

lab2

January 20, 2025

```
[1]: from sklearn.datasets import fetch_california_housing
import pandas as pd
data = fetch_california_housing()
df = pd.DataFrame(data.data, columns=data.feature_names)
df['MedHouseVal'] = data.target
```

```
[3]: print(df.shape)
```

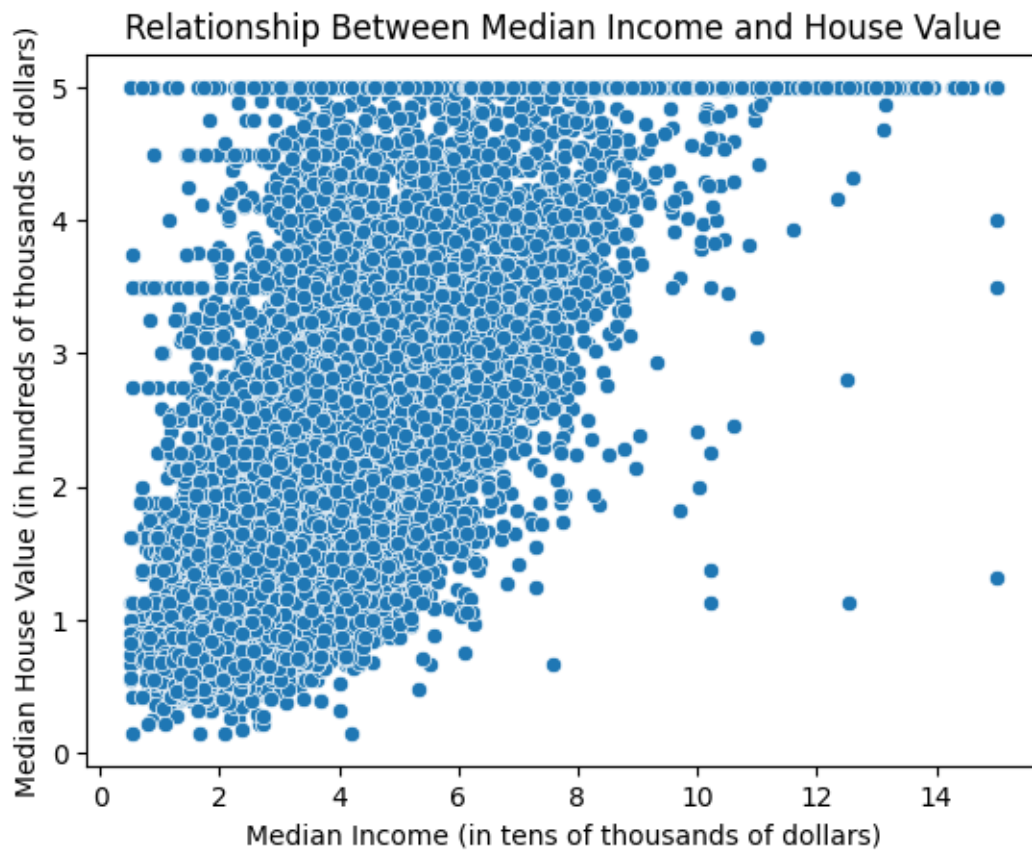
(20640, 9)

```
[5]: print(df.columns)
```

```
Index(['MedInc', 'HouseAge', 'AveRooms', 'AveBedrms', 'Population', 'AveOccup',
       'Latitude', 'Longitude', 'MedHouseVal'],
      dtype='object')
```

