

# Boston Housing Prices Summary

In this project I provide basic Exploratory Data Analysis for the Boston Housing Prices dataset. The packages used were Pandas, Pandas\_Profiling, NumPy, Matplotlib, and Seaborn. The numerical data was log scaled, and there were two categorical variables with over one hundred classes. The null values were imputed, either with zero, or with the mean after grouping by the number of bedrooms. Usage of groupby distributes the null values more evenly across the distribution. At least one column had multiple types of null values, and some of the numerical data was encoded as a string, with commas. The encoding of the variables was fixed for modelling. Some numerical features were in fact categorical features, and must be treated that way in the modelling process.