Student-Performance-Insights

April 28, 2025

[5]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
sns.set(style="whitegrid")

# Load the dataset
df = pd.read_csv("C://Users/Parvej//Downloads//student//student-mat.csv", sep=';
 ↪')

print("Dataset Info:")
print(df.info())

print("\nFirst 5 rows:")
print(df.head())

print("\nSummary Statistics:")
print(df.describe())

print("\nMissing Values:")
print(df.isnull().sum())

print("\nSchool distribution:")
print(df['school'].value_counts())

print("\nGender distribution:")
print(df['sex'].value_counts())

sns.countplot(x='sex', data=df)
plt.title('Gender Distribution')
plt.show()

sns.histplot(df['G3'], kde=True, bins=20)
plt.title('Distribution of Final Grades (G3)')
plt.xlabel('Final Grade')
```

```
plt.show()

sns.boxplot(x='sex', y='G3', data=df)
plt.title('Final Grade Distribution by Gender')
plt.show()

sns.scatterplot(x='studytime', y='G3', hue='sex', data=df)
plt.title('Study Time vs Final Grades')
plt.xlabel('Study Time (1: <2h, 2: 2-5h, 3: 5-10h, 4: >10h)')
plt.ylabel('Final Grade (G3)')
plt.show()

# Correlation Matrix (numeric columns only)
numeric_df = df.select_dtypes(include=[np.number])
corr = numeric_df.corr()

plt.figure(figsize=(14,10))
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()

selected_features = ['age', 'studytime', 'failures', 'absences', 'G1', 'G2',␣
 ↪'G3']
sns.pairplot(df[selected_features])
plt.suptitle('Pairplot of Selected Features', y=1.02)
plt.show()

sns.boxplot(x='school', y='G3', data=df)
plt.title('Final Grades by School')
plt.show()

sns.barplot(x='higher', y='G3', data=df)
plt.title('Aspiration for Higher Education vs Final Grades')
plt.show()

print("\nObservations:")
print("- Females slightly outperform males in final grades (G3).")
print("- More study time is associated with higher grades.")
print("- Students aspiring for higher education tend to have better final␣
 ↪grades.")
print("- G1 and G2 (previous grades) strongly predict G3 (final grade).")
print("- No missing data found.")

print("\nSummary of Findings:")
print("""
- Dataset contains 395 students with 33 features.
- Most students study between 2-5 hours per week.
```

```
- Previous period grades (G1 and G2) are highly correlated with final grade G3.
- Studytime, failures, and absences impact student performance.
- Female students and students aiming for higher education perform slightly␣
 ↪better.
""")
```

