Code :

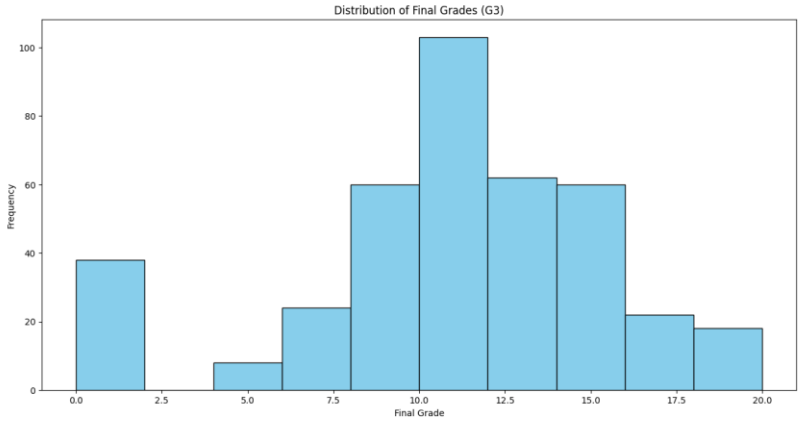
```
# Import Libraries

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import urllib.request
import zipfile

# Step 1: Download and Extract Dataset
url = "https://archive.ics.uci.edu/ml/machine-learning-
databases/00320/student.zip"
urllib.request.urlretrieve(url, "student.zip")
with zipfile.ZipFile("student.zip", "r") as zip_ref:
    zip_ref.extractall(".")
print("Dataset downloaded and extracted successfully!")

# Step 2: Load the Dataset
data = pd.read_csv("student-mat.csv", sep=";")
print("Data loaded successfully!")
print(data.head())  # Display the first few rows

# Step 3: Data Exploration
print("\nDataset Info:")
print(data.info())  # Dataset information (data types, missing values, etc.)
print("\nDataset Description:")
print(data.describe())  # Statistical summary of numeric columns
print("\nMissing Values:")
print(data.isnull().sum())  # Check for missing values
print("\nDataset Size (Rows, Columns):")
print(data.shape)  # Dataset size

# Step 4: Data Cleaning
# Remove rows with missing values (if any)
data = data.dropna()
# Remove duplicate rows
data = data.drop_duplicates()
print("\nDuplicate entries removed!")

# Step 5: Data Analysis
# Question 1: Average Math Score (G3)
average_score = data['G3'].mean()
print(f"\nAverage Math Score (G3): {average_score:.2f}")

# Question 2: Students Scoring Above 15 in Final Grade (G3)
students_above_15 = len(data[data['G3'] > 15])
print(f"Number of students scoring above 15: {students_above_15}")
```

```python
# Question 3: Correlation Between Study Time and Final Grade
correlation = data['studytime'].corr(data['G3'])
print(f"Correlation between study time and final grade: {correlation:.2f}")

# Question 4: Gender with Higher Average Final Grade
average_grade_by_gender = data.groupby('sex')['G3'].mean()
print("\nAverage Final Grade by Gender:")
print(average_grade_by_gender)

# Step 6: Data Visualization
# Histogram of Final Grades
plt.figure(figsize=(8, 5))
plt.hist(data['G3'], bins=10, color='skyblue', edgecolor='black')
plt.title("Distribution of Final Grades (G3)")
plt.xlabel("Final Grade")
plt.ylabel("Frequency")
plt.show()

# Scatter Plot: Study Time vs Final Grade
plt.figure(figsize=(8, 5))
sns.scatterplot(data=data, x='studytime', y='G3', hue='sex')
plt.title("Study Time vs Final Grade")
plt.xlabel("Study Time (hours)")
plt.ylabel("Final Grade")
plt.legend(title="Gender")
plt.show()

# Bar Chart: Average Scores by Gender
plt.figure(figsize=(8, 5))
average_grade_by_gender.plot(kind='bar', color=['blue', 'pink'])
plt.title("Average Final Grade by Gender")
plt.ylabel("Average Final Grade")
plt.xlabel("Gender")
plt.xticks(rotation=0)
plt.show()
```

