Case Study

2211cs010276 Group-4

K.Vinay Balaji Reddy

In [46]:
```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("class_marks.csv")
```

In [47]:
```
DF=df.sort_values("Total",ascending=True)
DF
```

Out[47]:

	Total	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7	Q5M1
69	3	1.0	NaN	1.0	NaN	NaN	NaN	1.0	NaN	Na
11	8	2.0	2.0	NaN	3.0	1.0	NaN	NaN	NaN	Na
23	9	4.0	3.0	NaN	NaN	NaN	NaN	NaN	NaN	Na
22	14	4.0	4.0	5.0	2.0	NaN	NaN	NaN	NaN	Na
57	17	3.0	NaN	NaN	4.0	NaN	NaN	3.0	7.0	Na
...	...	...	...	...	...	...	...	...	...	
73	40	4.0	6.0	NaN	NaN	5.0	5.0	3.0	NaN	10.
53	40	4.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN	10.
51	40	0.0	NaN	6.0	4.0	NaN	NaN	3.0	7.0	10.
33	40	NaN	NaN	6.0	4.0	5.0	5.0	3.0	7.0	Na
65	40	4.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN	10.

86 rows × 12 columns

TOTAL values are sorted in ascending order above.

In [49]:
```
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[49]:

	Total	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7	Q5M1
69	3	1.0	NaN	1.0	NaN	NaN	NaN	1.0	NaN	Na
11	8	2.0	2.0	NaN	3.0	1.0	NaN	NaN	NaN	Na
23	9	4.0	3.0	NaN	NaN	NaN	NaN	NaN	NaN	Na
22	14	4.0	4.0	5.0	2.0	NaN	NaN	NaN	NaN	Na
57	17	3.0	NaN	NaN	4.0	NaN	NaN	3.0	7.0	Na
...	...	...	...	...	...	...	...	...	...	
73	40	4.0	6.0	NaN	NaN	5.0	5.0	3.0	NaN	10.
53	40	4.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN	10.
51	40	0.0	NaN	6.0	4.0	NaN	NaN	3.0	7.0	10.
33	40	NaN	NaN	6.0	4.0	5.0	5.0	3.0	7.0	Na
65	40	4.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN	10.

86 rows × 17 columns

New Columns Q1,Q2,Q3,Q4,Q5,Q6 are created above to do analysis.

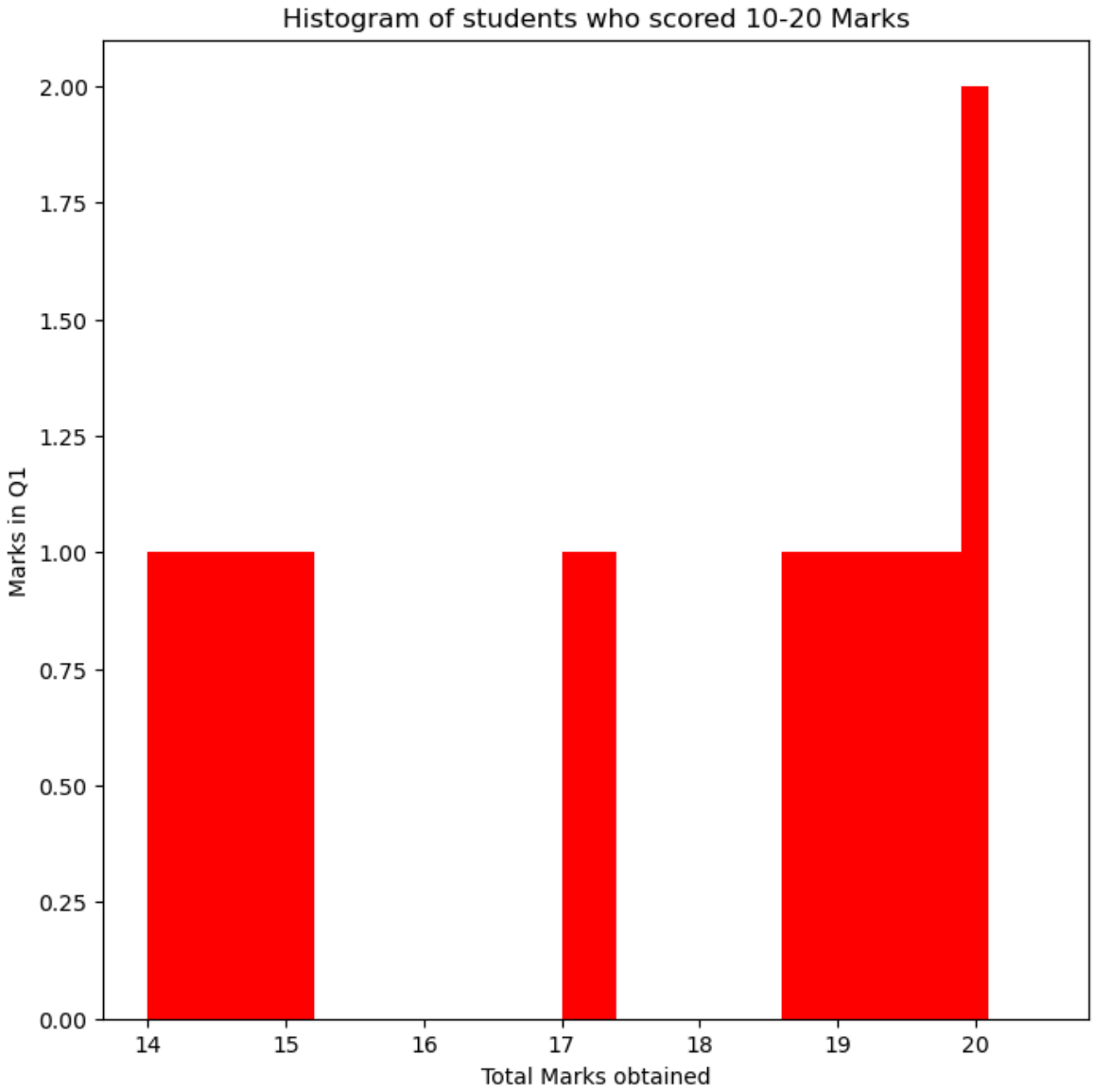
In [51]:
```
a=DF.loc[(DF['Total'] >= 10) & (DF['Total'] <= 20)]
a=a.reset_index()
a
```

Out[51]:

	index	Total	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7
0	22	14	4.0	4.0	5.0	2.0	NaN	NaN	NaN	NaN
1	57	17	3.0	NaN	NaN	4.0	NaN	NaN	3.0	7.0
2	76	17	2.0	3.0	4.0	2.0	4.0	2.0	NaN	NaN
3	63	18	4.0	NaN	4.0	2.0	NaN	NaN	NaN	NaN
4	34	19	2.0	3.0	3.0	1.0	2.0	3.0	NaN	NaN
5	68	20	4.0	6.0	6.0	4.0	NaN	NaN	NaN	NaN
6	5	20	4.0	6.0	6.0	4.0	NaN	NaN	NaN	NaN
7	60	20	2.0	5.0	3.0	2.0	NaN	NaN	NaN	NaN
8	30	20	4.0	4.0	4.0	4.0	5.0	NaN	NaN	NaN

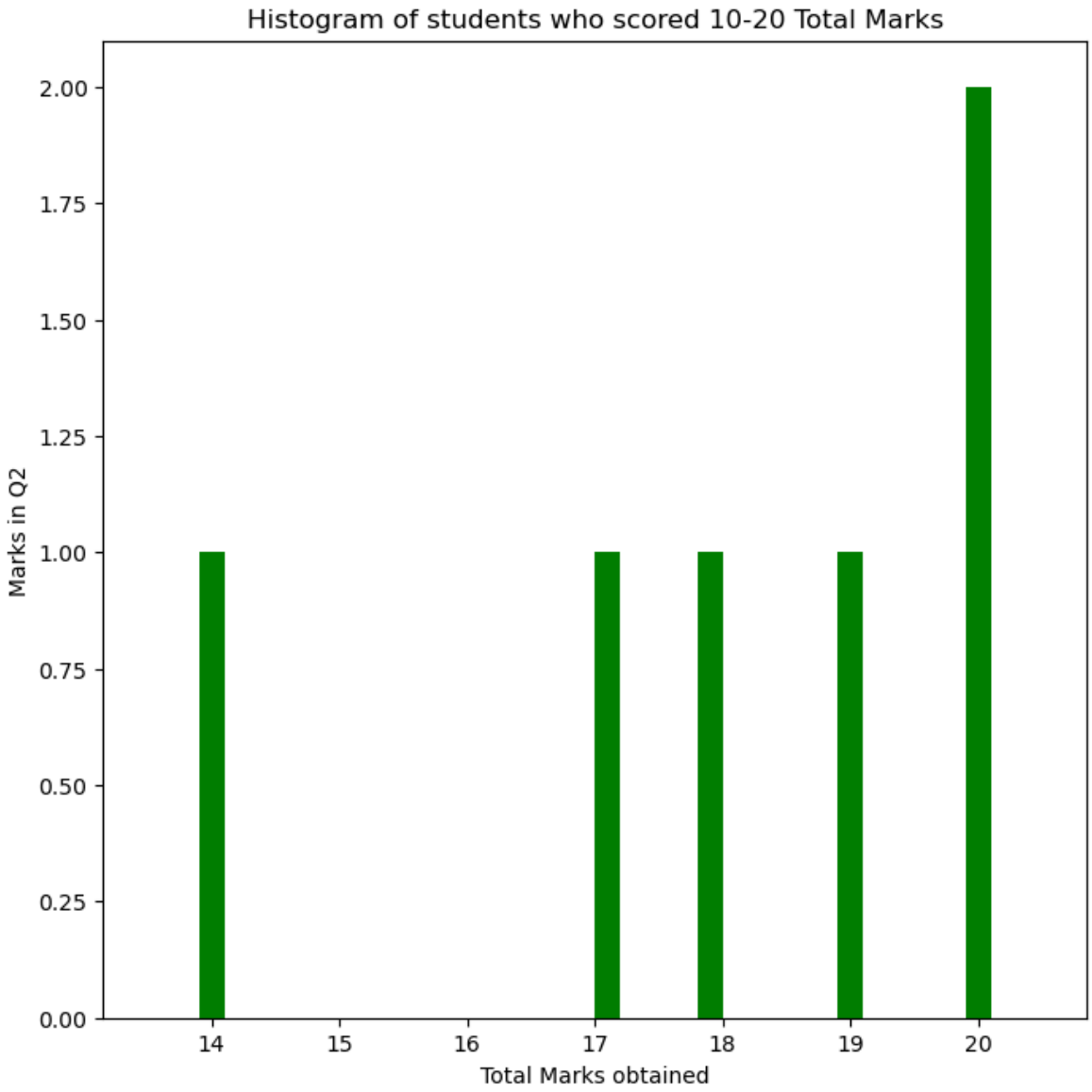
Total marks 10-20 is filtered above

In [53]:
```
a1=a.groupby('Q1')['Total']
a1.hist(color = 'red',figsize=[8,8],grid=False,bins=5)
plt.title("Histogram of students who scored 10-20 Marks")
plt.xlabel("Total Marks obtained")
plt.ylabel("Marks in Q1")
plt.show()
```



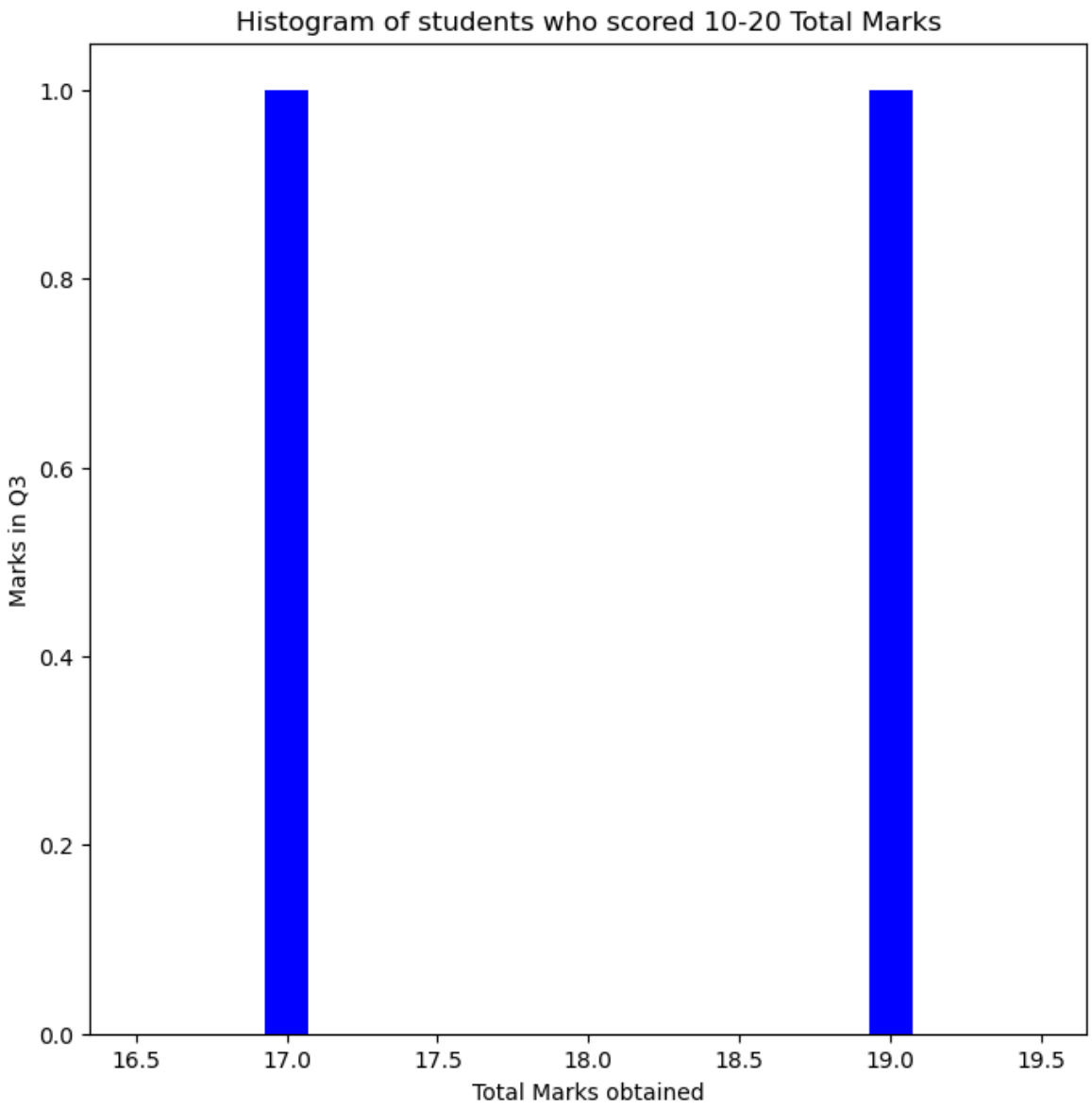
Most of the students scored 1 mark and maximum mark is 2, implying that all the students in this range secured very less marks in Q1

In [55]:
```
a2=a.groupby('Q2')['Total']
a2.hist(color='green',figsize=[8,8],grid=False,bins=5)
plt.title("Histogram of students who scored 10-20 Total Marks")
plt.xlabel("Total Marks obtained")
plt.ylabel("Marks in Q2")
plt.show()
```



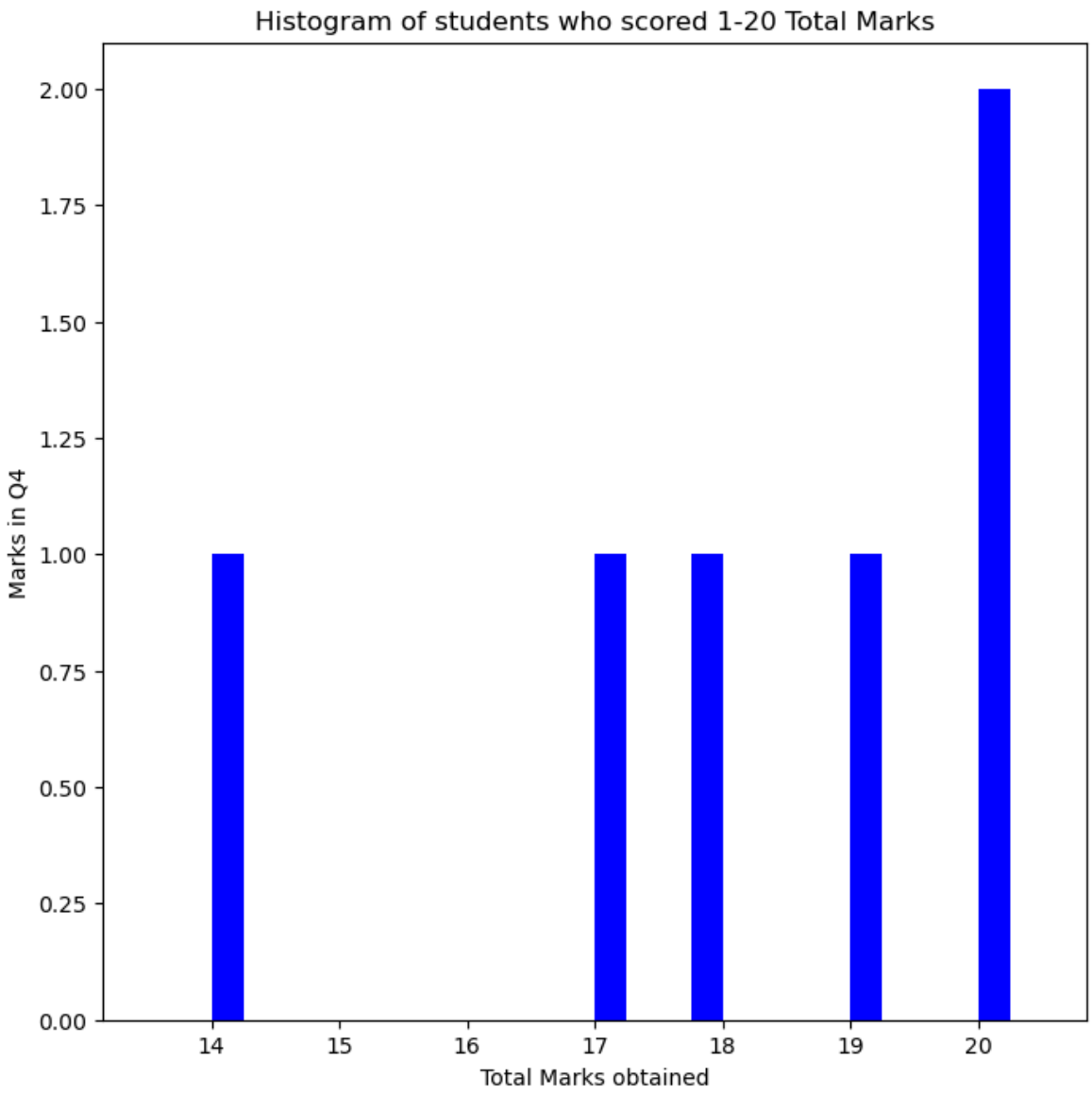
Very few students got high marks in Q2 and that's only 2, so overall performance in Q2 is not upto the mark

In [57]:
```
a3=a.groupby('Q3')['Total']
a3.hist(color='blue',figsize=[8,8],grid=False,bins=7)
plt.title("Histogram of students who scored 10-20 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

