**2311cs020302-day22**

**Using the same data set of Civil\_Engineering\_Regression\_Dataset.csv**

**Part 2: Simple Linear Regression**

1. Fit a simple linear regression model where **Building Height** is the independent variable and **Construction Cost** is the dependent variable.
2. What is the equation of the regression line?
3. Interpret the coefficient: How does Building Height impact Construction Cost?
4. Evaluate model performance using R-squared and Mean Squared Error (MSE).

Code:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, r2\_score

df = pd.read\_csv("Civil\_Engineering\_Regression\_Dataset.csv")

X = df[['Building Height']] # Independent variable

y = df['Construction Cost'] # Dependent variable

model = LinearRegression()

model.fit(X, y)

slope = model.coef\_[0]

intercept = model.intercept\_

print(f"Regression Equation: Construction Cost = {intercept:.2f} + {slope:.2f} \* Building Height")

print(f"Interpretation: A 1-unit increase in Building Height leads to an estimated {slope:.2f} increase in Construction Cost.")

y\_pred = model.predict(X)

r2 = r2\_score(y, y\_pred)

mse = mean\_squared\_error(y, y\_pred)

print(f"R-squared: {r2:.4f}")

print(f"Mean Squared Error: {mse:.4f}")

plt.scatter(X, y, color='blue', label="Actual Data")

plt.plot(X, y\_pred, color='red', linewidth=2, label="Regression Line")

plt.xlabel("Building Height")

plt.ylabel("Construction Cost")

plt.title("Simple Linear Regression: Building Height vs Construction Cost")

plt.legend()

plt.show()