DATA ANALYTICS ASSIGNMENT(1)ON MATPLOTLIB

M.AMITH YADAV

2211CS010360 group(4)

In [13]:
```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("CLASS MARKS.csv")
```

In [15]:
```
df
```

Out[15]:

	Total	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7	Q5M1
0	37	4.0	5.0	6.0	4.0	2.0	1.0	NaN	5.0	8.
1	32	4.0	3.0	4.0	3.0	NaN	NaN	3.0	6.0	9.
2	33	4.0	5.0	5.0	1.0	5.0	5.0	NaN	NaN	8.
3	24	4.0	6.0	6.0	3.0	2.0	2.0	NaN	NaN	Na
4	36	3.0	6.0	4.0	4.0	5.0	4.0	NaN	NaN	10.
...	...	...	...	...	...	...	...	...	...	
81	32	3.0	6.0	3.0	4.0	5.0	3.0	NaN	NaN	Na
82	27	2.0	2.0	5.0	3.0	NaN	NaN	NaN	NaN	7.
83	37	4.0	6.0	6.0	2.0	NaN	NaN	NaN	NaN	9.
84	28	4.0	NaN	5.0	4.0	5.0	4.0	NaN	NaN	6.
85	29	4.0	6.0	NaN	NaN	NaN	NaN	3.0	5.0	7.

86 rows × 12 columns

TOTAL values are sorted in ascending order above.

In [19]:
```
DF=df.sort_values("Total",ascending=True)
DF
```

Out[19]:

	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
76	17	2.0	3.0	4.0	2.0	4.0	2.0	NaN	NaN	Na
...	...	...	...	...	...	...	...	...	...	
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
53	40	4.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN	10.
73	40	4.0	6.0	NaN	NaN	5.0	5.0	3.0	NaN	10.
65	40	4.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN	10.