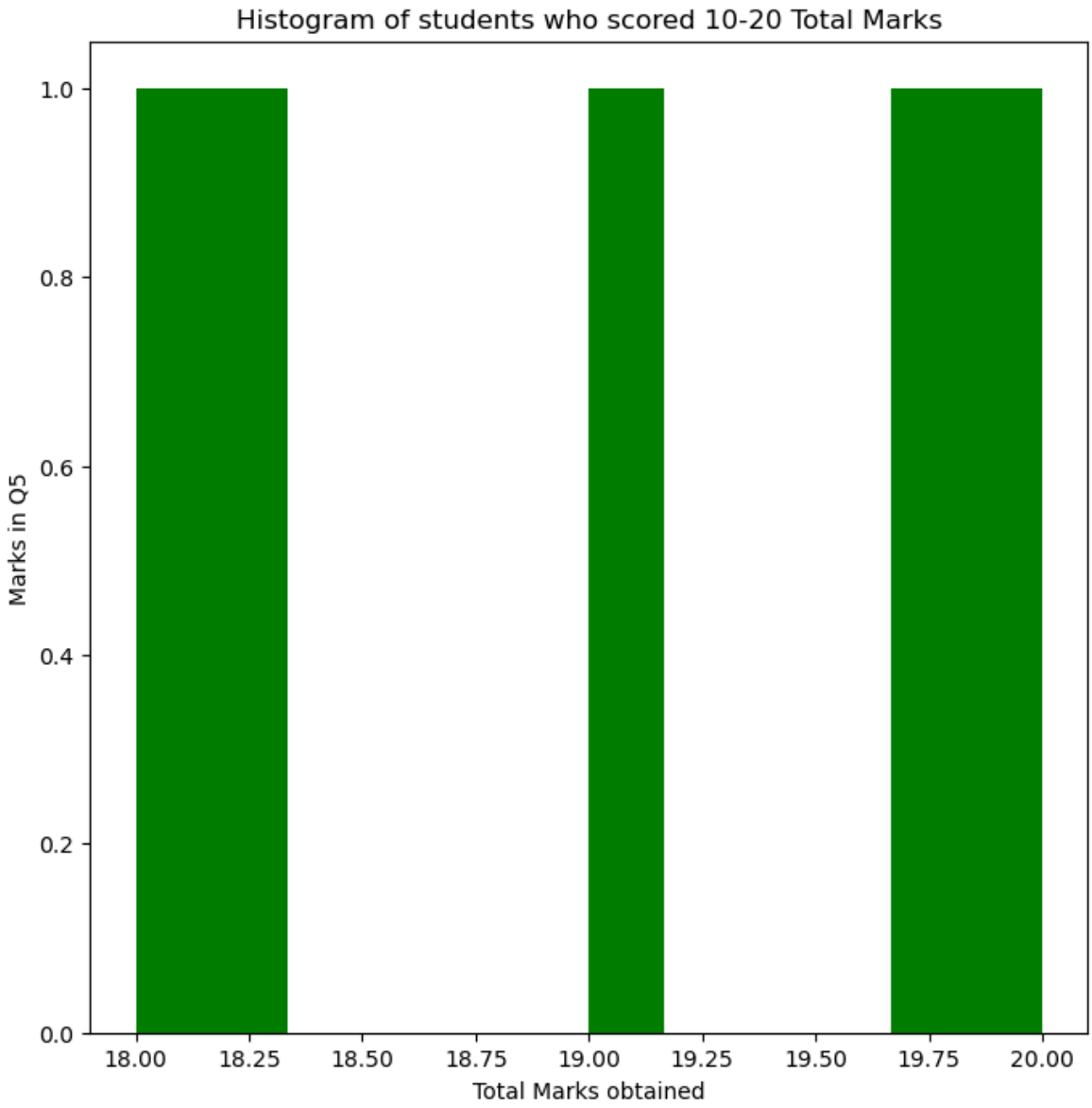
In [59]:
```
a4=a.groupby('Q4')['Total']
a2.hist(color='blue',figsize=[8,8],grid=False,bins=4)
plt.title("Histogram of students who scored 1-20 Total Marks")
plt.xlabel("Total Marks obtained")
plt.ylabel("Marks in Q4")
plt.show()
```



