**Name: Akshat Gandhi  
Roll no: 281061  
Batch: A3  
Assignment 6**

**Statement:**

In this assignment, we apply Linear Regression on a dataset containing month-wise temperatures of India. We will predict temperatures based on months and evaluate the model using various performance metrics.

Objective:

* Understand and apply Linear Regression for temperature prediction.
* Evaluate the model using MSE, MAE, and R-squared.
* Visualize the regression results to interpret the model's effectiveness.

Resources Used:

* **Software used:** Jupyter Notebook
* **Libraries used:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn

Introduction to Regression:

Regression is a fundamental machine learning technique used to predict continuous values based on given input features. In this assignment, we apply **Linear Regression** to predict monthly temperatures using historical data.

Dataset Description:

The dataset contains:

* **YEAR:** The year of observation.
* **JAN - DEC:** Average monthly temperatures recorded in Celsius.

Since the dataset is in a **wide format** (each month is a separate column), we reshape it using the **melt() function** to make it suitable for regression.

**Methodology:**

1. Data Collection and Exploration:

* Load the dataset into a Pandas DataFrame.
* Check for missing values and clean the data.
* Reshape the dataset to have columns: YEAR, Month, Temperature.

2. Data Preprocessing:

* Convert month names (JAN, FEB, etc.) into numerical values (1-12).
* Split the dataset into training and testing sets.

3. Model Implementation:

* Apply **Linear Regression** using Scikit-Learn.
* Train the model using Month as the input feature and Temperature as the target.

4. Model Evaluation:

* **Mean Squared Error (MSE):** Measures average squared difference between actual and predicted values.
* **Mean Absolute Error (MAE):** Measures the absolute average difference between actual and predicted values.
* **R-squared Score (R²):** Determines how well the model explains variability in the data.

5. Visualization:

* Scatter plot of **Actual vs. Predicted Temperatures**.
* Regression line to observe trends.

**Program Implementation:**

1. Importing Libraries:

* pandas for data handling
* numpy for numerical operations
* matplotlib & seaborn for visualization
* sklearn for machine learning (Linear Regression, Metrics, Train-Test Split)

2. Data Loading and Preprocessing:

* Read the CSV file into a Pandas DataFrame.
* Reshape the dataset using melt().
* Map month names to numeric values.
* Split into training and testing sets.

3. Applying Linear Regression:

* Train the model using LinearRegression().
* Make predictions on the test set.

4. Model Performance Metrics:

* Compute **MSE, MAE, and R²** values.

5. Visualization:

* Plot **scatterplot and regression line** to analyze predictions.

Results:

* **MSE:** *0.6311*
* **MAE:** *0.5945*
* **R-squared Score:** *0.9365*
* The visualization demonstrates a clear trend between months and temperatures, validating the effectiveness of the model.

Advantages:

* Predicts temperature trends efficiently.
* Helps in seasonal analysis for weather forecasting.
* Easy to interpret and implement.

Disadvantages:

* Assumes a **linear relationship**, which may not always hold.
* Sensitive to outliers, affecting prediction accuracy.

Conclusion:

This assignment provided hands-on experience with **Linear Regression** for predicting temperatures. By reshaping the dataset, training the model, and evaluating its performance, we gained insights into how regression can be used for **time-series forecasting**. The model successfully predicted temperature trends with reasonable accuracy, reinforcing the significance of regression analysis in real-world applications.