86 rows × 12 columns

In [21]:

```
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[21]:

	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
76	17	2.0	3.0	4.0	2.0	4.0	2.0	NaN	NaN	Na
...	...	...	...	...	...	...	...	...	...	
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
53	40	4.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN	10.
73	40	4.0	6.0	NaN	NaN	5.0	5.0	3.0	NaN	10.
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,Q6 are created above to do anlytics

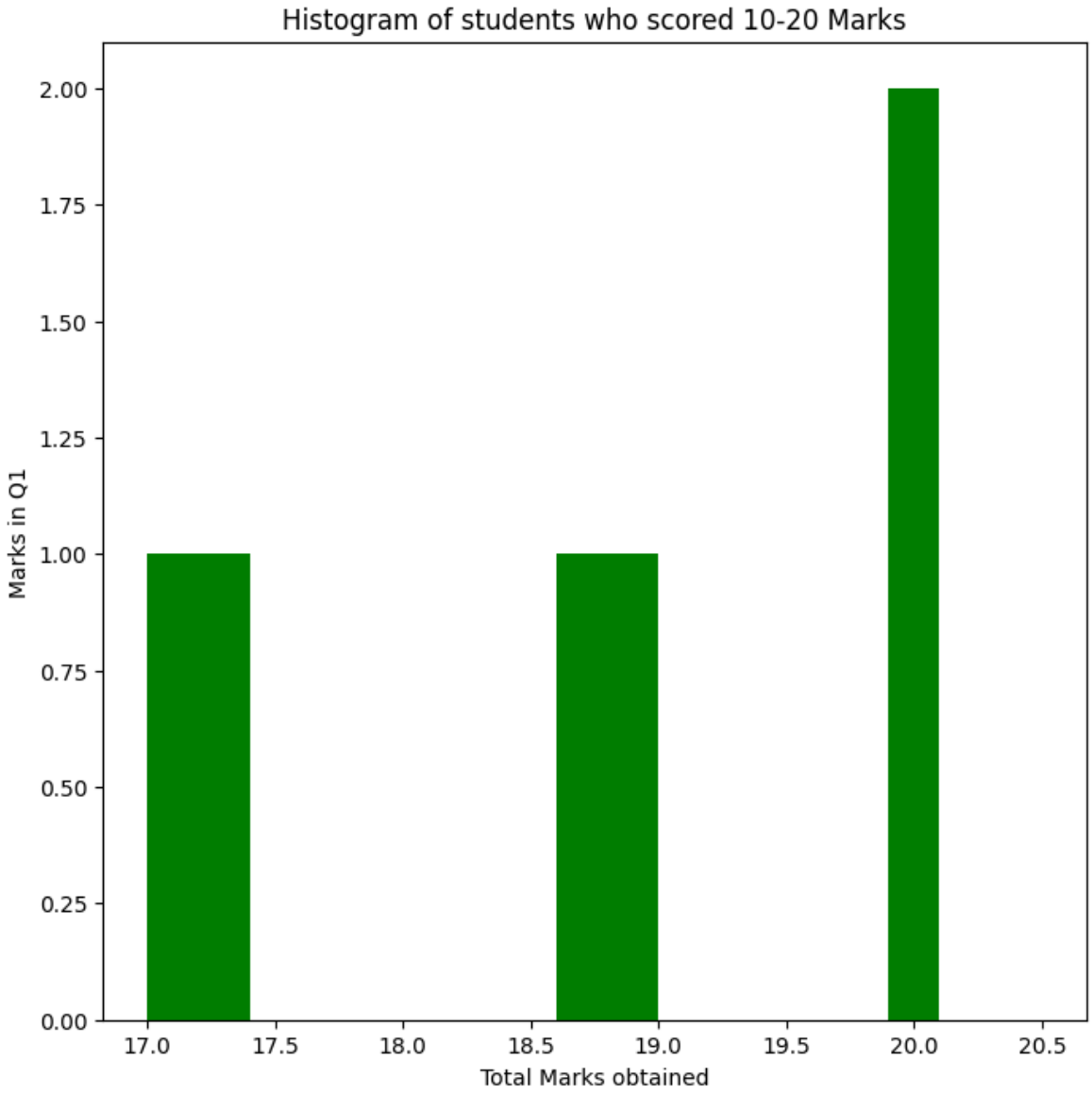
In [25]:
```
a=DF.loc[(DF['Total'] >= 15) & (DF['Total'] <= 20)]
a=a.reset_index()
a
```

Out[25]:

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

Total marks 10-20 is filtered above

In [102…
```
a1=a.groupby('Q1')['Total']
a1.hist(color = 'green',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()
```

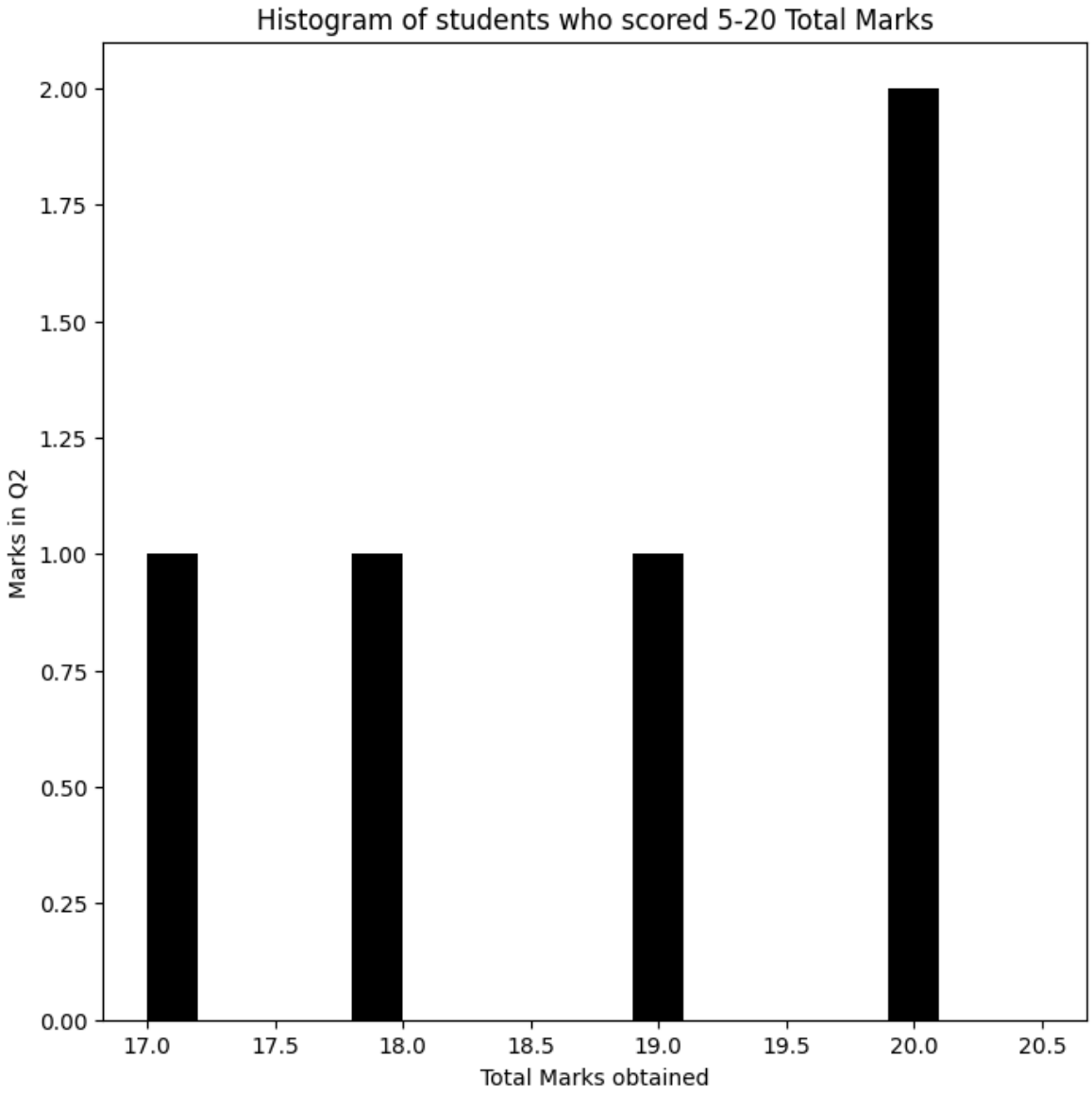


Most of the students scored 1 mark and maximum mark