```
Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 395 entries, 0 to 394
Data columns (total 33 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   school      395 non-null    object
 1   sex         395 non-null    object
 2   age         395 non-null    int64
 3   address     395 non-null    object
 4   famsize     395 non-null    object
 5   Pstatus     395 non-null    object
 6   Medu        395 non-null    int64
 7   Fedu        395 non-null    int64
 8   Mjob        395 non-null    object
 9   Fjob        395 non-null    object
 10  reason      395 non-null    object
 11  guardian    395 non-null    object
 12  traveltime  395 non-null    int64
 13  studytime   395 non-null    int64
 14  failures    395 non-null    int64
 15  schoolsup   395 non-null    object
 16  famsup      395 non-null    object
 17  paid        395 non-null    object
 18  activities  395 non-null    object
 19  nursery     395 non-null    object
 20  higher      395 non-null    object
 21  internet    395 non-null    object
 22  romantic    395 non-null    object
 23  famrel      395 non-null    int64
 24  freetime    395 non-null    int64
 25  goout       395 non-null    int64
 26  Dalc        395 non-null    int64
 27  Walc        395 non-null    int64
 28  health      395 non-null    int64
 29  absences    395 non-null    int64
 30  G1          395 non-null    int64
 31  G2          395 non-null    int64
 32  G3          395 non-null    int64
dtypes: int64(16), object(17)
memory usage: 102.0+ KB
None
```

First 5 rows:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	…	\
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	…	
1	GP	F	17	U	GT3	T	1	1	at_home	other	…	
