Multivariate Linear Regression House Price Prediction

This was a task issued by Sunayu as an evaluation of approach, style, ability, and performance. The R script will analyze a house price data set to predict the price of a house via different independent variables. It will move through data familiarization, preprocessing, splitting the training and testing data, exploratory data analysis, model evaluation, predictions, and results. The R Markdown report is the same script, maybe a little more refined, prettier, and a one stop shop to read the results.

Problem Statement

Housing markets, financial institutions, and real-estate agents want to predict housing prices for regions in the USA as accurately as possible. It aids in negotiation, marketing, and planning of future operations. You were given a data set to make housing price predictions upon. You have been tasked to perform a multivariate linear regression model on the data set. Your task is to create a model that will help in estimating the prices of housing and what they should expect to sell for.

Data set includes 5000 observations and 7 variable columns within a CSV file.

Variables include:

- *Average Income* of the house owner of the city that the house is located within.
- *Average Age* of the houses in the same city.
- *Average Number of Rooms* for houses in the same city.
- *Average Number of Bedrooms* for houses in the same city.
- *Area Population* of the city.
- *Price* that the house sold for.
- *Address* of the houses.