is 2, implying that all the students in this range secured very less marks in Q1

In [106…

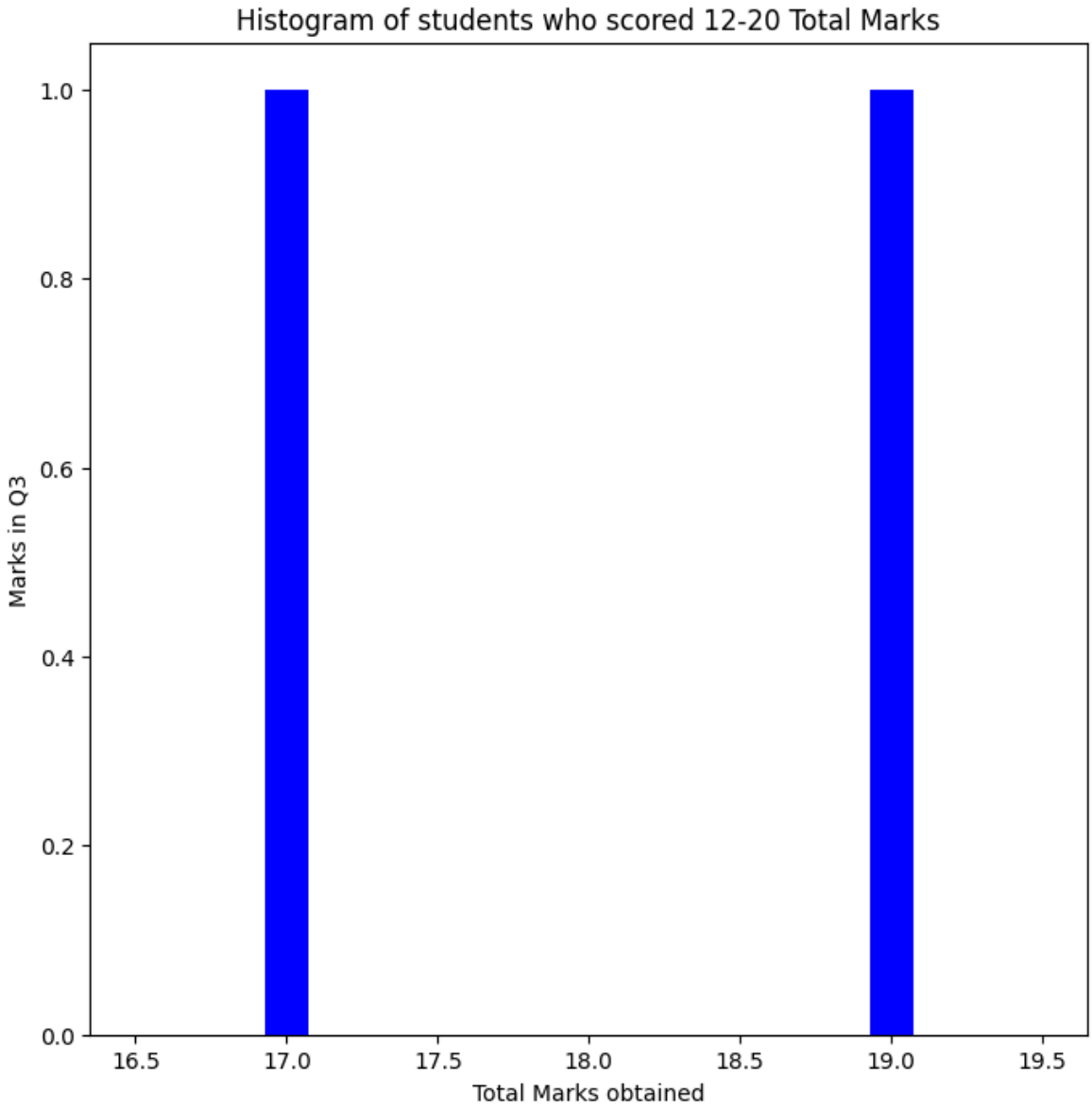
```
a2=a.groupby('Q2')['Total']
a2.hist(color='black',figsize=[8,8],grid=False,bins=5)
plt.title("Histogram of students who scored 5-20 Total Marks")
plt.xlabel("Total Marks obtained")
plt.ylabel("Marks in Q2")
plt.show()
```



Very few students got high marks in Q2 and that's only2, so overall performance in Q2 is not upto themark

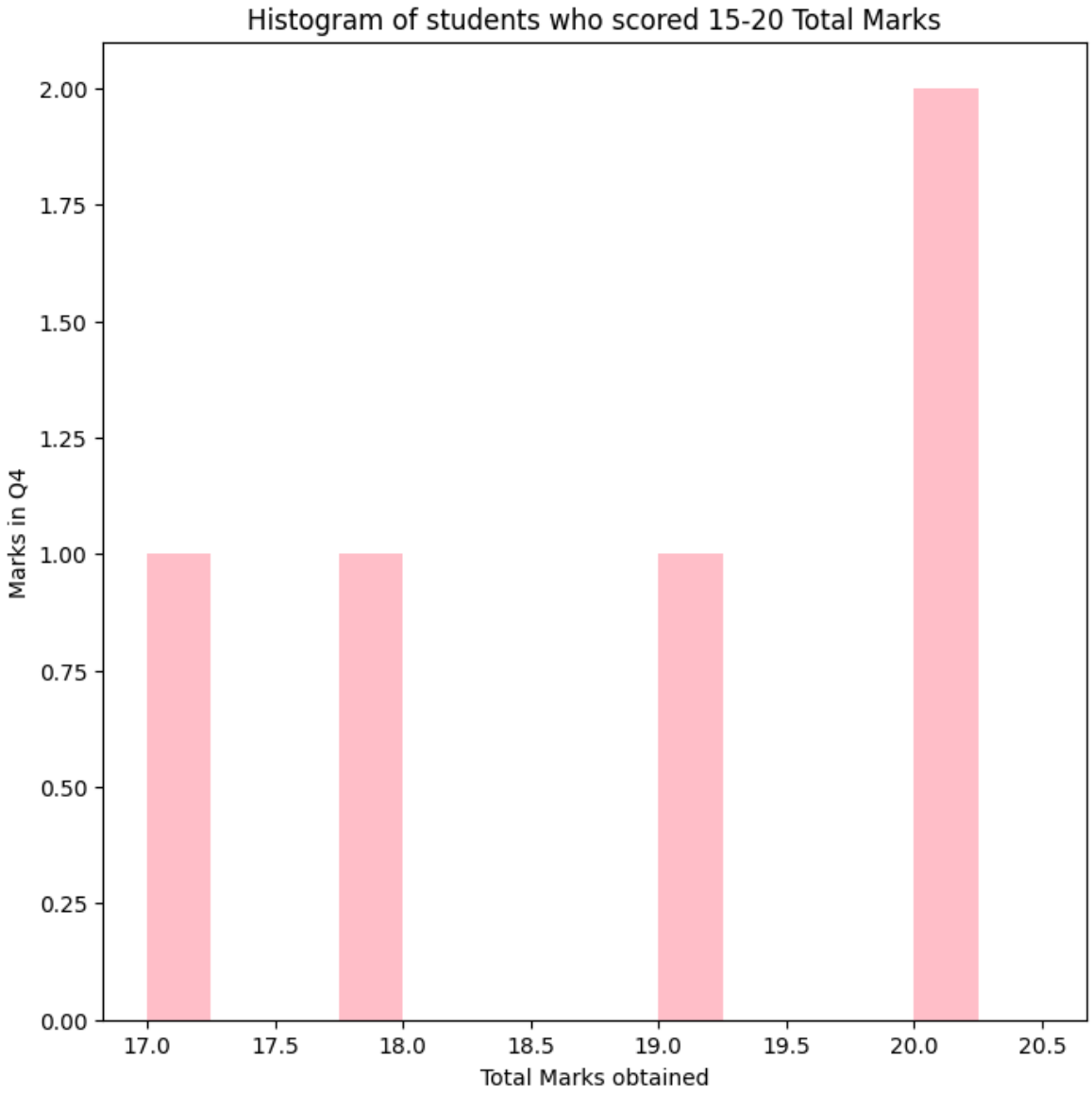
In [109…

