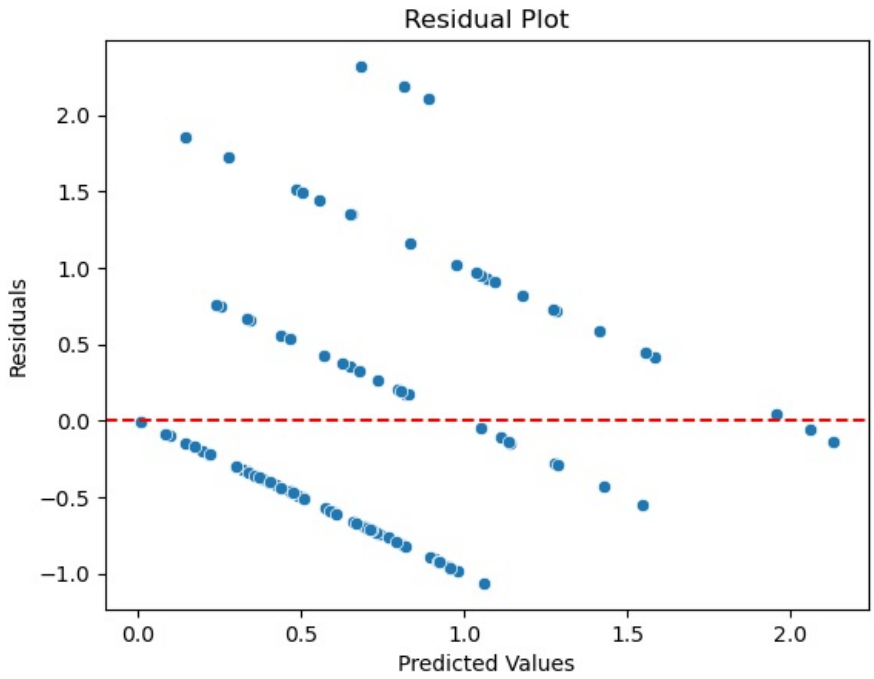
In [3]:

```
# ================================
# Task 3: Linear Regression (Auto-detect version)
# ================================

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

# ================================
# 1. Load Dataset
# ================================
# Replace with your dataset filename
df = pd.read_csv("Housing.csv")

print("Dataset Shape:", df.shape)
print("First 5 rows:\n", df.head())
print("\nAvailable columns:\n", df.columns)

# Drop missing values
df = df.dropna()

# ================================
# 2. Auto Feature & Target Selection
# ================================
# Keep only numeric columns
numeric_cols = df.select_dtypes(include=[np.number]).columns.tolist()
print("\nNumeric columns found:", numeric_cols)

# Assume last numeric column = target (Price)
target_col = numeric_cols[-1]
feature_cols = numeric_cols[:-1]

print("Selected Features:", feature_cols)
print("Selected Target:", target_col)

X = df[feature_cols]
y = df[target_col]

# ================================
# 3. Train-Test Split
# ================================
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

# ================================
# 4. Train Model
# ================================
model = LinearRegression()
model.fit(X_train, y_train)

# ================================
# 5. Predictions
# ================================
y_pred = model.predict(X_test)

# ================================
# 6. Evaluation
# ================================
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)

print("\nModel Performance:")
print("MAE:", mae)
print("MSE:", mse)
print("RMSE:", rmse)
print("R² Score:", r2)

# ================================
# 7. Coefficients
# ================================
print("\nIntercept:", model.intercept_)
coeff_df = pd.DataFrame({
    "Feature": feature_cols,
    "Coefficient": model.coef_
```

```python
})
print("\nFeature Coefficients:\n", coeff_df)

# ==============================
# 8. Plot Regression (if simple regression)
# ==============================
if len(feature_cols) == 1:
    plt.scatter(X_test, y_test, color="blue", label="Actual")
    plt.plot(X_test, y_pred, color="red", linewidth=2, label="Predicted")
    plt.xlabel(feature_cols[0])
    plt.ylabel(target_col)
    plt.title("Simple Linear Regression")
    plt.legend()
    plt.show()

# ==============================
# 9. Residual Plot (for checking assumptions)
# ==============================
residuals = y_test - y_pred
sns.scatterplot(x=y_pred, y=residuals)
plt.axhline(y=0, color="red", linestyle="--")
plt.xlabel("Predicted Values")
plt.ylabel("Residuals")
plt.title("Residual Plot")
plt.show()
```

```
Dataset Shape: (545, 13)
First 5 rows:
      price  area  bedrooms  bathrooms  stories mainroad guestroom basement  \
0  13300000  7420         4          2        3      yes        no       no
1  12250000  8960         4          4        4      yes        no       no
2  12250000  9960         3          2        2      yes        no      yes
3  12215000  7500         4          2        2      yes        no      yes
4  11410000  7420         4          1        2      yes       yes      yes

  hotwaterheating airconditioning  parking prefarea furnishingstatus
0              no             yes        2      yes        furnished
1              no             yes        3       no        furnished
2              no              no        2      yes   semi-furnished
3              no             yes        3      yes        furnished
4              no             yes        2       no        furnished

Available columns:
 Index(['price', 'area', 'bedrooms', 'bathrooms', 'stories', 'mainroad',
       'guestroom', 'basement', 'hotwaterheating', 'airconditioning',
       'parking', 'prefarea', 'furnishingstatus'],
      dtype='object')

Numeric columns found: ['price', 'area', 'bedrooms', 'bathrooms', 'stories', 'parking']
Selected Features: ['price', 'area', 'bedrooms', 'bathrooms', 'stories']
Selected Target: parking

Model Performance:
MAE: 0.6334119253074165
MSE: 0.6078468389300876
RMSE: 0.7796453289349508
R² Score: 0.2272813724236712

Intercept: -0.3348998368179409

Feature Coefficients:
      Feature   Coefficient
0      price  1.500301e-07
1       area  6.652296e-05
2   bedrooms  4.320597e-02
3  bathrooms  2.571284e-02
4    stories -1.058846e-01
```


Residual Plot
Residuals
Predicted Values
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
0.0
0.5
1.0
1.5
2.0

In [ ]: