Importing the required modules to do analysis

```
import pandas as pd
pd.set_option('display.max_columns', 50) # this is to specify to
display the no. of columns
import matplotlib.pyplot as plt
```

Data Loading

we need to load the data into jupyter notebook and it can be done by as follow

```
data = pd.read_csv("student-mat.csv", delimiter=';', quotechar='"')
#here delimiter and quotechar are used because of the format the data
is stored in

data.head() # displaying first five 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

   reason guardian  traveltime  studytime  failures schoolsup famsup
paid  \
0  course   mother           2          2         0       yes     no
no
1  course   father           1          2         0        no    yes
no
2   other   mother           1          2         3       yes     no
yes
3    home   mother           1          3         0        no    yes
yes
4    home   father           1          2         0        no    yes
yes

  activities nursery higher internet romantic  famrel  freetime  goout
Dalc  \
0         no     yes    yes       no       no       4         3      4
1
1         no      no    yes      yes       no       5         3      3
1
```

```
2         no     yes    yes      yes       no       4         3      2
2
3        yes     yes    yes      yes      yes       3         2      2
1
4         no     yes    yes       no       no       4         3      2
1

   Walc  health  absences  G1  G2  G3
0     1       3         6   5   6   6
1     1       3         4   5   5   6
2     3       3        10   7   8  10
3     1       5         2  15  14  15
4     2       5         4   6  10  10
```

Data Exploration

```
data.isnull().sum() #checking for null values in the data
```

```
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
```

seems like there are no null values present in the data.

```
data.dtypes #checking for the data types of columns
```

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

```
data.shape # shape of the data
```

```
(395, 33)
```

Data Cleaning

Since we don't have any missing values there is no scope to fill the missing values.

```
# let's check for duplicate rows
data.drop_duplicates()

    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
..     ...  ..  ...     ...     ...     ...   ...   ...       ...
...
390     MS   M   20       U     LE3       A     2     2  services
services
391     MS   M   17       U     LE3       T     3     1  services
services
392     MS   M   21       R     GT3       T     1     1     other
other
393     MS   M   18       R     LE3       T     3     2  services
other
394     MS   M   19       U     LE3       T     1     1     other
at_home

     reason guardian  traveltime  studytime  failures schoolsup famsup
paid  \
0    course   mother           2          2         0       yes     no
no
1    course   father           1          2         0        no    yes
no
2     other   mother           1          2         3       yes     no
yes
3      home   mother           1          3         0        no    yes
yes
4      home   father           1          2         0        no    yes
yes
..      ...      ...         ...        ...       ...       ...    ...
...
390  course    other           1          2         2        no    yes
yes
391  course   mother           2          1         0        no     no
no
```

```
392  course    other          1          1         3        no     no
no
393  course   mother          3          1         0        no     no
no
394  course   father          1          1         0        no     no
no

    activities nursery higher internet romantic  famrel  freetime
goout  \
0           no     yes    yes       no       no       4         3
4
1           no      no    yes      yes       no       5         3
3
2           no     yes    yes      yes       no       4         3
2
3          yes     yes    yes      yes      yes       3         2
2
4           no     yes    yes       no       no       4         3
2
..         ...     ...    ...      ...      ...     ...       ...    .
..
390         no     yes    yes       no       no       5         5
4
391         no      no    yes      yes       no       2         4
5
392         no      no    yes       no       no       5         5
3
393         no      no    yes      yes       no       4         4
1
394         no     yes    yes      yes       no       3         2
3

     Dalc  Walc  health  absences  G1  G2  G3
0       1     1       3         6   5   6   6
1       1     1       3         4   5   5   6
2       2     3       3        10   7   8  10
3       1     1       5         2  15  14  15
4       1     2       5         4   6  10  10
..    ...   ...     ...       ...  ..  ..  ..
390     4     5       4        11   9   9   9
391     3     4       2         3  14  16  16
392     3     3       3         3  10   8   7
393     3     4       5         0  11  12  10
394     3     3       5         5   8   9   9

[395 rows x 33 columns]
```

so the shape of the dateset does not changed that means there are no duplicates in the dataset.

Data Analysis Questions

1. What is the average score in math (G3)?

```
data['G3'].mean()

10.415189873417722
```

Therefore the mean is 10.415189873417722 for average score in math(G3)

2. How many students scored above 15 in their final grade (G3)?

```
data[data['G3'] > 15].shape[0]

40
```

Total 40 students secured marks more than 15 in G3

3. Is there a correlation between study time (study time) and the final grade (G3)?

```
correlation = data['studytime'].corr(data['G3'])
correlation

0.09781968965319633
```

0.09781968965319633 is near to 0 so that means there is no possible correlation between studytime and final grade(G3)

4. Which gender has a higher average final grade (G3)?

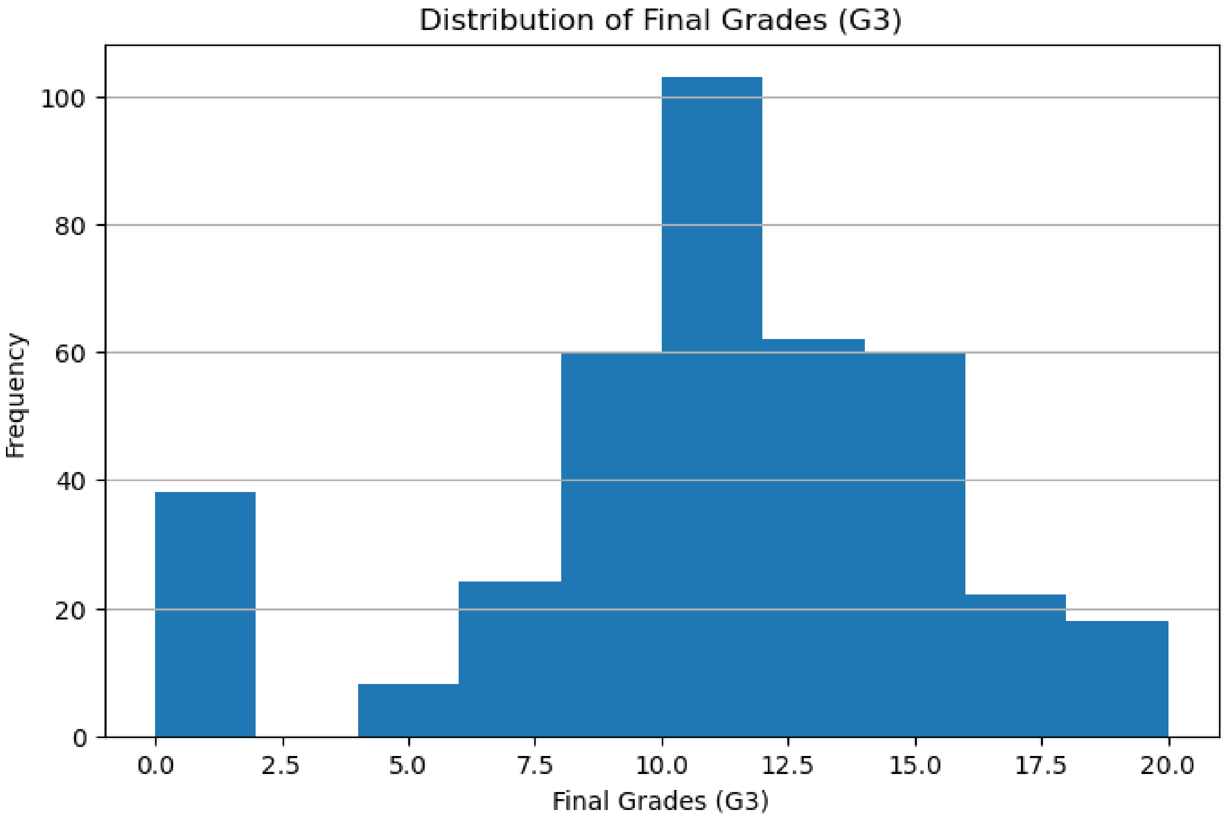
```
data.groupby('sex')['G3'].mean()

sex
F     9.966346
M    10.914439
Name: G3, dtype: float64
```