```
a3=a.groupby('Q3')['Total']
a3.hist(color='blue',figsize=[8,8],grid=False,bins=7)
plt.title("Histogram of students who scored 12-20 Total Marks")
plt.xlabel("Total Marks obtained")
plt.ylabel("Marks in Q3")
plt.show()
```



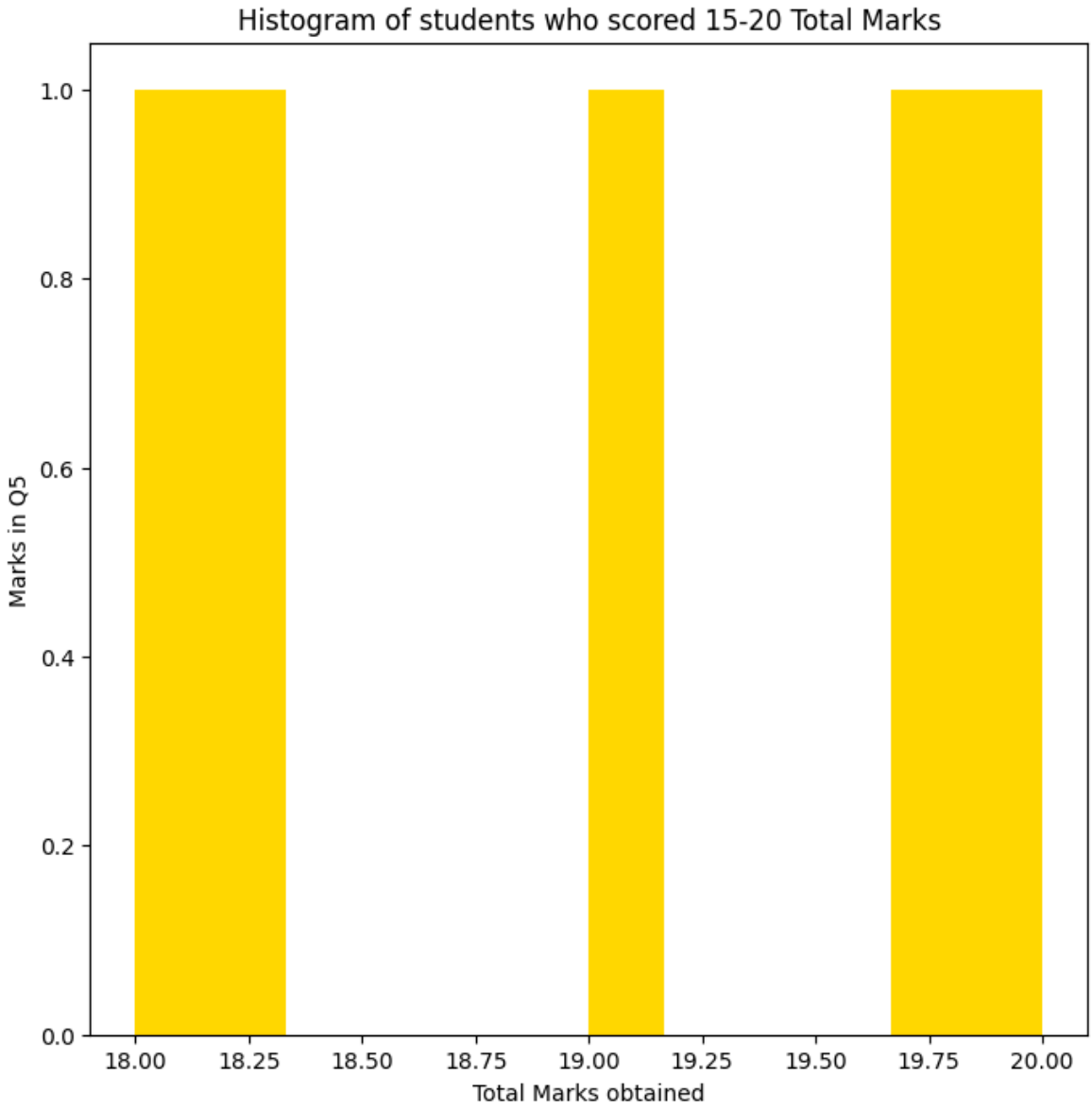
Only one student got 3 marks whereas others scoredbelow 2 marks, implying very less marks are secured inQ3 overall

In [33]:
```
a4=a.groupby('Q4')['Total']
a2.hist(color='pink',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()
```



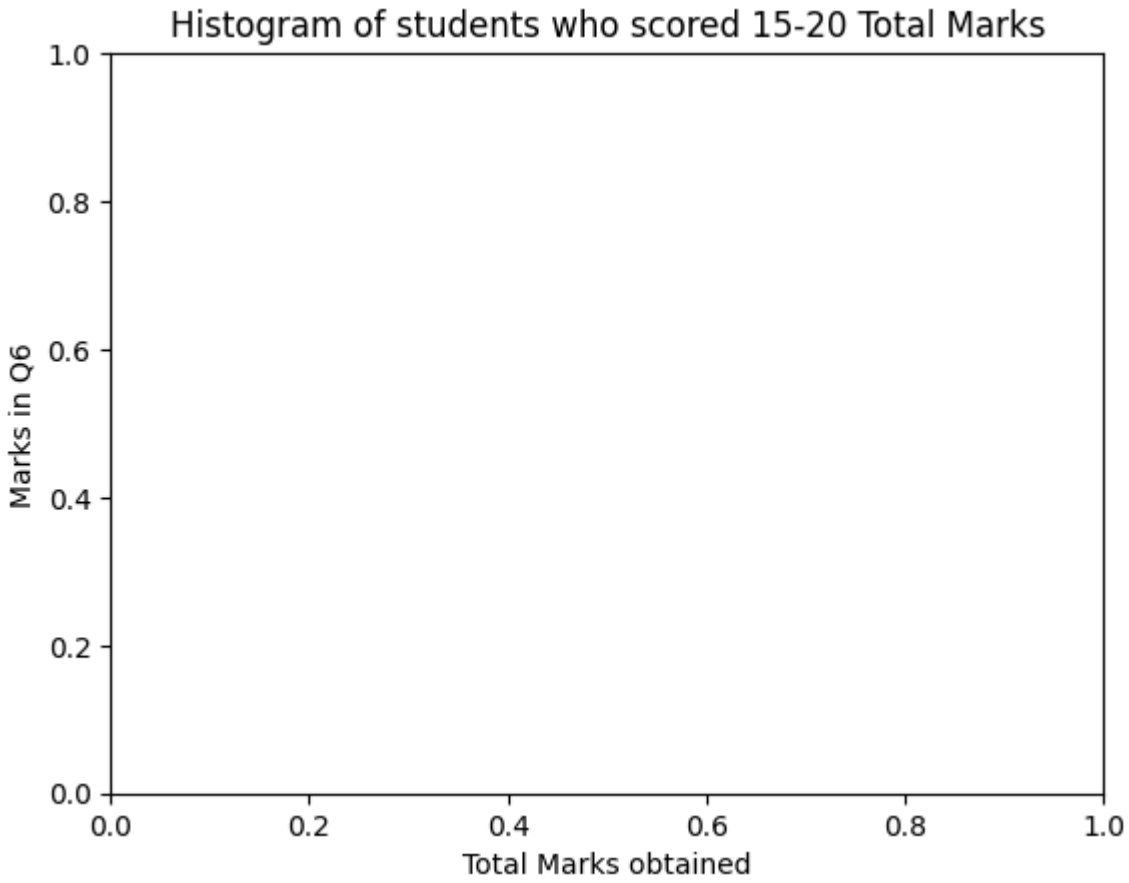
Many of these students who attempted Q4 got lessmarks and maximum mark is also just two only

In [35]:
```
a5=a.groupby('Q5M10')['Total']
a5.hist(color='gold',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 lessmarks and maximum mark is also just three only

