Overview

In this project I am going to be applying data science research method skills using Python.

Concepts covered:

* Data Science methodology
* Statistical research methods
* Statistical programming in Python

The Dataset

The dataset I’ll be using contains housing price data from Boston, Massachusetts, USA; a description can be found [here](https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html). This is a common data set and is available directly in R and in the *sklearn.datasets* module from scikit-learn in Python. The data set can be downloaded directly from [here](https://archive.ics.uci.edu/ml/machine-learning-databases/housing/).

Within the data set you’ll find variables such as per capita crime rate, average number of rooms, property tax rate, and more. Note that the variables are not for individual households, but for groups of households known as “census tracts.”

For this assignment, assume that the data is from a **sample** of census tracts in the Boston area, not all census tracts.