**Business Intelligence Lab**

**Experiment 11**

**Aim:**

To build Business Intelligence Mini Project

**Problem Statement:**

A jewelry firm wants to submit a bid to purchase a large collection of diamonds but is uncertain how much it should bid. A diamond distributor has recently decided to exit the market and has put up a collection of 1,000 diamonds up for auction. Seeing this as a great opportunity to expand its inventory, a jewelry firm is interested in making a bid. To determine how much to bid, the firm’s analytics department will use a large database of diamond prices to build a linear regression model to predict the price of a diamond based on its attributes.

**Problem Identification:**

The jewelry firm plans to use the results from a predictive model to make a recommendation on how much the jewelry company should bid for the diamonds. Thus our task will be to build the linear regression model and apply that model to make a recommendation for how much the company should bid for the entire collection of diamonds.

**Dataset:**

Diamonds dataset contains the prices and other attributes of almost 54,000 diamonds and this dataset is hosted on Kaggle. The dataset contains 53940 rows and 10 variables. Before jumping into building the model, let’s have a look into the variables & their definitions.

* Price is in US dollars
* Carat weight of the diamond
* 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))
* x length in mm
* y width in mm
* z depth in m
* depth: The height of a diamond
* table: The width of the diamond’s table expressed as a percentage of its average diameter