2	GP	F	15	U	LE3	T	1	1	at_home	other	…	
3	GP	F	15	U	GT3	T	4	2	health	services	…	
4	GP	F	16	U	GT3	T	3	3	other	other	…	

	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
0	4	3	4	1	1	3	6	5	6	6
1	5	3	3	1	1	3	4	5	5	6
2	4	3	2	2	3	3	10	7	8	10
3	3	2	2	1	1	5	2	15	14	15
4	4	3	2	1	2	5	4	6	10	10

[5 rows x 33 columns]

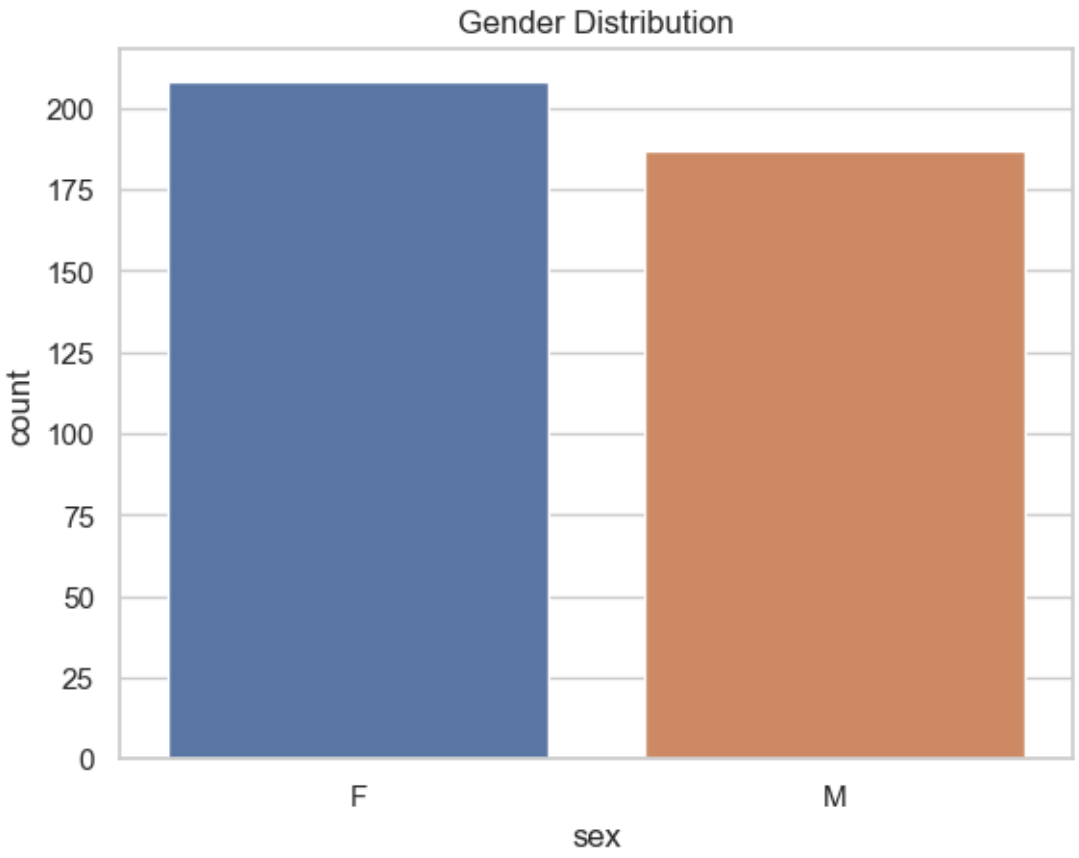
Summary Statistics:

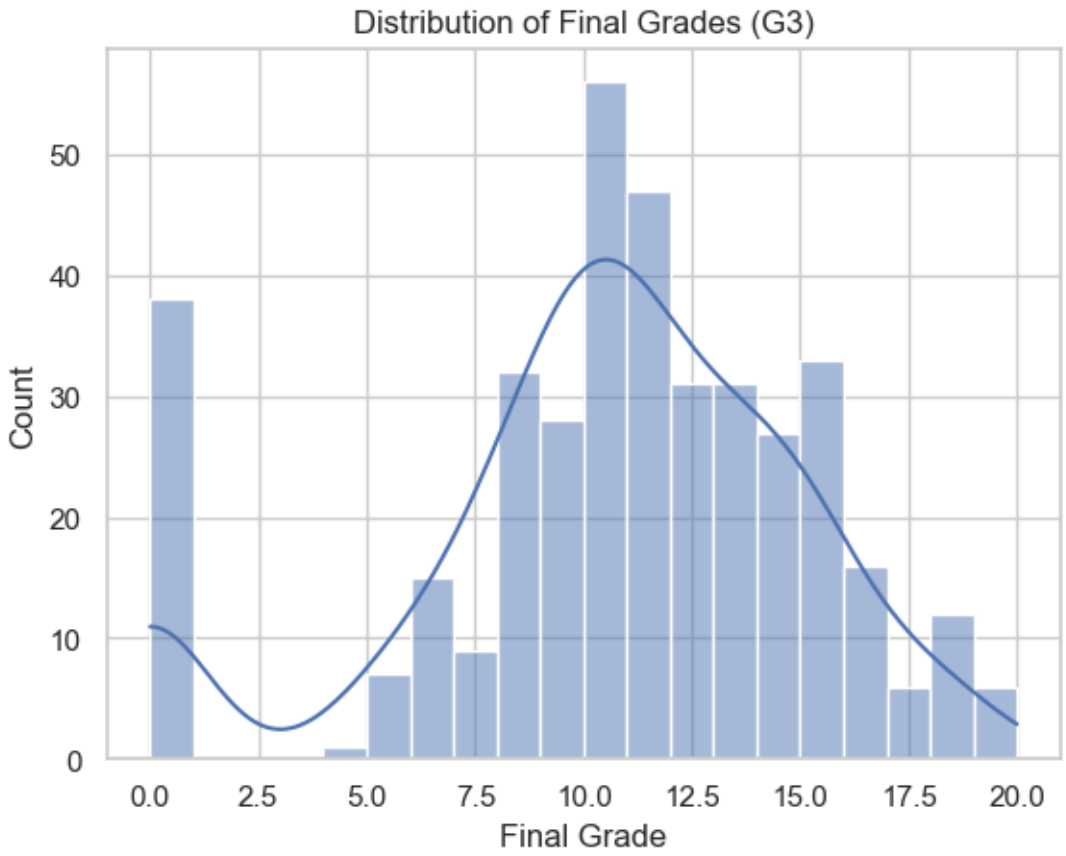
	age	Medu	Fedu	traveltime	studytime	failures	\
count	395.000000	395.000000	395.000000	395.000000	395.000000	395.000000	
mean	16.696203	2.749367	2.521519	1.448101	2.035443	0.334177	
std	1.276043	1.094735	1.088201	0.697505	0.839240	0.743651	
min	15.000000	0.000000	0.000000	1.000000	1.000000	0.000000	
25%	16.000000	2.000000	2.000000	1.000000	1.000000	0.000000	
50%	17.000000	3.000000	2.000000	1.000000	2.000000	0.000000	
