Step 1: Load Data
Houses: 506, Features: 13

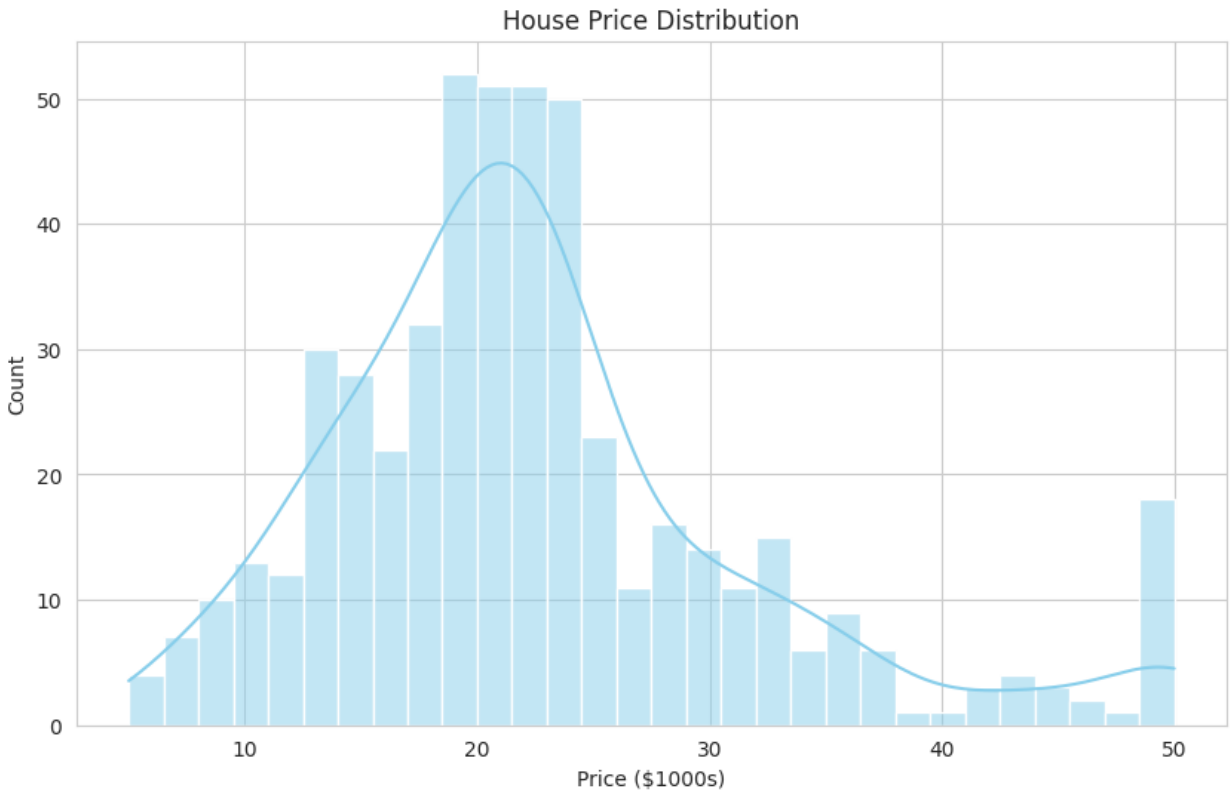
First 5 rows:

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296.0
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242.0
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242.0
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222.0
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222.0

	PTRATIO	B	LSTAT	PRICE
0	15.3	396.90	4.98	24.0
1	17.8	396.90	9.14	21.6
2	17.8	392.83	4.03	34.7
3	18.7	394.63	2.94	33.4
4	18.7	396.90	5.33	36.2

