# =========================

# Task 1: Student Performance Analysis

# =========================

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# -------------------------

# Step 1: Create Dataset (sample from student performance data)

# -------------------------

data = {

"sex": ["F","M","F","M","F","M","F","M","F","M"],

"studytime": [2,1,3,2,2,4,1,3,2,2],

"G3": [12,10,15,8,18,16,9,14,11,13]

}

df = pd.DataFrame(data)

print("✅ Sample dataset created")

print(df)

# -------------------------

# Step 2: Explore & Clean Data

# -------------------------

print("\nMissing values per column:\n", df.isnull().sum())

df = df.drop\_duplicates()

print("\nShape of dataset:", df.shape)

print("\nData types:\n", df.dtypes)

# -------------------------

# Step 3: Analysis Questions

# -------------------------

# (a) Average final grade

avg\_grade = df['G3'].mean()

print("\nAverage final grade (G3):", round(avg\_grade, 2))

# (b) Number of students scored above 15

above\_15 = (df['G3'] > 15).sum()

print("Number of students scoring above 15:", above\_15)

# (c) Correlation: Study Time vs Performance

correlation = df['studytime'].corr(df['G3'])

print("Correlation between study time and performance (G3):", round(correlation, 2))

# (d) Gender performance comparison

gender\_avg = df.groupby('sex')['G3'].mean()

print("\nAverage grades by gender:\n", gender\_avg)

# -------------------------

# Step 4: Visualizations

# -------------------------

plt.figure(figsize=(6,4))

plt.hist(df['G3'], bins=5, edgecolor="black")

plt.xlabel("Final Grade (G3)")

plt.ylabel("Frequency")

plt.title("Distribution of Final Grades")

plt.show()

plt.figure(figsize=(6,4))

sns.scatterplot(x="studytime", y="G3", data=df)

plt.title("Study Time vs Final Grade")

plt.xlabel("Study Time")

plt.ylabel("Final Grade (G3)")

plt.show()

plt.figure(figsize=(6,4))

gender\_avg.plot(kind='bar', color=["skyblue", "pink"])

plt.xlabel("Gender")

plt.ylabel("Average Final Grade")

plt.title("Average Grade by Gender")

plt.show()

# -------------------------

# Step 5: Conclusion

# -------------------------

print("\n📌 Conclusion:")

print(f"- The average student score is about {round(avg\_grade,2)}")

print(f"- {above\_15} students scored above 15")

print(f"- Study time correlation with performance is {round(correlation,2)}")

if gender\_avg['F'] > gender\_avg['M']:

print("- Female students perform slightly better on average")

else:

print("- Male students perform slightly better on average")

✅ Sample dataset created

sex studytime G3

0 F 2 12

1 M 1 10

2 F 3 15

3 M 2 8

4 F 2 18

5 M 4 16

6 F 1 9

7 M 3 14

8 F 2 11

9 M 2 13

Missing values per column:

sex 0

studytime 0

G3 0

dtype: int64

Shape of dataset: (10, 3)

Data types:

sex object

studytime int64

G3 int64

dtype: object

Average final grade (G3): 12.6

Number of students scoring above 15: 2

Correlation between study time and performance (G3): 0.63

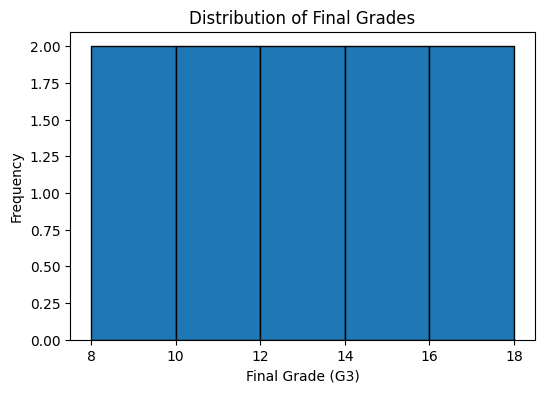
Average grades by gender:

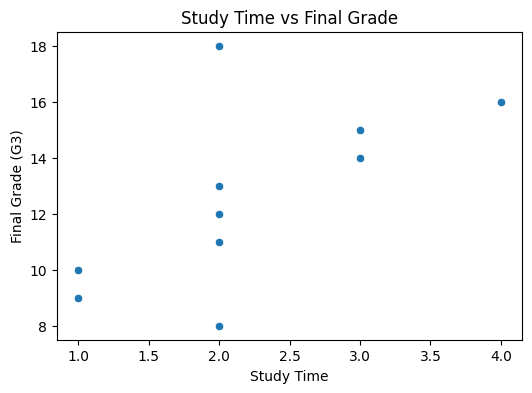
sex

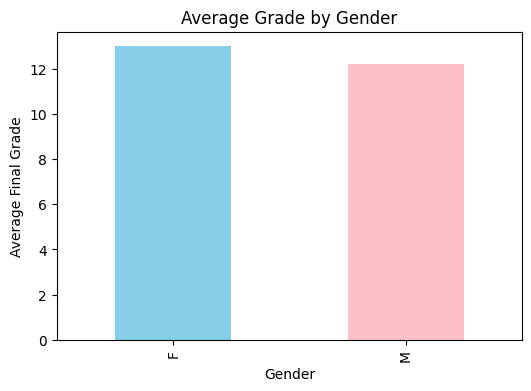
F 13.0

M 12.2

Name: G3, dtype: float64







📌 Conclusion:

- The average student score is about 12.6

- 2 students scored above 15

- Study time correlation with performance is 0.63

- Female students perform slightly better on average