```
import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
```

```
data = pd.read_csv('/content/sample_data/california_housing_train.csv')
```

```
X_train, X_test, y_train, y_test = train_test_split(data.drop('median_house_value', axis=1), data['median_house_value'], test_size=0
```

```
model = LinearRegression()
```

```
model.fit(X_train, y_train)
```

```
▾ LinearRegression
LinearRegression()
```

```
y_pred = model.predict(X_test)
```

```
print('Mean squared error:', mean_squared_error(y_test, y_pred))
print('Root mean squared error:', np.sqrt(mean_squared_error(y_test, y_pred)))
print('R-squared:', r2_score(y_test, y_pred))
```

```
Mean squared error: 4866883427.941616
Root mean squared error: 69763.05202570782
R-squared: 0.6387988057454814
```

```
import statsmodels.api as sm
import statsmodels.stats as sms
# Fit the model using statsmodels
model = sm.OLS(y_train, X_train)
results = model.fit()
print(results.summary())
```

```
                                 OLS Regression Results
================================================================================================
Dep. Variable:     median_house_value   R-squared (uncentered):                   0.900
Model:                            OLS   Adj. R-squared (uncentered):              0.900
Method:                 Least Squares   F-statistic:                          1.434e+04
Date:                Sat, 30 Dec 2023   Prob (F-statistic):                        0.00
Time:                        17:02:09   Log-Likelihood:                     -1.6120e+05
No. Observations:               12750   AIC:                                  3.224e+05
Df Residuals:                   12742   BIC:                                  3.225e+05
Df Model:                           8
Covariance Type:            nonrobust
================================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------
longitude          -2186.8871    136.970    -15.966      0.000   -2455.369   -1918.405
latitude           -8299.3461    434.304    -19.110      0.000   -9150.648   -7448.044
housing_median_age  1748.9889     57.034     30.666      0.000    1637.195    1860.783
total_rooms          -15.7229      1.062    -14.807      0.000     -17.804     -13.642
total_bedrooms        78.3352      9.398      8.335      0.000      59.913      96.757
population           -38.4413      1.435    -26.795      0.000     -41.253     -35.629
households           135.1659     10.121     13.354      0.000     115.326     155.006
median_income       4.579e+04    440.435    103.959      0.000    4.49e+04    4.67e+04
==============================================================================
Omnibus:                     2621.034   Durbin-Watson:                   1.977
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             9116.848
Skew:                           1.018   Prob(JB):                         0.00
Kurtosis:                       6.608   Cond. No.                     3.00e+03
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[3] The condition number is large,  3e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```