

## **Q1. Problem Statement: Linear Regression**

Load the housing\_price.csv dataset to a DataFrame and perform the following tasks: The housing\_price dataset contains all numeric data and the median\_house\_value column is our target variable, so with help of linear regression build a model that can predict accurate house prices.

Perform the below task and build a model.

1. Load the housing\_price dataset into DataFrame
2. Find the null value and drop it, if it is there
3. Split x and y into train and test data set based on test size as 0.2 and random\_state as 10
4. Call the LinearRegression model then fit the model using train data
5. Print R2 value, coefficient and intercept
6. Compare actual and predicted values.
7. Print the final summary

## Q2. Problem Statement: Logistic Regression

You are given a categorical dataset – “Heart\_Disease.csv.” Load the dataset into a DataFrame. Considering the “TenYearCHD” column as the target variable, perform the following tasks:

1. Explore the “Heart\_Disease.csv” dataset, identify the null values and fill them with the mean value of their respective columns
2. Split the data into test and train parts using `train_test_split()` function in 80:20 ratio (80% train, 20% test)
3. Perform scaling of numeric data using the `StandardScaler()` function
4. Build a Logistic regression model using the train dataset and test the model using the test dataset
5. Print the classification report of the model
6. Calculate the confusion matrix and plot the same using a heatmap
7. Calculate and print the accuracy score of the model
8. Print the decision boundary for  $\theta = 0$ ,  $\theta = 1$  and 2

### Q3. Problem Statement: Titanic Survival Prediction

You are given a categorical dataset – “titanic.csv.” Load the dataset into a DataFrame. Considering the “Survived” column as the target variable, perform the following tasks:

1. Explore the “titanic.csv” dataset, identify the null values and fill them with the mean value of their respective columns
2. Split the data into test and train parts using train\_test\_split() function in 80:20 ratio (80% train, 20% test)
3. Perform scaling of numeric data using the StandardScaler() function
4. Build a Logistic regression model using the train dataset and test the model using the test dataset
5. Print the classification report of the model
6. Calculate the confusion matrix and plot the same using a heatmap
7. Calculate and print the accuracy score of the model
8. Print the decision boundary for  $\theta = 0$ ,  $\theta = 1$  and 2