

Exercise0: Dataset Description

Ouassim Kiassa, 12113875, ouassim.kiassa@student.tuwien.ac.at

Antal Spilyka, 11907555, e11907555@student.tuwien.ac.at

Antoine Origer, XXXXX, XXXXXX

March 9, 2024

Dataset 1 Overview

This dataset consists of housing information extracted from the 1990 California census. It comprises demographic, geographic, and economic features pertaining to housing blocks, including location, age, number of rooms and bedrooms, population, household count, median income, median house value, and proximity to the ocean.

Source: 1990 California census data.

Attributes Description

The dataset comprises a total of 20,640 observations, each described by 10 distinct attributes. Among these attributes, nine are quantitative, embodying numeric data that can be either interval or ratio in nature. There is one categorical attribute, which is nominal, indicating classification without an inherent order. No attributes in the dataset are identified as ordinal, which would signify categorical data with a defined sequence or ranking.

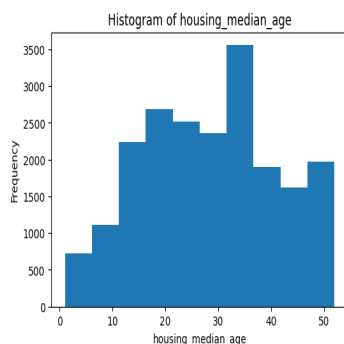


Figure 1: Distribution
Housing Median
Age

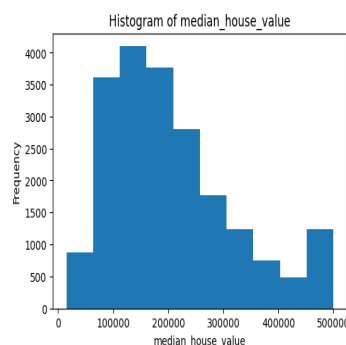


Figure 2: Distribution Me-
dia House Value

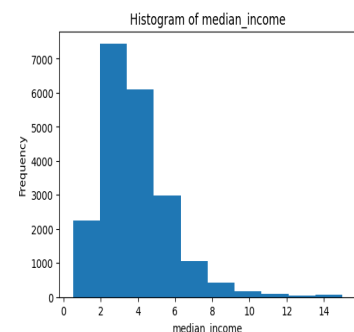


Figure 3: Distribution Me-
dian housing In-
come

Target Attribute

Objective: The dataset's primary application lies in the prediction of median housing values within Californian districts. This prediction is based on the analysis of the dataset's diverse features.

Count	20,640
Mean	206,855
Std	115,395
Min	14,999
25%	119,600
50%	179,700
75%	264,725
Max	500,001

Table 1: Descriptive statistics of the target attribute: Median House Value.