Name: Ayush Gupta

Roll No: 281049

Batch: A2

Assignment 6

Problem Statement:

Apply Linear Regression using a suitable library function to predict month-wise temperature and evaluate the model using performance metrics.

Objectives:

1) To apply regression techniques for predicting temperature trends.
2) To preprocess and analyze temperature data for better model performance.
3) To evaluate model performance using MSE, MAE, and R-Square metrics.
4) To visualize the regression model and interpret the results.

Resources used:

1) Software used: Visual Studio Code

2) Libraries used: Pandas, Matplotlib, Seaborn, SKLearn

Theory:

Regression is a supervised learning technique used to model relationships between a dependent variable (temperature) and one or more independent variables (month). Linear Regression assumes a linear relationship between these variables and fits a straight line that minimizes errors.

Linear Regression Formula:

$y = mx + c$

Where:

- y is the dependent variable (Temperature)
- x is the independent variable (Month)
- m is the slope (rate of change)
- c is the intercept (baseline value)

1. **Mean Squared Error (MSE):** Measures the average squared differences between actual and predicted values. Lower values indicate better performance.
2. **Mean Absolute Error (MAE):** Measures the average absolute differences between actual and predicted values.
3. **R-Square (R^2):** Represents the proportion of variance explained by the model. Closer to 1 indicates a better fit.

Methodology:

1. Data Preprocessing
- Load the dataset using Pandas.
- Handle missing values by imputation or removal.
- Convert categorical month values into numerical form (e.g., January = 1, February = 2, etc.).
- Split the dataset into training (80%) and testing (20%) sets.

2. Applying Linear Regression
- Use LinearRegression from sklearn.linear_model to fit the model.
- Train the model on the dataset.
- Predict temperature values for each month.

3. Model Evaluation
- Compute MSE, MAE, and R^2 scores to assess model performance.

4. Visualization
- Plot actual vs. predicted temperature values using Matplotlib/Seaborn.
- Display the regression line over the dataset.

Conclusion:

- The regression model effectively predicts monthly temperatures with a reasonable error margin.
- The evaluation metrics provide insight into model accuracy.