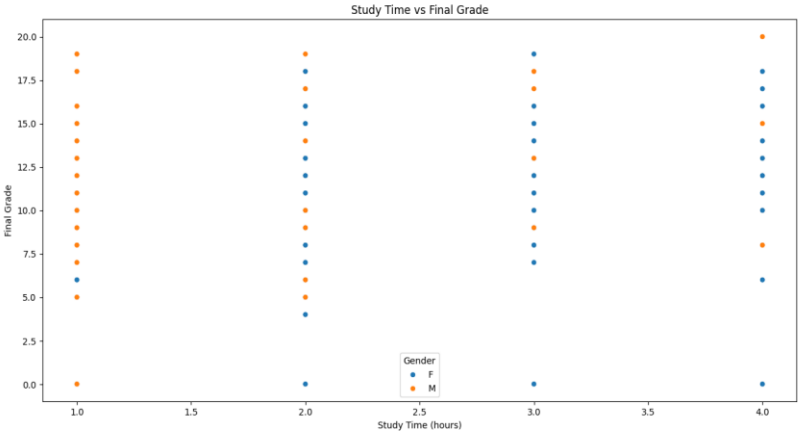
Output :

```
Dataset downloaded and extracted successfully!
Data loaded successfully!
  school sex  age address famsize Pstatus  Medu  Fedu     Mjob      Fjob  reason guardian  traveltime  ... higher internet romantic famrel freetime goout Dalc Walc health absences  G1  G2  G3
0     GP   F   18       U     GT3       A     4     4  at_home   teacher  course   mother           2  ...    yes       no       no      4        3     4    1    1      3        6   5   6   6
1     GP   F   17       U     GT3       T     1     1  at_home     other  course   father           1  ...    yes      yes       no      5        3     3    1    1      3        4   5   5   6
2     GP   F   15       U     LE3       T     1     1  at_home     other   other   mother           1  ...    yes      yes       no      4        3     2    2    3      3       10   7   8  10
3     GP   F   15       U     GT3       T     4     2   health  services    home   mother           1  ...    yes      yes      yes      3        2     2    1    1      5        2  15  14  15
4     GP   F   16       U     GT3       T     3     3    other     other    home   father           1  ...    yes       no       no      4        3     2    1    2      5        4   6  10  10

[5 rows x 33 columns]

 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

 Dataset Description:
              age        Medu        Fedu  traveltime   studytime    failures      famrel    freetime       goout        Dalc        Walc      health    absences          G1          G2          G3
count  395.000000  395.000000  395.000000  395.000000  395.000000  395.000000  395.000000  395.000000  395.000000  395.000000  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    3.944304    3.235443    3.108861    1.481013    2.291139    3.554430    5.708861   10.908861   10.713924   10.415190
std      1.276043    1.094735    1.088201    0.697505    0.839240    0.743651    0.896659    0.998862    1.113278    0.890741    1.287897    1.390303    8.003096    3.319195    3.761505    4.581443
min     15.000000    0.000000    0.000000    1.000000    1.000000    0.000000    1.000000    1.000000    1.000000    1.000000    1.000000    1.000000    0.000000    3.000000    0.000000    0.000000
25%     16.000000    2.000000    2.000000    1.000000    1.000000    0.000000    4.000000    3.000000    2.000000    1.000000    1.000000    3.000000    0.000000    8.000000    9.000000    8.000000
50%     17.000000    3.000000    2.000000    1.000000    2.000000    0.000000    4.000000    3.000000    3.000000    1.000000    2.000000    4.000000    4.000000   11.000000   11.000000   11.000000
75%     18.000000    4.000000    3.000000    2.000000    2.000000    0.000000    5.000000    4.000000    4.000000    2.000000    3.000000    5.000000    8.000000   13.000000   13.000000   14.000000
max     22.000000    4.000000    4.000000    4.000000    4.000000    3.000000    5.000000    5.000000    5.000000    5.000000    5.000000    5.000000   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
Dataset Size (Rows, Columns):
(395, 33)

Duplicate entries removed!

Average Math Score (G3): 10.42
Number of students scoring above 15: 40
Correlation between study time and final grade: 0.10

Average Final Grade by Gender:
sex
F     9.966346
M    10.914439
Name: G3, dtype: float64
PS C:\Users\sahil>
```

Histogram of Final Grades :



Scatter Plot: Study Time vs Final Grade :



Bar Chart: Average Scores by Gender