

In [1]:

```
# DV Case Study 4 on matplotlib
# AIML OMEGA
# 2111CS020505
# N.Shashank Sagar
```

In [2]:

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv(r"C:\Users\SHASHANK SAGAR\Desktop\DV_505\class_marks.csv")
```

In [3]:

```
DF=df.sort_values("TOTAL",ascending=True)
DF
```

Out[3]:

	TOTAL	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7	Q5M10	Q6aM4
69	3	1	0	1	0	0	0	1	0	0	0
11	8	2	2	0	3	1	0	0	0	0	0
23	9	4	3	0	0	0	0	0	0	0	1
22	14	4	4	5	2	0	0	0	0	0	0
57	17	3	0	0	4	0	0	3	7	0	4
...
73	40	4	6	0	0	5	5	3	0	10	4
53	40	4	6	6	4	5	5	0	0	10	0
51	40	0	0	6	4	0	0	3	7	10	0
33	40	0	0	6	4	5	5	3	7	0	4
65	40	4	6	6	4	5	5	0	0	10	0

86 rows × 12 columns



TOTAL values are sorted in ascending order above.

In [4]:

```
DF['Q1']=DF["Q1aM4"]+DF["Q1bM6"]
DF['Q2']=DF['Q2aM6']+DF['Q2bM4']
DF['Q3']=DF['Q3aM5']+DF['Q3bM5']
DF['Q4']=DF['Q4aM3']+DF['Q4bM7']
DF['Q6']=DF['Q6aM4']+DF['Q6bM6']
DF
```

Out[4]:

	TOTAL	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7	Q5M10	Q6aM4
69	3	1	0	1	0	0	0	1	0	0	0
11	8	2	2	0	3	1	0	0	0	0	0
23	9	4	3	0	0	0	0	0	0	0	1
22	14	4	4	5	2	0	0	0	0	0	0
57	17	3	0	0	4	0	0	3	7	0	4
...
73	40	4	6	0	0	5	5	3	0	10	4
53	40	4	6	6	4	5	5	0	0	10	0
51	40	0	0	6	4	0	0	3	7	10	0
33	40	0	0	6	4	5	5	3	7	0	4
65	40	4	6	6	4	5	5	0	0	10	0

86 rows × 17 columns



New Columns Q1,Q2,Q3,Q4,Q6 are created above to do analysis.

In [5]:

