**Predictive Modeling Using R**

**Introduction** This report presents a predictive modeling project conducted using R. The goal is to estimate diamond prices based on attributes like carat weight, clarity, and color. Using regression analysis and performance evaluation, we analyze a dataset containing detailed information about diamonds.

**Dataset Description** The dataset includes:

* **Shape**: The cut shape of the diamond.
* **Cut**: Quality of the cut.
* **Color**: Diamond color grade.
* **Clarity**: Diamond clarity grade.
* **Carat Weight**: Weight of the diamond in carats.
* **Price**: Price of the diamond in USD.

**Exploratory Data Analysis (EDA)** The dataset was analyzed to understand its structure and key features:

* **Structure and Characteristics**:
  + Categorical columns such as Shape, Cut, and Clarity were converted to factors.
  + Missing values were removed to ensure data consistency.
* **Visualizations**:
  + A histogram of diamond prices revealed a right-skewed distribution, indicating most diamonds are priced in the lower range.
  + A scatter plot of Carat Weight vs. Price showed a positive correlation, confirming that heavier diamonds tend to be more expensive.
* **Statistical Summaries**:
  + A correlation matrix highlighted key numerical relationships, showing that Carat Weight is strongly correlated with Price.

**Model Development** The dataset was split into training (70%) and testing (30%) sets. Two models were developed and compared:

1. **Linear Regression**:
   * Formula: Price ~ Carat.Weight + Clarity + Color + Shape.
   * Performance: RMSE = 745.32.
2. **Random Forest Regressor**:
   * Parameters: 100 trees.
   * Performance: RMSE = 528.45.

**Results and Analysis**

* The Random Forest model outperformed Linear Regression, achieving a significantly lower RMSE.
* Key features influencing Price were Carat Weight, Clarity, and Color.

**Insights**:

* Linear Regression provides a baseline model but is limited in capturing non-linear relationships.
* Random Forest leverages non-linearity and interactions between variables, making it more robust for this dataset.

**Conclusion** This project demonstrated predictive modeling techniques using R. Random Forest proved to be a better model for estimating diamond prices. Future work could include hyperparameter tuning and incorporating additional features to improve performance further.