75%	18.000000	4.000000	3.000000	2.000000	2.000000	0.000000	
max	22.000000	4.000000	4.000000	4.000000	4.000000	3.000000	

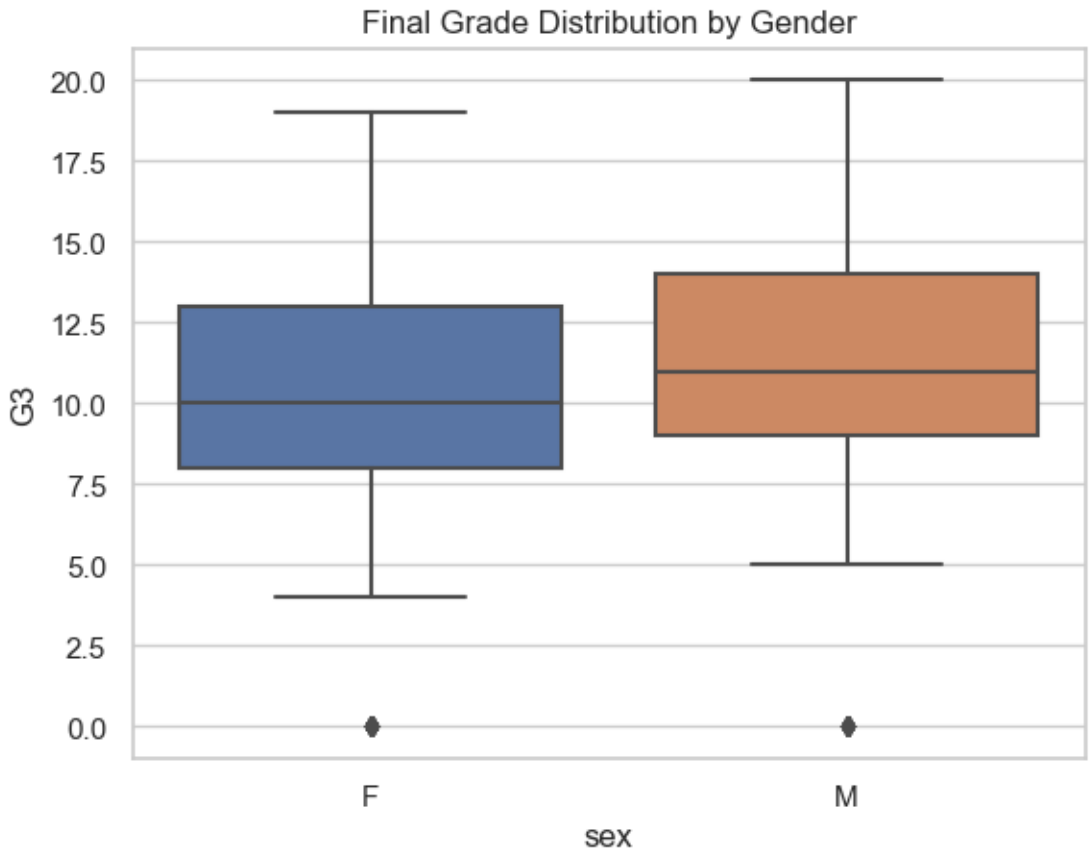
	famrel	freetime	goout	Dalc	Walc	health	\
count	395.000000	395.000000	395.000000	395.000000	395.000000	395.000000	
mean	3.944304	3.235443	3.108861	1.481013	2.291139	3.554430	
std	0.896659	0.998862	1.113278	0.890741	1.287897	1.390303	
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	
25%	4.000000	3.000000	2.000000	1.000000	1.000000	3.000000	
50%	4.000000	3.000000	3.000000	1.000000	2.000000	4.000000	
75%	5.000000	4.000000	4.000000	2.000000	3.000000	5.000000	
max	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	

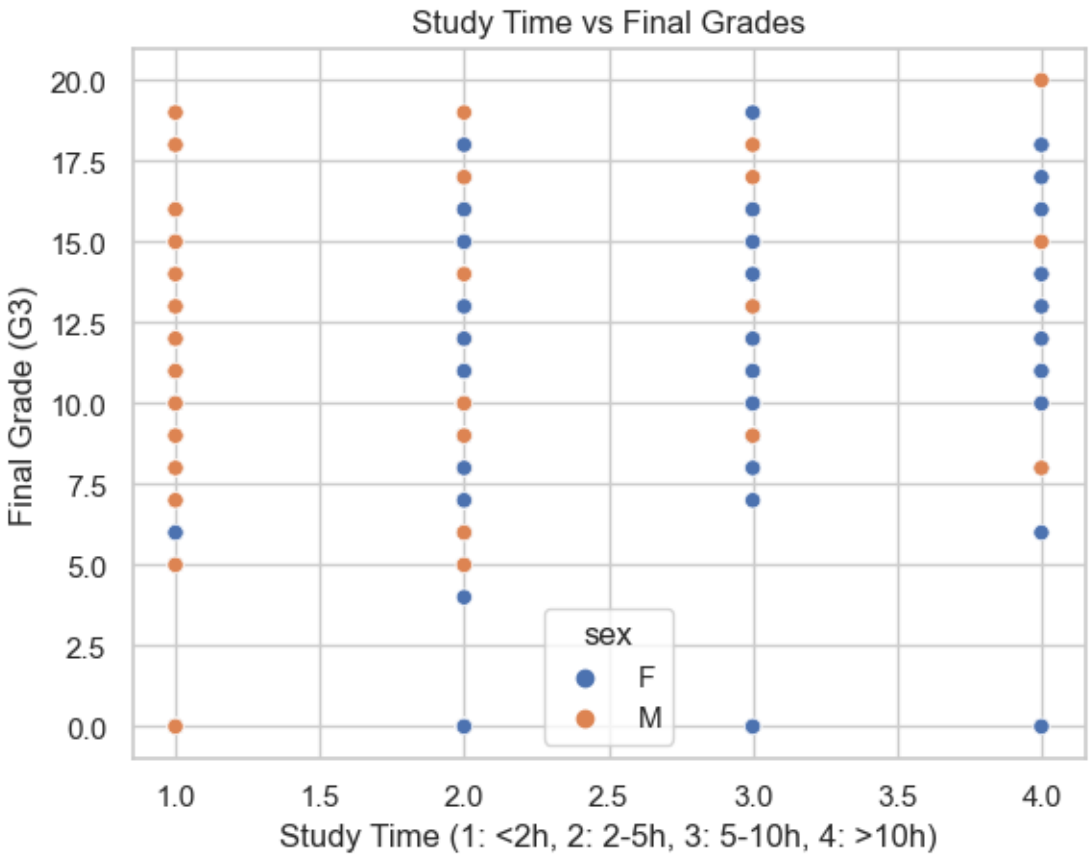
	absences	G1	G2	G3
count	395.000000	395.000000	395.000000	395.000000
mean	5.708861	10.908861	10.713924	10.415190