we can observe that Males got more average than Females in G3

Data Visualization

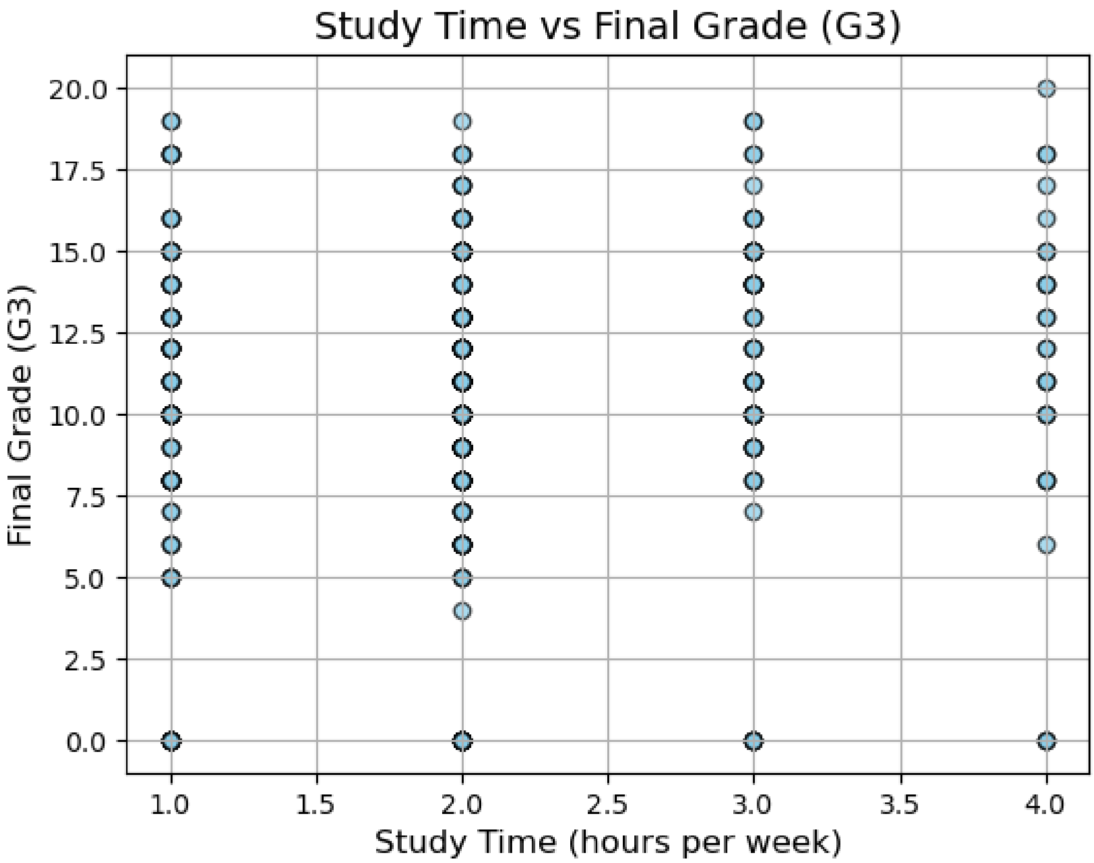
1. Plot a histogram of final grades (G3)

```
plt.figure(figsize=(8, 5))
plt.hist(data['G3'], bins=10)
plt.title('Distribution of Final Grades (G3)')
plt.xlabel('Final Grades (G3)')
plt.ylabel('Frequency')
plt.grid(axis='y')
plt.show()
```



