**Project Overview**

In this project, you will apply basic machine learning concepts on data collected for housing prices in the Boston, Massachusetts area to predict the selling price of a new home. You will first explore the data to obtain important features and descriptive statistics about the dataset. Next, you will properly split the data into testing and training subsets, and determine a suitable performance metric for this problem. You will then analyze performance graphs for a learning algorithm with varying parameters and training set sizes. This will enable you to pick the optimal model that best generalizes for unseen data. Finally, you will test this optimal model on a new sample and compare the predicted selling price to your statistics.

Prepare for this project with our supplementary course in [Model Evaluation and Validation](https://www.udacity.com/course/viewer#!/c-ud725-nd).

**Project Highlights**

This project is designed to get you acquainted to working with datasets in Python and applying basic machine learning techniques using NumPy and Scikit-Learn. Before being expected to use many of the available algorithms in the sklearn library, it will be helpful to first practice analyzing and interpreting the performance of your model.

Things you will learn by completing this project:

* How to use NumPy to investigate the latent features of a dataset.
* How to analyze various learning performance plots for variance and bias.
* How to determine the best-guess model for predictions from unseen data.
* How to evaluate a model’s performance on unseen data using previous data.