**Boston House Price Prediction**

**Problem Statement**

You want to be the best real estate agent out there. In order to compete with other agents in your area, you decide to use machine learning. You are going to use various statistical analysis tools to build the best model to predict the value of a given house. Your task is to find the best price your client can sell their house.

Build linear regression model with SGD to predict the median value of owner-occupied homes from the dataset having attributes such as: Average number of rooms, accessibility to radial highways etc.

**Data Description**

The Boston data frame has 506 rows and 14 columns.

This data frame contains the following columns:

* CRIM: Per capita crime rate by town
* ZN: Proportion of residential land zoned for lots over 25,000 sq. ft
* INDUS: Proportion of non-retail business acres per town
* CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
* NOX: Nitric oxide concentration (parts per 10 million)
* RM: Average number of rooms per dwelling
* AGE: Proportion of owner-occupied units built prior to 1940
* DIS: Weighted distances to five Boston employment centers
* RAD: Index of accessibility to radial highways
* TAX: Full-value property tax rate per $10,000
* PTRATIO: Pupil-teacher ratio by town
* B: 1000(Bk — 0.63) ², where Bk is the proportion of [people of African American descent] by town
* LSTAT: Percentage of lower status of the population
* MEDV: Median value of owner-occupied homes in $1000s

**Evaluation Parameters**

Evaluation will be based on:

* Model Fitting
* Model Comparison
* Model Selection

**Model Fitting**

Fit a Linear Regression with SGD model to predict the median value of owner-occupied homes.

**Model Comparison**

Change the hyperparameters of the linear regression with SGD model and compare results.

**Model Selection**

Select the best model. Model selection to be based on model MSE and R^2.

**Expected Outcome**

Low MSE is expected while predicting the outcome using test data and high coefficient of determination, R^2.