# Лабораторная работа №4
# Линейная регрессия

**1.** Загрузить датасет "Boston Housing" .

**2.** Изучите структуру и содержание данных. Вывести первые несколько строк и описание каждого столбца

**3.** Провести предварительный анализ данных, включая проверку наличия пропущенных значений, выбросов и корреляции между переменными. Если необходимо, то провести стандартизацию (нормализацию) данных, чтобы все признаки имели одинаковый масштаб.

**4.** Исследовательский анализ данных: построить матрицу корреляций, чтобы выявить сильные и слабые связи между признаками и целевой переменной (медианной стоимостью занимаемых домов); построить гистограммы и ящики с усами для каждого признака, чтобы изучить их распределение и выбросы; проверить наличие мультиколлинеарности между признаками и исключить коррелирующие признаки, если это необходимо

**5.** Построить графики, которые помогут понять данные *(можно выполнить не 4, а 2 подпункта):*

- создать гистограммы распределения каждого признака и целевой переменной;
- построить диаграммы рассеивания (scatter plots) между целевой переменной и каждым из признаков, чтобы выявить линейные и нелинейные зависимости;
- построить матрицу корреляций между всеми признаками, включая целевую переменную;
- создать box plots для исследования выбросов и распределения значений в признаках.

**6.** Разбить данные на обучающую и тестовую выборки в соотношении 70/30.

**7.** Обучить модель линейной регрессии на обучающей выборке.

**8.** Оценить качество модели на тестовой выборке с помощью таких метрик как средняя абсолютная ошибка (MAE), средняя квадратичная ошибка (MSE), корень среднеквадратичной ошибки (RMSE) и коэффициент детерминации (R^2).

**9.** Визуализируйте результаты предсказания, сравнив исходные значения цены недвижимости с предсказанными значениями.

**10.** Провести кросс-валидацию модели и оценить ее качество с помощью метрик MSE, RMSE и R^2.

**11.** Попробовать улучшить качество модели путем отбора наиболее значимых переменных, использования регуляризации или других методов.