

Laboratory assignment

Component 2

Authors: Ichim Stefan, Mirt Leonard

Group: 246/1

November 11, 2024

1 Data Analysis

First, to be able to take statistics from features, we made sure that all the features are in numeric form. We tranformed the ocean proximity feature into a numeric features as follows: '< 1H OCEAN': 0, 'INLAND': 1, 'ISLAND': 2, 'NEAR BAY': 3, 'NEAR OCEAN': 4.

Table 1: Statistical Summary of Features

Feature	Count	Mean	Std	Min	Max
Longitude	20640	-119.57	2.00	-124.35	-114.31
Latitude	20640	35.63	2.14	32.54	41.95
Housing Age	20640	28.64	12.59	1.00	52.00
Total Rooms	20640	2635.76	2181.62	2.00	39320.00
Total Bedrooms	20433	537.87	421.39	1.00	6445.00
Population	20640	1425.48	1132.46	3.00	35682.00
Households	20640	499.54	382.33	1.00	6082.00
Median Income	20640	3.87	1.90	0.50	15.00
Median House Value	20640	206855.82	115395.62	14999.00	500001.00
Ocean Proximity	20640	1.17	1.42	0.00	4.00

1.1 Pearson Test: Data Correlation and Independence

1.2 Statistical Tests: Feature Importance

Table 2: Statistical Tests for Feature Importance

Feature	Pearson	Spearman	F-Score	Mutual Information
Median Income	0.688	0.677	1.000	0.051
Latitude	0.144	0.166	0.024	0.049
Total Rooms	0.134	0.206	0.020	0.006
Housing Median Age	0.106	0.075	0.013	0.004
Ocean Proximity	0.082	0.133	0.007	0.028
Households	0.066	0.113	0.005	0.004
Total Bedrooms	0.049	0.086	0.003	0.003
Longitude	0.046	0.070	0.002	0.053
Population	0.025	0.004	0.001	0.003

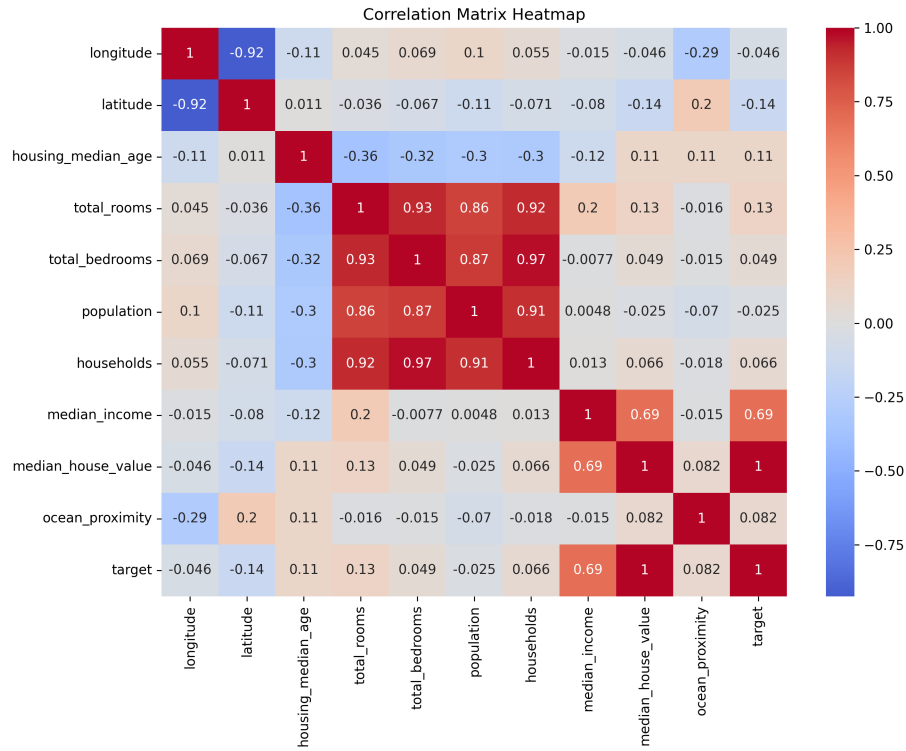


Figure 1: Pearson test results: 1 - strong positive, -1 - strong negative, 0 - no connection

