

Numeric dataset

General information:

Name: California Housing Data (1990)

Label: median_house_value

Training Samples: 15287

Test Samples: 3822

Implementation details:

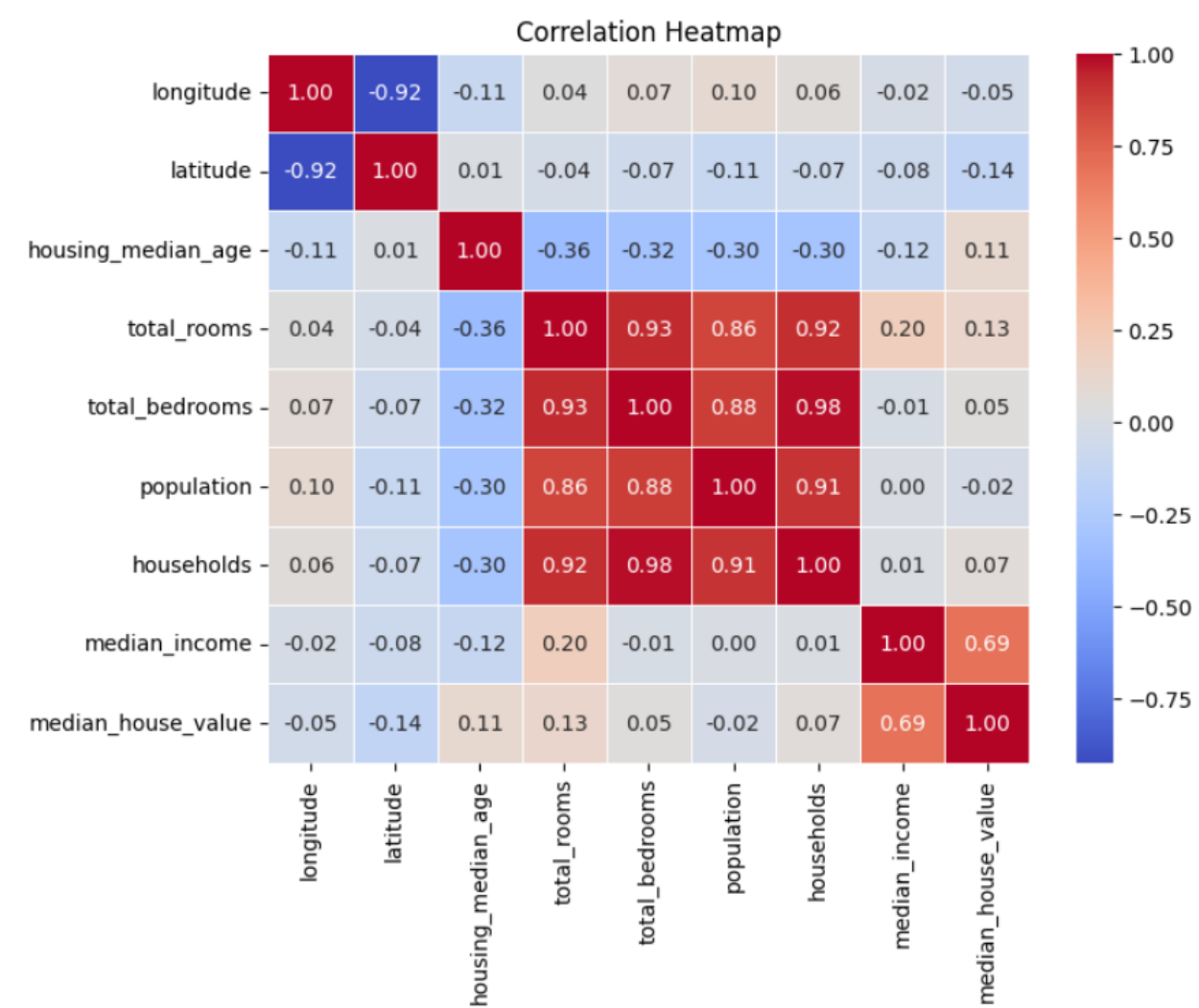
Feature Engineering:

New features: rooms_per_household , population_per_household

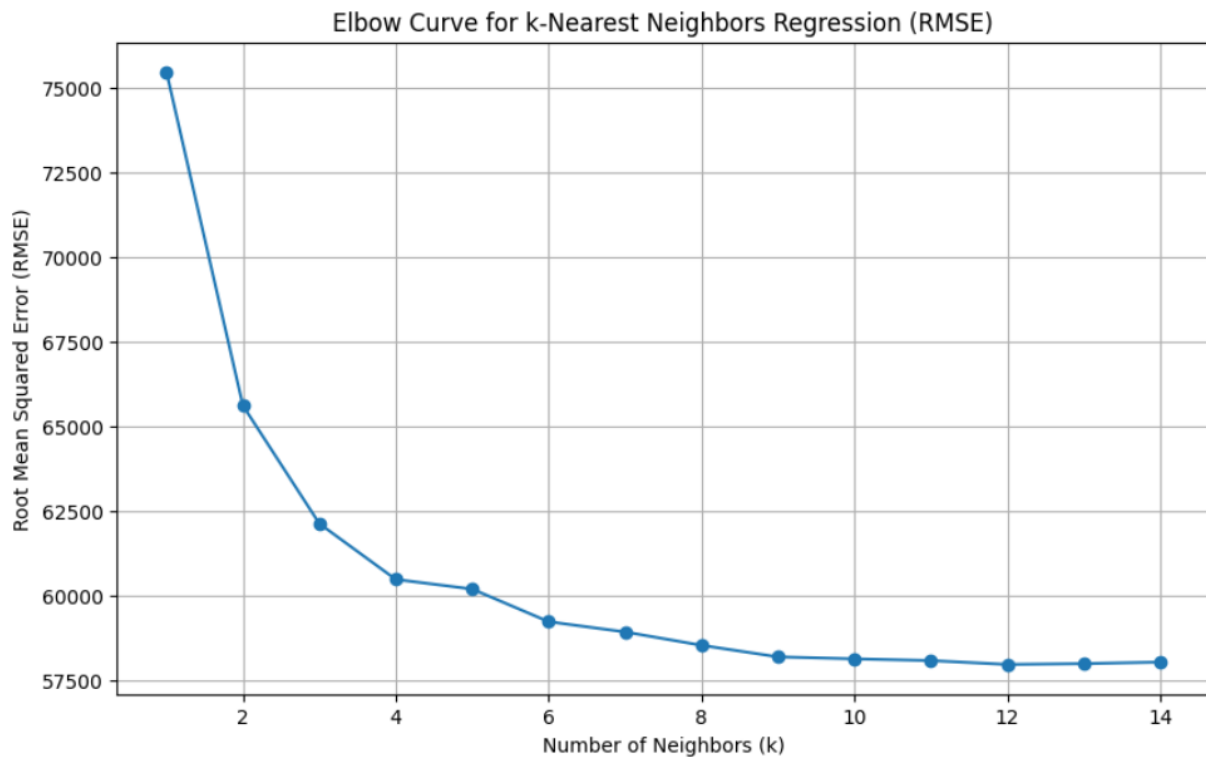
Cross Validation:

Number of folds: 5, **Ratio:** 16%

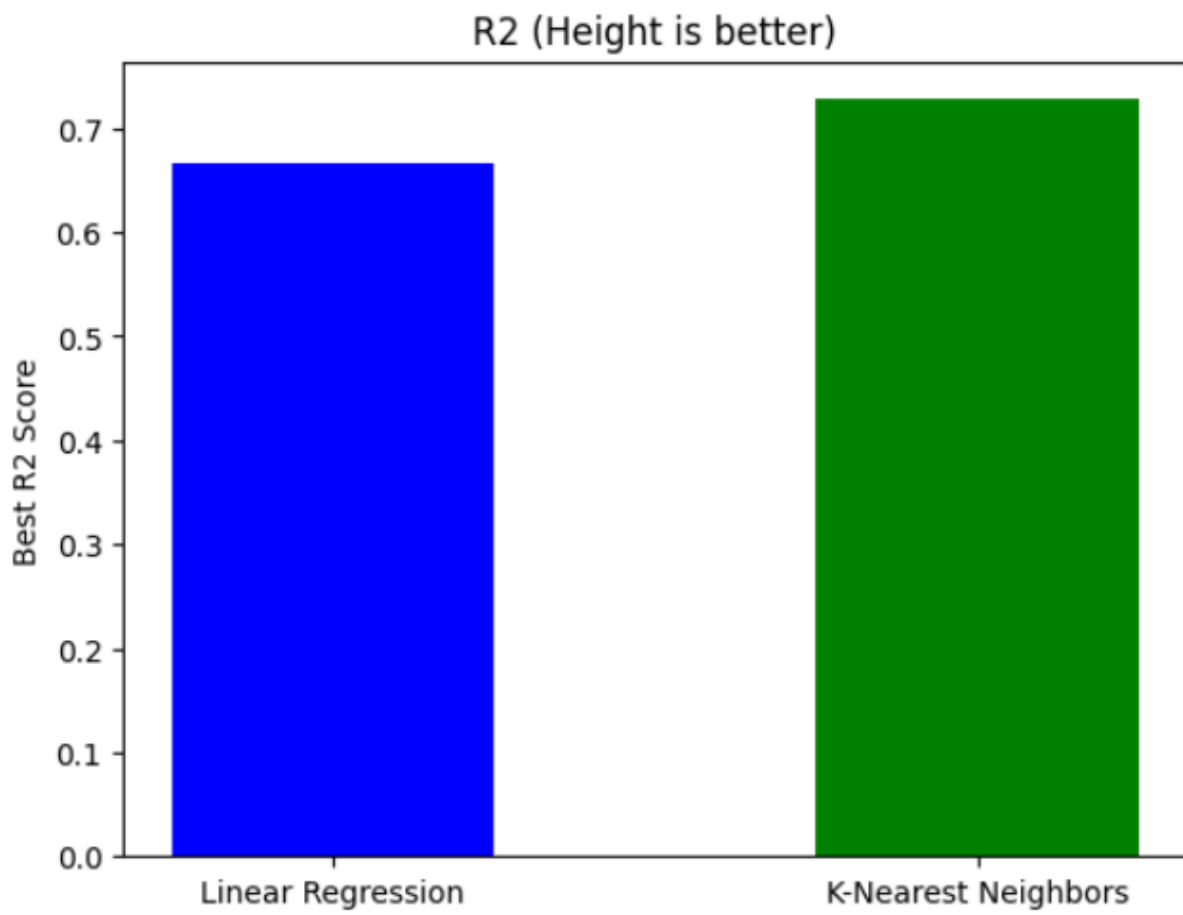
Correlation heatmap :



