```
In [3]: import pandas as pd
         import numpy as np
         from sklearn.model selection import train test split
         from sklearn.linear_model import LinearRegression
         from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error
         import matplotlib.pyplot as plt
         import seaborn as sns
 In [4]: data = pd.read_csv("C:/Users/Atharva/OneDrive/Desktop/House Rent Dataset.csv")
         print(data.head())
            Posted On BHK Rent Size Floor Area Type \
        0 2022-05-18 2 10000 1100 Ground out of 2 Super Area
        1 2022-05-13 2 20000 800 1 out of 3 Super Area
2 2022-05-16 2 17000 1000 1 out of 3 Super Area
                                                             Super Area
        3 2022-07-04 2 10000 800
                                              1 out of 2 Super Area
        4 2022-05-09 2 7500 850
                                              1 out of 2 Carpet Area
                       Area Locality
                                         City Furnishing Status Tenant Preferred \
                              Bandel Kolkata Unfurnished Bachelors/Family
        1 Phool Bagan, Kankurgachi Kolkata Semi-Furnished Bachelors/Family
        2
           Salt Lake City Sector 2 Kolkata Semi-Furnished Bachelors/Family
                       South Dum Dum Kolkata

South Dum Dum Kolkata
        3
        4
           Bathroom Point of Contact
        0
                  2
                       Contact Owner
        1
                  1
                       Contact Owner
        2
                  1
                      Contact Owner
        3
                  1
                       Contact Owner
        4
                  1
                       Contact Owner
 In [5]: # Pre-process dataset
         # Convert 'Posted On' to datetime, if necessary
         data['Posted On'] = pd.to_datetime(data['Posted On'])
 In [6]: # Extract numerical values from the 'Floor' column
         data['Floor Level'] = data['Floor'].str.extract(r'(\d+)').fillna(0).astype(int)
 In [7]: # One-hot encode categorical columns (e.g., 'Area Type', 'City', 'Furnishing Status')
         data = pd.get_dummies(data, columns=['Area Type', 'City', 'Furnishing Status', 'Tenant Preferred'], drop_first='
 In [8]: # Drop unnecessary columns
         data.drop(['Posted On', 'Floor', 'Area Locality', 'Point of Contact'], axis=1, inplace=True)
 In [9]: # Identify outliers based on 'Rent' (99th percentile)
         data = data[data['Rent'] < data['Rent'].quantile(0.99)]</pre>
In [10]: # Check correlation
         plt.figure(figsize=(10, 8))
         sns.heatmap(data.corr(), annot=True, fmt=".2f", cmap="coolwarm")
         plt.title("Correlation Matrix")
         plt.show()
```

```
Rent - 0.49 1.00 0.49 0.60 0.49 0.32 -0.32 -0.12 -0.02 -0.12 -0.17 0.50 0.05 -0.18 -0.12 0.12
                                                                                                                                                   - 0.75
                                         Size - 0.70 0.49 1.00 0.73 0.14 0.06 -0.06 0.06 -0.11 0.18 -0.09 -0.10 0.17 -0.20 -0.10 0.10
                                   Bathroom - 0.78 0.60 0.73 1.00 0.30 0.17 -0.17 0.03 -0.04 0.11 -0.21 0.15 0.14 -0.20 -0.12 0.14
                                                                                                                                                   - 0.50
                                  Floor Level - 0.19 0.49 0.14 0.30 1.00 0.24 -0.24 -0.13 -0.12 -0.08 -0.13 0.50 0.05 -0.09 -0.09 0.14
                      Area Type_Carpet Area - 0.14 0.32 0.06 0.17 0.24 1.00 -1.00 -0.12 0.03 -0.19 0.02 0.36 0.00 -0.03 -0.37 0.14
                                                                                                                                                   - 0.25
                       Area Type Super Area --0.14 -0.32 -0.06 -0.17 -0.24 -1.00 1.00 0.12 -0.03 0.19 -0.02 -0.36 0.00 0.03 0.37 -0.14
                                City Chennai - 0.04 -0.12 0.06 0.03 -0.13 -0.12 0.12 1.00 -0.19 -0.23 -0.17 -0.24 0.03 0.02 0.00 0.03
                                                                                                                                                   - 0.00
                                   City_Delhi - 0.02 -0.02 -0.11 -0.04 -0.12 0.03 -0.03 -0.19 1.00 -0.18 -0.14 -0.19 -0.01 -0.00 -0.01 -0.10
                             City Hyderabad - 0.08 -0.12 0.18 0.11 -0.08 -0.19 0.19 -0.23 -0.18 1.00 -0.17 -0.24 -0.01 0.02 0.06 0.01
                                                                                                                                                    -0.25
                                 City Kolkata --0.04 -0.17 -0.09 -0.21 -0.13 0.02 -0.02 -0.17 -0.14 -0.17 1.00 -0.18 -0.14 0.17 -0.00 -0.07
                                City Mumbai --0.02 0.50 -0.10 0.15 0.50 0.36 -0.36 -0.24 -0.19 -0.24 -0.18 1.00 -0.07 -0.03 -0.11 0.15
                                                                                                                                                   -0.50
           Furnishing Status Semi-Furnished - 0.14 0.05 0.17 0.14 0.05 0.00 0.00 0.03 -0.01 -0.01 -0.01 -0.14 -0.07 1.00 -0.75 0.04 0.04
               Furnishing Status_Unfurnished --0.16 -0.18 -0.20 -0.20 -0.09 -0.03 0.03 0.02 -0.00 0.02 0.17 -0.03 -0.75 1.00 -0.05 -0.05
                                                                                                                                                   - -0.75
          Tenant Preferred_Bachelors/Family --0.11 -0.12 -0.10 -0.12 -0.09 -0.37 0.37 0.00 -0.01 0.06 -0.00 -0.11 0.04 -0.05 1.00
                     Tenant Preferred_Family - 0.11 0.12 0.10 0.14 0.14 0.14 -0.14 0.03 -0.10 0.01 -0.07 0.15 0.04 -0.05 -0.54
                                                                                  Area Type_Super Area
                                                                                                                    -urnishing Status_Semi-Furnished
                                                                       Floor Level
                                                                            Area Type_Carpet Area
                                                                                       City_Chennai
                                                                                             City_Delhi
                                                                                                   City_Hyderabad
                                                                                                        City_Kolkata
                                                                                                                         Furnishing Status_Unfurnished
                                                                                                              City_Mumbai
                                                                                                                                     Tenant Preferred Family
                                                                 Bathroom
                                                                                                                               Tenant Preferred Bachelors/Family
In [11]: # Prepare data for training
            X = data.drop(['Rent'], axis=1)
            y = data['Rent']
            X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
In [12]: # Train linear regression model
            model = LinearRegression()
            model.fit(X_train, y_train)
Out[12]:
                 LinearRegression 🔍
            LinearRegression()
In [13]: # Evaluate model
            y pred = model.predict(X test)
            print("R2 Score:", r2_score(y_test, y_pred))
            print("RMSE:", np.sqrt(mean_squared_error(y_test, y_pred)))
            print("MAE:", mean_absolute_error(y_test, y_pred))
          R2 Score: 0.6593245415708426
          RMSE: 22942.337960853245
          MAE: 14322.210200018913
In [14]: # Plot actual vs. predicted values
            plt.figure(figsize=(10, 6))
```

