In [1]:
```
# importing libraries
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
```

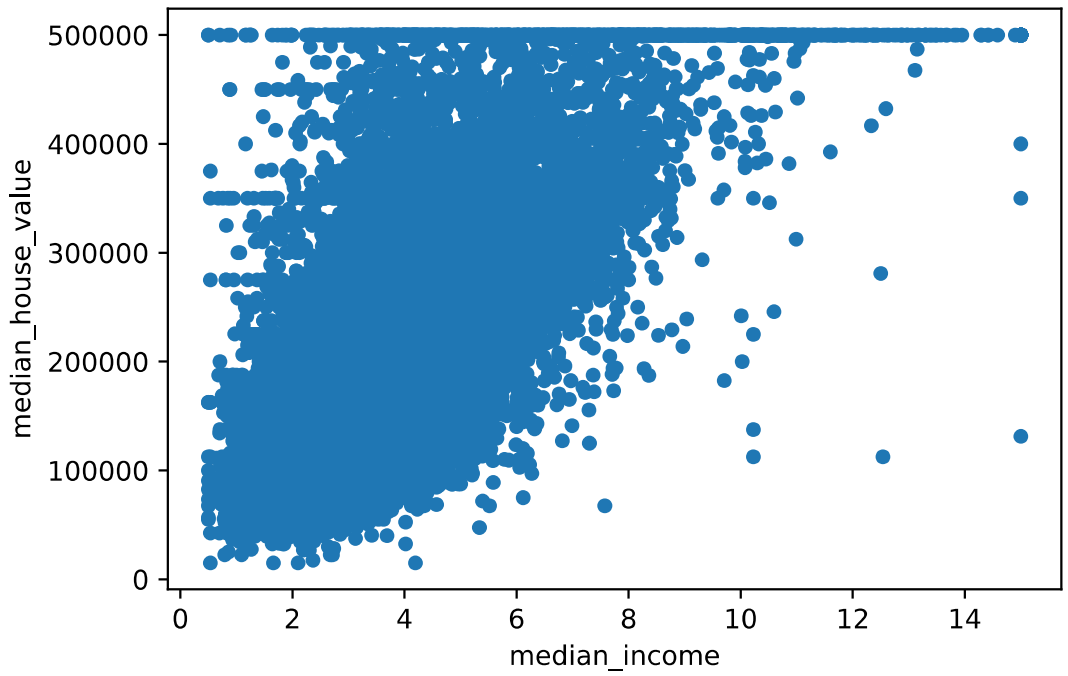
In [2]:
```
housing = pd.read_csv('housing.csv')
```

In [3]:
```
housing.shape
```

Out[3]:
```
(20640, 10)
```

In [4]:
```
housing.plot.scatter("median_income", "median_house_value")
```

Out[4]:
```
<AxesSubplot:xlabel='median_income', ylabel='median_house_value'>
```



In [5]:
```
x_train, x_test, y_train, y_test = train_test_split(housing.median_income, housing.medi
```

In [6]:
```
regr = LinearRegression()
regr.fit(np.array(x_train).reshape(-1,1), y_train)
preds = regr.predict(np.array(x_test).reshape(-1,1))
```

In [7]:
```
y_test.head()
```

Out[7]:
```
18294    500001.0
11748    233000.0
```

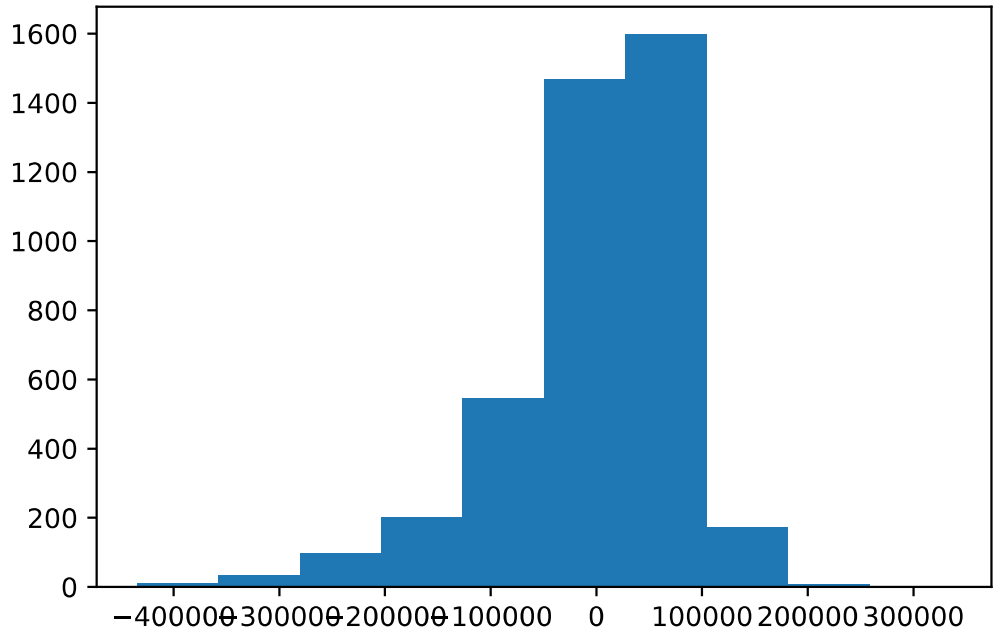
```
4953     117100.0
10043    153800.0
8267     382100.0
Name: median_house_value, dtype: float64
```

In [8]:
```
preds
```

Out[8]:
```
array([363775.25404594, 256941.06158928, 130212.24925173, ...,
       164636.55365647, 219809.1806042 , 289055.34702594])
```

In [9]:
```
residuals = preds - y_test
plt.hist(residuals)
```

Out[9]:
```
(array([   9.,   33.,   96.,  201.,  545., 1466., 1598.,  172.,    6.,
           2.]),
 array([-434310.25880956, -357356.53653961, -280402.81426965,
        -203449.09199969, -126495.36972973,  -49541.64745978,
          27412.07481018,  104365.79708014,  181319.5193501 ,
         258273.24162005,  335226.96389001]),
 <BarContainer object of 10 artists>)
```



In [10]:
```
mean_squared_error(y_test, preds) ** 0.5
```

Out[10]:
```
84657.4566938547
```