In [37]:
```
a6=a.groupby('Q6')['Total']
a6.hist(color='Red',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 rangeremains bad as well.

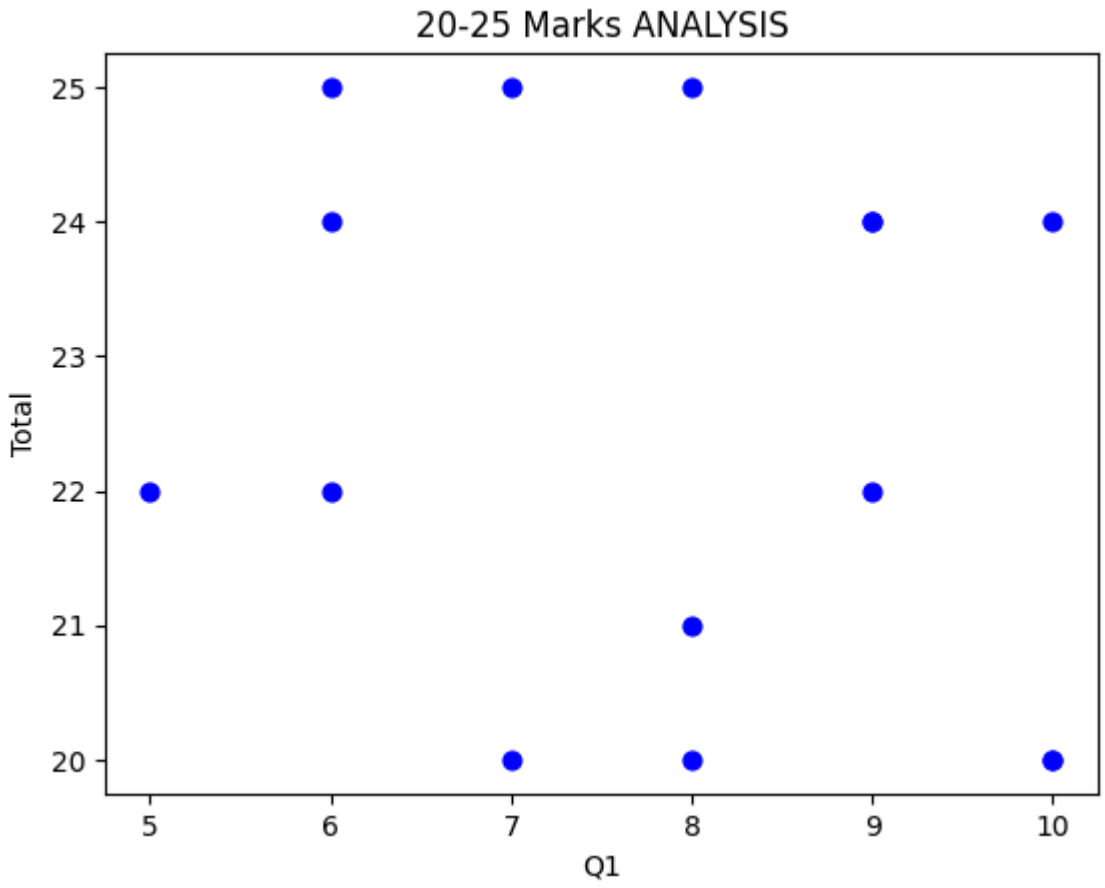
In [39]:
```
b=DF.loc[(DF['Total'] >= 20) & (DF['Total'] <= 25)]
b=b.reset_index()
b
```

Out[39]:

	index	Total	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7
0	60	20	2.0	5.0	3.0	2.0	NaN	NaN	NaN	NaN
1	68	20	4.0	6.0	6.0	4.0	NaN	NaN	NaN	NaN
2	30	20	4.0	4.0	4.0	4.0	5.0	NaN	NaN	NaN
3	5	20	4.0	6.0	6.0	4.0	NaN	NaN	NaN	NaN
4	75	21	4.0	NaN	6.0	1.0	1.0	1.0	NaN	NaN
5	54	21	2.0	6.0	NaN	NaN	5.0	5.0	3.0	NaN
6	39	22	3.0	2.0	NaN	1.0	3.0	3.0	NaN	NaN
7	25	22	4.0	2.0	5.0	2.0	4.0	3.0	2.0	NaN
8	77	22	4.0	5.0	NaN	NaN	3.0	2.0	2.0	NaN
9	61	24	4.0	5.0	6.0	4.0	NaN	5.0	NaN	NaN
10	3	24	4.0	6.0	6.0	3.0	2.0	2.0	NaN	NaN
11	50	24	4.0	5.0	NaN	NaN	5.0	5.0	NaN	NaN
12	18	24	3.0	3.0	5.0	3.0	NaN	NaN	2.0	1.0
13	6	25	3.0	4.0	NaN	2.0	5.0	5.0	NaN	NaN
14	72	25	2.0	NaN	4.0	4.0	NaN	5.0	3.0	7.0
15	48	25	2.0	6.0	NaN	NaN	NaN	NaN	3.0	6.0
16	74	25	1.0	5.0	6.0	4.0	NaN	NaN	NaN	NaN

Total marks 20-25 is filtered from the data set

In [41]:
```
b.plot.scatter(x='Q1',y='Total',color='blue',s=40)
plt.title("20-25 Marks ANALYSIS")
plt.show()
```



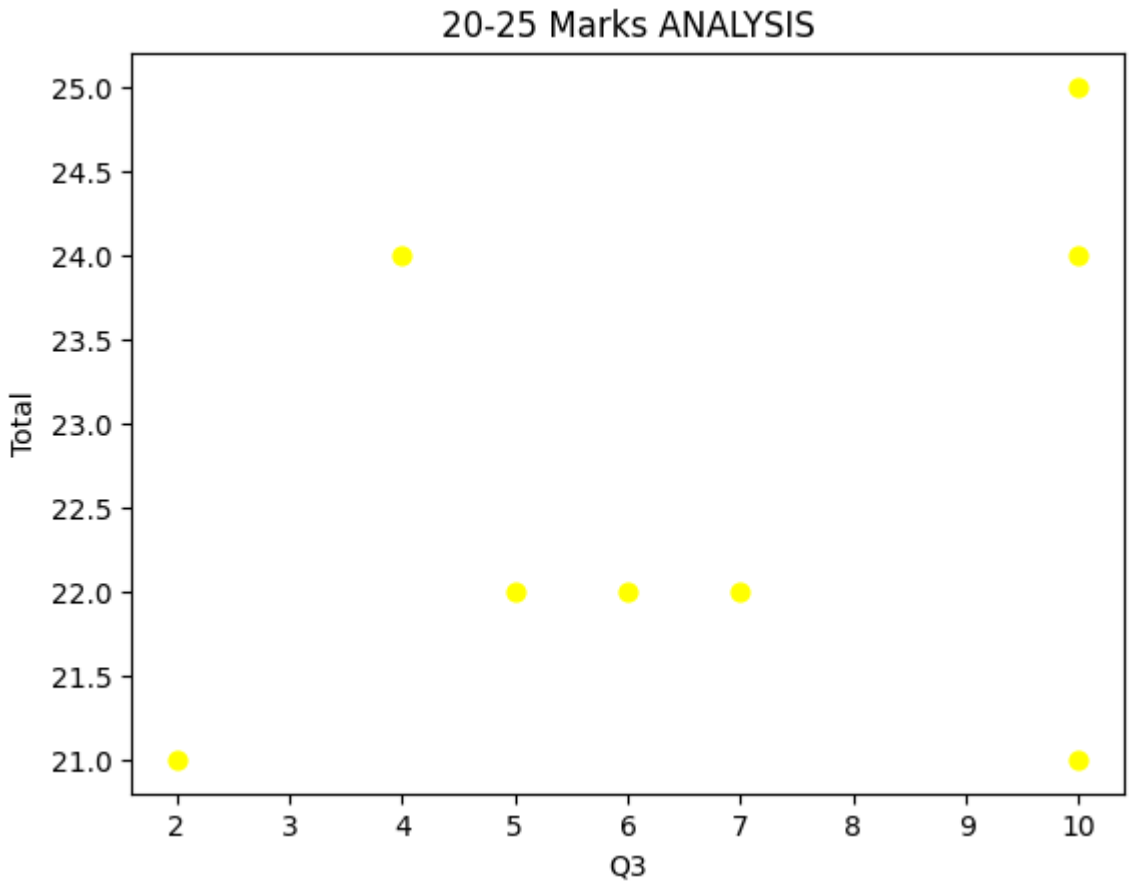
In [11]:
```
b.plot.scatter(x='Q2',y='Total',color='green',s=40)
plt.title("20-25 Marks ANALYSIS")
plt.show()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[11], line 1
----> 1 b.plot.scatter(x='Q2',y='Total',color='green',s=40)
      2 plt.title("20-25 Marks ANALYSIS")
      3 plt.show()