1.3 Data Visualization and Interpretation

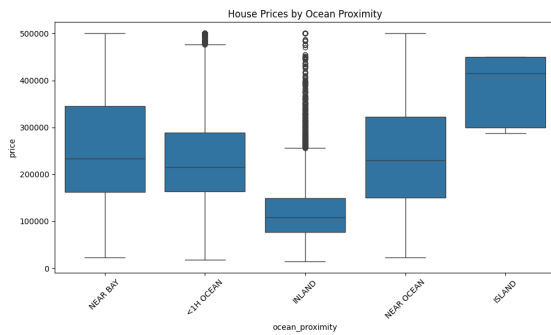


Figure 2: House prices variation by ocean proximity

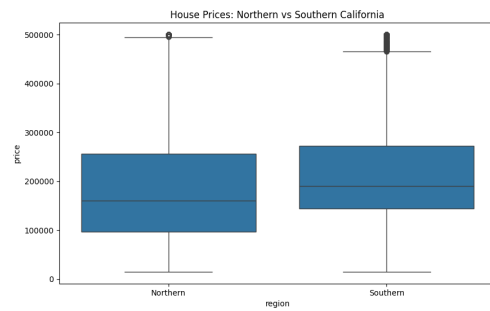


Figure 3: Price comparison: Northern vs Southern California

1.4 Feature distribution



Figure 4: House prices across different area types

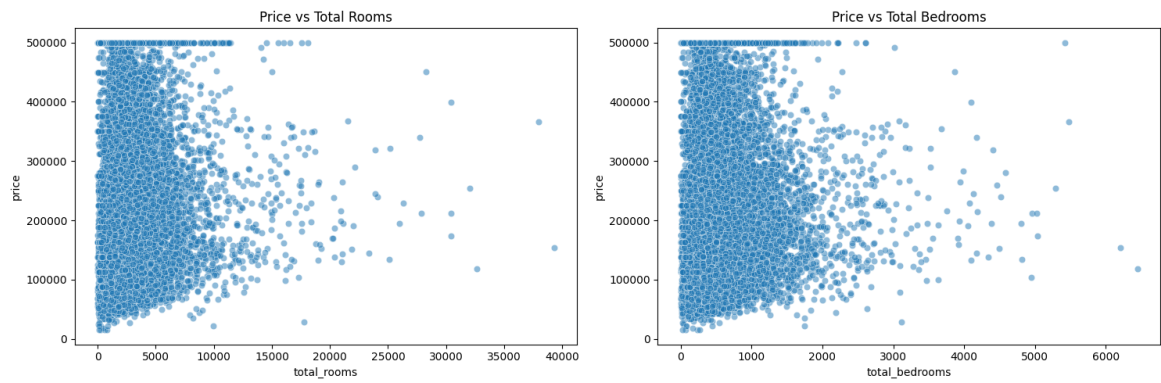


Figure 5: Relationship between house size and price

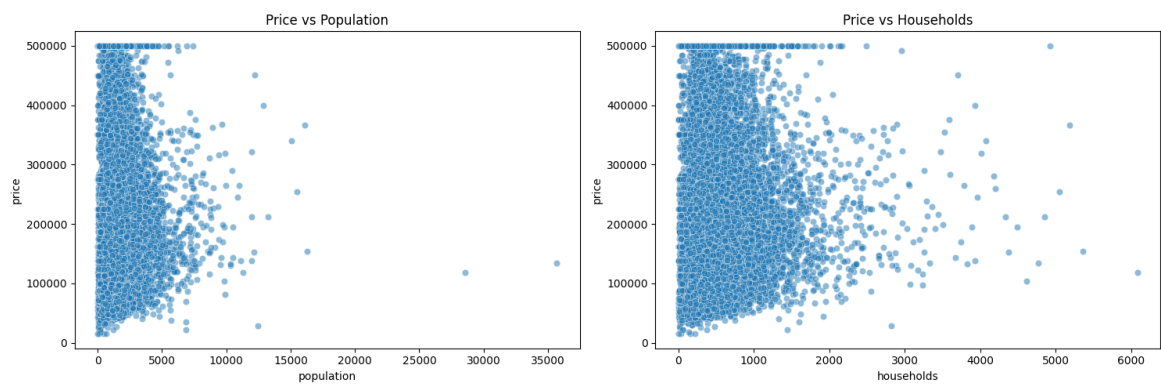


Figure 6: Impact of neighborhood characteristics on price

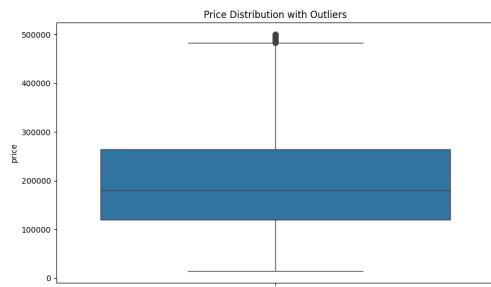