```
a=DF.loc[(DF['TOTAL'] >= 15) & (DF['TOTAL'] <= 20)]
a=a.reset_index()
a
```

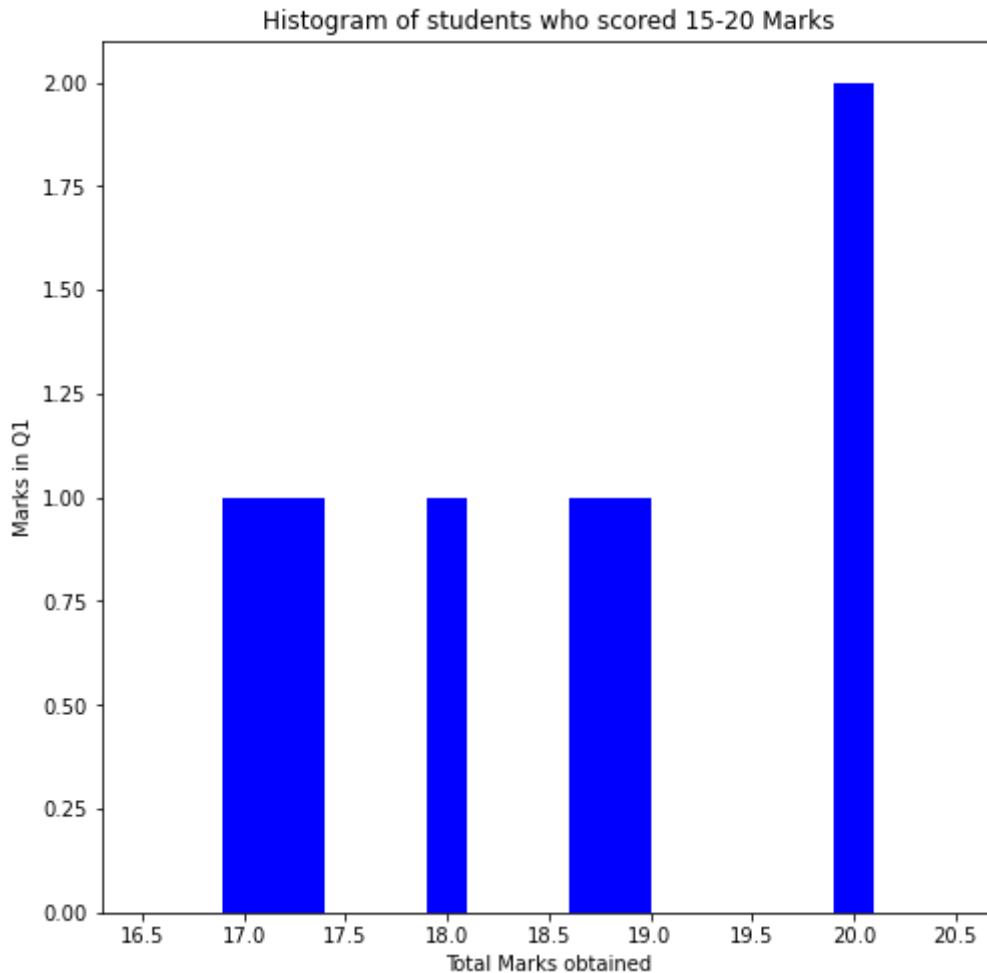
Out[5]:

	index	TOTAL	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7	Q5M10	C
0	57	17	3	0	0	4	0	0	3	7	0	
1	76	17	2	3	4	2	4	2	0	0	0	
2	63	18	4	0	4	2	0	0	0	0	8	
3	34	19	2	3	3	1	2	3	0	0	5	
4	68	20	4	6	6	4	0	0	0	0	0	
5	5	20	4	6	6	4	0	0	0	0	0	
6	60	20	2	5	3	2	0	0	0	0	8	
7	30	20	4	4	4	4	5	0	0	0	0	

Total marks 15-20 is filtered above

In [33]:

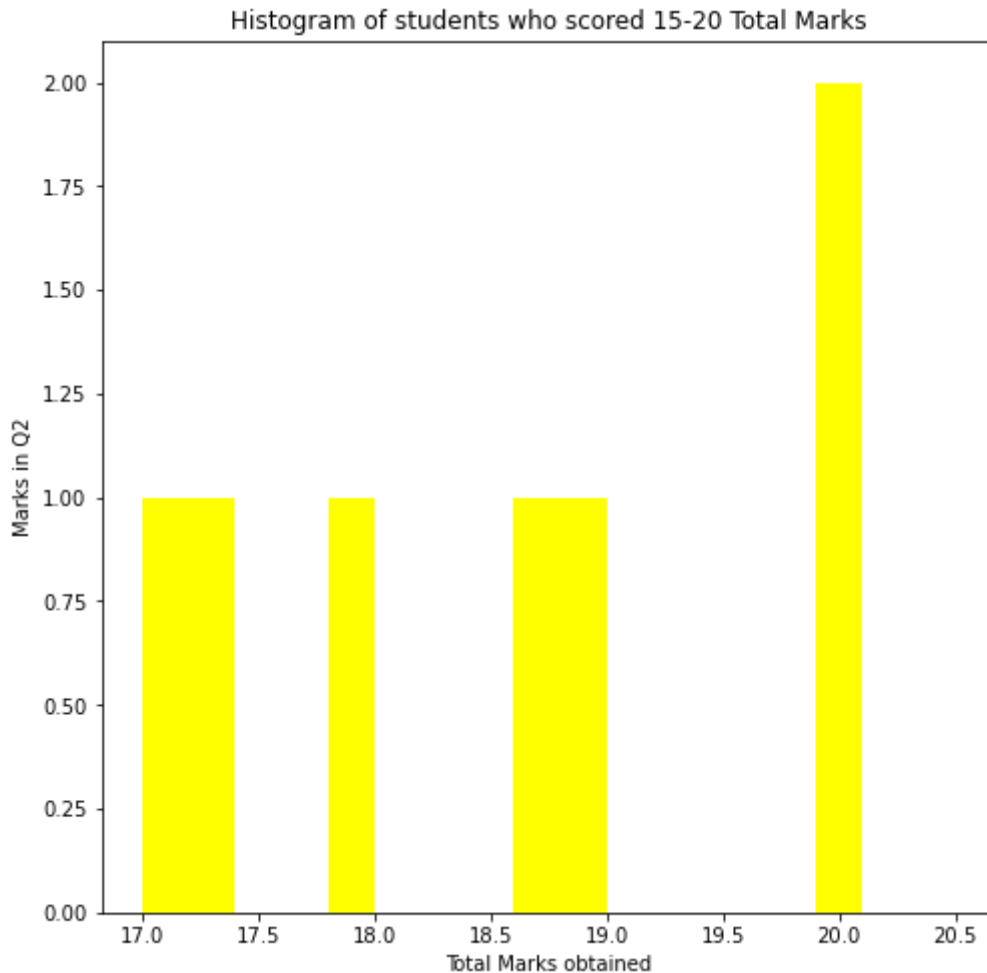
```
a1=a.groupby('Q1')['TOTAL']  
a1.hist(color = 'blue',figsize=[8,8],grid=False,bins=5)  
plt.title("Histogram of students who scored 15-20 Marks")  
plt.xlabel("Total Marks obtained")  
plt.ylabel("Marks in Q1")  
plt.show()
```



Most of the students scored 1 mark and maximum mark is 2, implying that all the students in this range secured very less marks in Q1

In [34]:

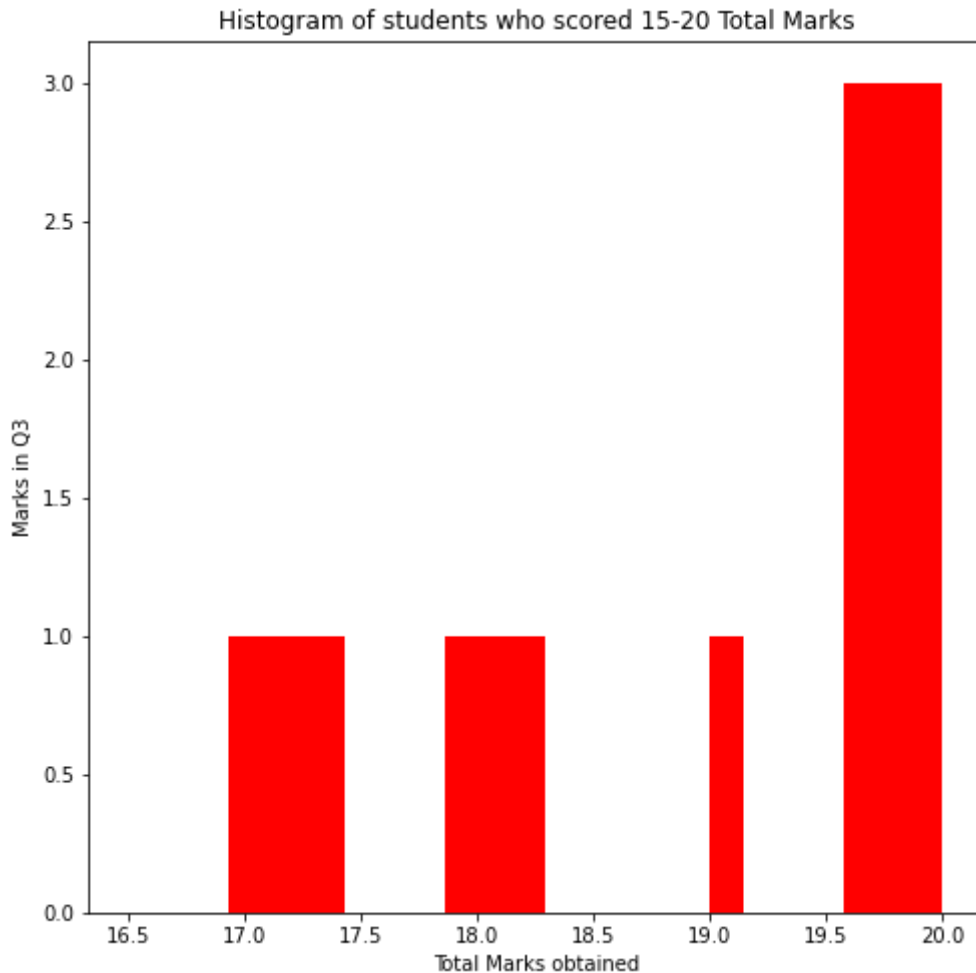
```
a2=a.groupby('Q2')['TOTAL']  
a2.hist(color='yellow',figsize=[8,8],grid=False,bins=5)  
plt.title("Histogram of students who scored 15-20 Total Marks")  
plt.xlabel("Total Marks obtained")  
plt.ylabel("Marks in Q2")  
plt.show()
```



Very few students got high marks in Q2 and that's only 2, so overall performance in Q2 is not upto the mark