NameError: name 'b' is not defined
```

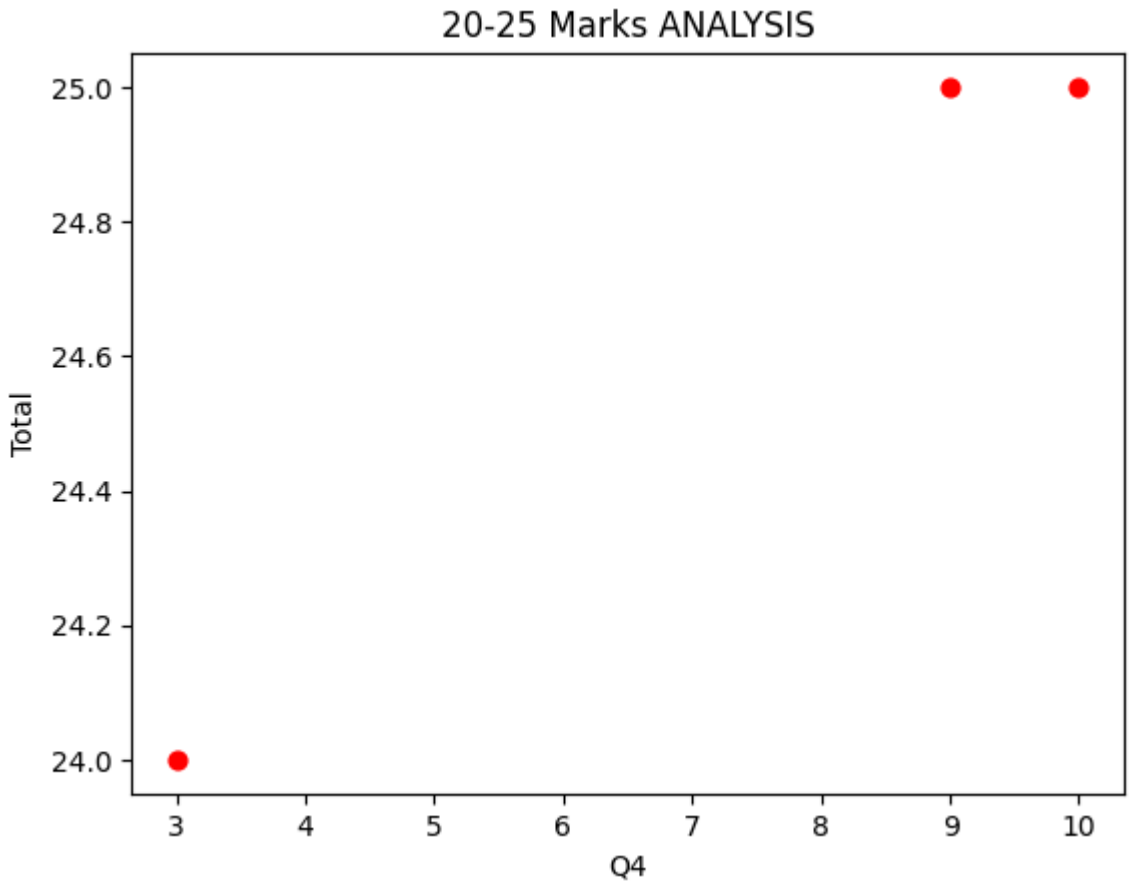
Majority of the students in this range scored marksbetween 6-10 in Q1 and very few, just 3 students scoredbelow 5 marks, maximum mark scored is 10 by twostudents.

In [45]:
```
b.plot.scatter(x='Q3',y='Total',color='yellow',s=40)
plt.title("20-25 Marks ANALYSIS")
plt.show()
```



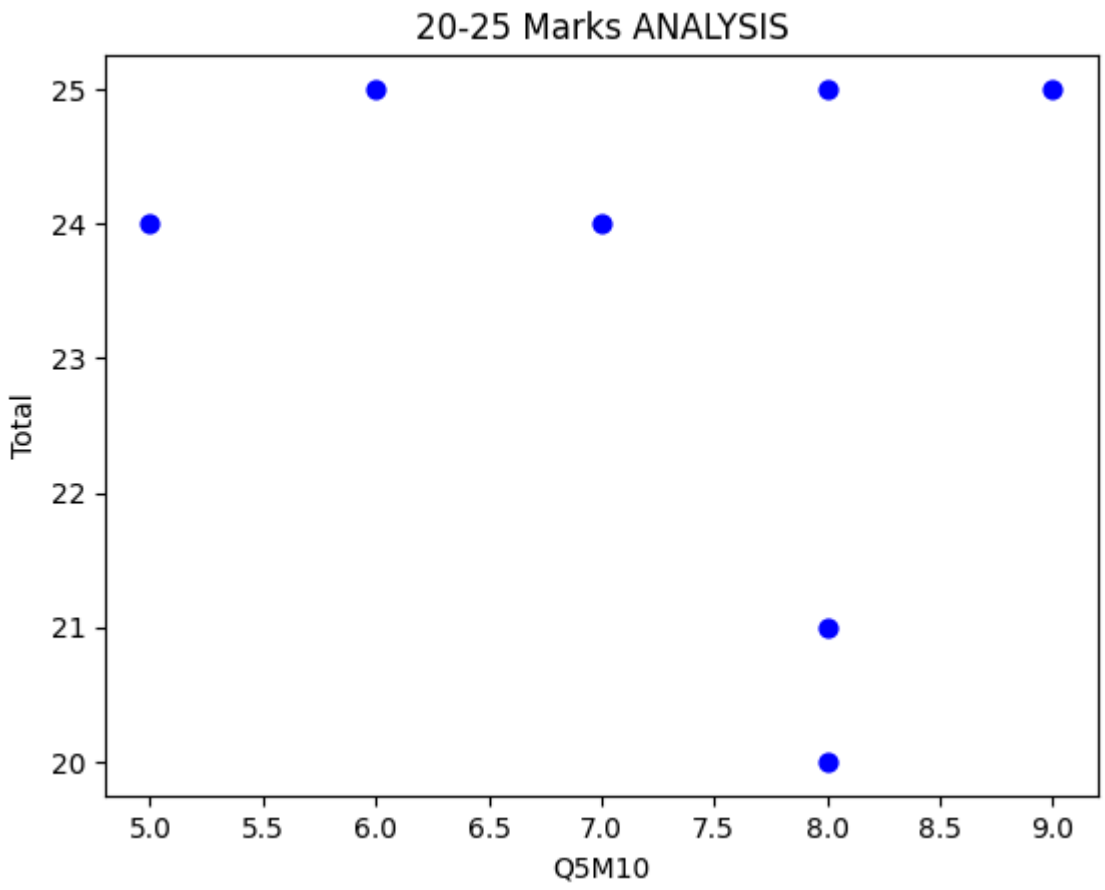
Three students haven't attempted the question andmost of them who attempted scored 4-7 marks andmaximum mark is 10 scored by three students.

In [47]:
```
b.plot.scatter(x='Q4',y='Total',color='red',s=40)
plt.title("20-25 Marks ANALYSIS")
plt.show()
```



Most of the students in this range haven't attempted

In [49]:
```
b.plot.scatter(x='Q5M10',y='Total',color='blue',s=40)
plt.title("20-25 Marks ANALYSIS")
plt.ylabel("Total")
plt.show()
```



Majority of the students who attempted scored marksbetween 5 and 9, also some scored zero, meaning fivestudents left the question unattempted

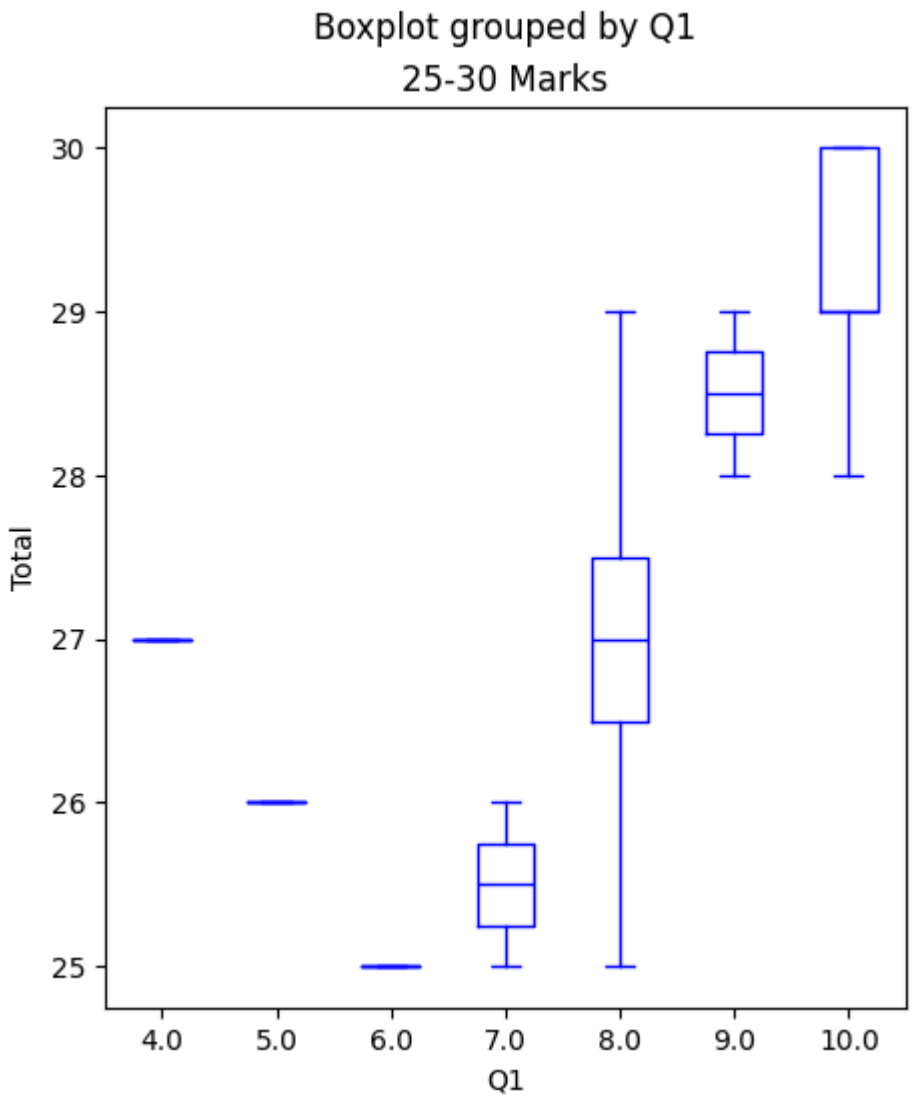
In [51]:
```
c=DF.loc[(DF['Total'] >= 25) & (DF['Total'] <= 30)]
