

Output Explanation - California Housing Price Prediction Using Linear and Ridge Regression

1. Correlation Matrix (Heatmap)

The correlation heatmap shows how strongly each feature is related to other features.

Values range from -1 to +1

+1 (red) → strong positive correlation

-1 (blue) → strong negative correlation

0 (white/light) → weak or no correlation

From the heatmap

total_rooms, total_bedrooms, population, households are highly positively correlated with each other (dark red blocks)

2. Linear Regression

Linear Regression: RMSE $\approx 70,803$ $R^2 \approx 0.6257$

RMSE (Root Mean Squared Error)

→ Average prediction error in house price

→ Lower RMSE = better model

R^2 Score (Coefficient of Determination)

→ Model explains 62.57% of the variance in house prices

→ Remaining variation is due to other factors not in the model

→ this shows moderate performance of Linear Regression

3. Ridge Regression

Ridge Regression: RMSE $\approx 70,828$ $R^2 \approx 0.6258$

RMSE is slightly different from Linear Regression

R^2 score is slightly improved

Ridge Regression adds L2 regulation

→ Penalizes large coefficients

→ Handles multicollinearity better

→ Even though improvement is small, Ridge gives More stable model Better generalization

4. Comparison

Model	RMSE	R^2 Score	Interpretation
Linear Regression	~70,803	0.6257	Affected by multicollinearity
Ridge Regression	~70,828	0.6258	Handles multicollinearity better

5. Conclusion

The correlation matrix shows strong multicollinearity among housing features. Ridge Regression slightly improves model stability over Linear Regression by applying regularization, making it more suitable for this dataset