House Price Summary:
Min: $5,000.00
Max: $50,000.00
Mean: $22,532.81

Step 2: Explore Data



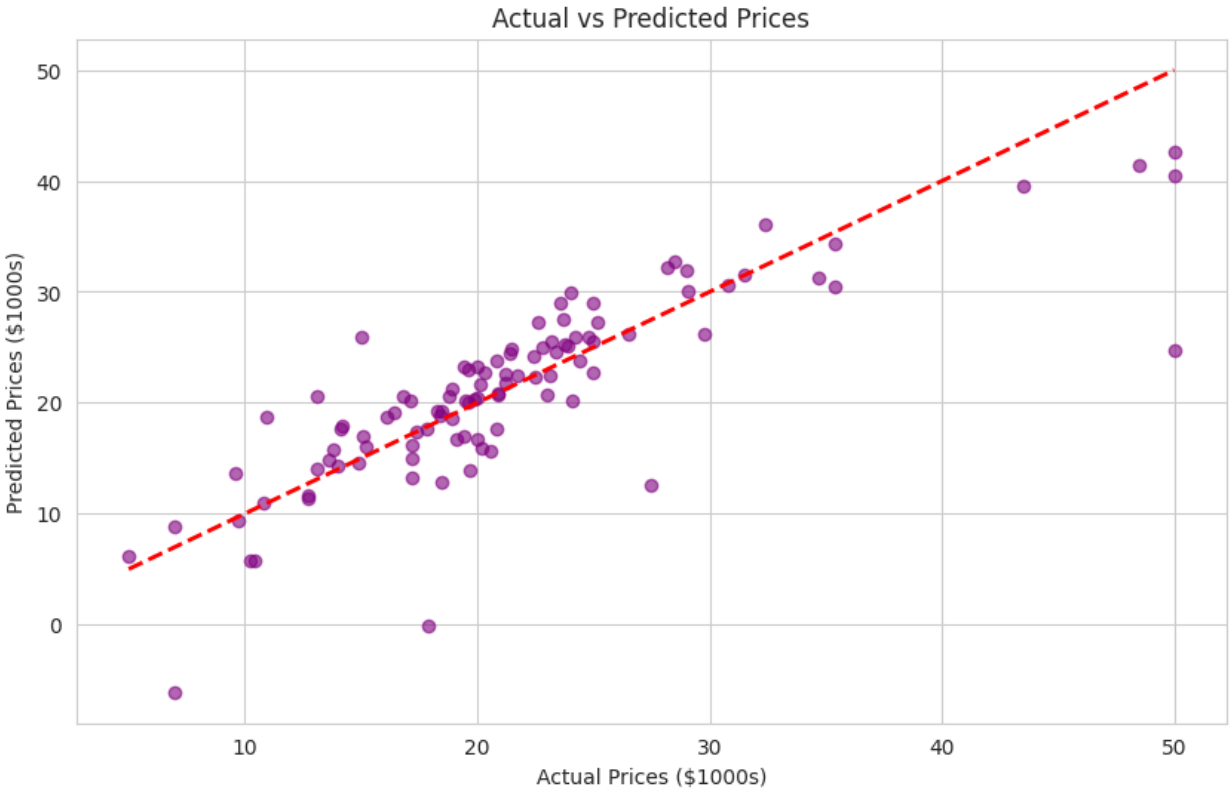

```
Feature Correlation with Price:
PRICE      1.000000
RM         0.695360
ZN         0.360445
B          0.333461
DIS        0.249929
CHAS       0.175260
AGE       -0.376955
RAD       -0.381626
CRIM      -0.388305
NOX       -0.427321
TAX       -0.468536
INDUS     -0.483725
PTRATIO   -0.507787
LSTAT     -0.737663
Name: PRICE, dtype: float64

Most correlated feature: RM
```

Observation: Price distribution is right-skewed. Most houses are mid-priced, few are expensive.

Step 3: Train Model
Train samples: 404, Test samples: 102
Model trained successfully!

Model Performance:
R^2 Score: 0.6688
RMSE: 4.9286 ($1000s)



Step 4: Analysis

R^2 score: 0.6688 → Model explains about 66.9% of variance.

Model predicts mid-range houses well, but underestimates very expensive ones.