INTERNSHIP TASK - 3

Q1. Loading data in pandas data frame .

[2]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

[3]:
```
df=pd.read_csv('Student_Marks.csv')
df.head()
```

[3]:

	number_courses	time_study	Marks
0	3	4.508	19.202
1	4	0.096	7.734
2	4	3.133	13.811
3	6	7.909	53.018
4	8	7.811	55.299

Q2. Basic Statistical Analysis.

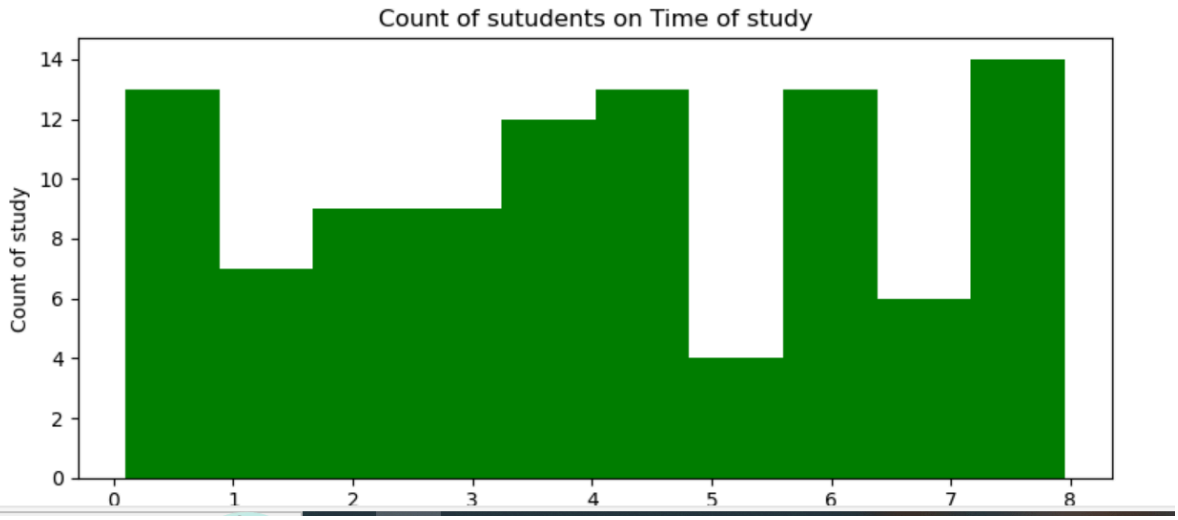
[7]:
```
df.describe()
```

[7]:

	number_courses	time_study	Marks
count	100.000000	100.000000	100.000000
mean	5.290000	4.077140	24.417690
std	1.799523	2.372914	14.326199
min	3.000000	0.096000	5.609000
25%	4.000000	2.058500	12.633000
50%	5.000000	4.022000	20.059500
75%	7.000000	6.179250	36.676250
max	8.000000	7.957000	55.299000

