House Price Prediction

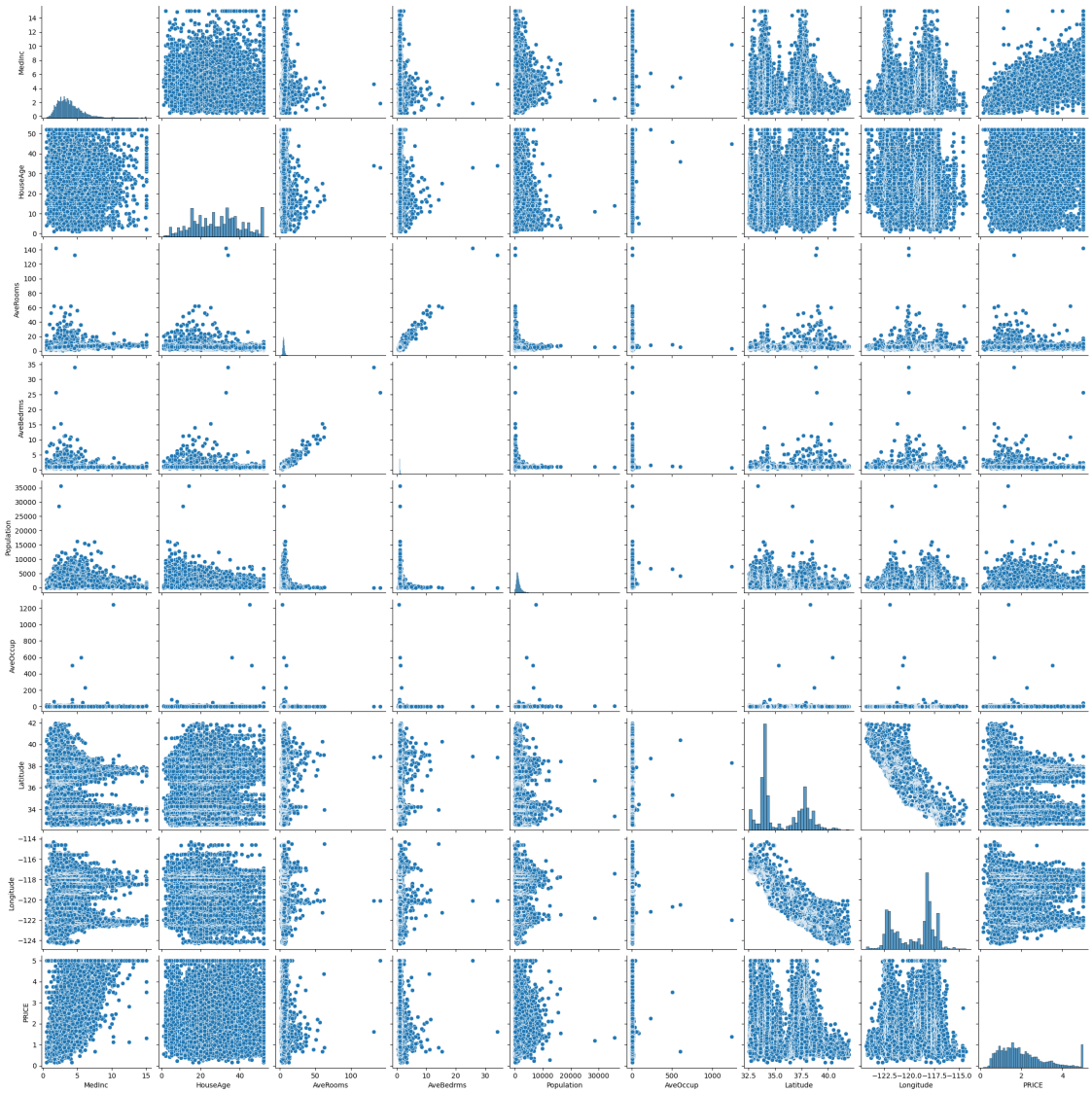
July 11, 2024

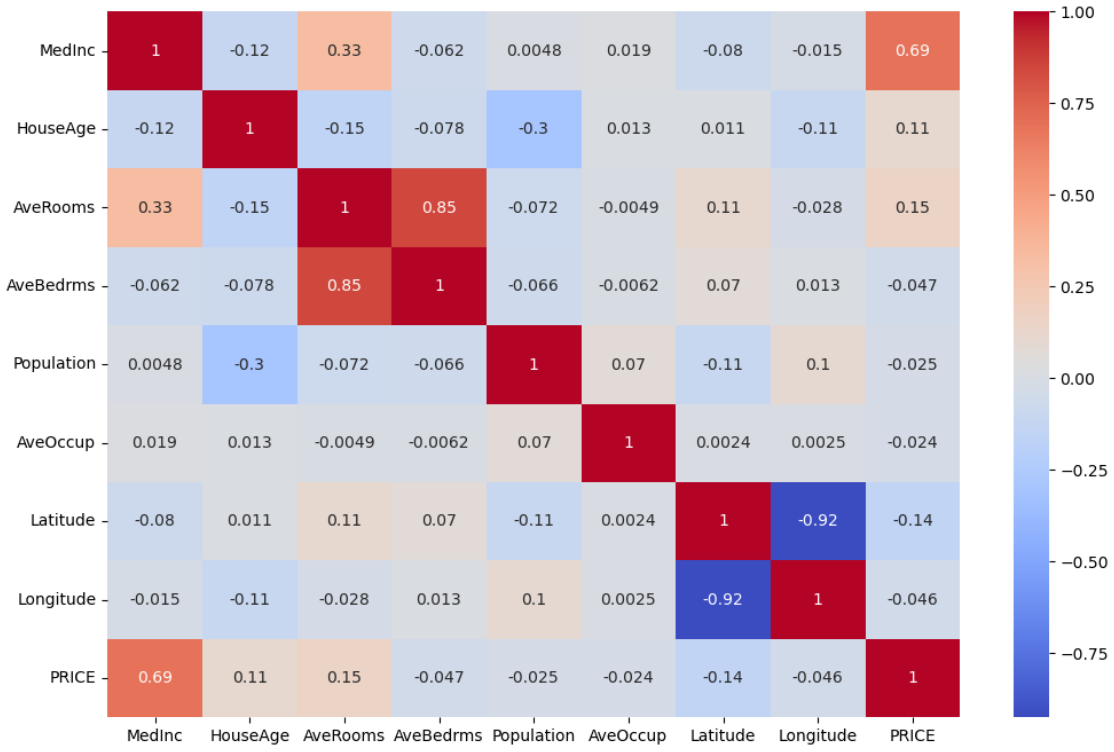
[5]:
```
# 1. Setting up the Environment
#pip install pandas scikit-learn matplotlib seaborn

# 2. Loading the Dataset
import pandas as pd
from sklearn.datasets import fetch_california_housing

# Load the dataset
california = fetch_california_housing()
df = pd.DataFrame(data=california.data, columns=california.feature_names)
df['PRICE'] = california.target

print(df.head())

# 3. Exploratory Data Analysis (EDA)
import seaborn as sns
import matplotlib.pyplot as plt

# Pairplot
sns.pairplot(df)
plt.show()

# Correlation heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
plt.show()

# 4. Preparing the Data
from sklearn.model_selection import train_test_split

X = df.drop('PRICE', axis=1)
y = df['PRICE']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,␣
 ↪random_state=42)

# 5. Training the Model
```

```python
from sklearn.linear_model import LinearRegression

# Initialize the model
model = LinearRegression()

# Train the model
model.fit(X_train, y_train)

# 6. Making Predictions

# Make predictions on the test set
y_pred = model.predict(X_test)


# 7. Evaluating the Model

from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

# MAE
mae = mean_absolute_error(y_test, y_pred)
print(f'Mean Absolute Error: {mae:.2f}')

# MSE
mse = mean_squared_error(y_test, y_pred)
print(f'Mean Squared Error: {mse:.2f}')

# R-squared
r2 = r2_score(y_test, y_pred)
print(f'R-squared: {r2:.2f}')

#8. Visualizing the Results

plt.figure(figsize=(10, 6))
plt.scatter(y_test, y_pred)
plt.plot([min(y_test), max(y_test)], [min(y_test), max(y_test)], color='red')
plt.xlabel('Actual Prices')
plt.ylabel('Predicted Prices')
plt.title('Actual vs Predicted Prices')
plt.show()
```