In [35]:

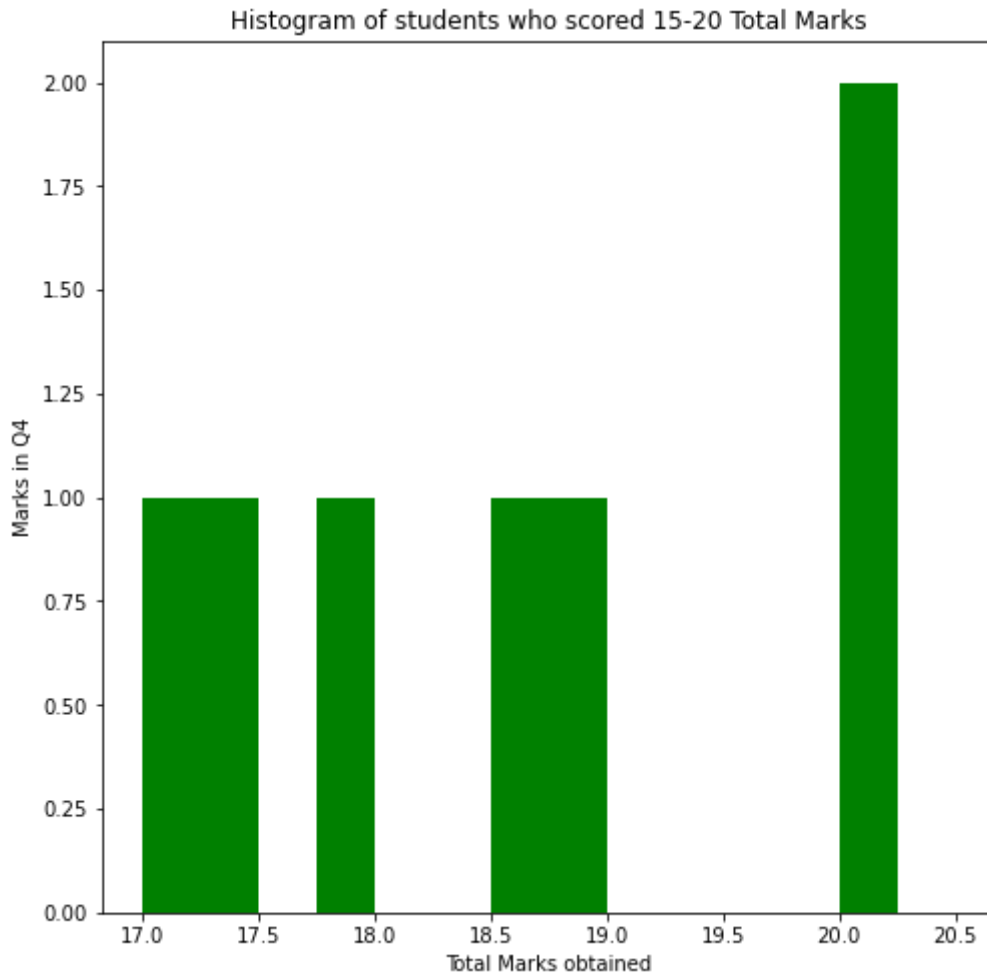
```
a3=a.groupby('Q3')['TOTAL']  
a3.hist(color='RED',figsize=[8,8],grid=False,bins=7)  
plt.title("Histogram of students who scored 15-20 Total Marks")  
plt.xlabel("Total Marks obtained")  
plt.ylabel("Marks in Q3")  
plt.show()
```



Only one student got 3 marks whereas others scored below 2 marks, implying very less marks are secured in Q3 overall

In [37]:

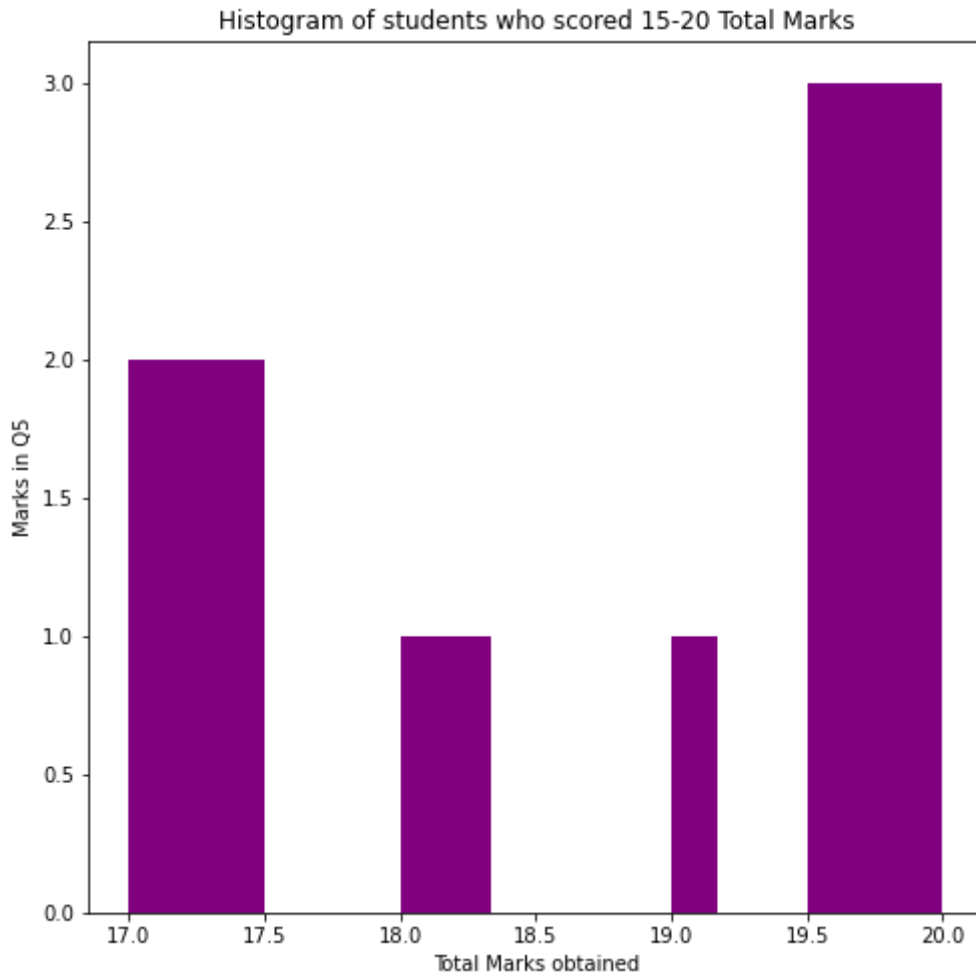
```
a4=a.groupby('Q4')['TOTAL']  
a2.hist(color='green',figsize=[8,8],grid=False,bins=4)  
plt.title("Histogram of students who scored 15-20 Total Marks")  
plt.xlabel("Total Marks obtained")  
plt.ylabel("Marks in Q4")  
plt.show()
```



Many of these students who attempted Q4 got less marks and maximum mark is also just two only

In [38]:

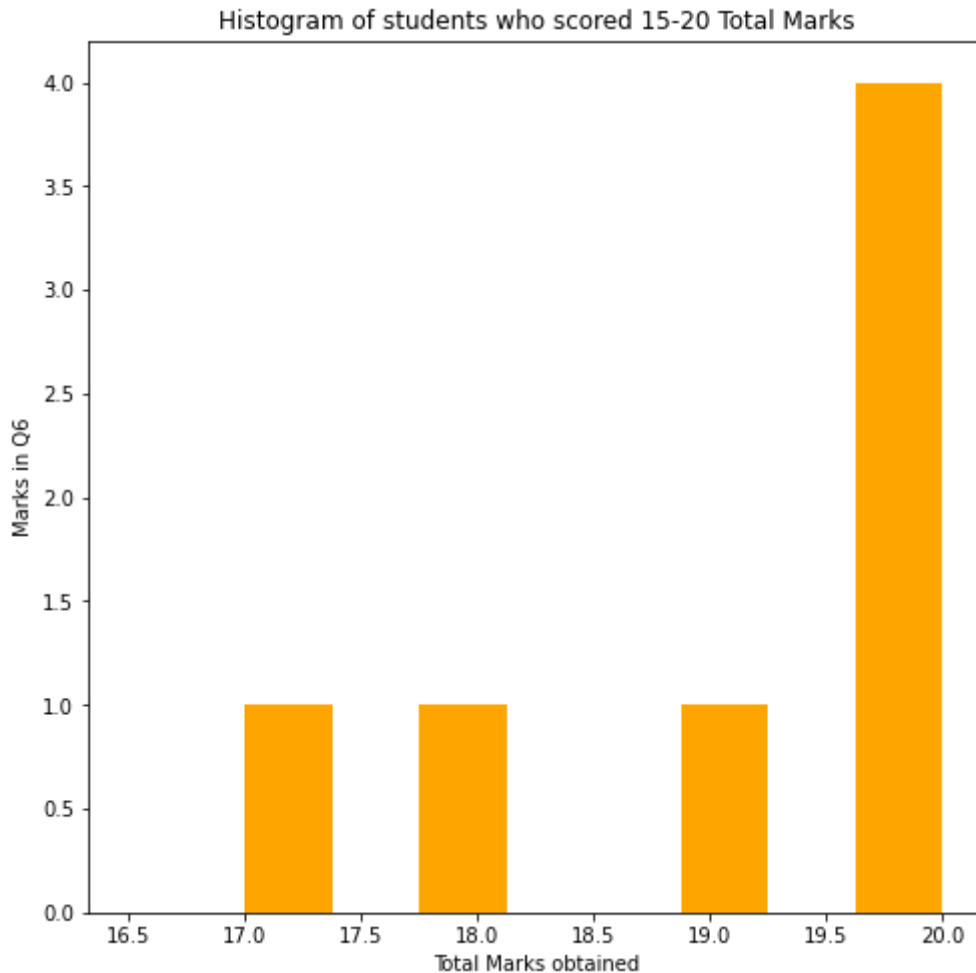
```
a5=a.groupby('Q5M10')['TOTAL']  
a5.hist(color='purple',figsize=[8,8],grid=False,bins=6)  
plt.title("Histogram of students who scored 15-20 Total Marks")  
plt.xlabel("Total Marks obtained")  
plt.ylabel("Marks in Q5")  
plt.show()
```



Majority of these students who attempted Q5 got less marks and maximum mark is also just three only

In [39]:

```
a6=a.groupby('Q6')['TOTAL']  
a6.hist(color='orange',figsize=[8,8],grid=False,bins=8)  
plt.title("Histogram of students who scored 15-20 Total Marks")  
plt.xlabel("Total Marks obtained")  
plt.ylabel("Marks in Q6")  
plt.show()
```



The maximum mark scored in this question is 4 marks, meaning the performance of students in this range remains bad as well.

In [12]:

```
b=DF.loc[(DF['TOTAL'] >= 20) & (DF['TOTAL'] <= 25)]
b=b.reset_index()
b
```

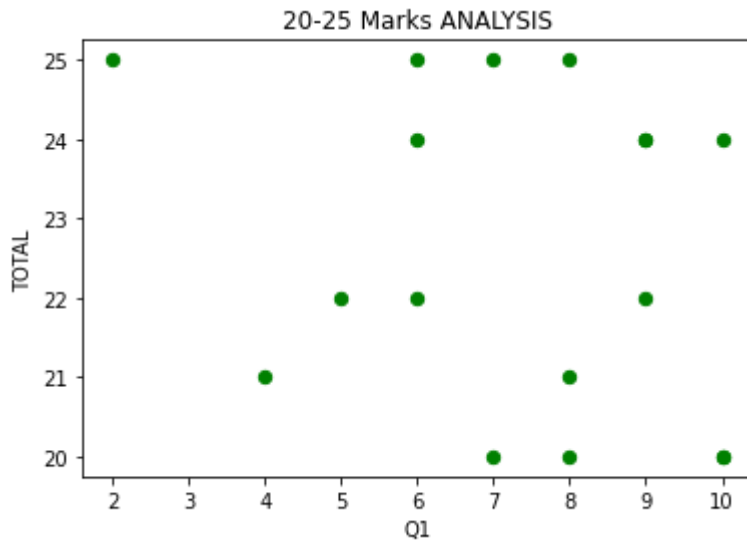
Out[12]:

	index	TOTAL	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7	Q5M10
0	68	20	4	6	6	4	0	0	0	0	0
1	5	20	4	6	6	4	0	0	0	0	0
2	60	20	2	5	3	2	0	0	0	0	8
3	30	20	4	4	4	4	5	0	0	0	0
4	54	21	2	6	0	0	5	5	3	0	0
5	75	21	4	0	6	1	1	1	0	0	8
6	77	22	4	5	0	0	3	2	2	0	0
7	39	22	3	2	0	1	3	3	0	0	0
8	25	22	4	2	5	2	4	3	2	0	0
9	61	24	4	5	6	4	0	5	0	0	0
10	18	24	3	3	5	3	0	0	2	1	7
11	50	24	4	5	0	0	5	5	0	0	5
12	3	24	4	6	6	3	2	2	0	0	0
13	72	25	2	0	4	4	0	5	3	7	0
14	48	25	2	6	0	0	0	0	3	6	8
15	6	25	3	4	0	2	5	5	0	0	6
16	74	25	1	5	6	4	0	0	0	0	9

Total marks 20-25 is filtered from the data set

In [13]:

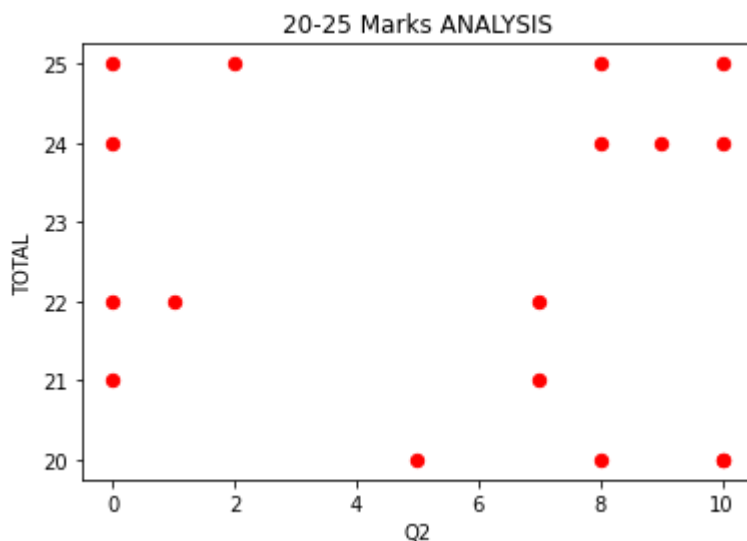
```
b.plot.scatter(x='Q1',y='TOTAL',color='green',s=40)
plt.title("20-25 Marks ANALYSIS")
plt.show()
```



Majority of the students in this range scored marks between 6-10 in Q1 and very few, just 3 students scored below 5 marks, maximum mark scored is 10 by two students.

In [14]:

```
b.plot.scatter(x='Q2',y='TOTAL',color='red',s=40)
plt.title("20-25 Marks ANALYSIS")
plt.show()
```

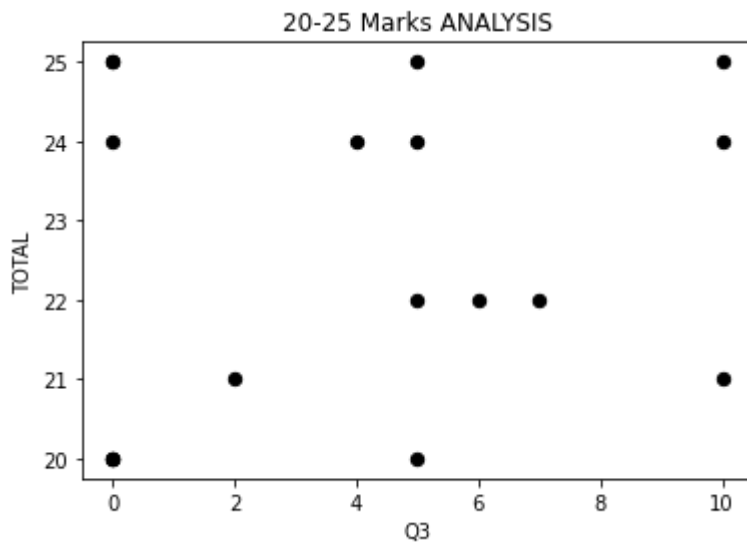


Some students have not attempted this question whereas those who have attempted scored marks

between 7 and 10 overall. Maximum mark is 10 scored by three students.

In [15]:

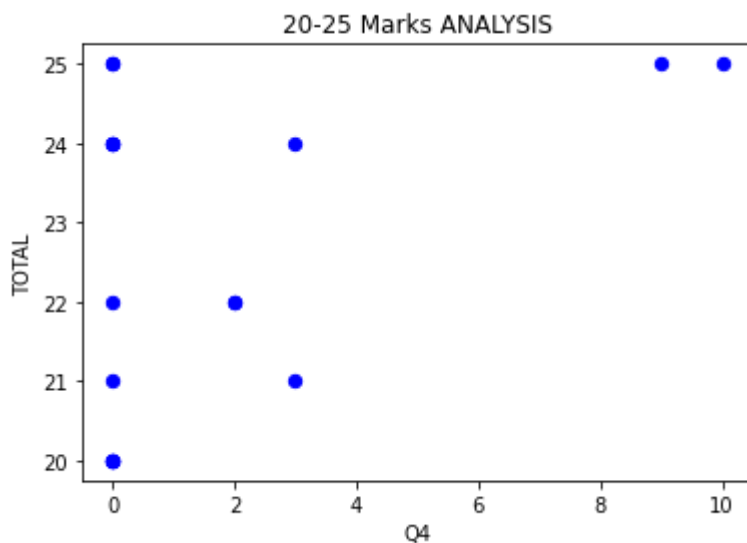
```
b.plot.scatter(x='Q3',y='TOTAL',color='black',s=40)
plt.title("20-25 Marks ANALYSIS")
plt.show()
```



Three students haven't attempted the question and most of them who attempted scored 4-7 marks and maximum mark is 10 scored by three students.

In [16]:

```
b.plot.scatter(x='Q4',y='TOTAL',color='blue',s=40)
plt.title("20-25 Marks ANALYSIS")
plt.show()
```

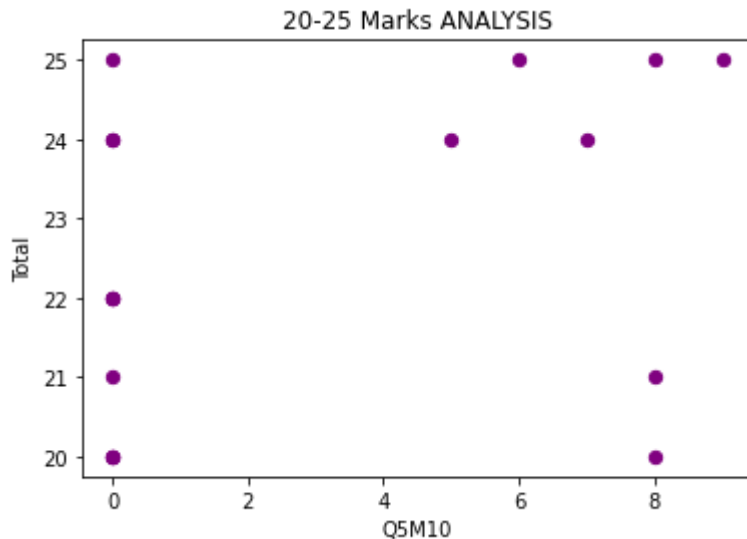


Most of the students in this range haven't attempted

this question, others got 2-4 marks and maximum mark is 10 by one student.

In [17]:

```
b.plot.scatter(x='Q5M10',y='TOTAL',color='purple',s=40)
plt.title("20-25 Marks ANALYSIS")
plt.ylabel("Total")
plt.show()
```



Majority of the students who attempted scored marks between 5 and 9, also some scored zero, meaning five students left the question unanswered

In [18]:

```
c=DF.loc[(DF['TOTAL'] >= 25) & (DF['TOTAL'] <= 30)]
c=c.reset_index()
c
```

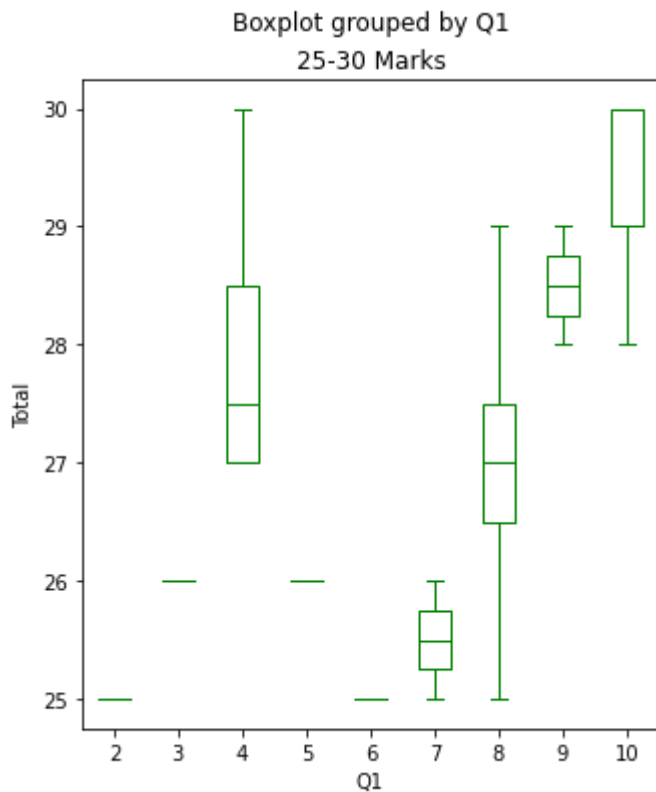
Out[18]:

	index	TOTAL	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7	Q5M10
0	72	25	2	0	4	4	0	5	3	7	0
1	48	25	2	6	0	0	0	0	3	6	8
2	6	25	3	4	0	2	5	5	0	0	6
3	74	25	1	5	6	4	0	0	0	0	9
4	41	26	2	3	4	3	4	3	0	3	4
5	19	26	3	0	6	4	0	2	2	1	7
6	31	26	3	4	6	2	2	0	1	0	0
7	8	27	3	5	5	0	0	0	0	0	9
8	79	27	2	6	0	3	2	5	0	0	9
9	29	27	4	0	6	1	0	0	0	7	0
10	82	27	2	2	5	3	0	0	0	0	7
11	38	28	4	5	6	4	5	3	1	0	0
12	67	28	4	6	4	4	0	0	0	0	10
13	84	28	4	0	5	4	5	4	0	0	6
14	40	29	4	6	6	4	0	0	1	1	7
15	20	29	2	6	2	2	5	5	0	0	7
16	52	29	4	5	4	3	0	0	3	6	4
17	85	29	4	6	0	0	0	0	3	5	7
18	35	30	4	6	6	4	0	1	0	0	8
19	14	30	4	6	6	2	4	5	3	0	0
20	16	30	4	0	6	4	5	2	0	0	8

Total marks 25-30 is filtered from the data set

In [19]:

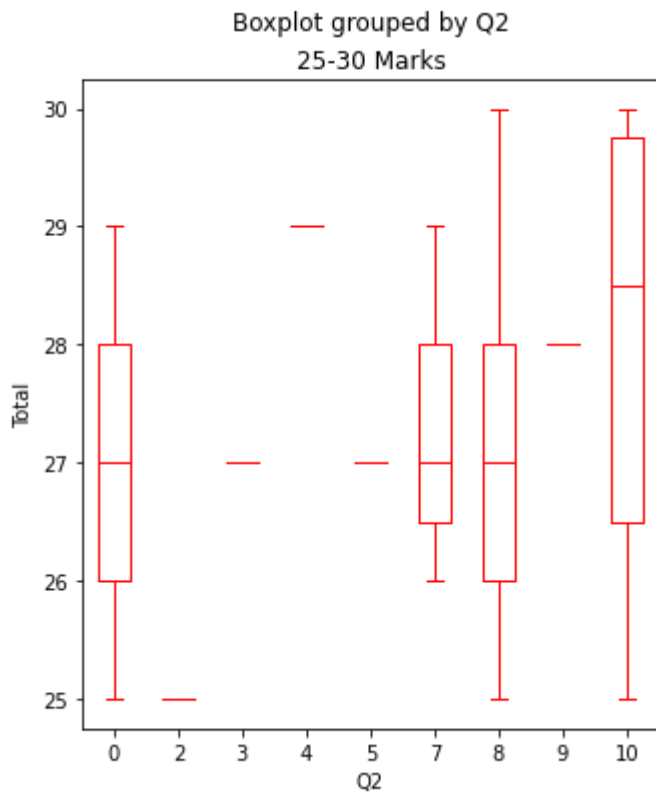
```
c.boxplot(by='Q1', column=['TOTAL'], grid=False, color='Green', figsize=[5,6])  
plt.title("25-30 Marks")  
plt.ylabel("Total")  
plt.show()
```



Many of the students in this range got marks between 8 and 10, the maximum mark is 10 and minimum mark is 0.

In [20]:

```
c.boxplot(by='Q2', column=['TOTAL'], grid=False, color='red', figsize=[5,6])  
plt.title("25-30 Marks")  
plt.ylabel("Total")  
plt.show()
```

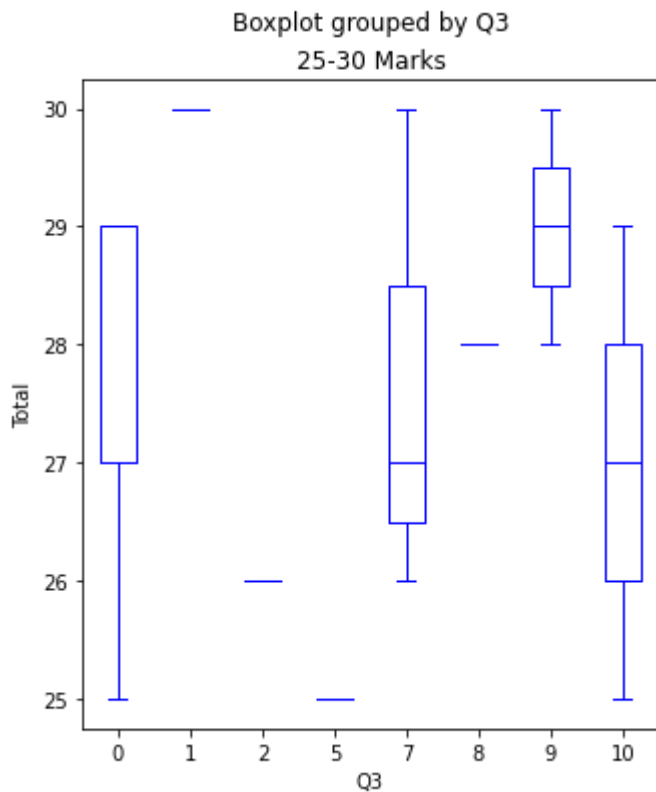


Majority of the students who attempted the question in this range scored marks between 7 and 10

The maximum mark is 10 whereas minimum mark is 0.

In [21]:

```
c.boxplot(by='Q3', column=['TOTAL'], grid=False, color='blue', figsize=[5,6])  
plt.title("25-30 Marks")  
plt.ylabel("Total")  
plt.show()
```

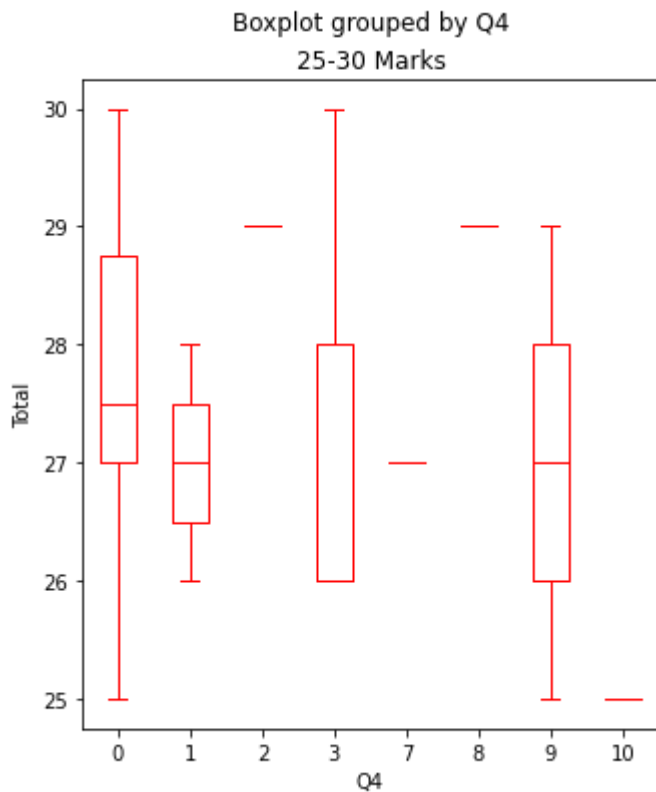


The minimum mark in this question is 0 and maximum mark is 10

Most of the students have secured marks between 7 and 9

In [22]:

```
c.boxplot(by='Q4', column=['TOTAL'], grid=False, color='red', figsize=[5,6])  
plt.title("25-30 Marks")  
plt.ylabel("Total")  
plt.show()
```

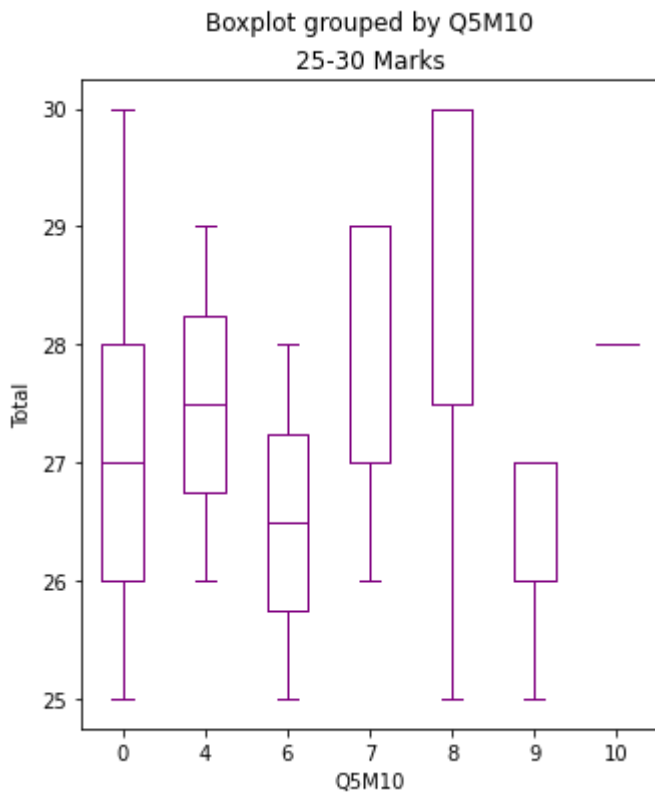


No student in this range secured full mark for this question

Minimum mark is 0 and the ones who attempted secured low marks.

In [23]:

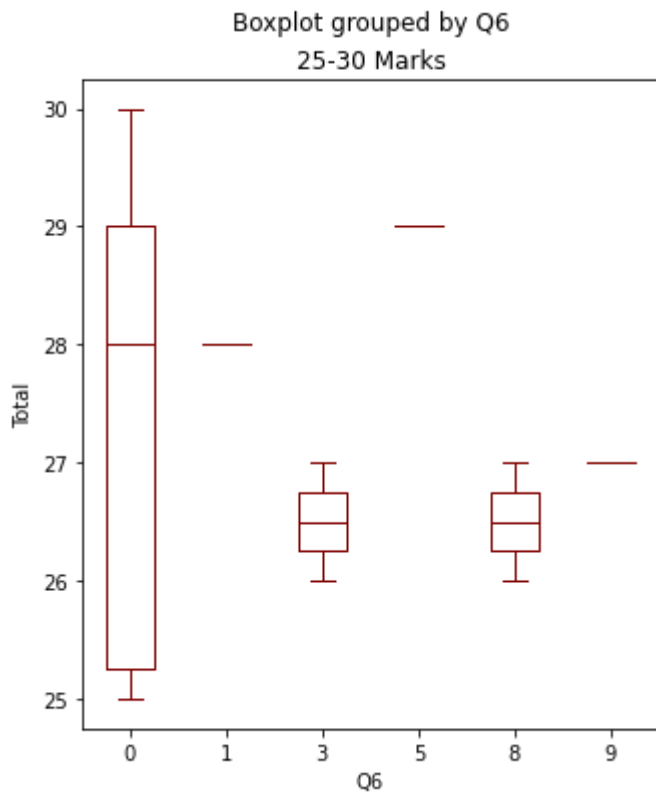
```
c.boxplot(by='Q5M10', column=['TOTAL'], grid=False, color='purple', figsize=[5,6])  
plt.title("25-30 Marks")  
plt.ylabel("Total")  
plt.show()
```



It can be observed that many students scored marks between 4 and 9, the minimum mark remains 0 and none of the students scored full marks in this question.

In [24]:

```
c.boxplot(by='Q6', column=['TOTAL'], grid=False, color='maroon', figsize=[5,6])  
plt.title("25-30 Marks")  
plt.ylabel("Total")  
plt.show()
```



None of the students who attempted the question scored full mark

Many of them didn't attempt this question so minimum mark remains 0.

In [25]:

```
d=DF.loc[(DF['TOTAL']>=30) & (DF['TOTAL']<=35)]
d=d.reset_index()
d
```

Out[25]:

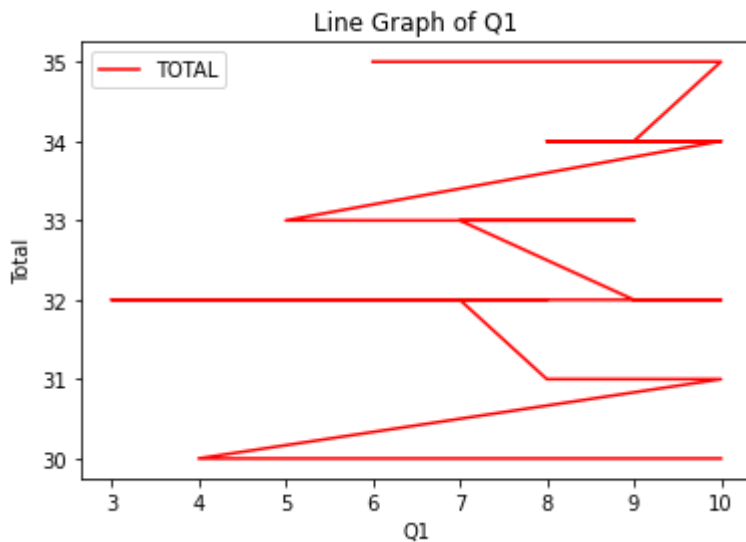
	index	TOTAL	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7	Q5M10
0	35	30	4	6	6	4	0	1	0	0	8
1	14	30	4	6	6	2	4	5	3	0	0
2	16	30	4	0	6	4	5	2	0	0	8
3	80	31	4	6	6	2	2	5	0	0	6
4	66	31	4	5	5	2	5	3	1	5	7
5	37	31	4	4	6	4	0	0	0	0	10
6	1	32	4	3	4	3	0	0	3	6	9
7	15	32	3	0	2	1	5	5	3	7	0
8	27	32	2	6	6	1	5	5	3	3	0
9	13	32	3	3	6	4	3	5	0	0	9
10	32	32	4	6	6	4	2	0	0	0	10
11	81	32	3	6	3	4	5	3	0	0	0
12	24	33	1	6	6	3	5	5	3	3	7
13	2	33	4	5	5	1	5	5	0	0	8
14	43	33	4	5	0	0	0	0	3	4	8
15	78	33	2	3	6	4	5	5	0	0	8
16	21	34	4	6	5	3	5	5	0	3	0
17	45	34	2	6	6	4	5	5	0	0	6
18	7	34	4	6	6	4	0	0	2	0	0
19	12	34	4	4	5	3	2	2	2	1	9
20	58	34	4	5	6	3	0	0	3	0	6
21	70	35	4	6	6	4	5	5	0	0	5
22	56	35	2	6	0	0	0	0	3	7	7
23	9	35	2	4	5	4	5	5	0	0	0



Total marks 30-35 is filtered from the data set

In [26]:

```
d.plot.line(x='Q1',y='TOTAL',color='red')  
plt.title("Line Graph of Q1")  
plt.ylabel("Total")  
plt.show()
```

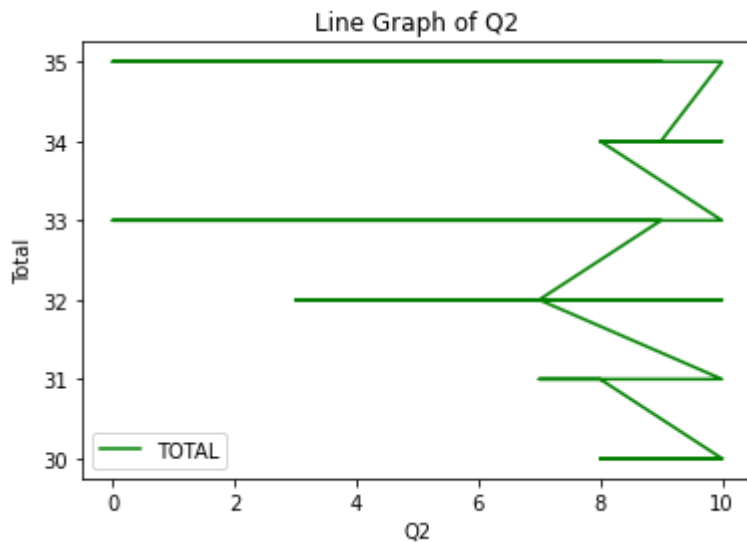


It can be deduced from the above line graph that majority of the students in this range scored marks between 7 and 10.

The maximum mark is 10 and minimum mark is 3 meaning none of the students left the question unattempted.

In [27]:

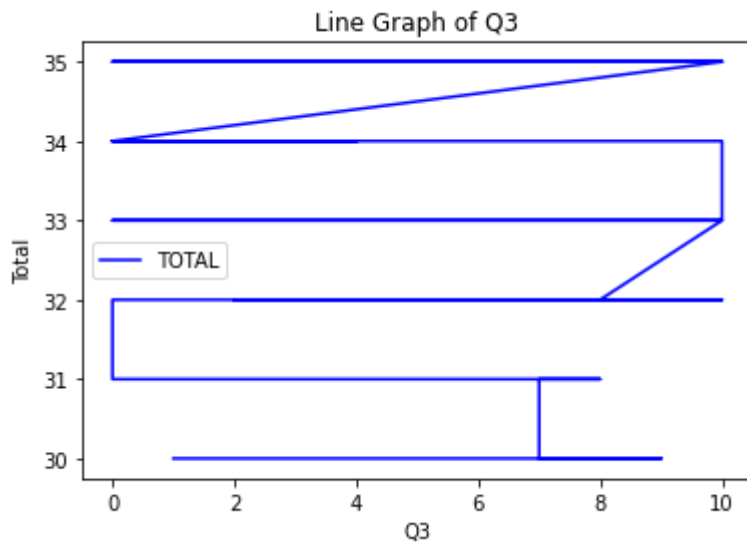
```
d.plot.line(x='Q2',y='TOTAL',color='green')  
plt.title("Line Graph of Q2")  
plt.ylabel("Total")  
plt.show()
```



The students who attempted this question scored marks between 8 and 10 overall, also the minimum mark is zero whereas the maximum mark is 10.

In [28]:

```
d.plot.line(x='Q3',y='TOTAL',color='blue')  
plt.title("Line Graph of Q3")  
plt.ylabel("Total")  
plt.show()
```

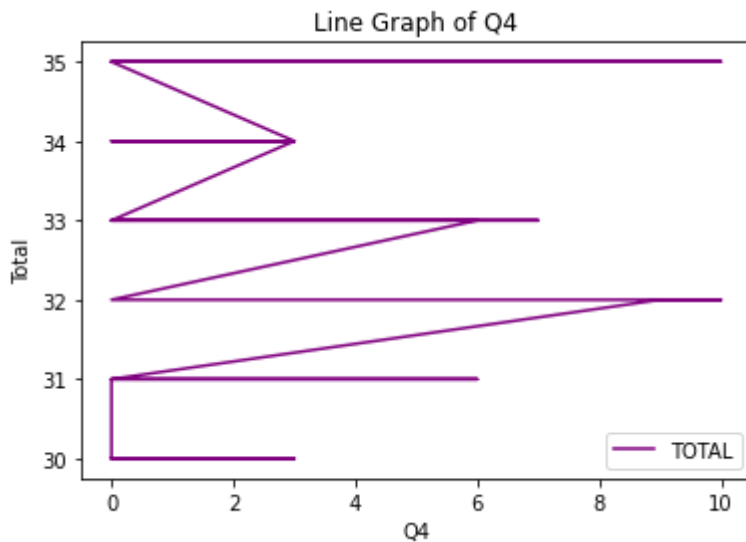


The students in this range scored 6-10 marks on an average

The maximum mark is 10 whereas minimum mark is 0, meaning some students did not attempt this question.

In [29]:

```
d.plot.line(x='Q4',y='TOTAL',color='purple')  
plt.title("Line Graph of Q4")  
plt.ylabel("Total")  
plt.show()
```

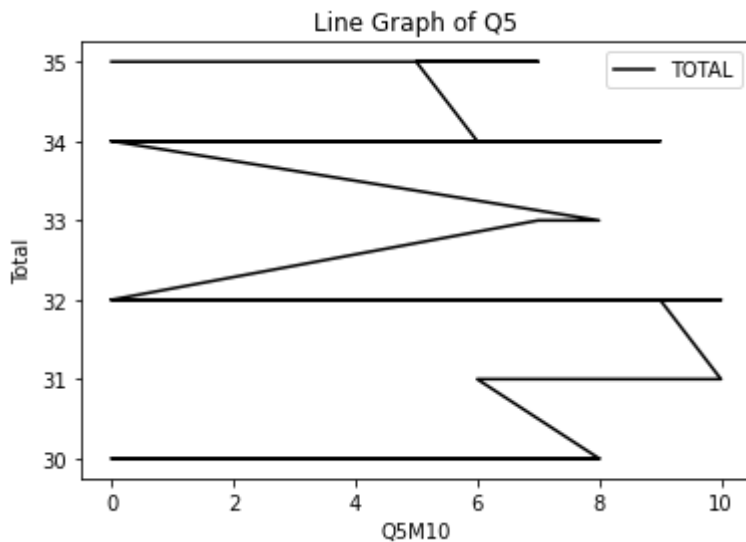


Most of the students who have attempted the question scored marks between 2 and 8 on an average.

The minimum mark for the question is 0 whereas highest mark scored is 10

In [30]:

```
d.plot.line(x='Q5M10',y='TOTAL',color='black')  
plt.title("Line Graph of Q5")  
plt.ylabel("Total")  
plt.show()
```



Majority of the students who attempted this question scored marks between 6 and 9 overall.

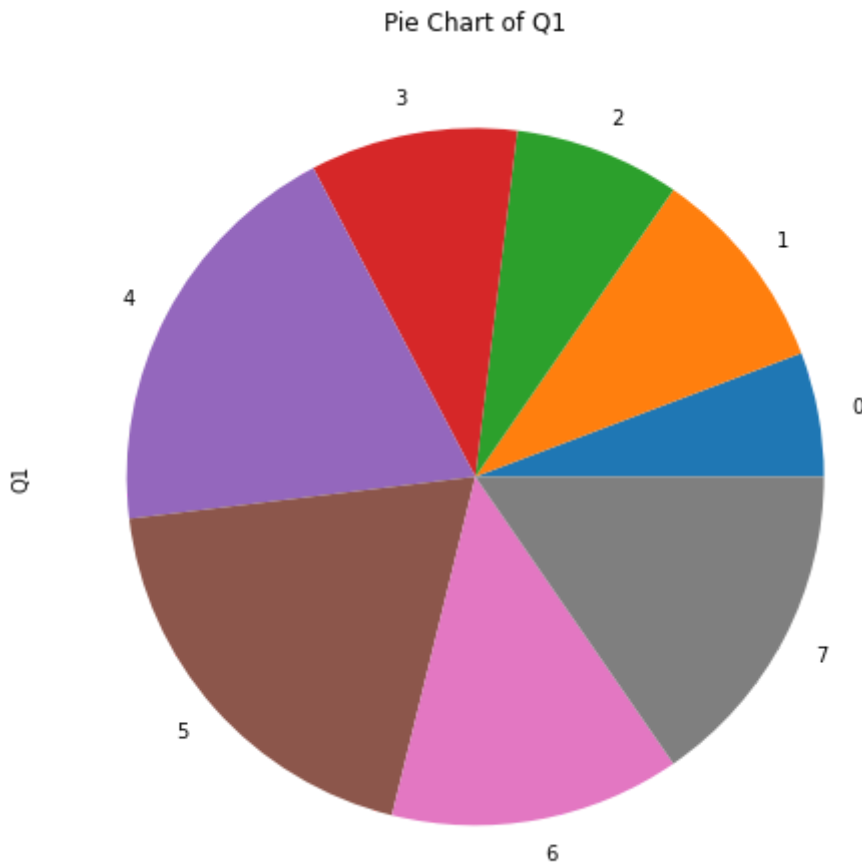
The maximum mark is 10 whereas minimum mark is 0, meaning some students left the question unattempted.

In [31]:

```
a['Q1'].plot(kind='pie',subplots=True,figsize=(8,8))  
plt.title("Pie Chart of Q1")
```

Out[31]:

Text(0.5, 1.0, 'Pie Chart of Q1')



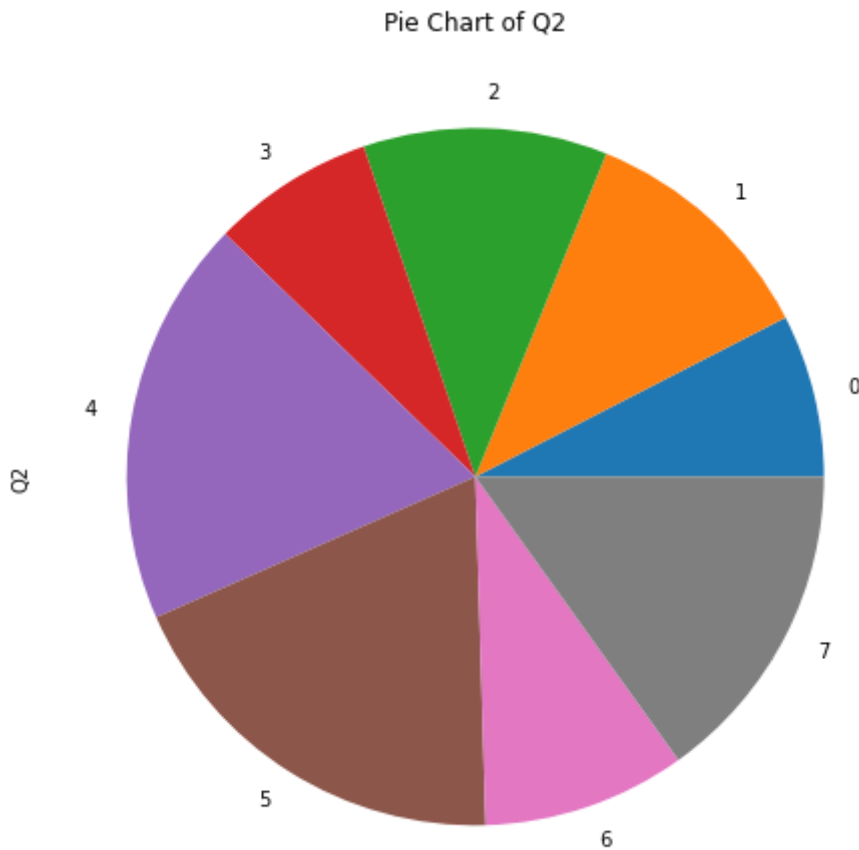
From the above pie chart we can deduce that most of the students who attempted the question scored marks between 4 and 7, the minimum mark is 0

In [32]:

```
a['Q2'].plot(kind='pie',subplots=True,figsize=(8,8))  
plt.title("Pie Chart of Q2")
```

Out[32]:

```
Text(0.5, 1.0, 'Pie Chart of Q2')
```



Majority of the students who attempted the question scored better marks and very few scored 0 or left unattempted

THANK YOU