# Boston house prices dataset

The Boston house-price data has been used in many machine learning papers that address regression problems. It consists of price of houses in various areas in Boston. Alongside with price, the dataset also provide information about the specific area.

## Dataset characteristics

**Number of Instances**: 506

**Number of Attributes**: 13 numeric/categorical predictive. Median Value (MEDV, attribute 14) is usually the target.

**Attribute Information** (in order):
1. CRIM per capita crime rate by town
2. ZN proportion of residential land zoned for lots over 25,000 sq.ft.
3. INDUS proportion of non-retail business acres per town
4. CHAS Charles River dummy variable (= 1 if tract bounds river; 0 5. otherwise)
5. NOX nitric oxides concentration (parts per 10 million)
6. RM average number of rooms per dwelling
7. AGE proportion of owner-occupied units built prior to 1940
8. DIS weighted distances to five Boston employment centres
9. RAD index of accessibility to radial highways
10. TAX full-value property-tax rate per $10,000
11. PTRATIO pupil-teacher ratio by town
12. B 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town
13. LSTAT % lower status of the population
14. MEDV Median value of owner-occupied homes in $1000's

**Missing Attribute Values**: None

**Creator**: Harrison, D. and Rubinfeld, D.L.

**References**:
- Belsley, Kuh & Welsch, 'Regression diagnostics: Identifying Influential Data and Sources of Collinearity', Wiley, 1980. 244-261.

- Quinlan,R. (1993). Combining Instance-Based and Model-Based Learning. In Proceedings on the Tenth International Conference of Machine Learning, 236-243, University of Massachusetts, Amherst. Morgan Kaufmann.

- **https://archive.ics.uci.edu/ml/machine-learning-databases/housing/**

- **https://medium.com/@haydar_ai/learning-data-science-day-9-linear-regression-on-boston-housing-dataset-cd62a80775ef**

- **https://www.kaggle.com/c/boston-housing**