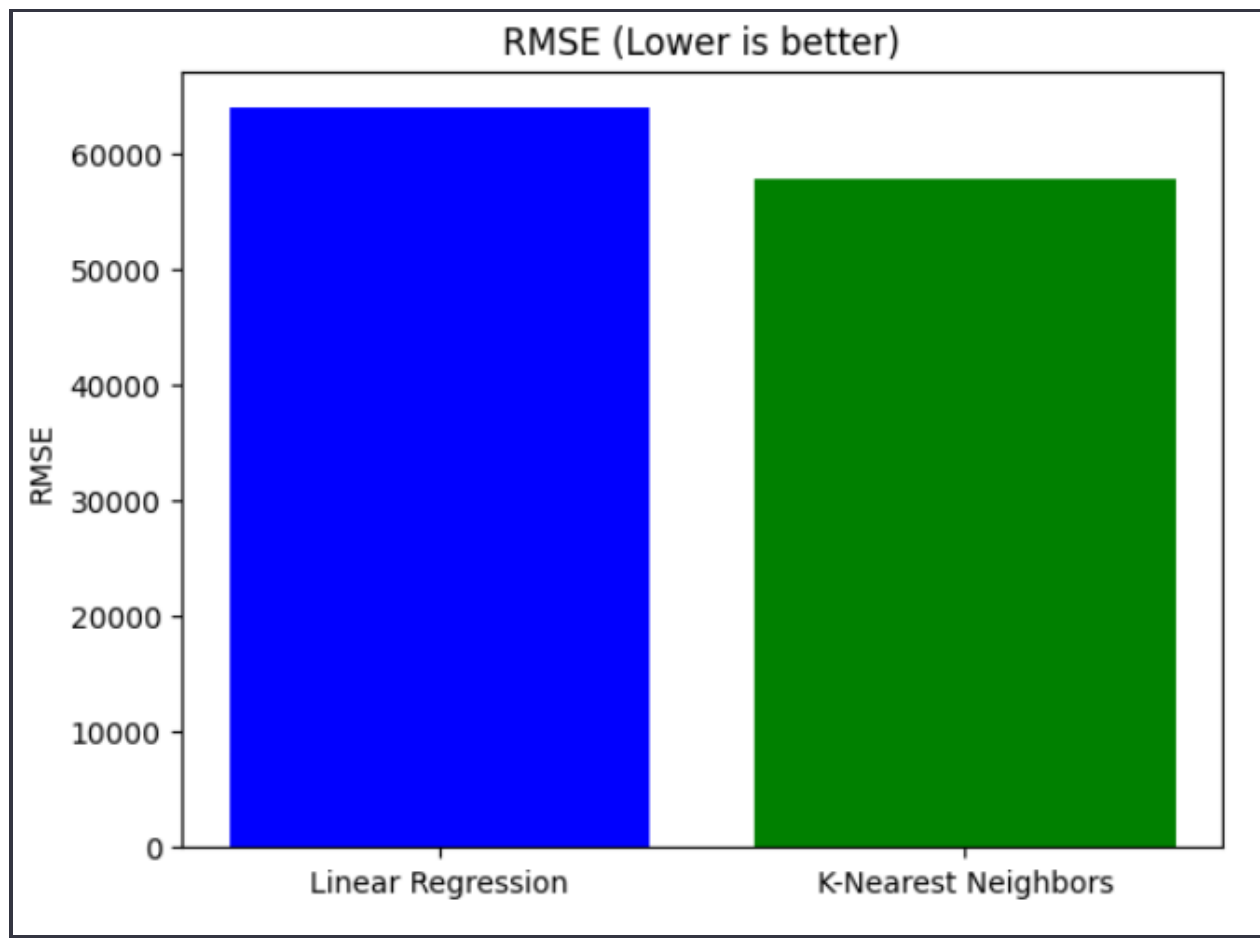
Best K for KNN :



R_squared between KNN and linear :



RSME between KNN and linear :



Accuracy of linear :

RMSE: 65664.14
R2 Score: 0.64
Accuracy: 0.66

Accuracy of knn :

K-Nearest Neighbors RMSE: 58518.26
K-Nearest Neighbors R2 Score: 0.72
Accuracy: 0.78
