

## Assignment 6

Name: Yashraj Nimbalkar

Roll No: 281039

Batch: A2

### Problem Statement

The goal of this assignment is to use Linear Regression to predict temperature patterns over a span of months using a provided dataset. The steps include importing data, applying the linear regression algorithm, visualizing the relationship, and evaluating the model using standard performance metrics such as Mean Squared Error, Mean Absolute Error, and  $R^2$  Score.

### Objectives

- Use Python to compute various statistical measures from the dataset.
- Display distributions of data attributes using histograms.
- Prepare the data for analysis by handling missing values and cleaning irrelevant fields.
- Develop and evaluate a Linear Regression model to predict monthly temperature values.

### Tools and Libraries Used

The implementation was done using Visual Studio Code. Python libraries like Pandas were used for data handling, Matplotlib and Seaborn for visualizations, and Scikit-learn for building and testing the machine learning model.

### Overview of Linear Regression and Pandas

Pandas is a Python tool designed for manipulating and analyzing structured datasets. It makes data wrangling efficient and intuitive. Linear Regression, on the other hand, is a fundamental algorithm used in predictive modeling to establish a straight-line relationship between dependent and independent variables. It's often used when a trend or correlation needs to be identified.

### Workflow and Steps Followed

1. Data Loading and Inspection  
The dataset was imported using Pandas. Basic inspection was done to understand the structure, detect missing values, and review column formats.
2. Data Preprocessing  
Missing entries were filled using statistical values like the median. Unnecessary columns were removed to streamline the analysis.

### 3. Statistical Analysis

Summary statistics such as the mean, standard deviation, minimum, maximum, and percentiles were generated to better understand the data.

### 4. Visualization

Histograms and line plots were created to observe data distribution and detect patterns. Outliers and skewness in the data were also examined visually.

### 5. Model Building

The dataset was split into training and test sets. A Linear Regression model was trained on the training portion, and predictions were made on the test data.

### 6. Model Evaluation

The model's performance was checked using three key metrics:

- Mean Squared Error
- Mean Absolute Error
- $R^2$  Score

Visual comparison of predicted vs. actual values was also performed.

## Advantages

- Streamlined data analysis using Pandas.
- Quick model building using Linear Regression.
- Easy visualization of trends using Seaborn and Matplotlib.

## Limitations

- Linear Regression works best for linear relationships; it might not perform well on non-linear data.
- Handling very large datasets can become slow and resource-intensive.

## Conclusion

Through this assignment, I was able to understand and apply the concepts of Linear Regression effectively. I worked with real-world temperature data, cleaned and analyzed it using Python, and built a predictive model. This hands-on experience enhanced my understanding of regression modeling and data-driven forecasting.