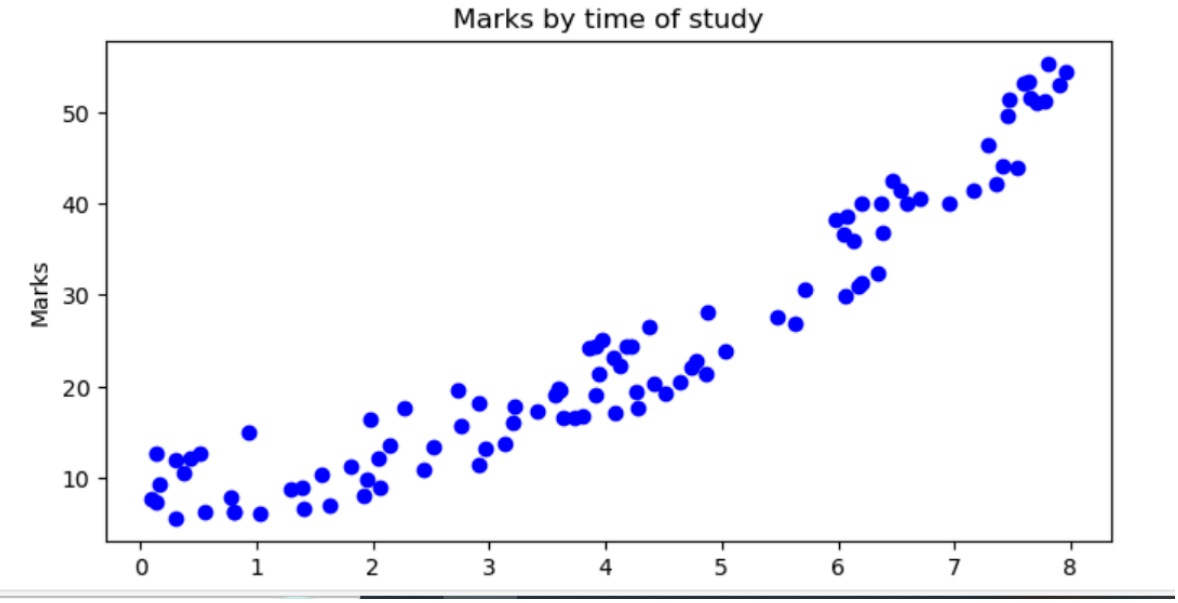
Histogram of count of students on the basis of number of study hours

[44]:
```
plt.figure(figsize=(8,4))
plt.hist(df['time_study'],color='green')
plt.xlabel('Time of study')
plt.ylabel('Count of study')
plt.title('Count of sutudents on Time of study')
plt.tight_layout()
plt.show()
```



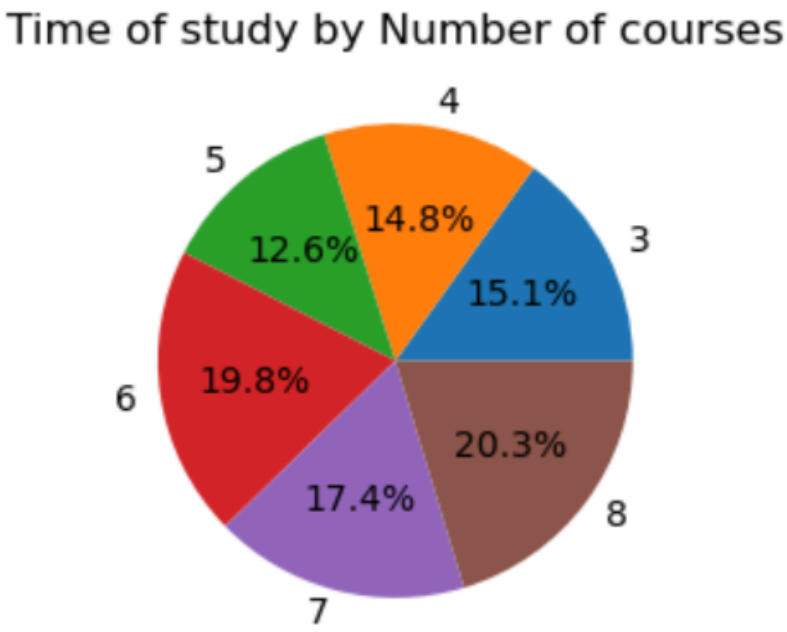
Scatter plot of study hours and Mar

[30]:
```
plt.figure(figsize=(8,4))
plt.scatter(df['time_study'],df['Marks'],color='blue')
plt.xlabel('Time of study')
plt.ylabel('Marks')
plt.title('Marks by time of study')
plt.show()
```



Bar plot of Number of hours of study and number of courses

```
[40]: plt.figure(figsize=(3,3))
      plt.pie(group,labels=group.index,autopct='%1.1f%%')
      plt.title('Time of study by Number of courses')
      plt.show()
```



Conclusion 🙂

1. The average marks of students is 24.4
2. 50 percent students have marks less then 20
3. Maximun marks obtained by student are 55.2
4. Hours of study and marks are highly correlated
5. Maximum number courses are 8.