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

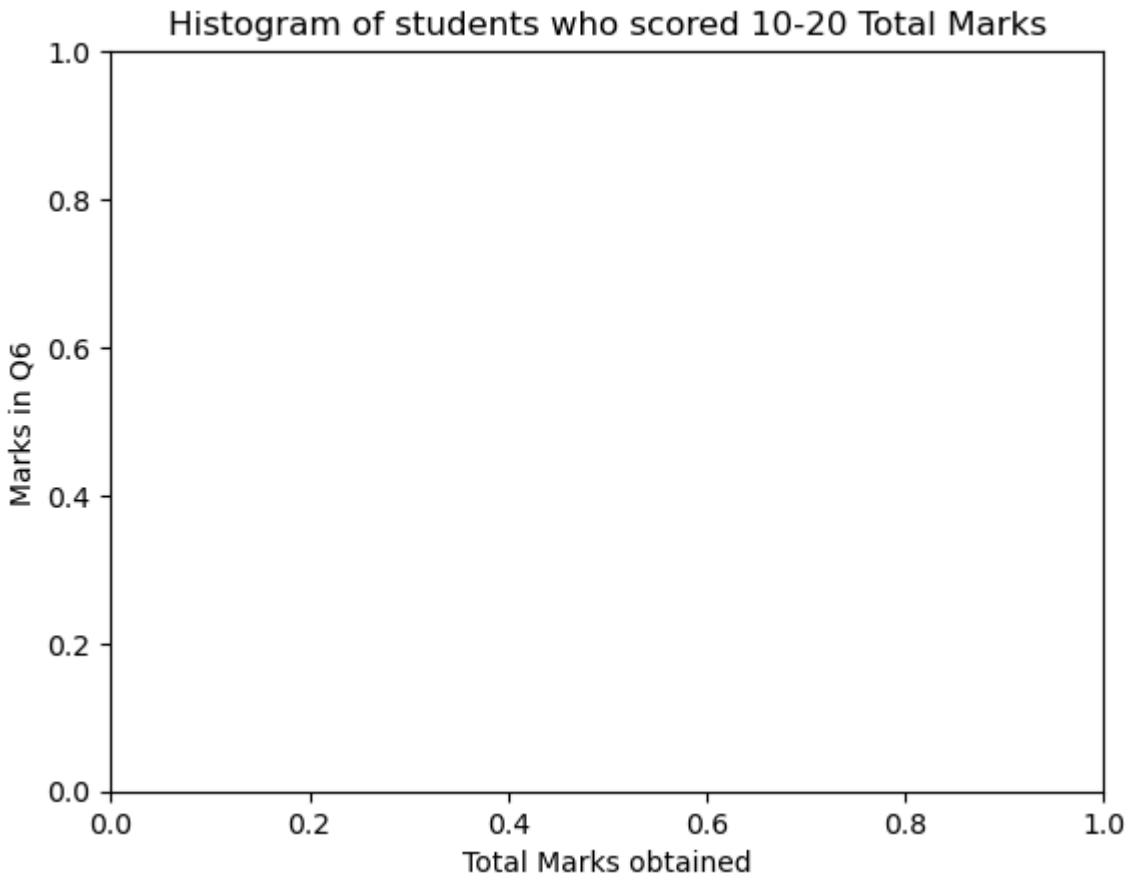
In [148…

```
a5=a.groupby('Q5M10')['Total']
a5.hist(color='green',figsize=[8,8],grid=False,bins=6)
plt.title("Histogram of students who scored 10-20 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 [63]:
```
a6=a.groupby('Q6')['Total']
a6.hist(color='red',figsize=[8,8],grid=False,bins=8)
plt.title("Histogram of students who scored 10-20 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 [65]:
```
b=DF.loc[(DF['Total'] >= 20) & (DF['Total'] <= 30)]
b=b.reset_index()
b
```

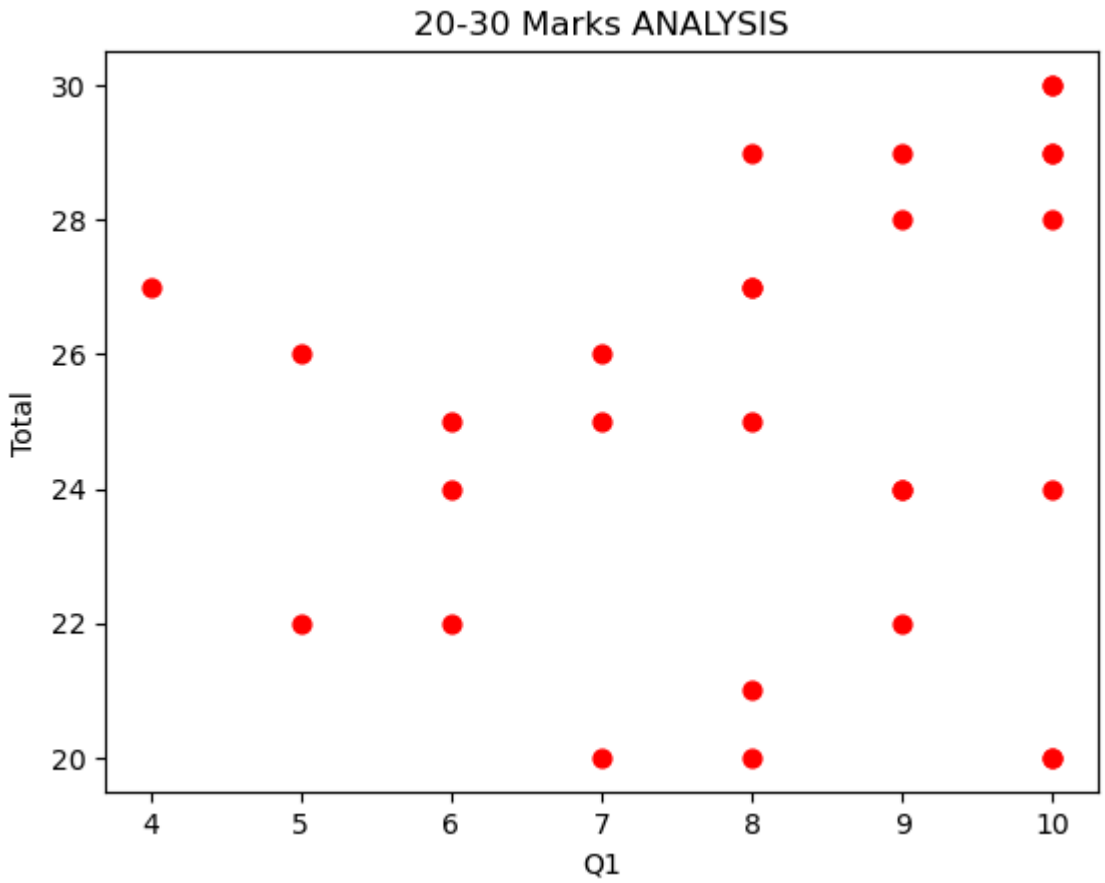
Out[65]:

	index	Total	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7
0	68	20	4.0	6.0	6.0	4.0	NaN	NaN	NaN	NaN
1	5	20	4.0	6.0	6.0	4.0	NaN	NaN	NaN	NaN
2	60	20	2.0	5.0	3.0	2.0	NaN	NaN	NaN	NaN
3	30	20	4.0	4.0	4.0	4.0	5.0	NaN	NaN	NaN
4	54	21	2.0	6.0	NaN	NaN	5.0	5.0	3.0	NaN
5	75	21	4.0	NaN	6.0	1.0	1.0	1.0	NaN	NaN
6	77	22	4.0	5.0	NaN	NaN	3.0	2.0	2.0	NaN
7	39	22	3.0	2.0	NaN	1.0	3.0	3.0	NaN	NaN
8	25	22	4.0	2.0	5.0	2.0	4.0	3.0	2.0	NaN
9	61	24	4.0	5.0	6.0	4.0	NaN	5.0	NaN	NaN
10	18	24	3.0	3.0	5.0	3.0	NaN	NaN	2.0	1.0
11	50	24	4.0	5.0	NaN	NaN	5.0	5.0	NaN	NaN
12	3	24	4.0	6.0	6.0	3.0	2.0	2.0	NaN	NaN
13	72	25	2.0	NaN	4.0	4.0	NaN	5.0	3.0	7.0
14	48	25	2.0	6.0	NaN	NaN	NaN	NaN	3.0	6.0
15	6	25	3.0	4.0	NaN	2.0	5.0	5.0	NaN	NaN
16	74	25	1.0	5.0	6.0	4.0	NaN	NaN	NaN	NaN
17	41	26	2.0	3.0	4.0	3.0	4.0	3.0	NaN	3.0
18	19	26	3.0	NaN	6.0	4.0	NaN	2.0	2.0	1.0
19	31	26	3.0	4.0	6.0	2.0	2.0	NaN	1.0	NaN
20	8	27	3.0	5.0	5.0	NaN	NaN	NaN	NaN	NaN
21	79	27	2.0	6.0	NaN	3.0	2.0	5.0	NaN	NaN
22	29	27	4.0	NaN	6.0	1.0	NaN	NaN	NaN	7.0
23	82	27	2.0	2.0	5.0	3.0	NaN	NaN	NaN	NaN
24	38	28	4.0	5.0	6.0	4.0	5.0	3.0	1.0	NaN
25	67	28	4.0	6.0	4.0	4.0	NaN	NaN	NaN	NaN
26	84	28	4.0	NaN	5.0	4.0	5.0	4.0	NaN	NaN
27	40	29	4.0	6.0	6.0	4.0	NaN	NaN	1.0	1.0
28	20	29	2.0	6.0	2.0	2.0	5.0	5.0	NaN	NaN
29	52	29	4.0	5.0	4.0	3.0	NaN	NaN	3.0	6.0
30	85	29	4.0	6.0	NaN	NaN	NaN	NaN	3.0	5.0
31	35	30	4.0	6.0	6.0	4.0	NaN	1.0	NaN	NaN
32	14	30	4.0	6.0	6.0	2.0	4.0	5.0	3.0	NaN

	index	Total	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7
33	16	30	4.0	NaN	6.0	4.0	5.0	2.0	NaN	NaN

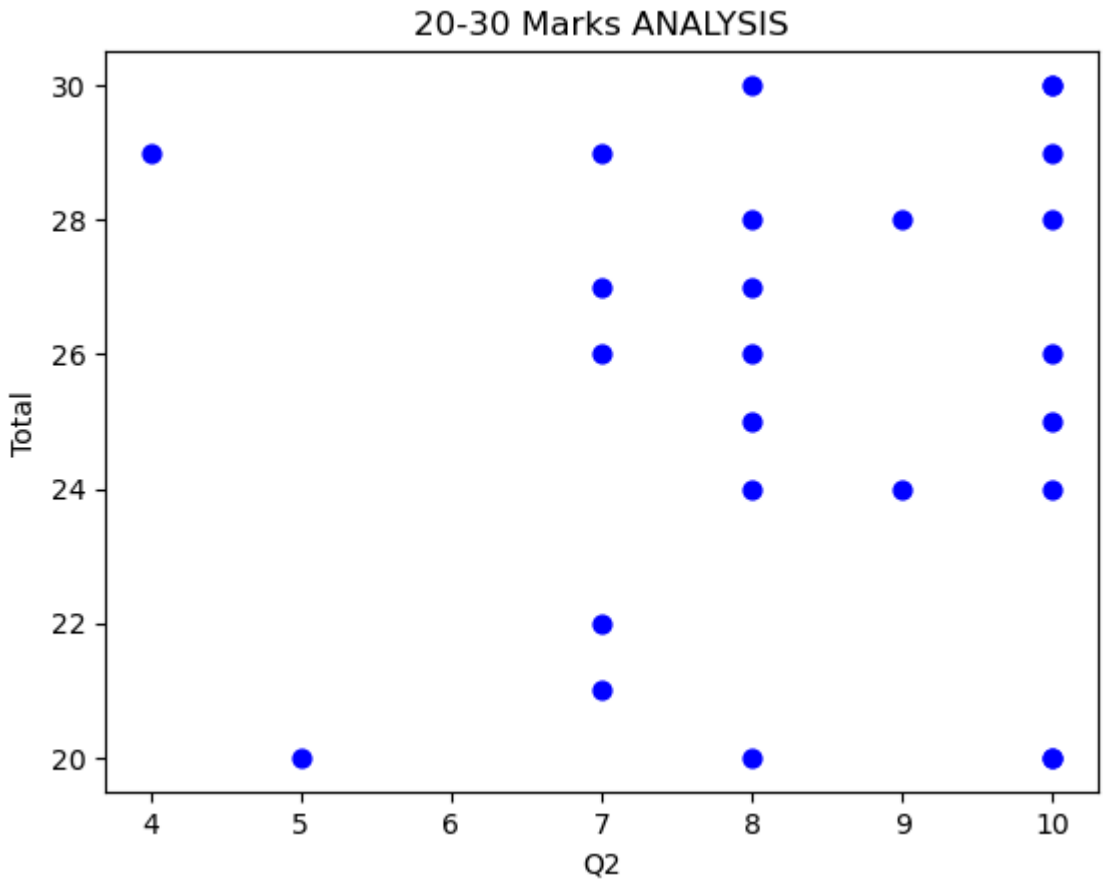
Total marks 20-30 is filtered from the data set

In [67]:
```
b.plot.scatter(x='Q1',y='Total',color='red',s=40)
plt.title("20-30 Marks ANALYSIS")
plt.show()
```



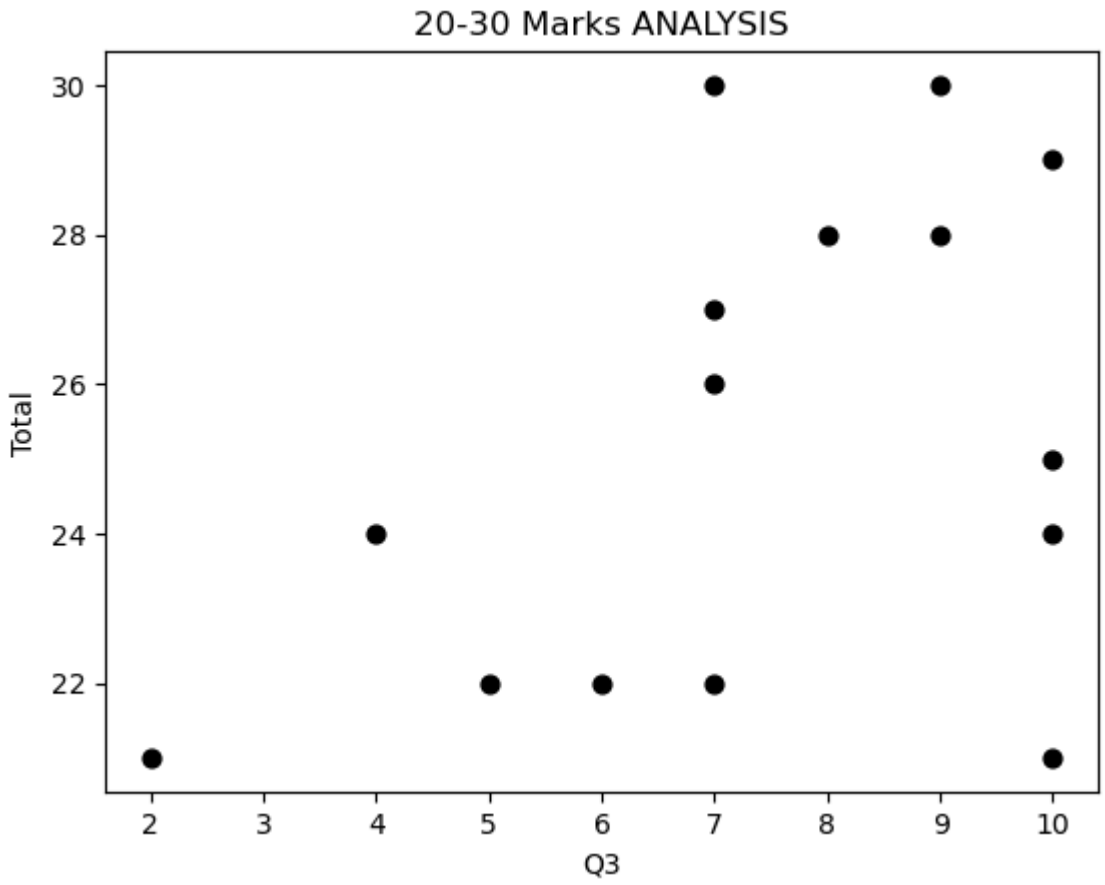
Majority of the students in this range scored marks between 6-10 in Q1 and very few, just 3 students scored below 5 marks, maximum mark scored is 10 by two students.

In [152…
```
b.plot.scatter(x='Q2',y='Total',color='blue',s=40)
plt.title("20-30 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.

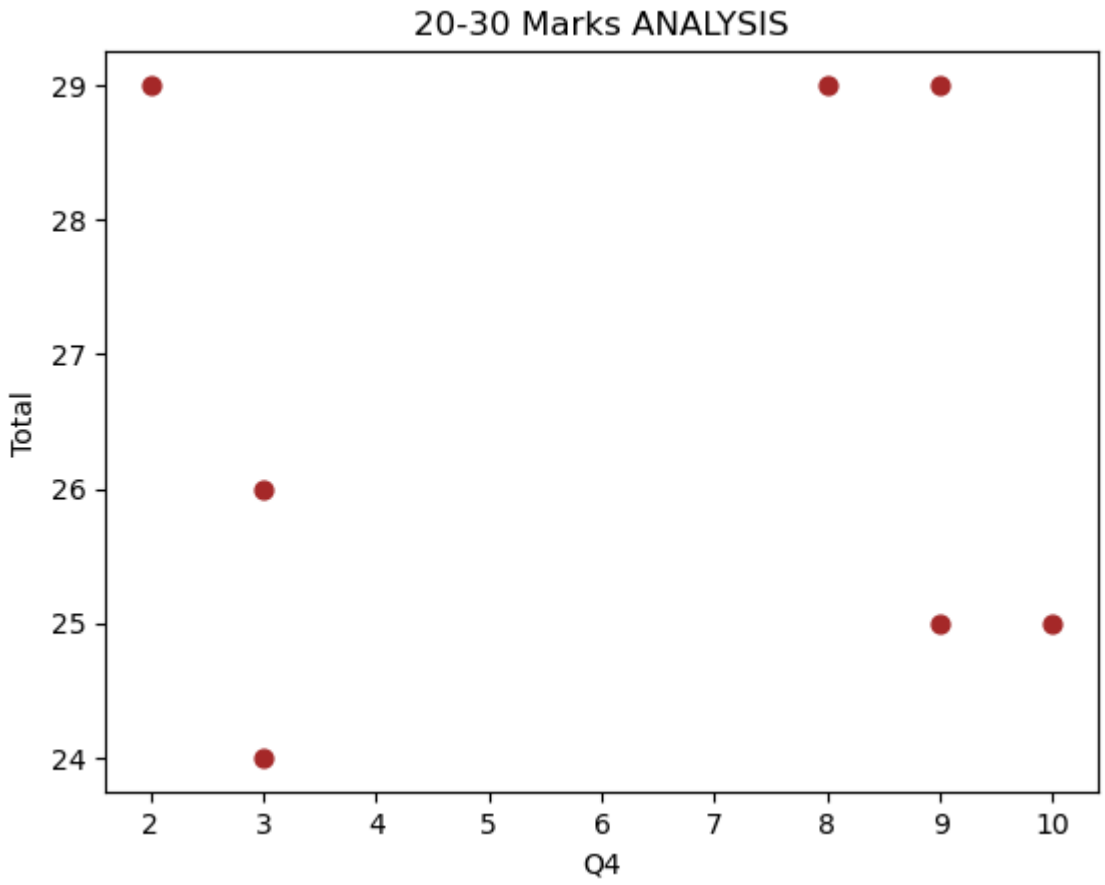
In [71]:
```
b.plot.scatter(x='Q3',y='Total',color='black',s=40)
plt.title("20-30 Marks ANALYSIS")
plt.show()
```



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 [154…

```
b.plot.scatter(x='Q4',y='Total',color='brown',s=40)
plt.title("20-30 Marks ANALYSIS")
plt.show()
```

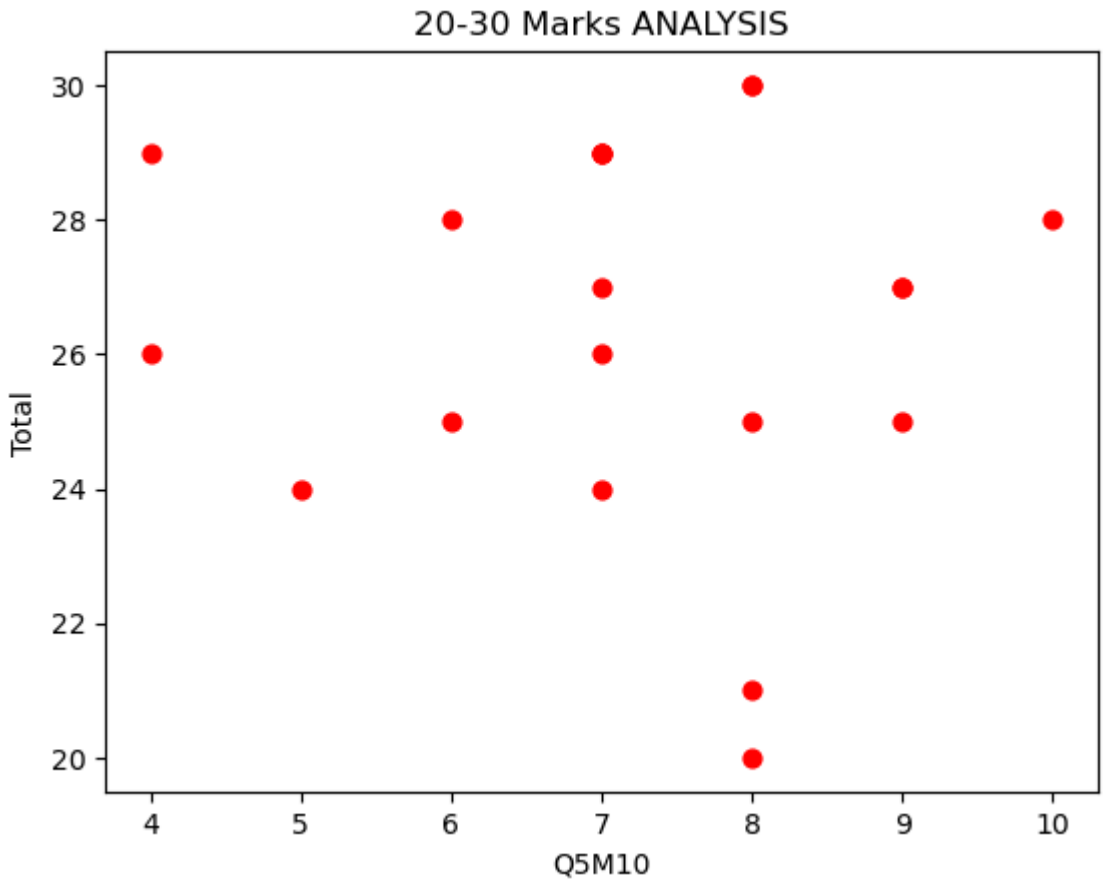


Most of the students in this range haven't attempted

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

In [156…

```
b.plot.scatter(x='Q5M10',y='Total',color='red',s=40)
plt.title("20-30 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 unattempted

In [78]:
```
c=DF.loc[(DF['Total'] >= 25) & (DF['Total'] <= 35)]
c=c.reset_index()
c
```

Out[78]:

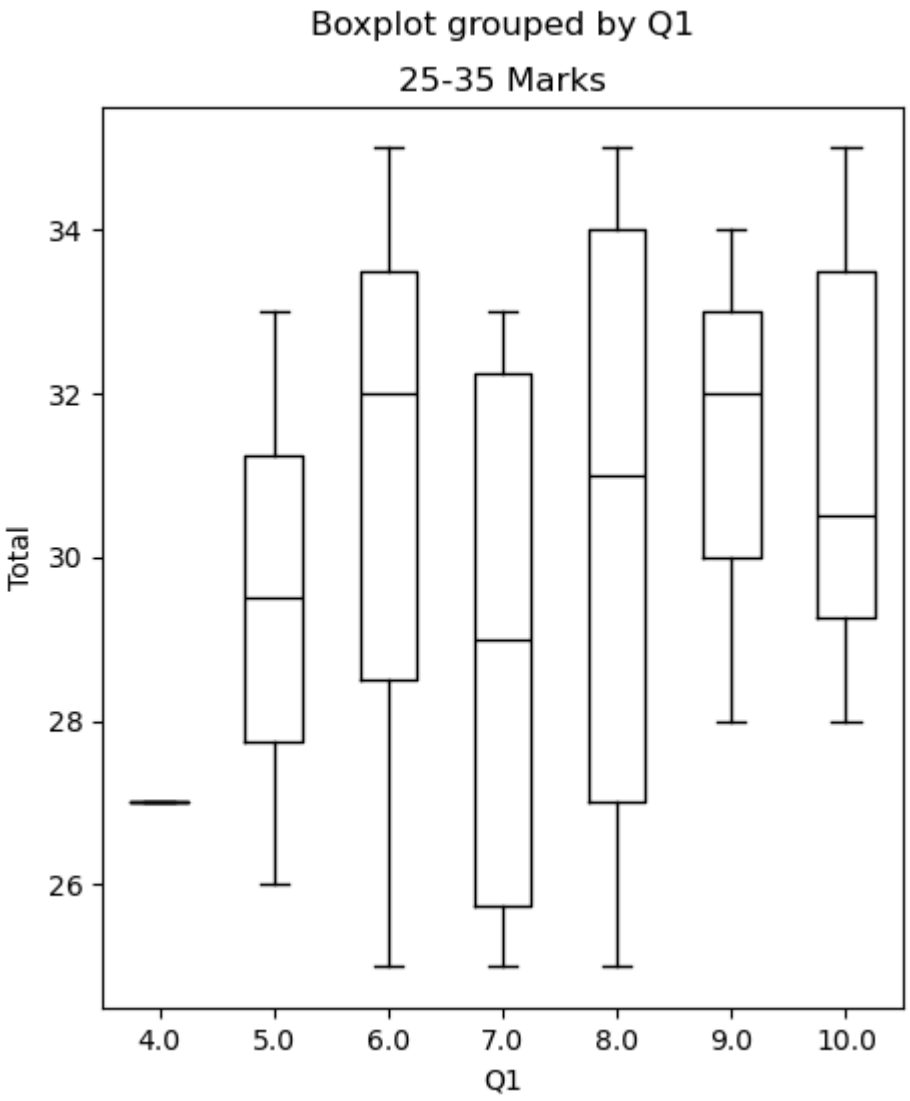
	index	Total	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7
0	72	25	2.0	NaN	4.0	4.0	NaN	5.0	3.0	7.0
1	48	25	2.0	6.0	NaN	NaN	NaN	NaN	3.0	6.0
2	6	25	3.0	4.0	NaN	2.0	5.0	5.0	NaN	NaN
3	74	25	1.0	5.0	6.0	4.0	NaN	NaN	NaN	NaN
4	41	26	2.0	3.0	4.0	3.0	4.0	3.0	NaN	3.0
5	19	26	3.0	NaN	6.0	4.0	NaN	2.0	2.0	1.0
6	31	26	3.0	4.0	6.0	2.0	2.0	NaN	1.0	NaN
7	8	27	3.0	5.0	5.0	NaN	NaN	NaN	NaN	NaN
8	79	27	2.0	6.0	NaN	3.0	2.0	5.0	NaN	NaN
9	29	27	4.0	NaN	6.0	1.0	NaN	NaN	NaN	7.0
10	82	27	2.0	2.0	5.0	3.0	NaN	NaN	NaN	NaN
11	38	28	4.0	5.0	6.0	4.0	5.0	3.0	1.0	NaN
12	67	28	4.0	6.0	4.0	4.0	NaN	NaN	NaN	NaN
13	84	28	4.0	NaN	5.0	4.0	5.0	4.0	NaN	NaN
14	40	29	4.0	6.0	6.0	4.0	NaN	NaN	1.0	1.0
15	20	29	2.0	6.0	2.0	2.0	5.0	5.0	NaN	NaN
16	52	29	4.0	5.0	4.0	3.0	NaN	NaN	3.0	6.0
17	85	29	4.0	6.0	NaN	NaN	NaN	NaN	3.0	5.0
18	35	30	4.0	6.0	6.0	4.0	NaN	1.0	NaN	NaN
19	14	30	4.0	6.0	6.0	2.0	4.0	5.0	3.0	NaN
20	16	30	4.0	NaN	6.0	4.0	5.0	2.0	NaN	NaN
21	80	31	4.0	6.0	6.0	2.0	2.0	5.0	NaN	NaN
22	66	31	4.0	5.0	5.0	2.0	5.0	3.0	1.0	5.0
23	37	31	4.0	4.0	6.0	4.0	NaN	NaN	NaN	NaN
24	1	32	4.0	3.0	4.0	3.0	NaN	NaN	3.0	6.0
25	15	32	3.0	NaN	2.0	1.0	5.0	5.0	3.0	7.0
26	27	32	2.0	6.0	6.0	1.0	5.0	5.0	3.0	3.0
27	13	32	3.0	3.0	6.0	4.0	3.0	5.0	NaN	NaN
28	32	32	4.0	6.0	6.0	4.0	2.0	NaN	NaN	NaN
29	81	32	3.0	6.0	3.0	4.0	5.0	3.0	NaN	NaN
30	24	33	1.0	6.0	6.0	3.0	5.0	5.0	3.0	3.0
31	2	33	4.0	5.0	5.0	1.0	5.0	5.0	NaN	NaN
32	43	33	4.0	5.0	NaN	NaN	NaN	NaN	3.0	4.0

	index	Total	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7
33	78	33	2.0	3.0	6.0	4.0	5.0	5.0	NaN	NaN
34	21	34	4.0	6.0	5.0	3.0	5.0	5.0	NaN	3.0
35	45	34	2.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN
36	7	34	4.0	6.0	6.0	4.0	NaN	NaN	2.0	NaN
37	12	34	4.0	4.0	5.0	3.0	2.0	2.0	2.0	1.0
38	58	34	4.0	5.0	6.0	3.0	NaN	NaN	3.0	NaN
39	70	35	4.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN
40	56	35	2.0	6.0	NaN	NaN	NaN	NaN	3.0	7.0
41	9	35	2.0	4.0	5.0	4.0	5.0	5.0	NaN	NaN

Total marks 25-35 is filtered from the data set

In [158…

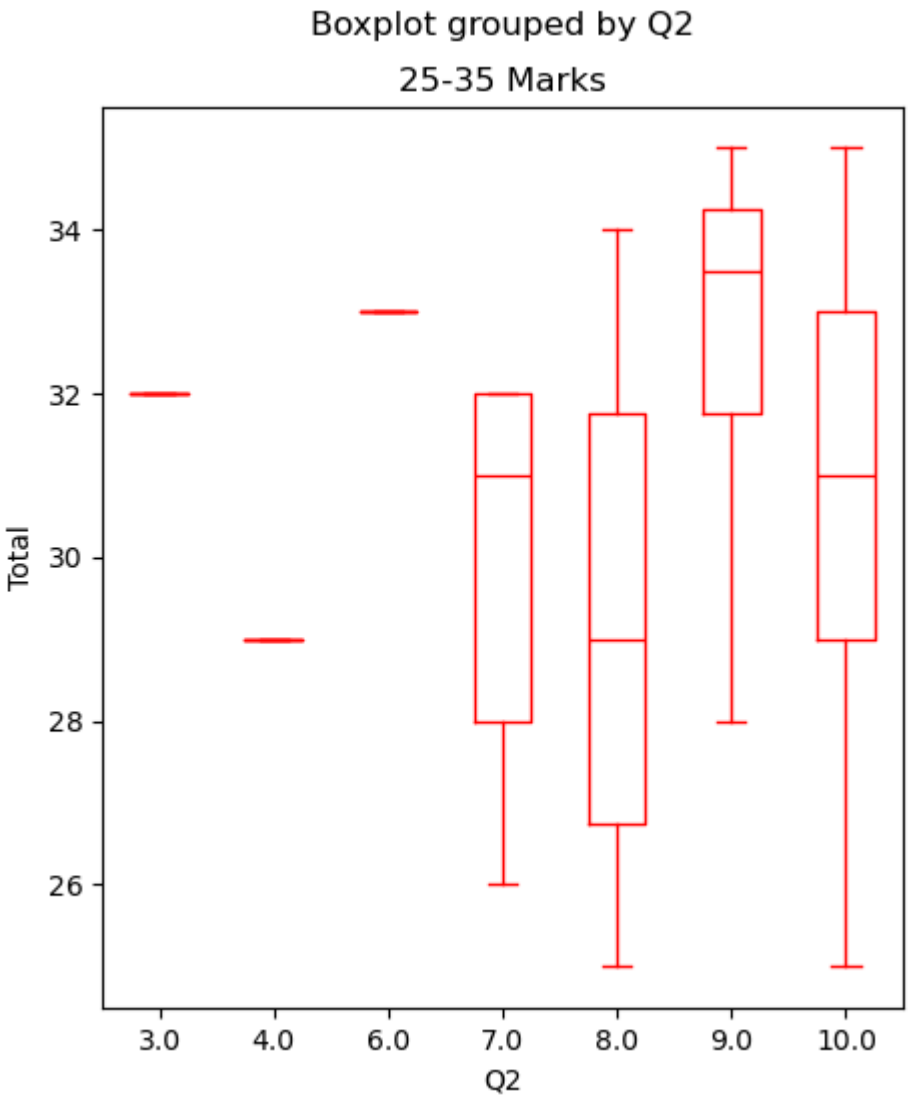
```
c.boxplot(by='Q1', column =['Total'], grid = False,color='black',figsize=[5,6])
plt.title("25-35 Marks")
plt.ylabel("Total")
plt.show()
```



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

In [160…

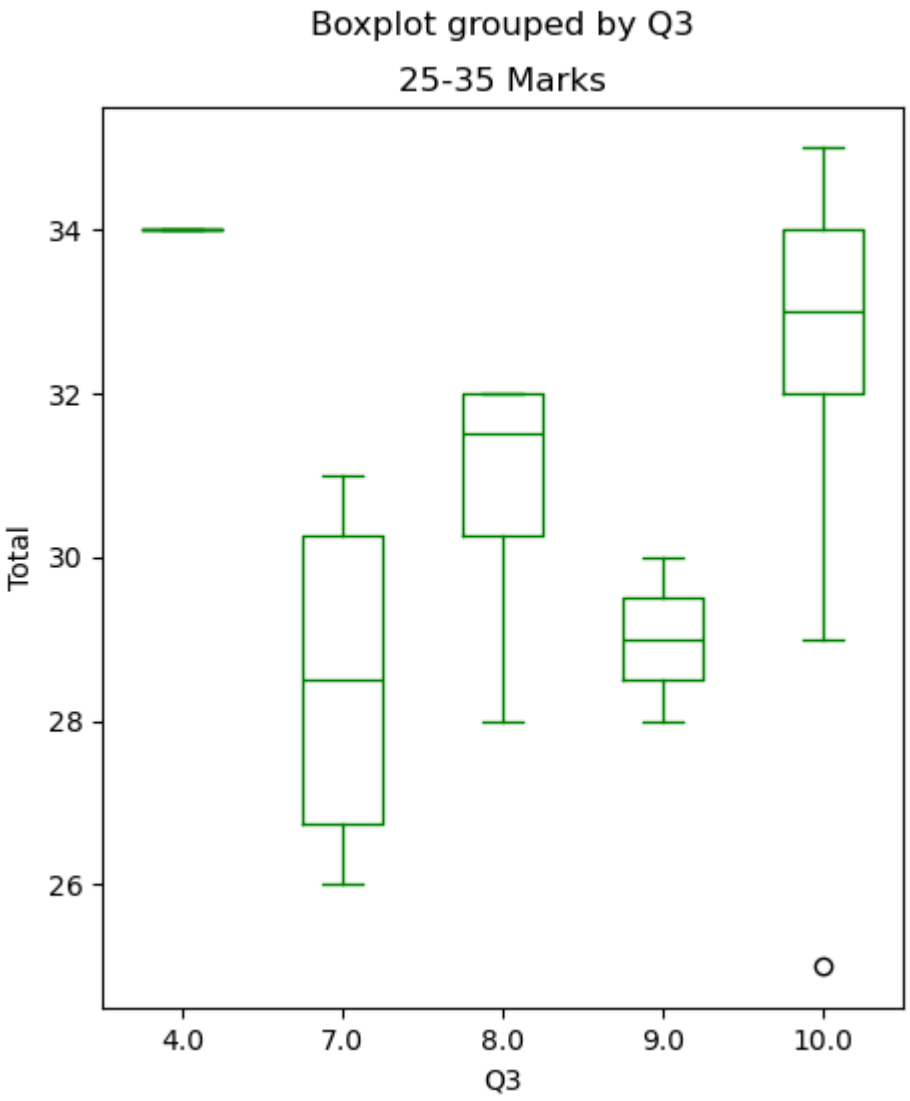
```
c.boxplot(by='Q2', column =['Total'], grid = False,color='red',figsize=[5,6])
plt.title("25-35 Marks")
plt.ylabel("Total")
plt.show()
```



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.

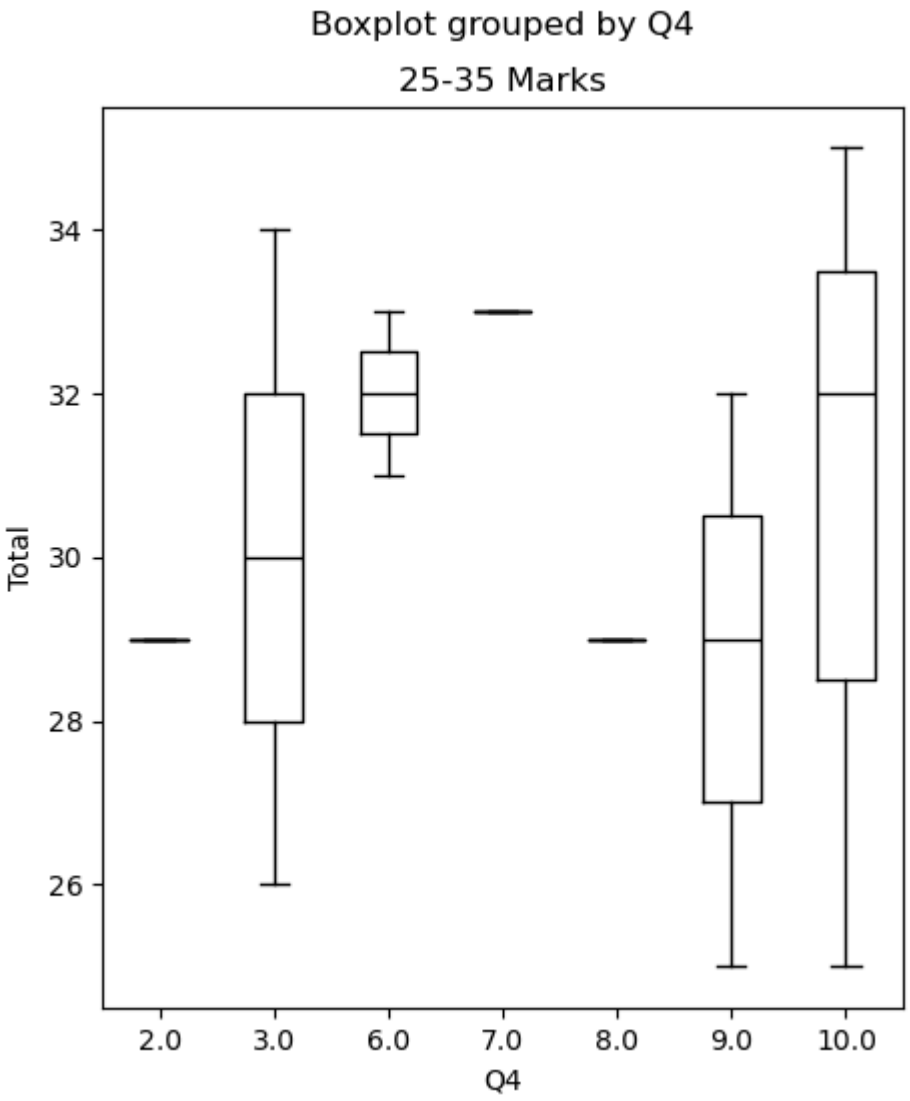
In [85]:
```
c.boxplot(by='Q3', column =['Total'], grid = False,color='green',figsize=[5,6])
plt.title("25-35 Marks")
plt.ylabel("Total")
plt.show()
```



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

Most of the students have secured marks between 7 and 9

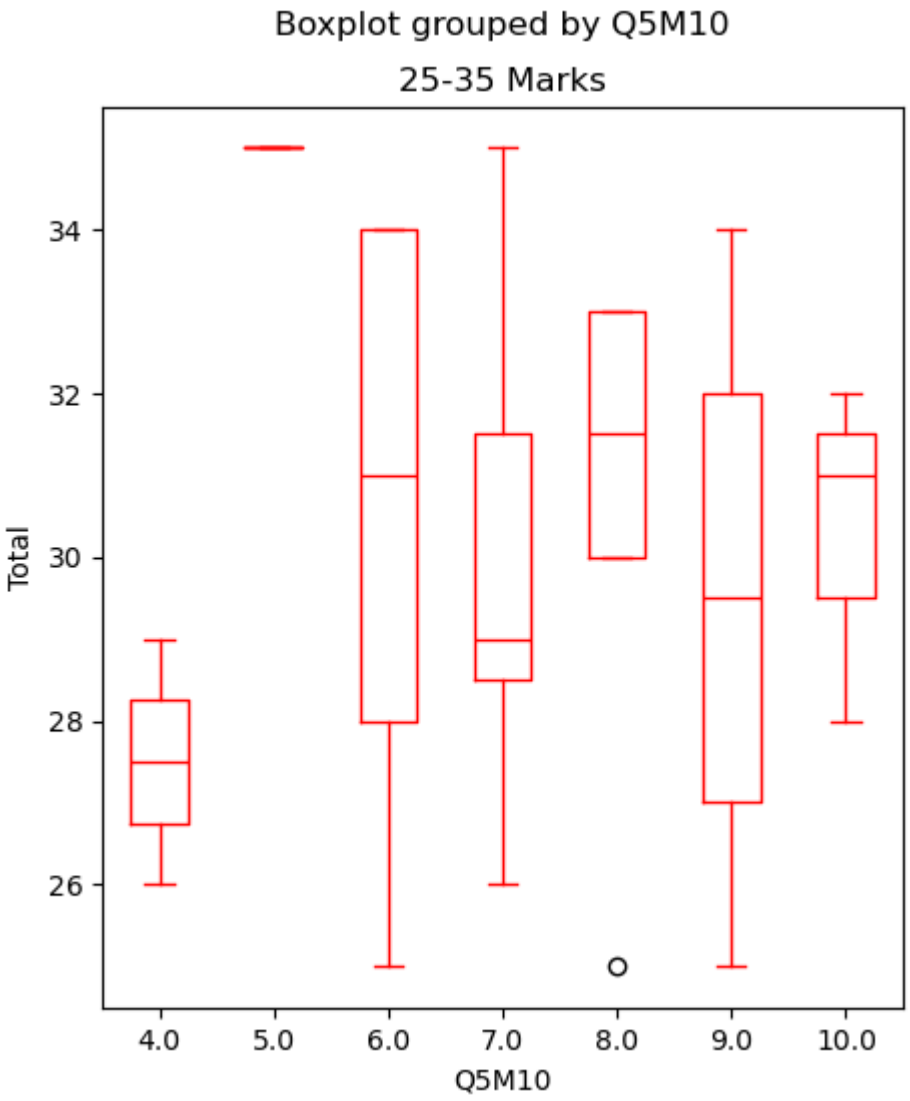
In [88]:
```
c.boxplot(by='Q4', column =['Total'], grid = False,color='black',figsize=[5,6])
plt.title("25-35 Marks")
plt.ylabel("Total")
plt.show()
```



No student in this range secured full mark for this question

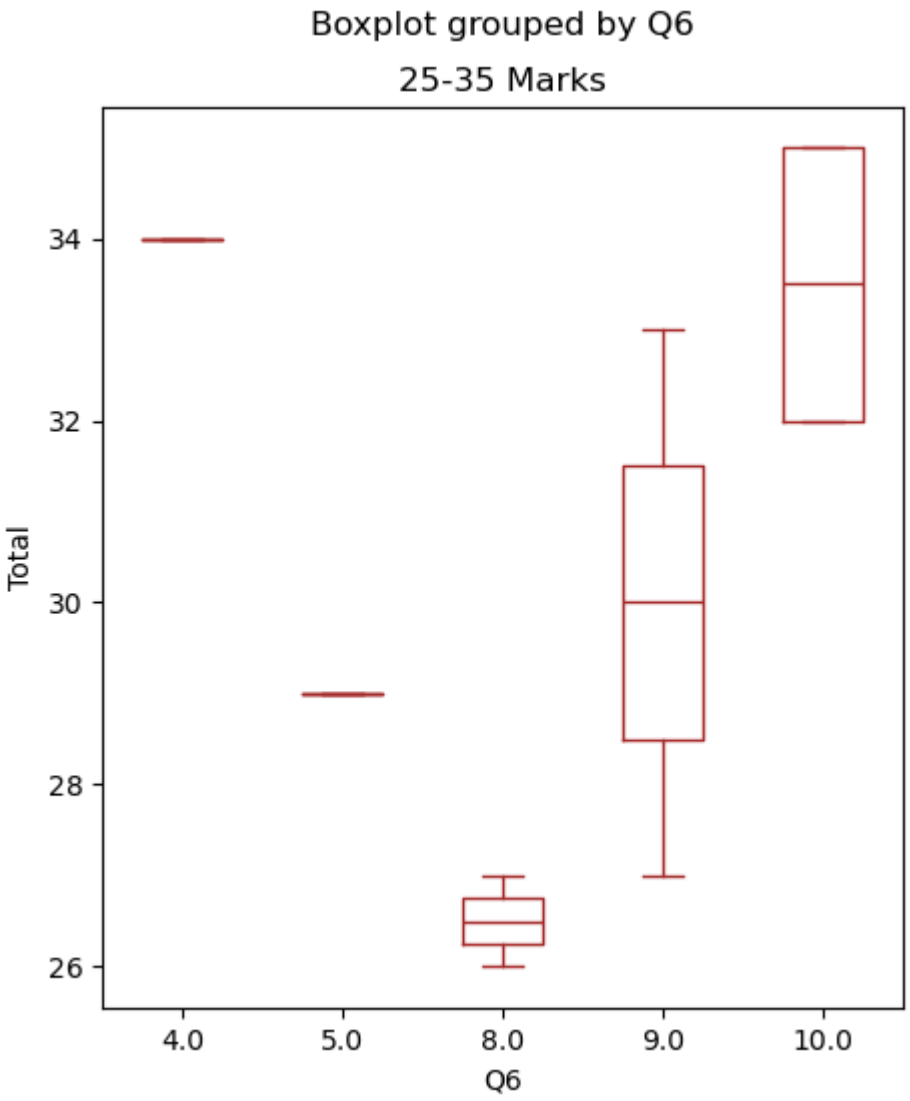
Minimum mark is 0 and the ones who attempted secured low marks.

In [91]:
```
c.boxplot(by='Q5M10', column =['Total'], grid = False,color='red',figsize=[5,6]
plt.title("25-35 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 [93]:
```
c.boxplot(by='Q6', column =['Total'], grid = False,color='brown',figsize=[5,6])
plt.title("25-35 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 [96]:
```
d=DF.loc[(DF['Total']>=35) & (DF['Total']<=40)]
d=d.reset_index()
d
```

Out[96]:

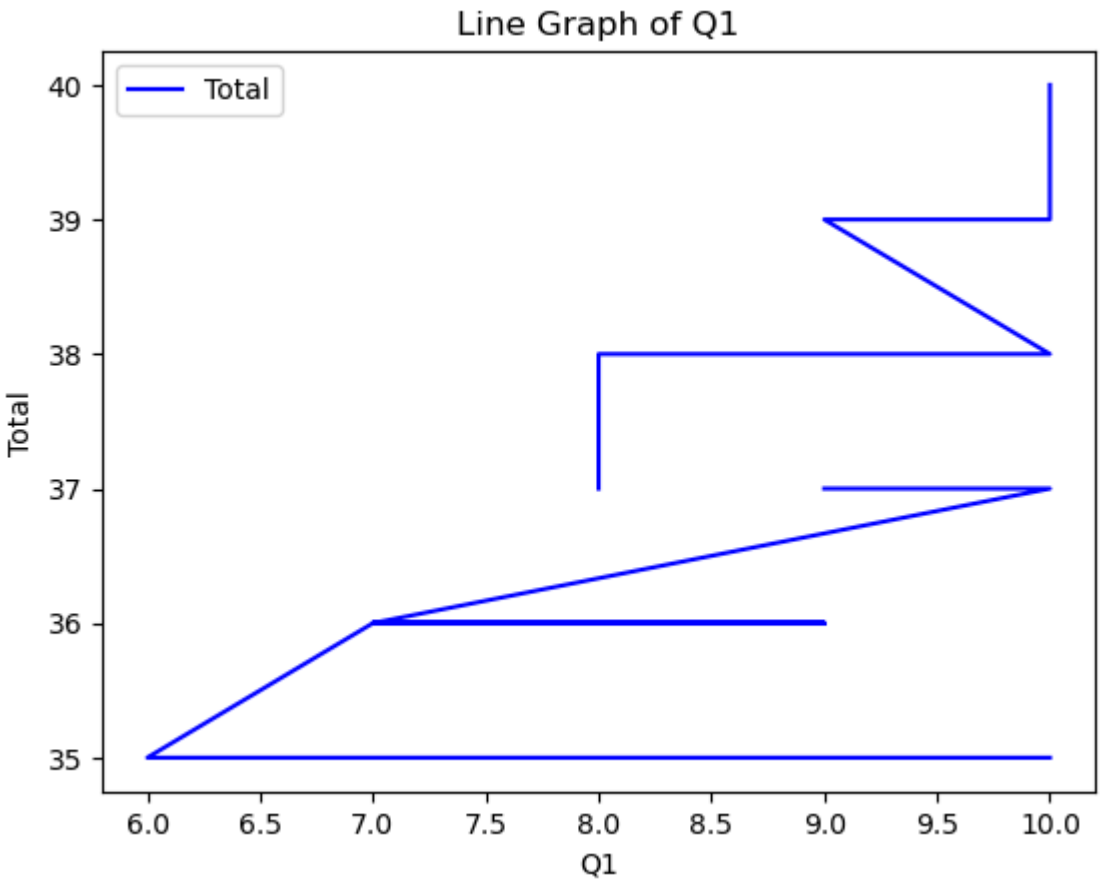
	index	Total	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7
0	70	35	4.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN
1	56	35	2.0	6.0	NaN	NaN	NaN	NaN	3.0	7.0
2	9	35	2.0	4.0	5.0	4.0	5.0	5.0	NaN	NaN
3	62	36	3.0	4.0	6.0	4.0	5.0	5.0	NaN	NaN
4	4	36	3.0	6.0	4.0	4.0	5.0	4.0	NaN	NaN
5	64	36	1.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN
6	71	36	4.0	5.0	6.0	4.0	5.0	5.0	NaN	NaN
7	44	36	3.0	6.0	6.0	2.0	NaN	NaN	2.0	7.0
8	46	36	4.0	5.0	6.0	4.0	5.0	5.0	NaN	NaN
9	17	36	3.0	4.0	6.0	4.0	NaN	NaN	NaN	NaN
10	83	37	4.0	6.0	6.0	2.0	NaN	NaN	NaN	NaN
11	0	37	4.0	5.0	6.0	4.0	2.0	1.0	NaN	5.0
12	36	37	2.0	NaN	6.0	4.0	5.0	5.0	NaN	NaN
13	10	37	3.0	5.0	6.0	4.0	NaN	NaN	3.0	6.0
14	59	38	2.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN
15	55	38	3.0	5.0	6.0	4.0	NaN	NaN	NaN	NaN
16	47	38	2.0	6.0	6.0	4.0	5.0	5.0	3.0	7.0
17	28	38	2.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN
18	42	38	4.0	6.0	6.0	4.0	5.0	5.0	3.0	5.0
19	49	39	3.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN
20	26	39	4.0	6.0	6.0	3.0	4.0	NaN	NaN	NaN
21	73	40	4.0	6.0	NaN	NaN	5.0	5.0	3.0	NaN
22	53	40	4.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN
23	51	40	0.0	NaN	6.0	4.0	NaN	NaN	3.0	7.0
24	33	40	NaN	NaN	6.0	4.0	5.0	5.0	3.0	7.0
25	65	40	4.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN

Total marks 35-40 is filtered from the data set

In [162…

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

```
plt.ylabel("Total")
plt.show()
```

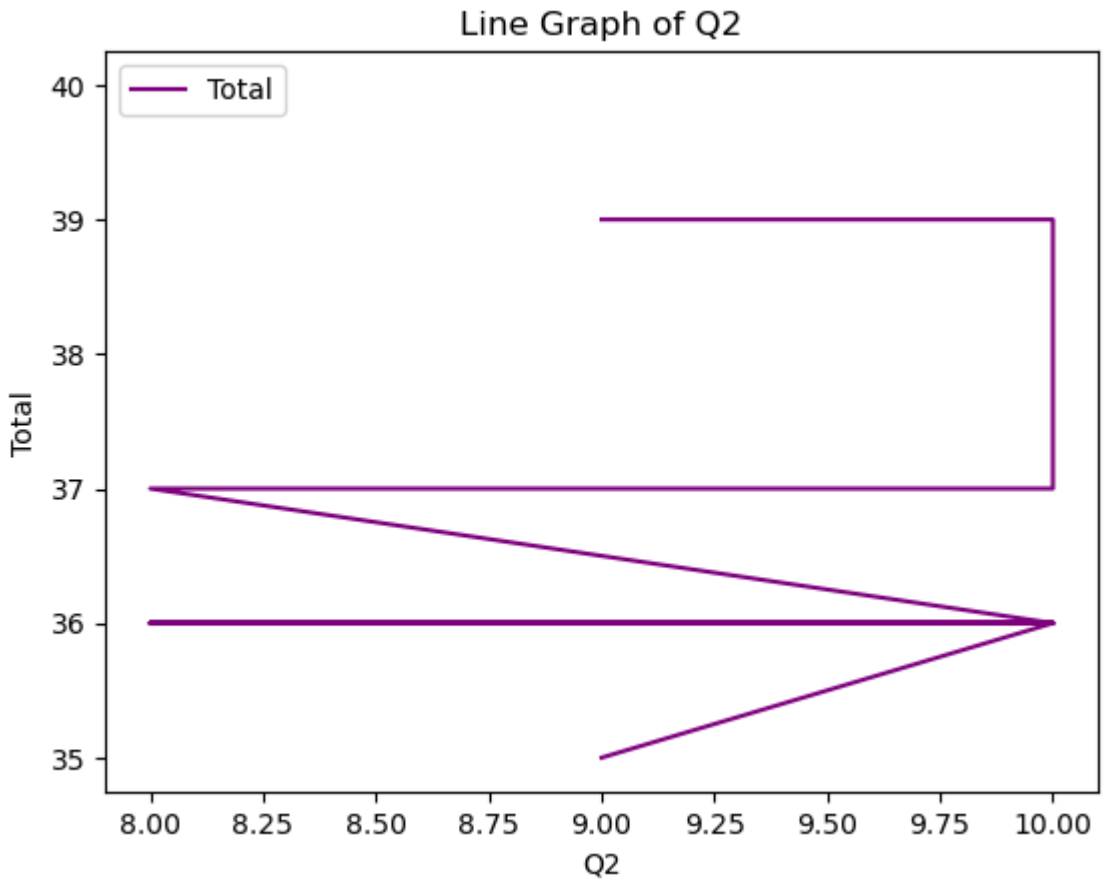


It can be deduced from the above line graph that majority of the students in this range scored marks between 7 and 10.

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

In [164…

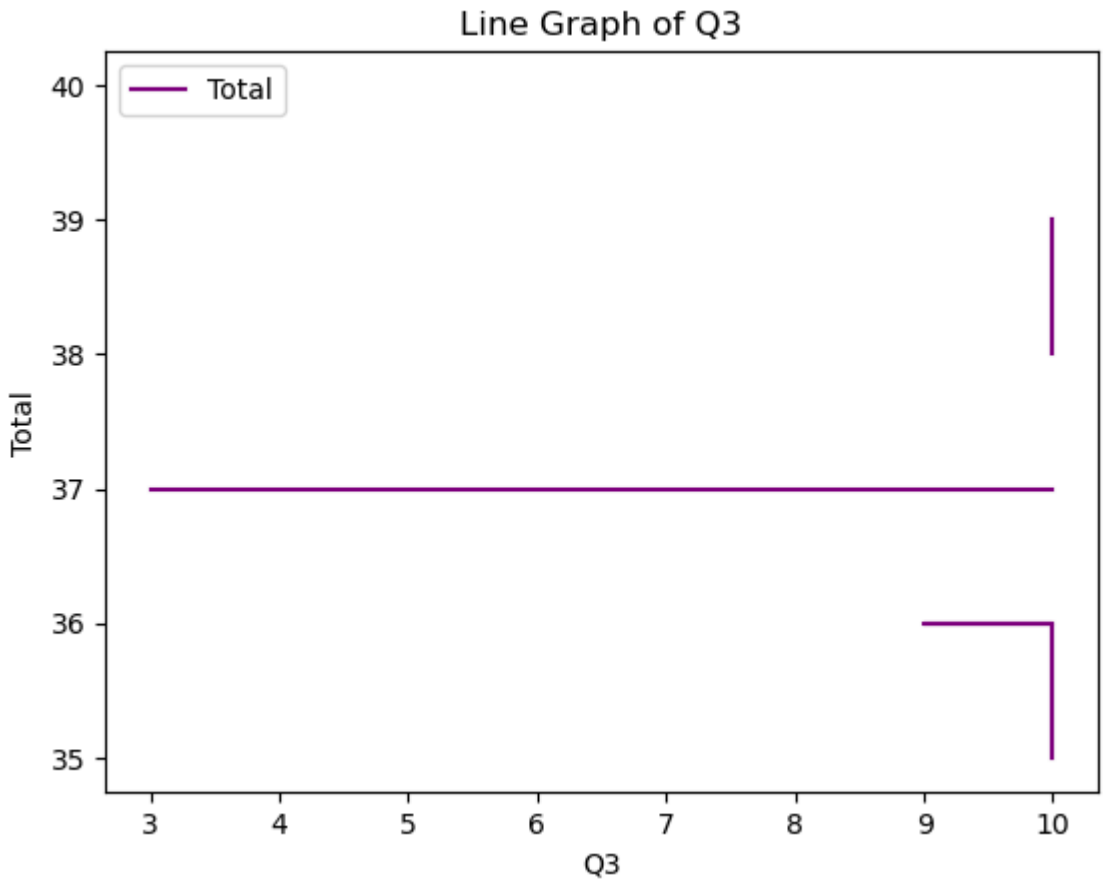
```
d.plot.line(x='Q2',y='Total',color='purple')
plt.title("Line Graph of Q2")
plt.ylabel("Total")
plt.show()
```



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 [166…

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

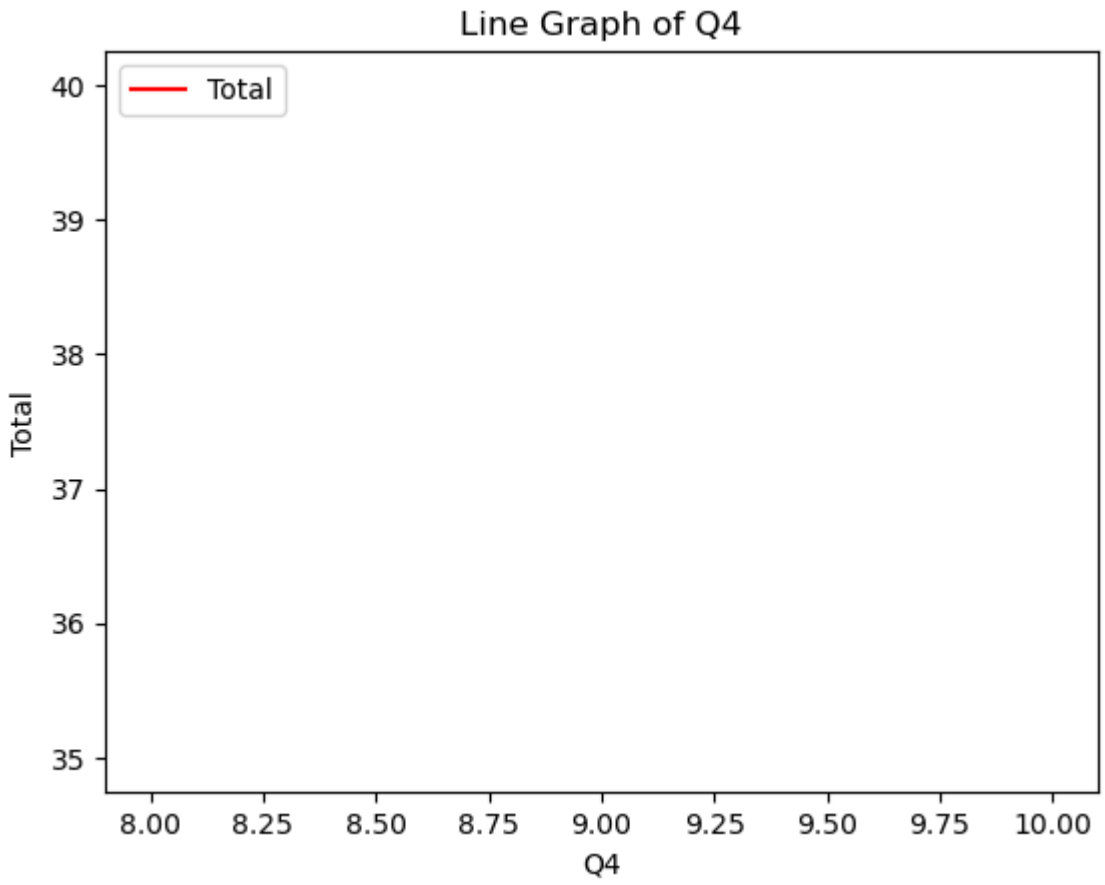


The students in this range scored 6-10 marks on an average

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

In [168…

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

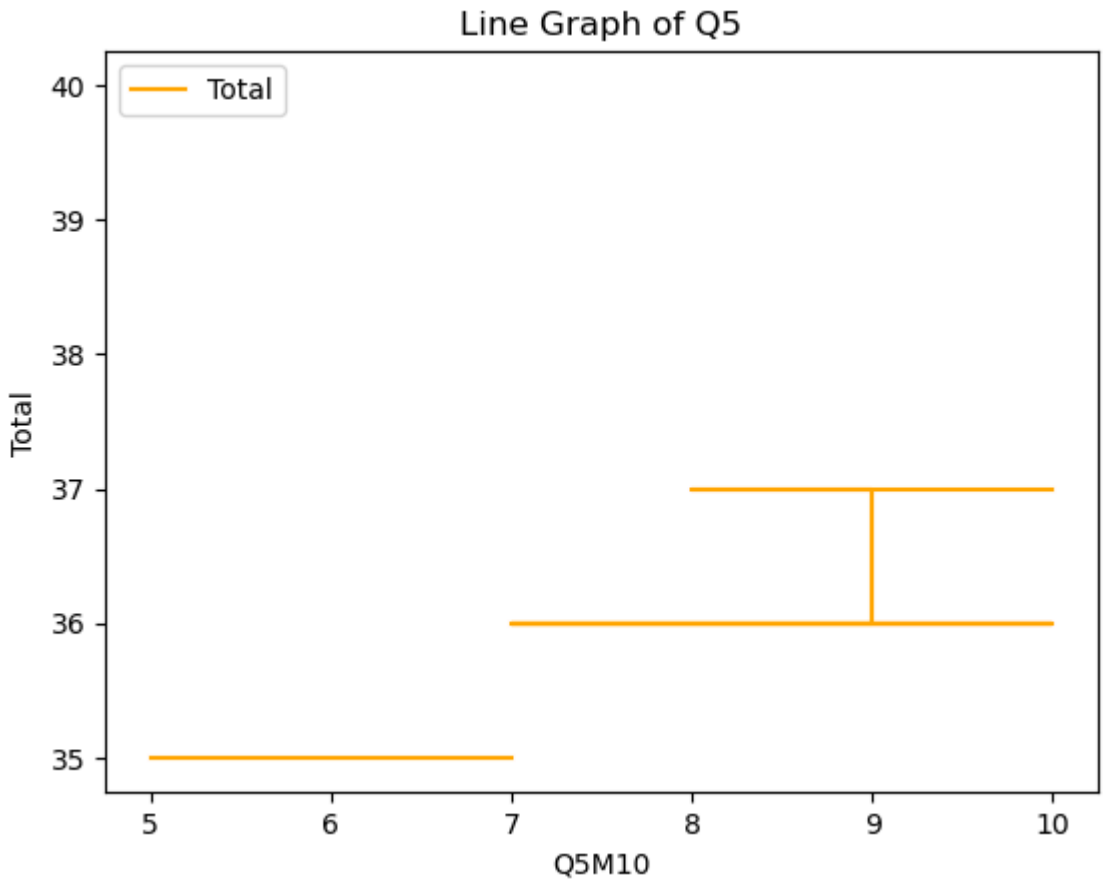


Most of the students who have attempted the question scored marks between 2 and 8 on an average.

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

In [170…

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

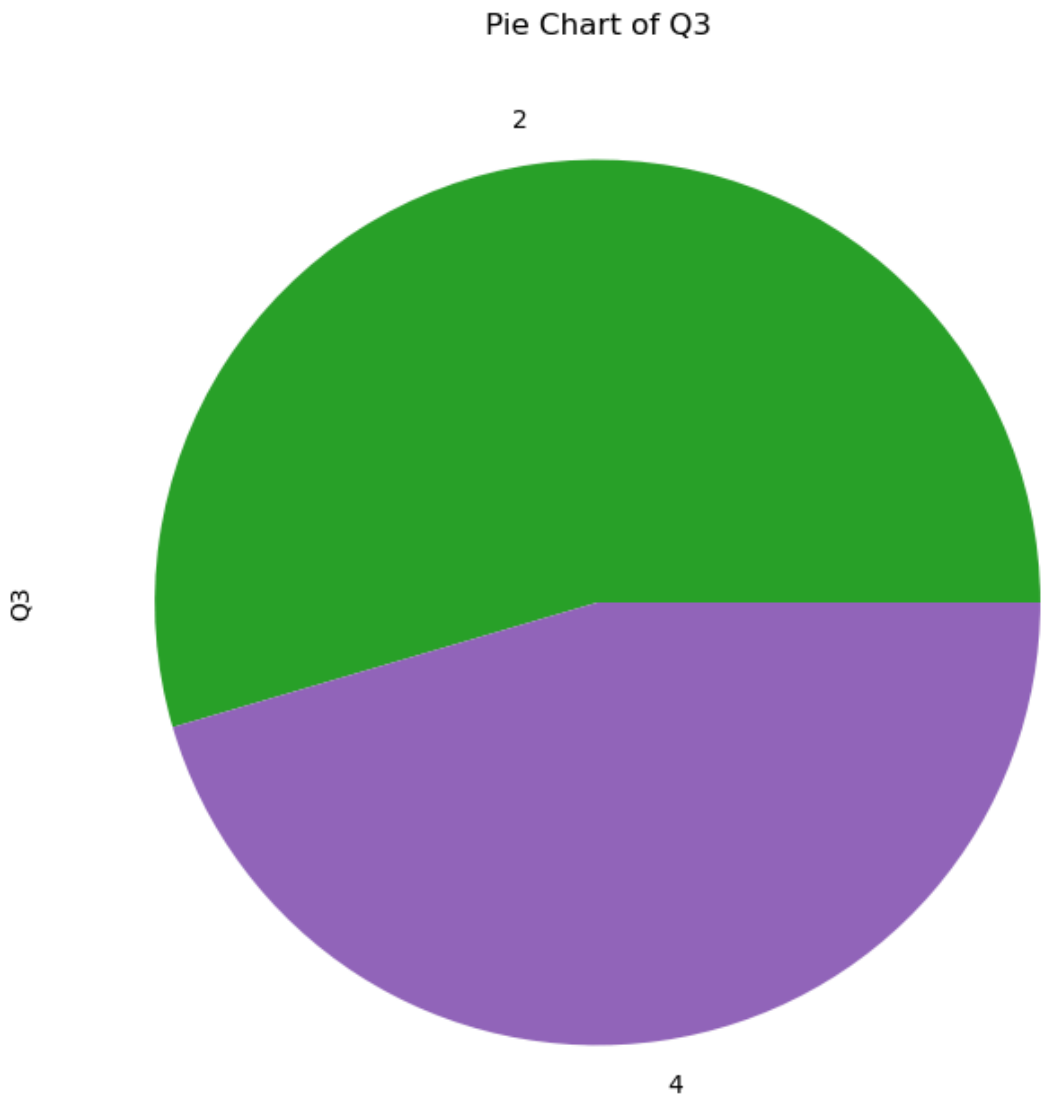


Majority of the students who attempted this question scored marks between 6 and 9 overall.

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

In [112…
```
a['Q3'].plot(kind='pie',subplots=True,figsize=(8,8))
plt.title("Pie Chart of Q3")
```

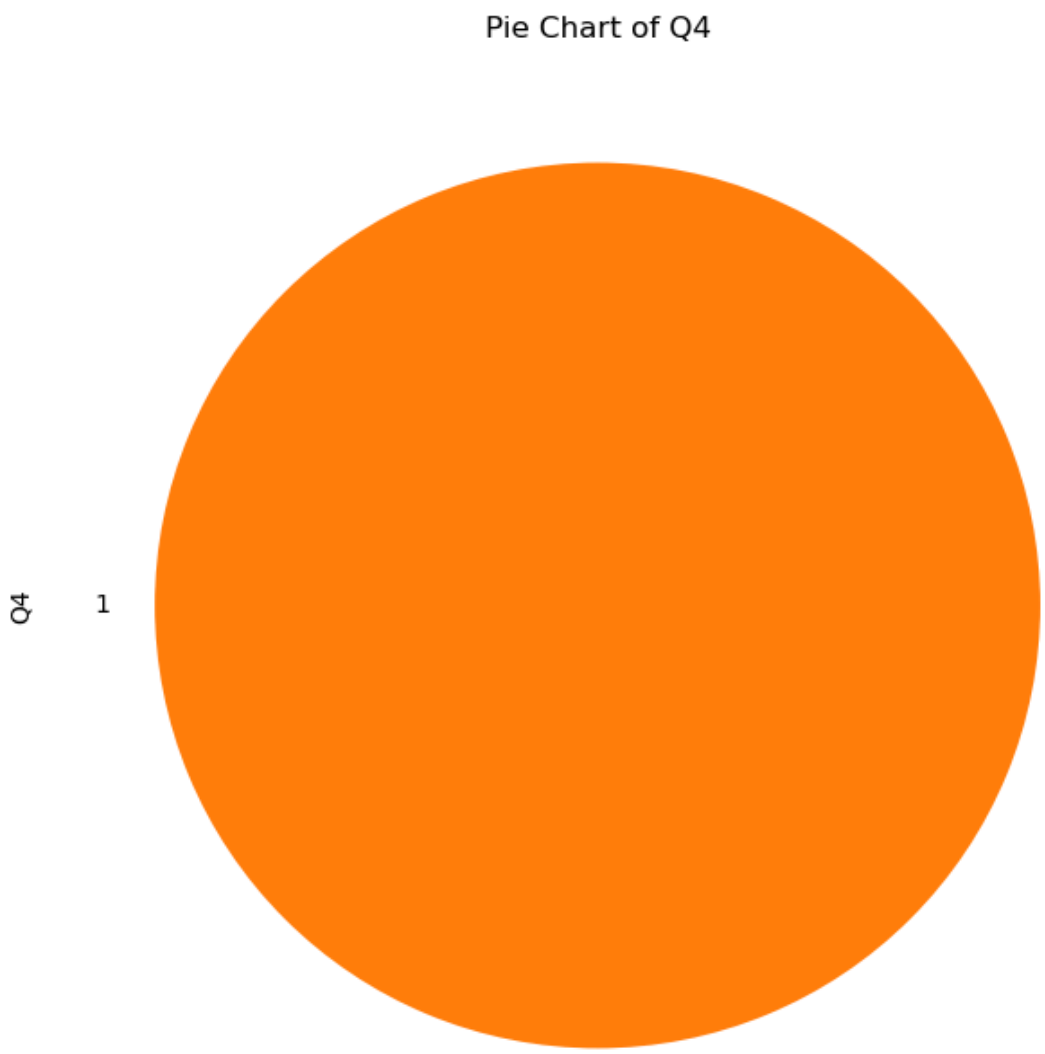
Out[112…
```
Text(0.5, 1.0, 'Pie Chart of Q3')
```



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 [114…
```
a['Q4'].plot(kind='pie',subplots=True,figsize=(8,8))
plt.title("Pie Chart of Q4")
```

Out[114…
```
Text(0.5, 1.0, 'Pie Chart of Q4')
```



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

THANK YOU