sns.regplot(x=y\_test, y=y\_pred, line\_kws={'color': 'red'})
plt.xlabel("Actual Rent")

plt.ylabel("Predicted Rent")

plt.show()

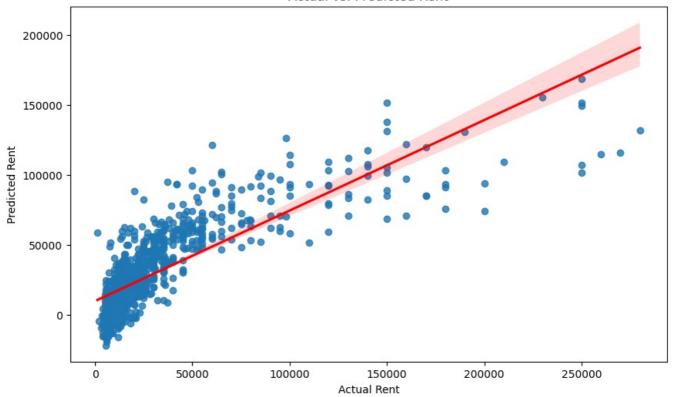
plt.title("Actual vs. Predicted Rent")

Correlation Matrix

BHK - 1.00 0.49 0.70 0.78 0.19 0.14 -0.14 0.04 0.02 0.08 -0.04 -0.02 0.14 -0.16 -0.11 0.11

1.00

## Actual vs. Predicted Rent



In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js