std	8.003096	3.319195	3.761505	4.581443
min	0.000000	3.000000	0.000000	0.000000
25%	0.000000	8.000000	9.000000	8.000000
50%	4.000000	11.000000	11.000000	11.000000
75%	8.000000	13.000000	13.000000	14.000000
max	75.000000	19.000000	19.000000	20.000000

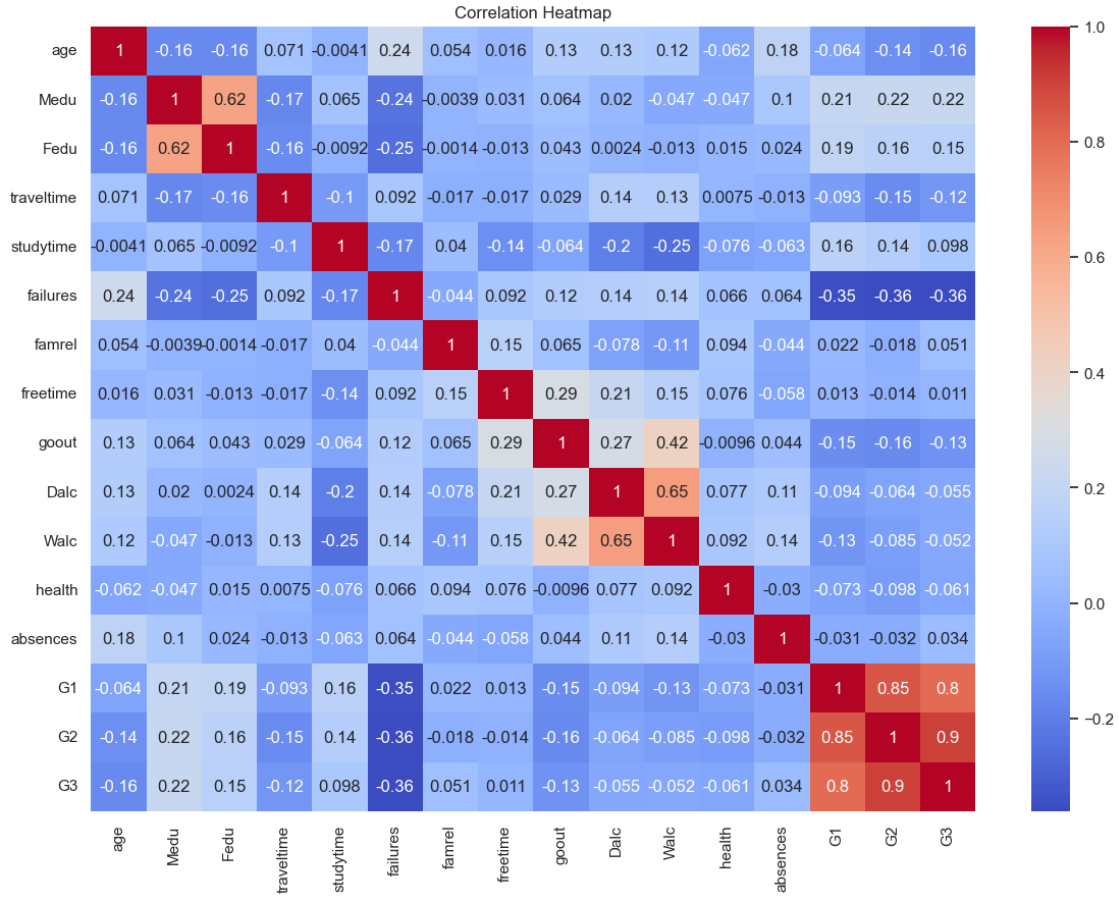
```
Missing Values:
school        0
sex           0
age           0
address       0
famsize       0
Pstatus       0
Medu          0
Fedu          0
Mjob          0
Fjob          0
reason        0
guardian      0
traveltime    0
studytime     0
failures      0
schoolsup     0
famsup        0
paid          0
activities    0
nursery       0
higher        0
internet      0
romantic      0
famrel        0
freetime      0
goout         0
Dalc          0
Walc          0
health        0
absences      0
G1            0
G2            0
G3            0
dtype: int64

