

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
data = pd.read_csv('student/student-mat.csv')

# load the data into dataframe
data = pd.read_csv('student/student-mat.csv', sep=';')
print("Data loaded successfully.")

# Data exploration
print(data.head())
print(data.info())
print("\nDataset information:")
```

File Edit Selection View Go Run ... ← → Intership data analysis

EXPLORER

INTERSHIP DATA ANALYSIS

student

student-mat - Copy.csv

student-mat.csv

student-merge.R

student-por.csv

student.txt

assignment.ipynb

student-mat.csv student-por.csv assignment.ipynb

assignment.ipynb > import pandas as pd

Generate + Code + Markdown | Run All Restart Clear All Outputs Go To Jupyter Variables

```
print("\nMissing values:")
print(data.isnull().sum())

# Data cleaning
print("\nMissing values:")
print(data.isnull().sum())

#Remove duplicates
data = data.drop_duplicates()
print("\nDuplicates removed.")
print(data.columns)

# Data analysis
avg_grades = data['G3'].mean()
print(f"\nAverage final grade (G3): {avg_grades}")

student_above_15 = len(data[data['G3'] > 15])
print(f"Number of students with final grade above 15: {student_above_15}")

correlation = data['studytime'].corr(data['G3'])
print(f"\nCorrelation between study time and final grade: {correlation:.2f}")

average_grade_by_age = data.groupby('age')['G3'].mean()
print("\nAverage final grade by age:")
print(average_grade_by_age)
```

```

# Data visualization
plt.figure(figsize=(8, 5))
plt.hist(data['G3'],bins=10,color='orange',edgecolor='black')
plt.title("Distribution of final Grades(G3)")
plt.xlabel("Final Grade")
plt.ylabel("Frequency")
plt.show()

plt.figure(figsize=(8, 5))
sns.scatterplot(data=data, x='studytime', y='G3', hue='age')
plt.title("Study Time vs Final Grade")
plt.xlabel("Study Time (hours)")
plt.ylabel("Final Grade")
plt.legend(title="Age") # Rename the legend title
plt.show()

