

## **Beginner Level - Linear Regression**

### **Assignment: Predicting House Prices**

**Task:** Given a dataset containing information about houses (e.g., size, number of bedrooms, neighborhood), ask candidates to build a linear regression model to predict house prices.

**Dataset:** You can use a dataset like the Boston Housing Prices dataset available in Scikit-

Learn or a similar one from Kaggle.

**Evaluation:** Assess their ability to pre-process data, create a linear regression model, and evaluate its performance.

### **GOOGLE COLAB NOTEBOOK**

[https://colab.research.google.com/drive/1nqyMToznzzdKwg7o7zJhvZZBIZ7jc\\_Jj?usp=sharing](https://colab.research.google.com/drive/1nqyMToznzzdKwg7o7zJhvZZBIZ7jc_Jj?usp=sharing)

# **Predictive Analysis of Housing Prices Using Linear Regression**

## **Abstract:**

The objective of this project is to construct a predictive model that can accurately estimate the prices of houses based on a set of features such as size, number of bedrooms, and neighborhood characteristics. The model employs linear regression, a statistical approach for modeling the relationship between a dependent variable and one or more independent variables.

## **Introduction:**

The housing market is influenced by various factors that affect the prices of homes. Understanding these factors and quantifying their impact on housing prices can aid potential buyers, sellers, and investors in making informed decisions. This project utilizes the Boston Housing Prices dataset, a widely recognized dataset in machine learning, which provides comprehensive data on various housing attributes.

## **Methodology:**

The methodology encompasses several stages:

### **1. Data Collection:**

The Boston Housing Prices dataset from Scikit-learn library is used, which includes information on 506 census tracts of Boston from the 1970 census.

### **2. Data Pre-processing:**

This stage involves cleaning the data, handling missing values, normalizing data, and encoding categorical variables to prepare the dataset for the model.

### **3. Model Building:**

A linear regression model is constructed using the pre-processed data. The model is trained to understand the relationship between the house features (independent variables) and the house prices (dependent variable).

### **4. Model Evaluation:**

The model's performance is evaluated using appropriate metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared value. Cross-validation techniques are also employed to ensure the model's robustness.

### **Results:**

The linear regression model demonstrates a significant ability to predict housing prices when tested against a validation set. The performance metrics indicate a satisfactory level of accuracy, suggesting that the model can be a useful tool for estimating house prices.

### **Conclusion:**

This project illustrates the effectiveness of linear regression in predicting housing prices. The model's success is contingent upon the quality of data pre-processing and the selection of relevant features. Future work may involve exploring more complex models and incorporating additional data sources to enhance predictive accuracy.

SAMPLE OUTPUT