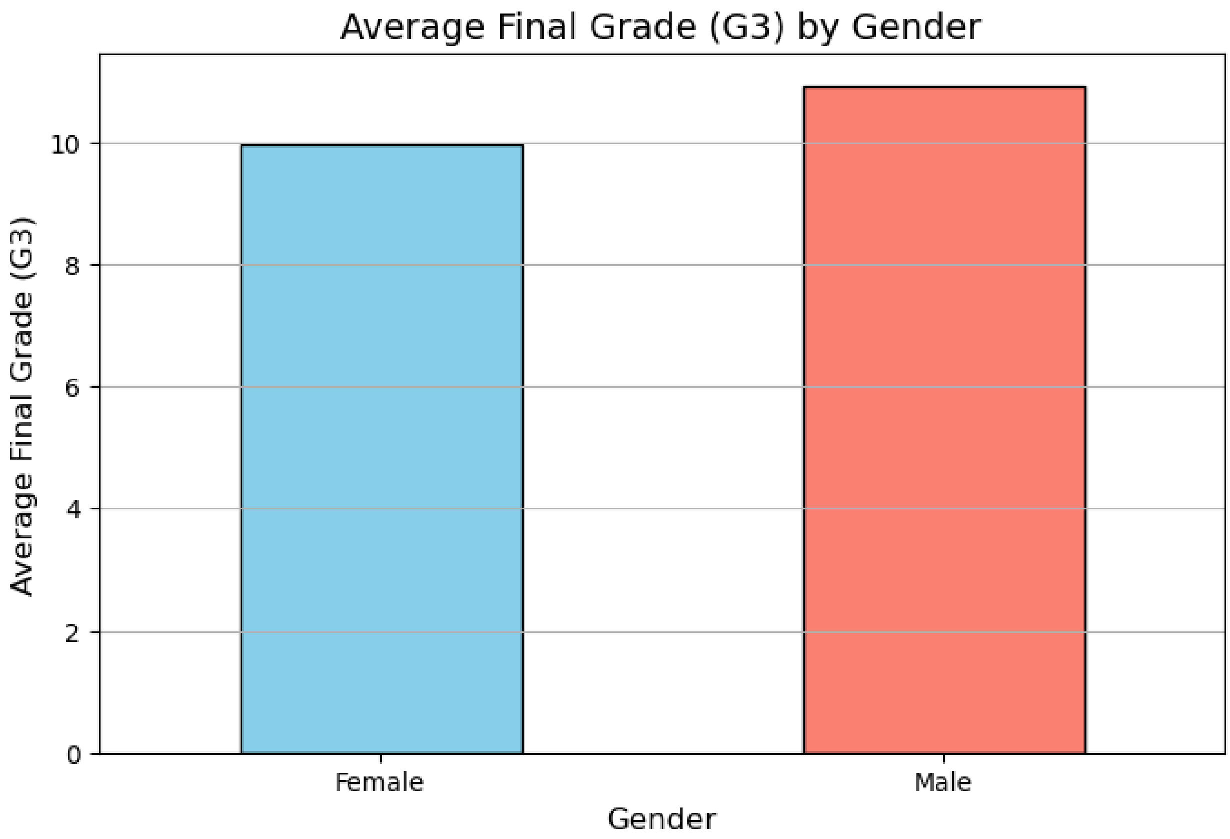
2. Create a scatter plot between study time (study time) and final grade (G3)

```
plt.scatter(data['studytime'], data['G3'], color='skyblue', alpha=0.7,
edgecolor='black')

# Adding labels and title for the plot
plt.title('Study Time vs Final Grade (G3)', fontsize=14)
plt.xlabel('Study Time (hours per week)', fontsize=12)
plt.ylabel('Final Grade (G3)', fontsize=12)
plt.grid()

plt.show()
```



3. Create a bar chart comparing the average scores of male and female students.

```python
plt.figure(figsize=(8, 5))
data.groupby('sex')['G3'].mean().plot(kind='bar', color=['skyblue', 'salmon'], edgecolor='black')

plt.title('Average Final Grade (G3) by Gender', fontsize=14)
plt.xlabel('Gender', fontsize=12)
plt.ylabel('Average Final Grade (G3)', fontsize=12)
plt.xticks(ticks=[0, 1], labels=['Female', 'Male'], rotation=0) # used to specify the marks on the x axis
plt.grid(axis='y')
plt.show()
```


Average Final Grade (G3) by Gender
Average Final Grade (G3)
0
2
4
6
8
10
Female
Male
Gender