School distribution:
school
GP    349
MS     46
Name: count, dtype: int64

Gender distribution:
sex
F    208
M    187
Name: count, dtype: int64
```


Gender Distribution
200
175
150
125
100
75
50
25
0
count
F
M
sex

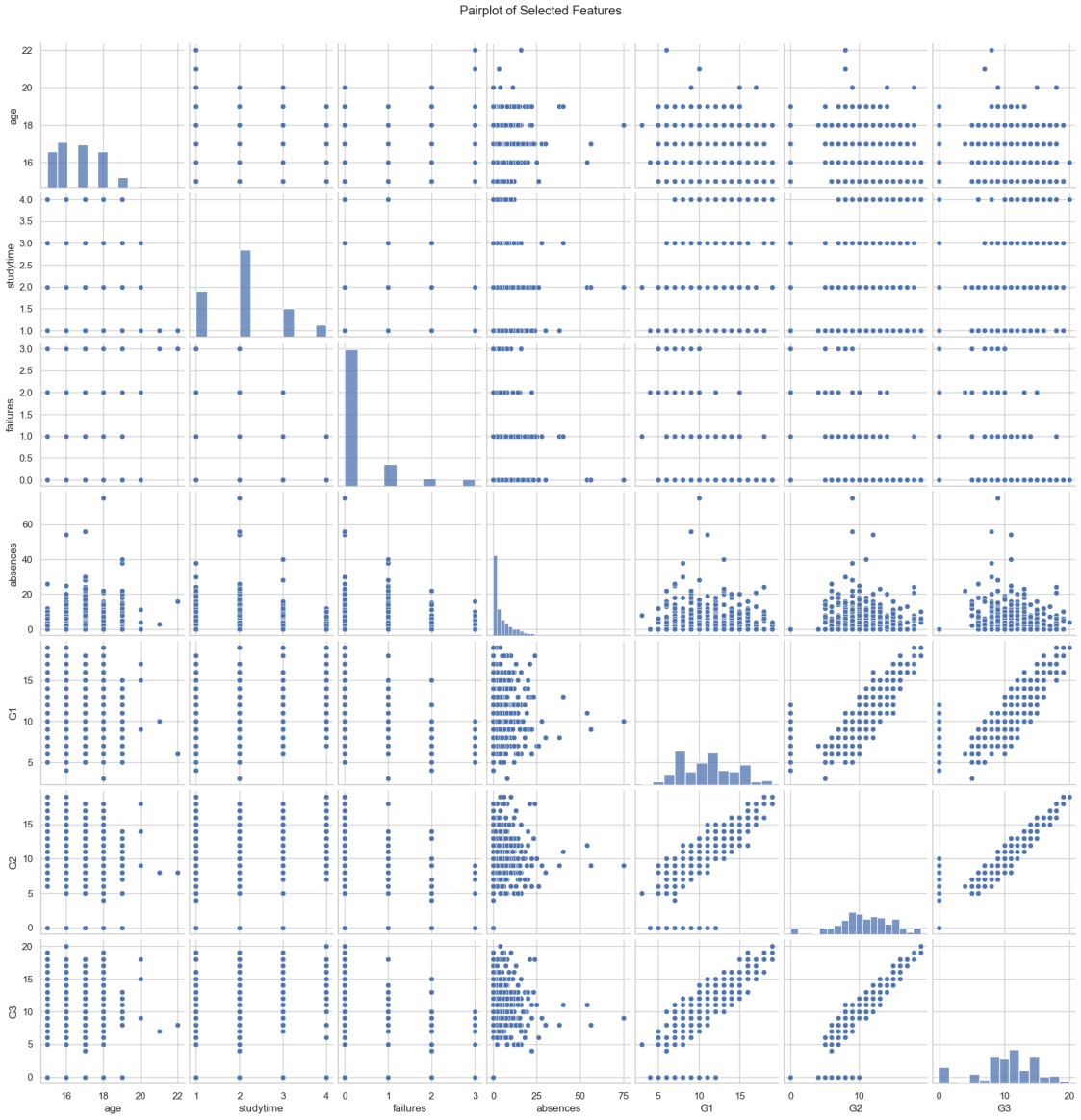

Distribution of Final Grades (G3)
Count
0
10
20
30
40
50
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Final Grade

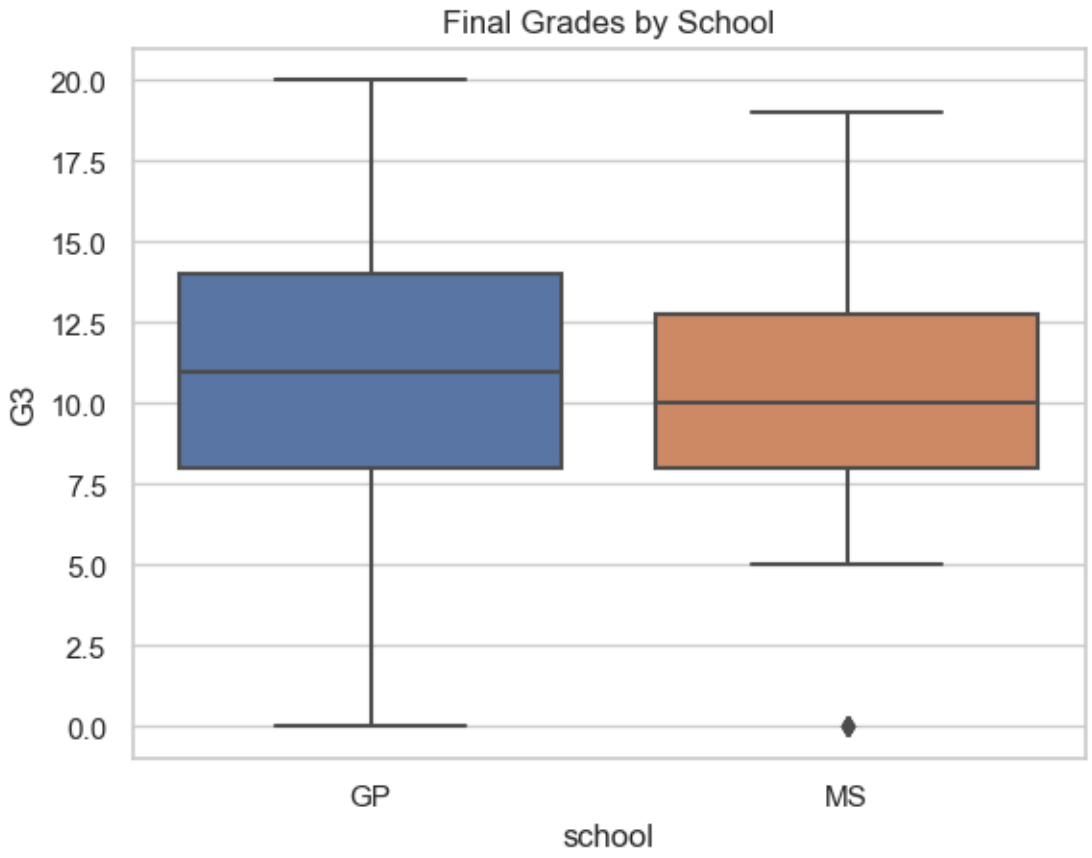

Final Grade Distribution by Gender
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
G3
F
M
sex

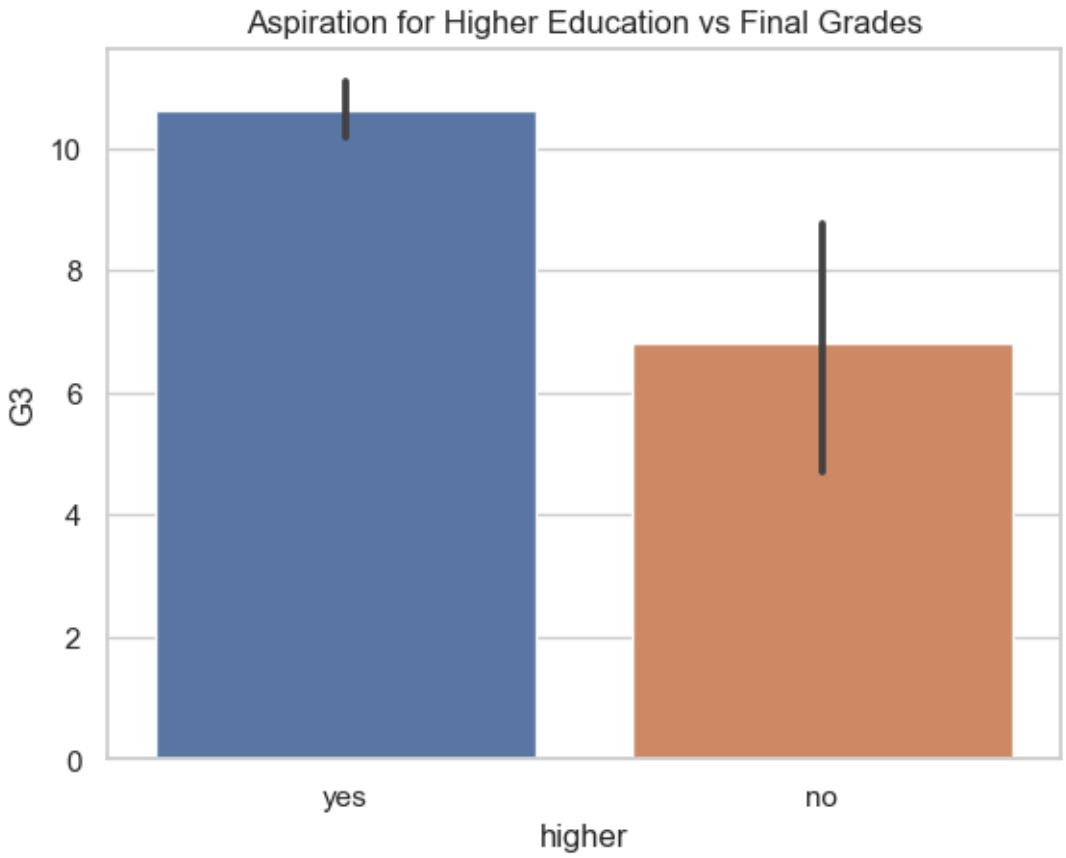

Study Time vs Final Grades
Final Grade (G3)
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Study Time (1: <2h, 2: 2-5h, 3: 5-10h, 4: >10h)
sex
F
M




```
C:\Users\Parvej\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118:
UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```


Pairplot of Selected Features
age
studytime
failures
absences
G1
G2
G3


Final Grades by School
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
G3
GP
MS
school




```
Observations:
- Females slightly outperform males in final grades (G3).
- More study time is associated with higher grades.
- Students aspiring for higher education tend to have better final grades.
- G1 and G2 (previous grades) strongly predict G3 (final grade).
- No missing data found.

Summary of Findings:

- Dataset contains 395 students with 33 features.
- Most students study between 2-5 hours per week.
- Previous period grades (G1 and G2) are highly correlated with final grade G3.
- Studytime, failures, and absences impact student performance.
- Female students and students aiming for higher education perform slightly
better.
```

[ ]: