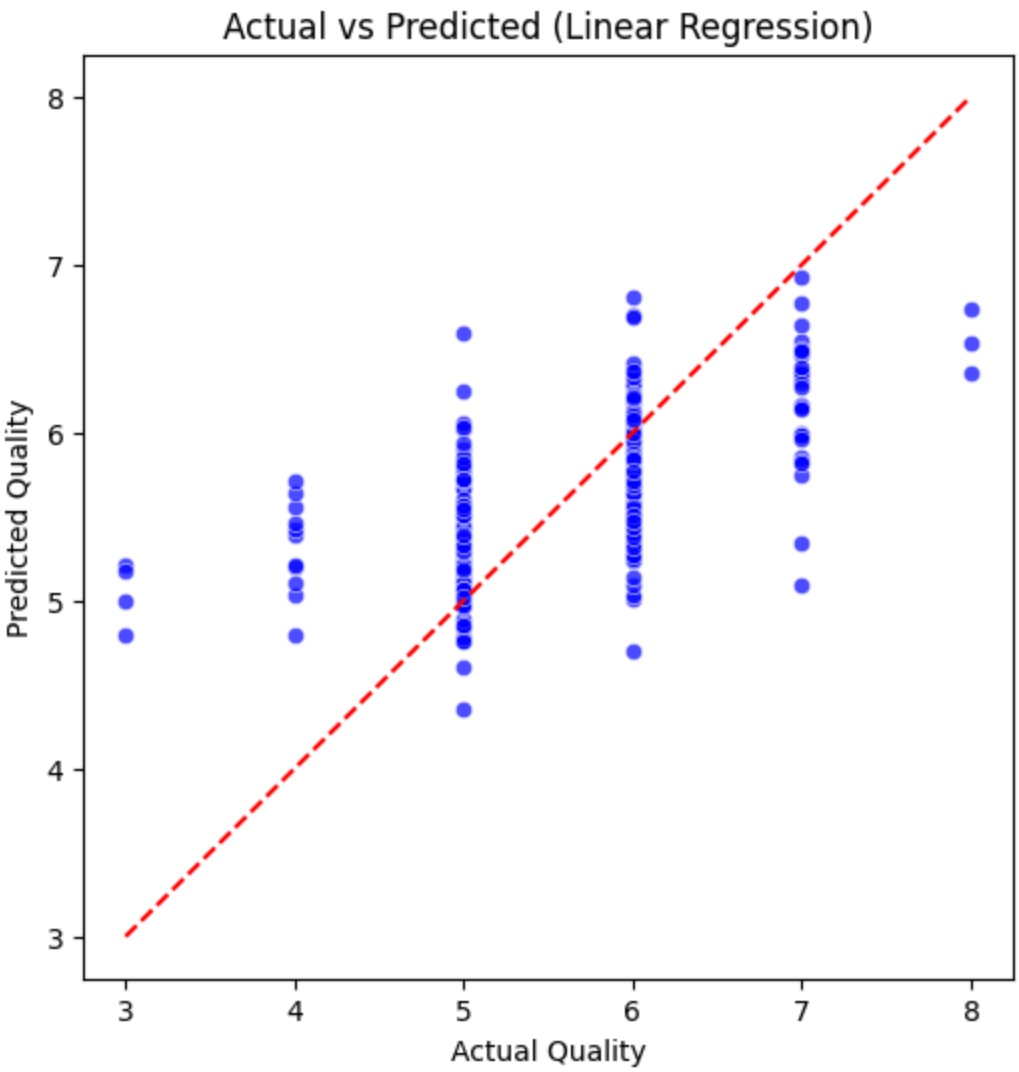
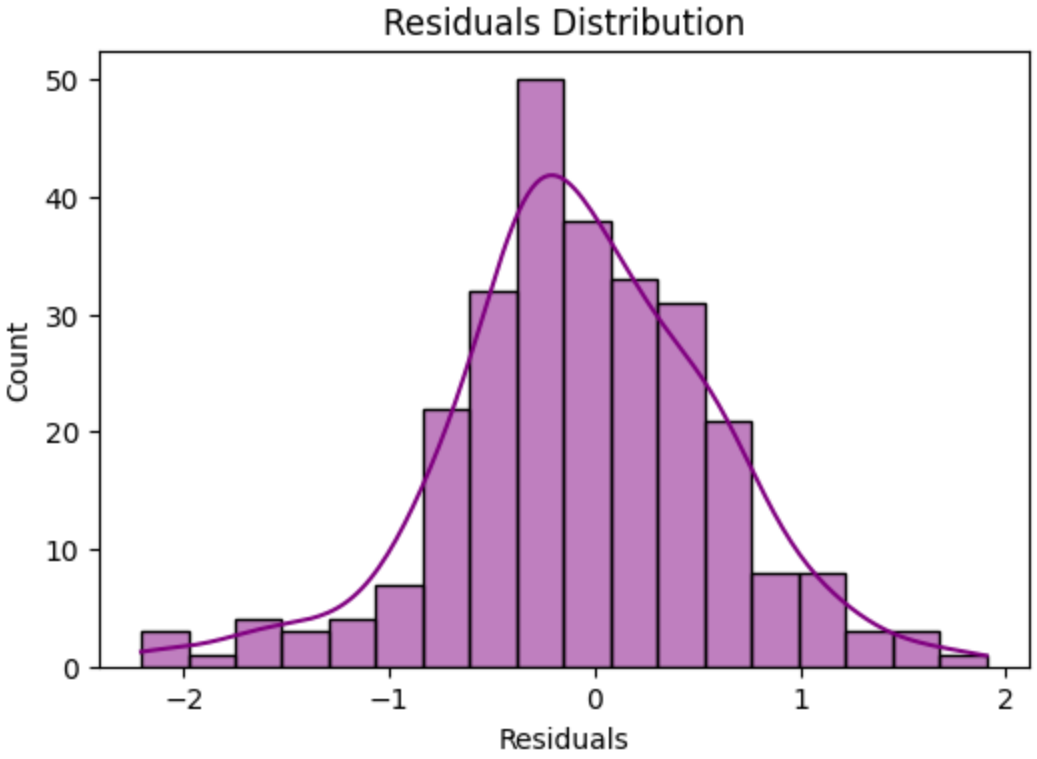
In [30]:

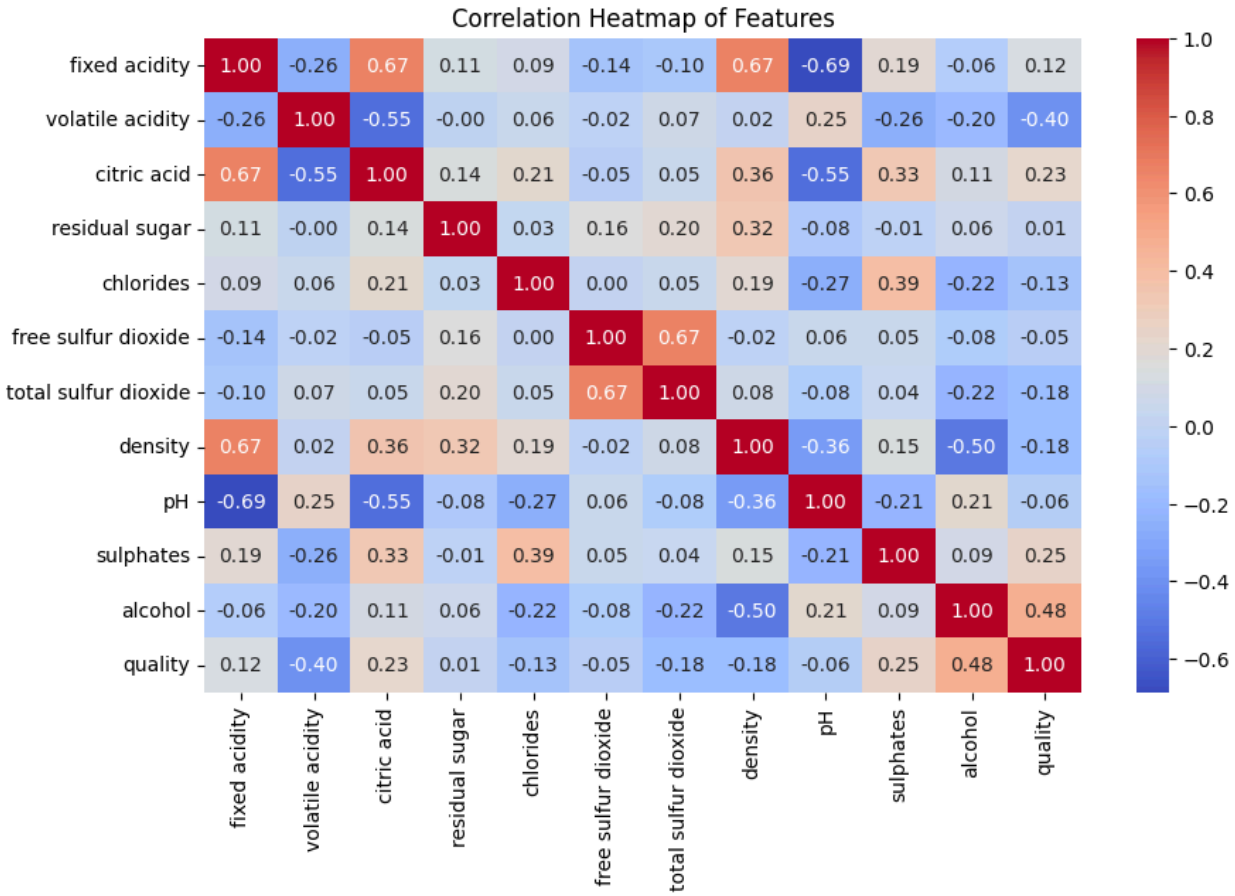
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