average_grade_by_age.plot(kind='bar', color=['purple','yellow'], figsize=(8, 5))
plt.title("Average Final Grade by Age")
plt.xlabel("Age")
plt.ylabel("Average Final Grade")
plt.show()

```



EXPLORE

INTENSHIP DATA ANALYSIS

student

student-mat - Copy.csv

student-mat.csv

student-merge.R

student-por.csv

student.txt

assignment.ipynb

student-mat.csv

student-por.csv

assignment.ipynb

assignment.ipynb > import pandas as pd

Generate

+ Code

+ Markdown

Run All

Restart

Clear All Outputs

Go To

Jupyter Variables

...

Data loaded successfully.

	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]

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 395 entries, 0 to 394

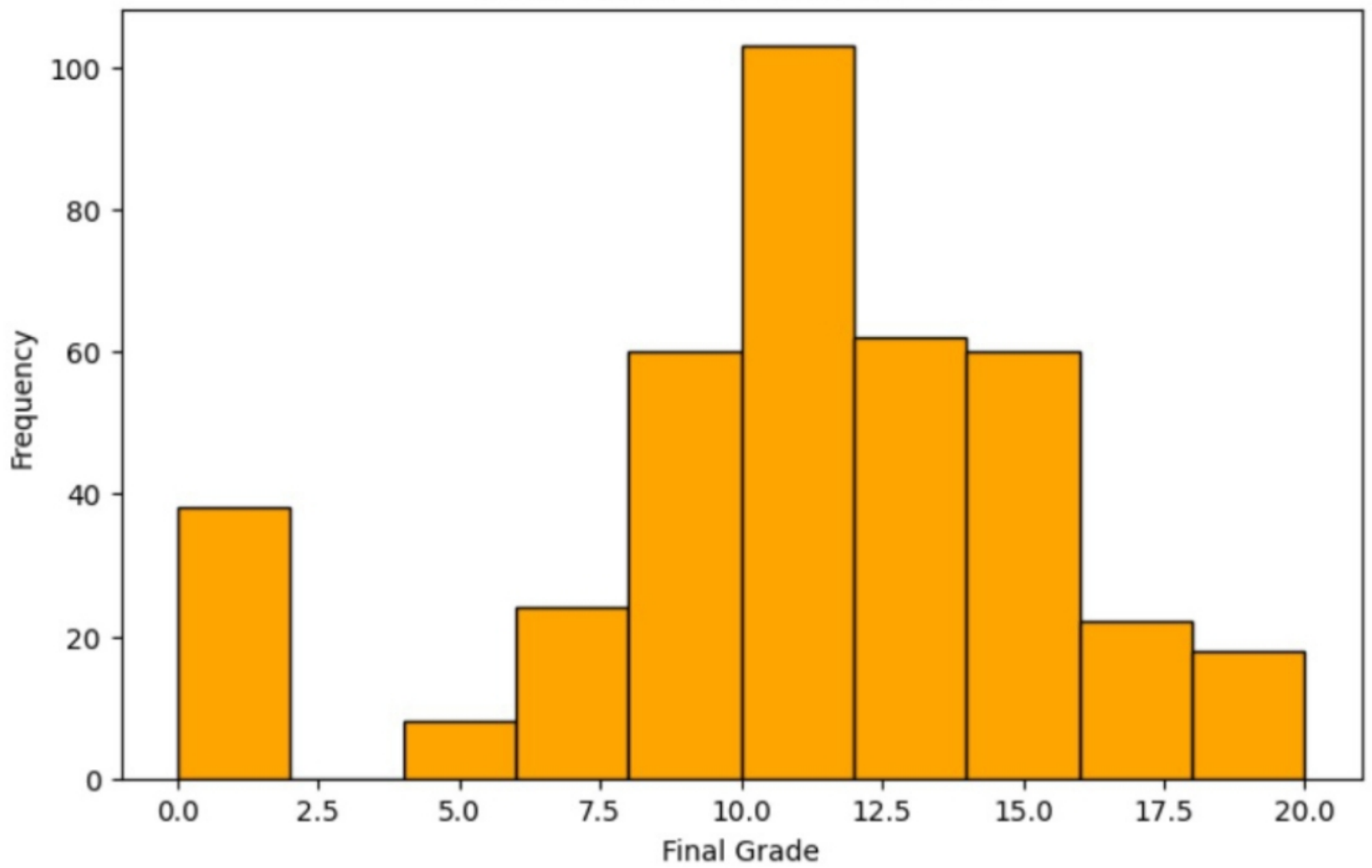
Data columns (total 33 columns):

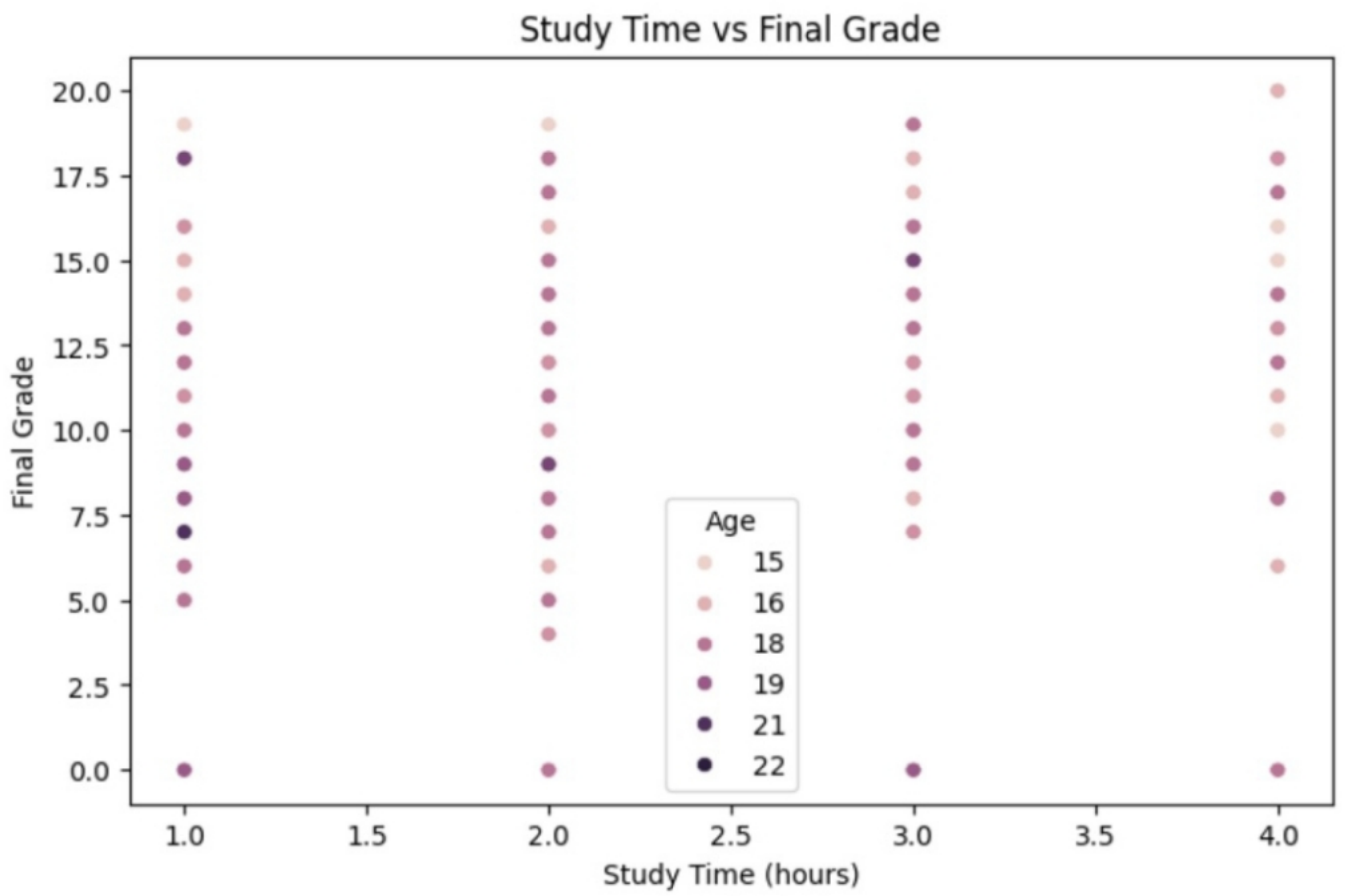
#	Column	Non-Null Count	Dtype
0	school	395 non-null	object
1	sex	395 non-null	object
2	age	395 non-null	int64
3	address	395 non-null	object
...			
20		14.000000	
21		7.000000	

OUTLINE

TIMELINE

Distribution of final Grades(G3)





Average Final Grade by Age

