mids4b

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import LabelEncoder, StandardScaler

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, r2\_score

df = pd.read\_csv("abalone.csv")

df['Sex'] = LabelEncoder().fit\_transform(df['Sex'])

df['Age'] = df['Rings'] + 1.5

X = df.drop(['Rings', 'Age'], axis=1)

y = df['Age']

scaler = StandardScaler()

X\_scaled = scaler.fit\_transform(X)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X\_scaled, y, test\_size=0.2, random\_state=42)

model = LinearRegression()

model.fit(X\_train, y\_train)

y\_pred = model.predict(X\_test)

plt.figure(figsize=(8, 6))

plt.scatter(y\_test, y\_pred, alpha=0.6, color='blue', label='Predicted vs Actual')

plt.plot([y\_test.min(), y\_test.max()], [y\_test.min(), y\_test.max()], 'r--', label='Ideal Fit')

plt.xlabel('Actual Age')

plt.ylabel('Predicted Age')

plt.title('Linear Regression: Actual vs Predicted Age of Abalone')

plt.legend()

plt.grid(True)

plt.tight\_layout()

plt.show()

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}")