Figure 7: Distribution of price outliers

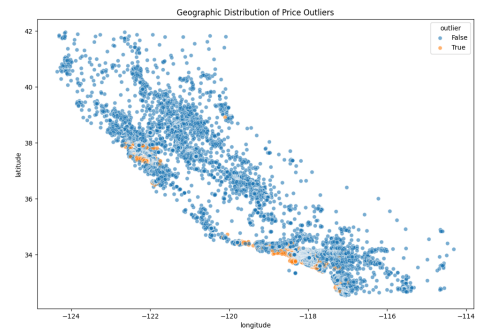


Figure 8: Geographic distribution of outliers

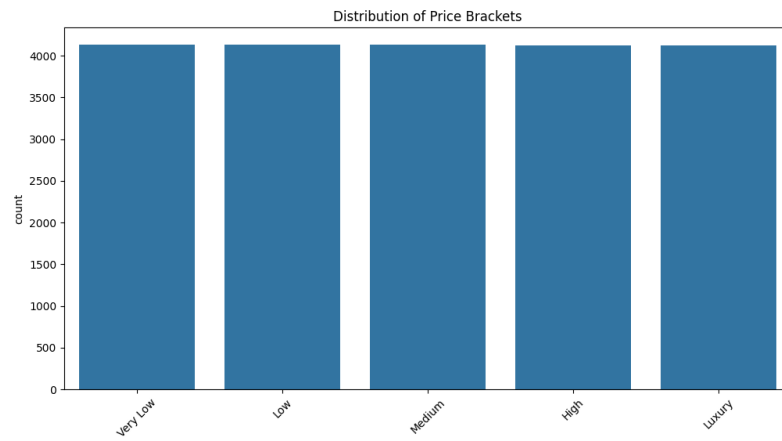


Figure 9: Distribution of properties across price brackets

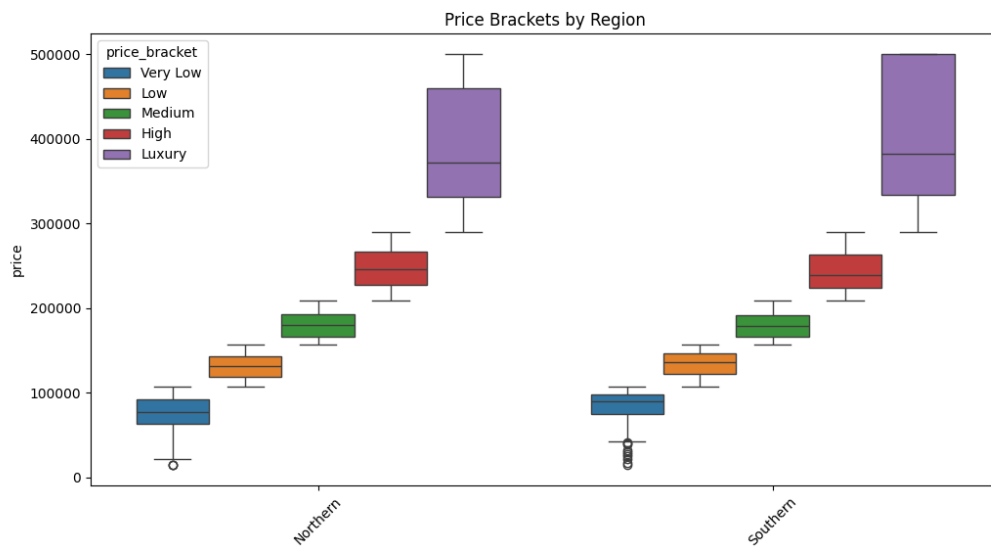


Figure 10: Regional market characteristics by price bracket

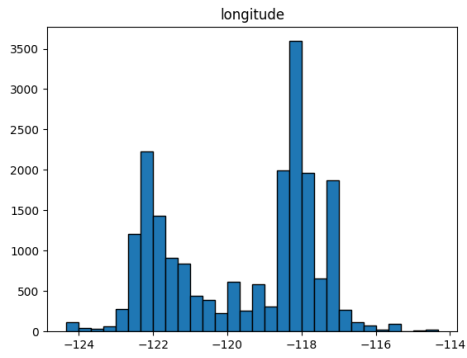


Figure 11: Longitude distribution

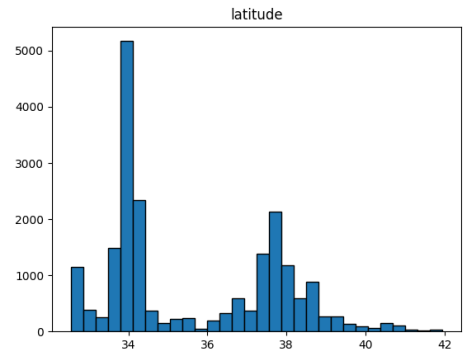


Figure 12: Latitude distribution

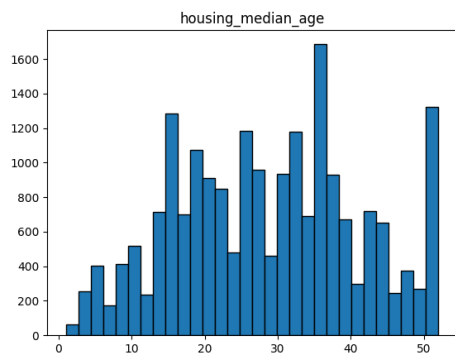


Figure 13: Housing median age distribution

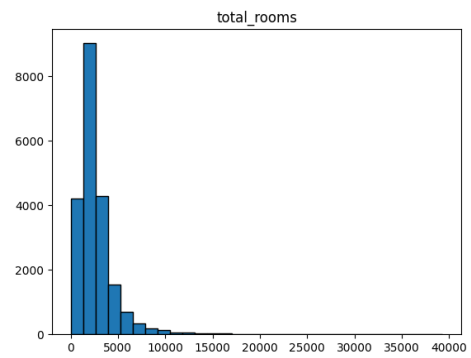


Figure 14: Total rooms distribution

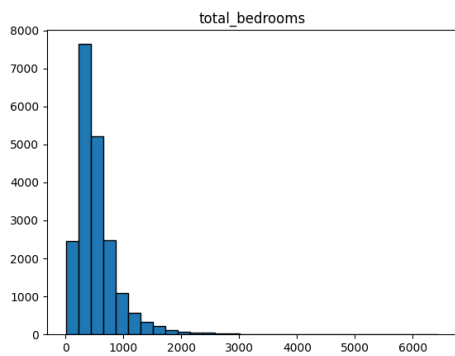


Figure 15: Total bedrooms distribution

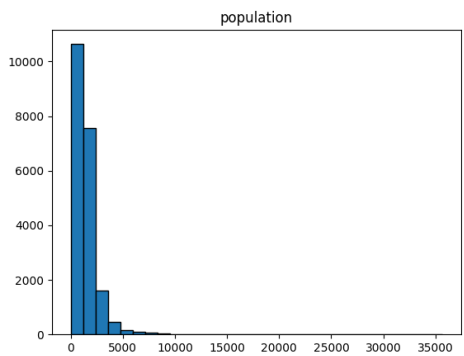


Figure 16: Population distribution

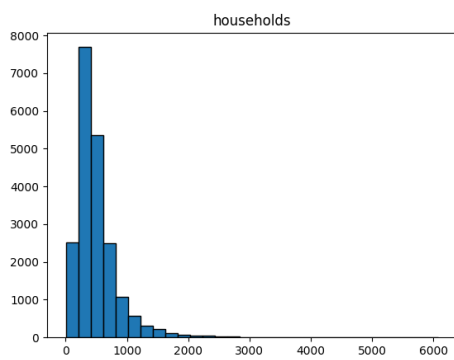


Figure 17: Households distribution

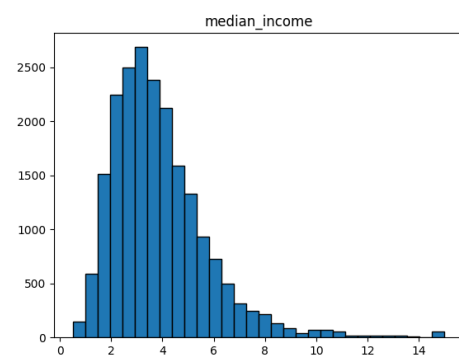


Figure 18: Median income distribution

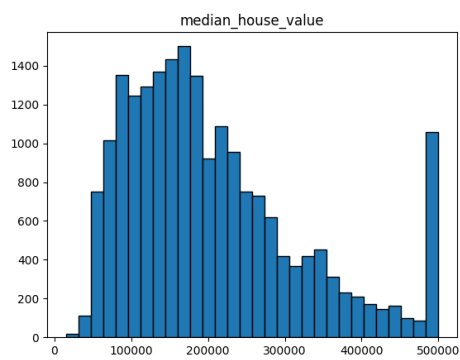


Figure 19: Median house value distribution

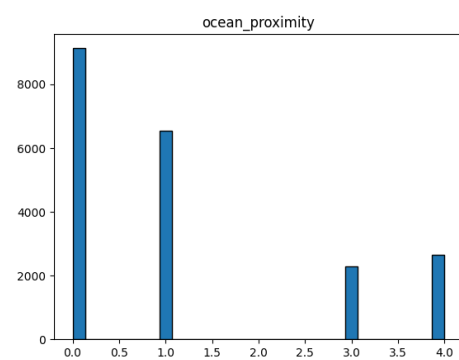


Figure 20: Ocean proximity distribution