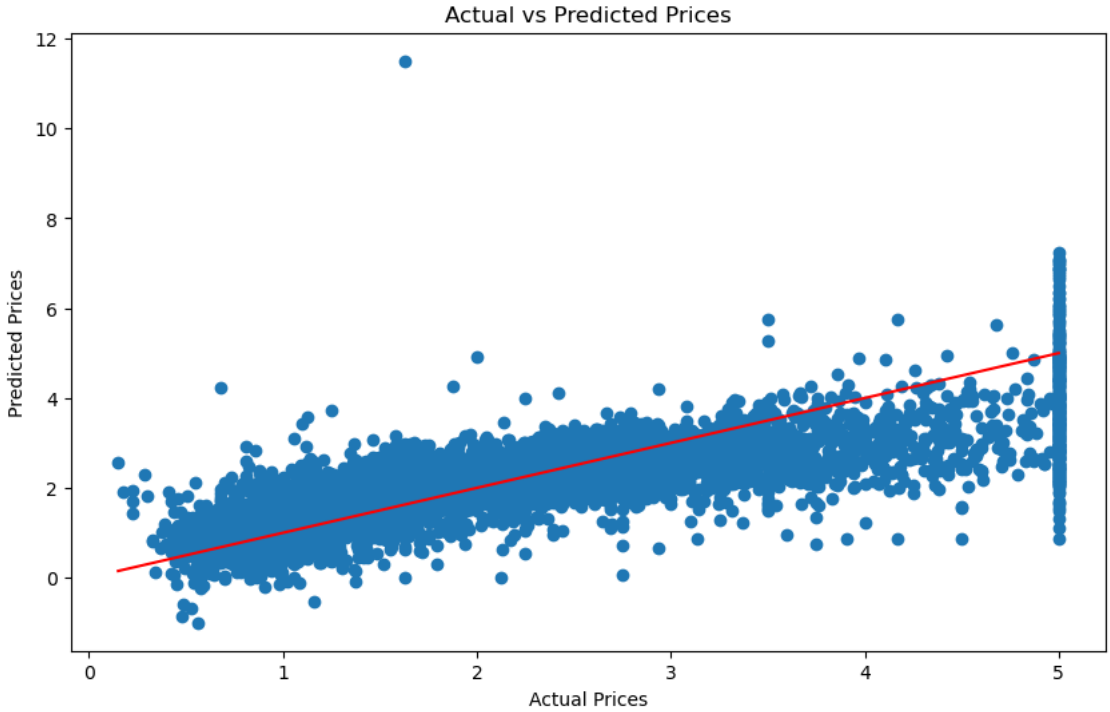
```
   MedInc  HouseAge  AveRooms  AveBedrms  Population  AveOccup  Latitude  \
0  8.3252      41.0  6.984127   1.023810       322.0  2.555556     37.88
1  8.3014      21.0  6.238137   0.971880      2401.0  2.109842     37.86
2  7.2574      52.0  8.288136   1.073446       496.0  2.802260     37.85
3  5.6431      52.0  5.817352   1.073059       558.0  2.547945     37.85
4  3.8462      52.0  6.281853   1.081081       565.0  2.181467     37.85
```

```
   Longitude  PRICE
0    -122.23  4.526
1    -122.22  3.585
2    -122.24  3.521
3    -122.25  3.413
4    -122.25  3.422
```






```
Mean Absolute Error: 0.53
Mean Squared Error: 0.53
R-squared: 0.60
```


Actual vs Predicted Prices
Predicted Prices
Actual Prices
0
1
2
3
4
5
0
2
4
6
8
10
12

[ ]: