**Name:** Aziz Sayyad
**Roll No:** 281069
**Batch:** A3

# Practical 6

## Problem Statement:

a) Apply Linear Regression using suitable library function and predict the Month-wise temperature.

b) Assess the performance of regression models using MSE, MAE and R-Square metrics

c) Visualize simple regression model.

## Dataset:

Download temperature data from the following link:
Temperature Data - Kaggle
This dataset consists of average monthly temperatures recorded across various locations in India. The temperature values are recorded in Celsius.

## Objectives:

1. Implement Linear Regression using a suitable library function.
2. Predict Month-wise temperatures based on historical data.
3. Assess the performance of the regression model using MSE, MAE, and R-Square metrics.
4. Visualize the simple regression model.

## Resources Used:

- **Software:** Visual Studio Code
- **Libraries:** NumPy, Matplotlib, Scikit-learn

## Theory:

### 1. Linear Regression

Linear Regression is a statistical technique used for forecasting analysis. It predicts continuous, real, or numerical values such as sales, earnings, age, and product prices.

Linear Regression models the relationship between one or more independent variables (X) and a dependent variable (Y). It helps determine how changes in the independent variable(s) influence the dependent variable.

**Types of Linear Regression:**

1. **Simple Linear Regression:** Uses one independent variable to predict a dependent variable.
2. **Multiple Linear Regression:** Uses multiple independent variables to predict a dependent variable.

# Applications of Simple Linear Regression:

1. Predicting student grades based on hours studied.
2. Estimating agricultural yield using rainfall data.
3. Predicting salary based on years of experience.

# Limitations of Simple Linear Regression:

1. **Assumption of Linearity:** Assumes a straight-line relationship between variables, which may not always hold.
2. **Sensitive to Outliers:** Outliers can disproportionately affect the regression model.
3. **Does Not Establish Causation:** A strong correlation does not imply causation.

# Methodology:

## 1. Data Preprocessing

- Load the dataset using Pandas.
- Handle missing values (if any).
- Analyze and visualize data distribution.
- Split the dataset into training and testing sets (e.g., 80% training, 20% testing).

## 2. Model Training

- Implement Linear Regression using Scikit-learn.
- Train the model using the training dataset.
- Predict Month-wise temperatures using the trained model.

## 3. Performance Evaluation

Assess model performance using the following metrics:

- **Mean Squared Error (MSE):** Measures the average squared difference between actual and predicted values.
- **Mean Absolute Error (MAE):** Measures the average absolute difference between actual and predicted values.

- **R-Square Score:** Indicates how well the model explains variance in the dependent variable.

### 4. Model Visualization

- Plot the actual vs. predicted temperature values using Matplotlib.
- Visualize the regression line to understand model fit.

## Conclusion:

- Successfully implemented Linear Regression to predict Month-wise temperatures.
- Evaluated model performance using MSE, MAE, and R-Square metrics.
- Visualized the regression model to understand its effectiveness.
- Potential improvements include feature engineering and experimenting with polynomial regression for better accuracy.