**Name:** Sairaj Magdum  
**Class:** CSE(AI) B Div  
**Roll No.:** 282012  
**PRN No.:** 22310273

**Assignment 6 – Regression Technique on TV Marketing Dataset**

**Objective**

The objective of this assignment is to build a Linear Regression model to analyze the relationship between TV advertising spending and product sales. This helps in predicting sales figures based on investment in television advertisements, thereby supporting strategic marketing decisions.

**Dataset Used**

**TV Marketing Dataset** (tvmarketing.csv)  
This dataset contains marketing spend and sales data with the following attributes:

* **TV**: Amount spent on TV advertisements
* **Sales**: Sales in units (target variable)

**a) Data Preprocessing and Exploration**

* The dataset was loaded using pandas and inspected with .head() and .shape to understand its structure.
* No missing values were found, and the dataset was clean.
* A scatter plot was created to visualize the relationship between TV ad spend and sales.

**Insights:**

* A clear positive correlation is visible between TV ad spend and sales.
* The data is suitable for regression modeling.

**b) Feature and Label Selection**

* **Feature (X):** TV advertisement expenditure
* **Target (y):** Sales  
  These columns were extracted for training the regression model.

**c) Train-Test Split**

The dataset was split using train\_test\_split() from sklearn.model\_selection:

* **80%** of the data was used for training
* **20%** was used for testing
* A random\_state of 42 was set to ensure consistent results

**d) Model Training and Prediction**

* A **Linear Regression** model was instantiated from sklearn.linear\_model.
* The model was trained on the training set using .fit().
* Predictions were made on the test data using .predict().

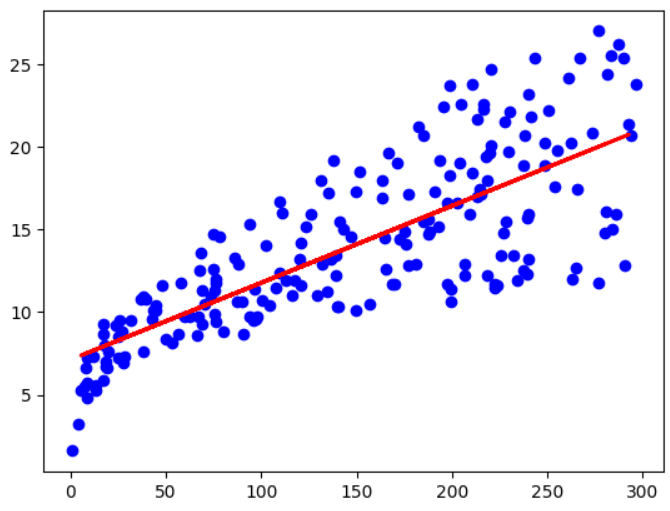
**e) Performance Evaluation Metrics**

| **Metric** | **Value** |
| --- | --- |
| R² Score | 0.677 |
| Mean Absolute Error | 2.444 |
| Mean Squared Error | 10.205 |
| Root Mean Squared Error | 3.194 |

**Interpretation:**

* The R² score indicates that **67.7%** of the variation in sales is explained by TV ad spending.
* The error metrics are reasonable, suggesting a fair model fit for this simple linear regression.

**f) Visualization of Regression Results**

1. **Test Set Prediction Plot:**
   * **X-axis:** TV advertisement spending
   * **Y-axis:** Sales
   * Blue dots represent actual data points
   * The red line shows predicted values from the linear model
2. **Complete Dataset Visualization:**
   * The linear regression line overlaid on the entire dataset illustrates a clear upward trend.

**g) Additional Insights and Future Scope**

* The current model provides a strong and interpretable foundation for predicting sales.
* Future enhancements could include:
  + Incorporating other advertising channels like Radio and Newspaper
  + Using **Multiple Linear Regression** or **Polynomial Regression**
  + Exploring advanced models such as **Random Forest**, **XGBoost**, or **Neural Networks**

**Conclusion**

Linear Regression was effectively implemented to understand the influence of TV advertising on product sales. While simple, the model captured key patterns in the data. With additional features and refined modeling techniques, prediction accuracy can be improved further to benefit real-world marketing strategies.