Title:1 Optimizing Linear Regression Model to Predict Medical Insurance Charges **Objective:** To build and optimize a **Linear Regression** model that predicts a person's **medical insurance charges** based on demographic and health attributes, ensuring: **Maximum R^2 Score** (better prediction accuracy) **Minimum Mean Squared Error (MSE)** (lower prediction error)
Dataset: File: insurance.csv

The goal is to **predict medical insurance charges** using a **Linear Regression model** based on personal and health attributes like age, gender, BMI, smoking status, number of children, and region.

- The **BMI** values are first **categorized** into four groups: *Underweight*, *Normal weight*, *Overweight*, and *Obese*.

Different **test sizes** (0.1, 0.2, 0.3, 0.4, 0.5) and **random states** (1 to 49) are tried to find: The model with the **highest R^2 score** (better prediction accuracy) The model with the **lowest Mean Squared Error (MSE)** (lower prediction error)

- After finding the best models, we **predict the insurance charges** for a **new individual** with specific given features (age 34, male, BMI 26, 3 children, smoker, southwest region).

Title:2 ProblemDefinition: The objective is to predict ice cream sales based on temperature using **polynomial regression**. The code identifies the best polynomial degree (from 1 to 5) by evaluating models through the **R-squared (R^2) score** to ensure optimal balance between accuracy and complexity. After finding the best degree, the final model is trained and used to **predict ice cream sales for a temperature of 4°C ("Don't split data)**

Title:3 Problem Definition: The goal is to predict hotel booking status using the K-Nearest Neighbors (KNN) classification algorithm. The dataset is preprocessed by one-hot encoding categorical features. The data is split into training and testing sets (80%-20%). The model is optimized by trying different values of 'k' (1 to 20), selecting the one that gives the highest accuracy. Finally, the best KNN model is trained and evaluated to predict booking status effectively.

Title: 4 Problem Definition:¶ The objective is to predict hotel booking status using a Decision Tree classifier. The dataset undergoes preprocessingone-hot encoding for categorical features. The data is split into training and testing sets (80%-20%). The model is optimized by trying different values of max_depth (1–20), selecting the depth that achieves the highest accuracy. The final model is trained with the best max_depth, and its performance is evaluated using accuracy score, classification report, and confusion matrix.

"Hotel Reservations.csv" for problem 3 and problem 4

Prdict for problem 3 and 4

'no_of_adults': 2, 'no_of_children': 1, 'no_of_weekend_nights': 2, 'no_of_week_nights': 3, 'type_of_meal_plan': 'Meal Plan 1', 'required_car_parking_space': 0, 'room_type_reserved': 'Room_Type 1', 'lead_time': 120, 'arrival_year': 2023 'arrival_month': 8 'arrival_date': 15, 'market_segment_type': 'Online', 'repeated_guest': 0, 'no_of_previous_cancellations': 0, 'no_of_previous_bookings_not_canceled': 1, 'avg_price_per_room': 110.5, 'no_of_special_requests': 1