California Housing Price

Context

The data contains information from the 1990 California census. So, although it may not help you with predicting current housing prices, it does provide an accessible introduction dataset for teaching people about the basics of machine learning.

Features

The data pertains to the houses found in a given California district and some summary stats based on the 1990 census data. Be warned that the data aren't cleaned, so some preprocessing steps are required! The columns are as follows, and their names are pretty self-explanatory:

- longitude
- latitude
- Housing *median* age
- total_rooms
- total_bedrooms
- population
- households
- median_income
- median *house* value
- ocean_proximity

Acknowledgments

This data was initially featured in the following paper:

Pace, R. Kelley, and Ronald Barry. "Sparse spatial autoregressions." Statistics & Probability Letters 33.3 (1997): 291-297.