df = pd.read_csv("winedata.csv", sep=';')

print(" Correctly loaded shape:", df.shape)
print("Columns:", df.columns.tolist())
print(df.head())

df = df.drop_duplicates()
df = df.dropna(how='any')

if 'quality' in df.columns:
    X = df.drop('quality', axis=1)
    y = df['quality']
else:
    X = df.iloc[:, :-1]
    y = df.iloc[:, -1]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, randc

model = LinearRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

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("\n📊 Evaluation Metrics:")
print(f"MAE:  {mae:.4f}")
print(f"MSE:  {mse:.4f}")
print(f"RMSE: {rmse:.4f}")
print(f"R²:   {r2:.4f}")

plt.figure(figsize=(6,6))
sns.scatterplot(x=y_test, y=y_pred, alpha=0.7, color='blue')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], color='re
plt.xlabel("Actual Quality")
plt.ylabel("Predicted Quality")
plt.title("Actual vs Predicted (Linear Regression)")
plt.show()

residuals = y_test - y_pred
plt.figure(figsize=(6,4))
sns.histplot(residuals, kde=True, color='purple')
```

```
plt.title("Residuals Distribution")
plt.xlabel("Residuals")
plt.show()

plt.figure(figsize=(10,6))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Heatmap of Features")
plt.show()
```

```
 Correctly loaded shape: (1599, 12)
Columns: ['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density', 'pH', 's
ulphates', 'alcohol', 'quality']
   fixed acidity  volatile acidity  citric acid  residual sugar  chlorides  \
0            7.4              0.70         0.00             1.9      0.076
1            7.8              0.88         0.00             2.6      0.098
2            7.8              0.76         0.04             2.3      0.092
3           11.2              0.28         0.56             1.9      0.075
4            7.4              0.70         0.00             1.9      0.076

   free sulfur dioxide  total sulfur dioxide  density    pH  sulphates  \
0                 11.0                  34.0   0.9978  3.51       0.56
1                 25.0                  67.0   0.9968  3.20       0.68
2                 15.0                  54.0   0.9970  3.26       0.65
3                 17.0                  60.0   0.9980  3.16       0.58
4                 11.0                  34.0   0.9978  3.51       0.56

   alcohol  quality
0      9.4        5
1      9.8        5
2      9.8        5
3      9.8        6
4      9.4        5

📊 Evaluation Metrics:
MAE:  0.5041
MSE:  0.4310
RMSE: 0.6565
R²:   0.3915
```


Actual vs Predicted (Linear Regression)
Predicted Quality
Actual Quality
3
4
5
6
7
8


Residuals Distribution
Count
Residuals
50
40
30
20
10
0
−2
−1
0
1
2


Correlation Heatmap of Features
fixed acidity 1.00 -0.26 0.67 0.11 0.09 -0.14 -0.10 0.67 -0.69 0.19 -0.06 0.12
volatile acidity -0.26 1.00 -0.55 -0.00 0.06 -0.02 0.07 0.02 0.25 -0.26 -0.20 -0.40
citric acid 0.67 -0.55 1.00 0.14 0.21 -0.05 0.05 0.36 -0.55 0.33 0.11 0.23
residual sugar 0.11 -0.00 0.14 1.00 0.03 0.16 0.20 0.32 -0.08 -0.01 0.06 0.01
chlorides 0.09 0.06 0.21 0.03 1.00 0.00 0.05 0.19 -0.27 0.39 -0.22 -0.13
free sulfur dioxide -0.14 -0.02 -0.05 0.16 0.00 1.00 0.67 -0.02 0.06 0.05 -0.08 -0.05
total sulfur dioxide -0.10 0.07 0.05 0.20 0.05 0.67 1.00 0.08 -0.08 0.04 -0.22 -0.18
density 0.67 0.02 0.36 0.32 0.19 -0.02 0.08 1.00 -0.36 0.15 -0.50 -0.18
pH -0.69 0.25 -0.55 -0.08 -0.27 0.06 -0.08 -0.36 1.00 -0.21 0.21 -0.06
sulphates 0.19 -0.26 0.33 -0.01 0.39 0.05 0.04 0.15 -0.21 1.00 0.09 0.25
alcohol -0.06 -0.20 0.11 0.06 -0.22 -0.08 -0.22 -0.50 0.21 0.09 1.00 0.48
quality 0.12 -0.40 0.23 0.01 -0.13 -0.05 -0.18 -0.18 -0.06 0.25 0.48 1.00
fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
free sulfur dioxide
total sulfur dioxide
density
pH
sulphates
alcohol
quality
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6