**Name: Gaurav Yadav**

**Roll no:** 282042

**Batch:** B2

**Assignment-6**

**Problem Statement:**

*To analyze and model the month-wise temperature trends in India using regression techniques. The dataset contains the average temperature values recorded across various locations in India for each month. The temperature values are measured in degrees Celsius.*

**Objectives:**

1. To introduce and apply basic regression techniques in Python using suitable libraries such as Scikit-learn.
2. To implement a Linear Regression model for predicting month-wise temperature trends based on historical average temperature data of India.
3. To evaluate the performance of the regression model using statistical metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-Squared Score (R²).
4. To visualize the regression results through appropriate plots, including scatter plots and regression lines, for better interpretation and insights.

**Resources used:**

1. Software used: Visual Studio Code
2. Libraries used: Pandas, Matplotlib, Seaborn,numpy

**Theory:**

* Pandas:Pandas is a powerful and flexible open-source data analysis and manipulation library for Python. It provides data structures like Series (1D) and DataFrame (2D) that are designed to work seamlessly with structured data. Pandas is widely used in data science, machine learning, and data analysis for tasks such as data cleaning, transformation, aggregation, and visualization.

Key features:  
• Easy handling of missing data and data filtering

● Seaborn: Seaborn is a powerful Python visualization library based on Matplotlib, designed to simplify the process of creating informative and attractive statistical graphics. It provides a high-level interface for drawing various types of plots with simple functions, while also integrating with Pandas DataFrames for seamless data handling.

Key features:

* Built-in themes and color palettes.
* Integration with Pandas for DataFrame compatibility.
* More attractive default plots, including categorical plots and regression plots.

● Matplotlib: Matplotlib is one of the most widely used Python libraries for creating static, animated, and interactive visualizations. It provides a low-level interface for creating basic plots and offers extensive customization options.

Key features:

* Extensive support for various plot types such as line, scatter, bar, and pie charts.
* Highly customizable, allowing fine control over figure aesthetics.
* Integration with NumPy and Pandas.

● **Panda:**Pandas is a powerful and flexible open-source data analysis and manipulation library for Python. It provides data structures like Series (1D) and DataFrame (2D) that are designed to work seamlessly with structured data. Pandas is widely used in data science, machine learning, and data analysis for tasks such as data cleaning, transformation, aggregation, and visualization

Key features:

* Easy handling of missing data and data filtering
* Powerful group-by functionality for aggregating and summarizing data

**Methodology:**

**1. Data Preprocessing**

• Load dataset using Pandas.  
• Handle missing values (if any).  
• Encode month names numerically.• Normalize data if needed using MinMaxScaler/StandardScaler.  
• Split data into training and testing sets (e.g., 80/20 split).

**2. Model Selection & Training**

**•** Use Linear Regression as the predictive model.  
• Train the model on the training set.  
• Fit the model to predict temperatures based on month.

**3. Evaluation Metrics**

**• Predict using the test data.  
• Evaluate model performance using:**a)MSE (Mean Squared Error)  
 b) MAE (Mean Absolute Error)  
 c) R² Score (Coefficient of Determination)

**4. Visualization**

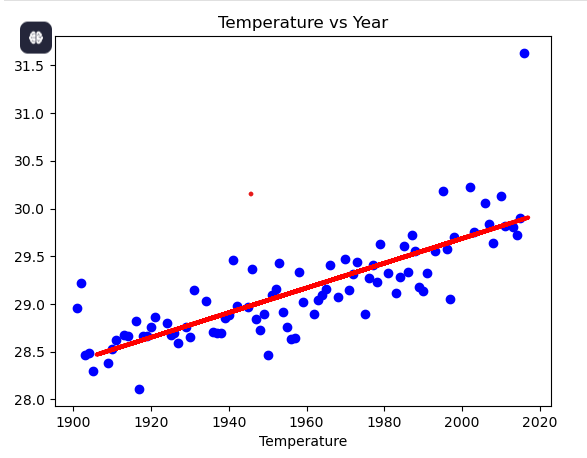
* Plot scatter plot with regression line.  
  • Use line charts or residual plots for additional insights.

**Results:**

**MSE :** 0.13680734009425016

**MAE :** 0.25660011136179517

**R-square :** 0.623947314874312



**Conclusion:**

The Linear Regression model effectively predicted month-wise average temperatures in India. Evaluation metrics (MSE, MAE, R²) indicated a good model fit. Visualizations confirmed seasonal trends and model accuracy, demonstrating the usefulness of regression for climate data analysis.