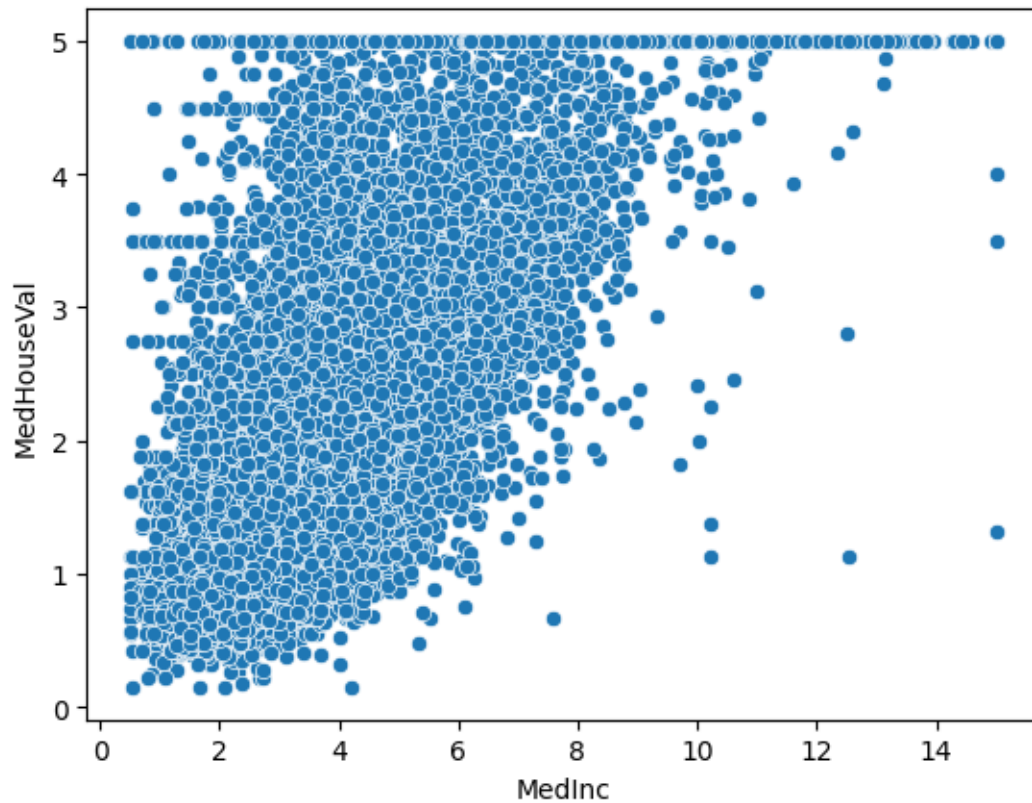
```
[10]: import seaborn as sns
import matplotlib.pyplot as plt

sns.scatterplot(x=df['MedInc'], y=df['MedHouseVal'])
plt.xlabel("Median Income (in tens of thousands of dollars)")
plt.ylabel("Median House Value (in hundreds of thousands of dollars)")
plt.title("Relationship Between Median Income and House Value")
plt.show()
```



```
[11]: sns.scatterplot(x=df['MedInc'], y=df['MedHouseVal'])
```

```
[11]: <Axes: xlabel='MedInc', ylabel='MedHouseVal'>
```



```
[17]: from sklearn.model_selection import train_test_split
      from sklearn.linear_model import LinearRegression
      from sklearn.metrics import mean_squared_error, r2_score
      import matplotlib.pyplot as plt
```

```
[18]: X = df[['MedInc']]
      y = df['MedHouseVal']

      X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
      ↪random_state=42)
```

```
[19]: X = df[['MedInc']]
      y = df['MedHouseVal']
```

```
[20]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
      ↪random_state=42)
```

```
[21]: model = LinearRegression()
      model.fit(X_train, y_train)
```

```
[21]: LinearRegression()
```

```
[22]: y_pred = model.predict(X_test)
```

```
[23]: mse = mean_squared_error(y_test, y_pred)

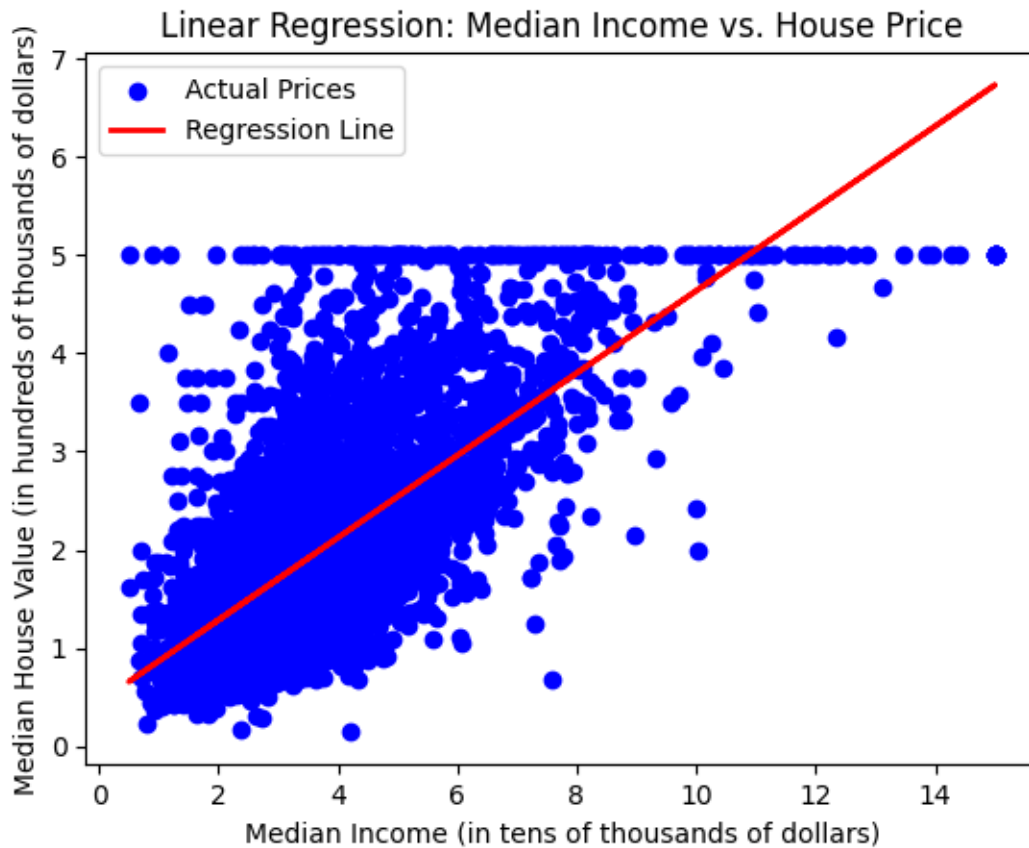
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse:.2f}")
print(f"R2 Score: {r2:.2f}")
```

Mean Squared Error: 0.71

R² Score: 0.46

```
[24]: plt.scatter(X_test, y_test, color='blue', label='Actual Prices')
plt.plot(X_test, y_pred, color='red', linewidth=2, label='Regression Line')
plt.xlabel("Median Income (in tens of thousands of dollars)")
plt.ylabel("Median House Value (in hundreds of thousands of dollars)")
plt.title("Linear Regression: Median Income vs. House Price")
plt.legend()
plt.show()
```



```
[25]: new_data = pd.DataFrame([[6]], columns=['MedInc'])

predicted_price = model.predict(new_data)

print(f"Predicted House Price for a Median Income of $60,000:␣
↪${predicted_price[0] * 100_000:.2f}")
```

Predicted House Price for a Median Income of \$60,000: \$296062.83