b. i. 4.

|  |  |
| --- | --- |
| **Feature Name** | **Pearson Correlation** |
| housing\_median\_age | 0.106432 |
| total\_rooms | 0.133294 |
| total\_bedrooms | 0.049686 |
| population | -0.025300 |
| households | 0.064894 |
| median\_income | 0.688355 |

d. ii.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **MAPE** | **RMSE** | **R2** |
| Linear Regression | 0.29% | 265.06 | 0.6373 |

e. Discussion

i.

Feature Coefficient

7 median\_income 0.642888

4 total\_bedrooms 0.326248

6 households 0.220022

2 housing\_median\_age 0.111307

9 ocean\_proximity\_ISLAND 0.012391

11 ocean\_proximity\_NEAR OCEAN 0.009728

10 ocean\_proximity\_NEAR BAY -0.017095

3 total\_rooms -0.097549

8 ocean\_proximity\_INLAND -0.156682

5 population -0.410426

0 longitude -0.488375

1 latitude -0.496330

ii.

The magnitude and direction of the pair-wised correlation results between features and the output value can provide a more detailed interpretation of how the affect in the linear regression model. Features with higher coefficients in the linear regression model, are also the most influential for the output value.

iii.

The result of MAPE is 0.29%. MAPE measures the accuracy of predictions. A MAPE of 0.29% indicates that the model's predictions deviate by only 0.29% from the actual data and it is quite accurate.

The result of RMSE is 265.06. RMSE measures the average difference between values predicted by a model and the actual values. A RMSE of 265.06 indicates the average difference is 265.06 units. The result shows that the model's predictions are relatively close to the actual values on average.

The result of R2 is 0.6373. R2 is the proportion of the variation in the dependent variable that is predictable from the independent variables. An R2 of 0.6373 means that approximately 63.73% of the variability of the independent variable in the data set is accounted for.

iv.

For centering, centering before computing PCA can eliminate the influence cause by the mean values of variables.

For scaling, scaling before computing PCA can ensure all the variables have the same influence for PCA. It prevents the variables that have different scales will affect the PCA disproportionate.