes. Diamonds are priced based on factors like carat weight, cut quality, color, and clarity.

#### **Factors Influencing Diamond Prices**

The price of a diamond is primarily influenced by the **4Cs**:

* **Carat:** Weight of the diamond. Larger carat typically increases the price.
* **Cut:** Quality of the diamond’s cut (e.g., Ideal, Premium, Good, Fair, Poor), affecting brilliance.
* **Color:** Graded from D (colorless) to Z (light yellow or brown). Less color usually means a higher price.
* **Clarity:** Measures internal flaws (inclusions) and external blemishes, graded from Flawless (FL) to Included (I3).

Other factors include **depth**, **table size**, and physical dimensions (x, y, z).

#### **Linear Regression for Price Prediction**

Linear Regression is a statistical method that models the relationship between a dependent variable (diamond price) and one or more independent variables (features like carat, cut, color, clarity).

**Steps in Diamond Price Prediction**

**Data Collection:** Collect data containing diamond features and their corresponding prices.

**Data Pre-processing:** Clean the data (remove missing values, outliers), encode categorical variables (e.g., cut, color, clarity).

**Feature Selection:** Choose relevant features that impact the price.

**Model Training:** Fit a linear regression model using the training data.

**Prediction:** Predict the price of diamonds using the model.

1. **Step 1: Load Dataset**
2. **Step 2: Encode Data (One-Hot Encoding)**
3. **Step 3: Split Data into Training and Testing Sets**
4. **Step 4: Train the Linear Regression Model**
5. **Step 5: Predict with Custom Input**

**Program:**

Take printout. (program and csv file)

**Conclusion:**