c=c.reset_index()
c
```

Out[51]:

	index	Total	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7
0	6	25	3.0	4.0	NaN	2.0	5.0	5.0	NaN	NaN
1	72	25	2.0	NaN	4.0	4.0	NaN	5.0	3.0	7.0
2	48	25	2.0	6.0	NaN	NaN	NaN	NaN	3.0	6.0
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	79	27	2.0	6.0	NaN	3.0	2.0	5.0	NaN	NaN
8	29	27	4.0	NaN	6.0	1.0	NaN	NaN	NaN	7.0
9	8	27	3.0	5.0	5.0	NaN	NaN	NaN	NaN	NaN
10	82	27	2.0	2.0	5.0	3.0	NaN	NaN	NaN	NaN
11	67	28	4.0	6.0	4.0	4.0	NaN	NaN	NaN	NaN
12	84	28	4.0	NaN	5.0	4.0	5.0	4.0	NaN	NaN
13	38	28	4.0	5.0	6.0	4.0	5.0	3.0	1.0	NaN
14	52	29	4.0	5.0	4.0	3.0	NaN	NaN	3.0	6.0
15	85	29	4.0	6.0	NaN	NaN	NaN	NaN	3.0	5.0
16	20	29	2.0	6.0	2.0	2.0	5.0	5.0	NaN	NaN
17	40	29	4.0	6.0	6.0	4.0	NaN	NaN	1.0	1.0
18	35	30	4.0	6.0	6.0	4.0	NaN	1.0	NaN	NaN
19	16	30	4.0	NaN	6.0	4.0	5.0	2.0	NaN	NaN
20	14	30	4.0	6.0	6.0	2.0	4.0	5.0	3.0	NaN

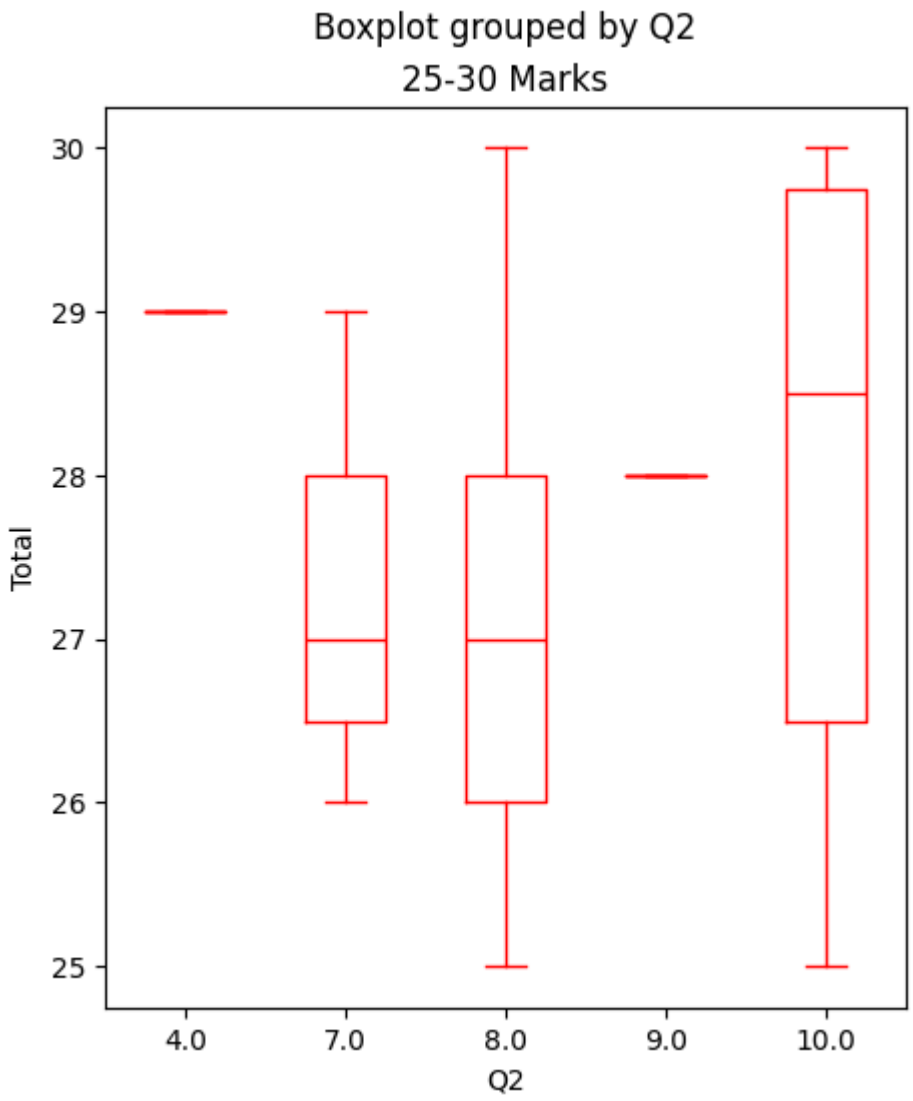
Total marks 25-30 is filtered from the data set

In [53]:
```
c.boxplot(by='Q1', column =['Total'], grid = False,color='blue',figsize=[5,6])
plt.title("25-30 Marks")
plt.ylabel("Total")
plt.show()
```

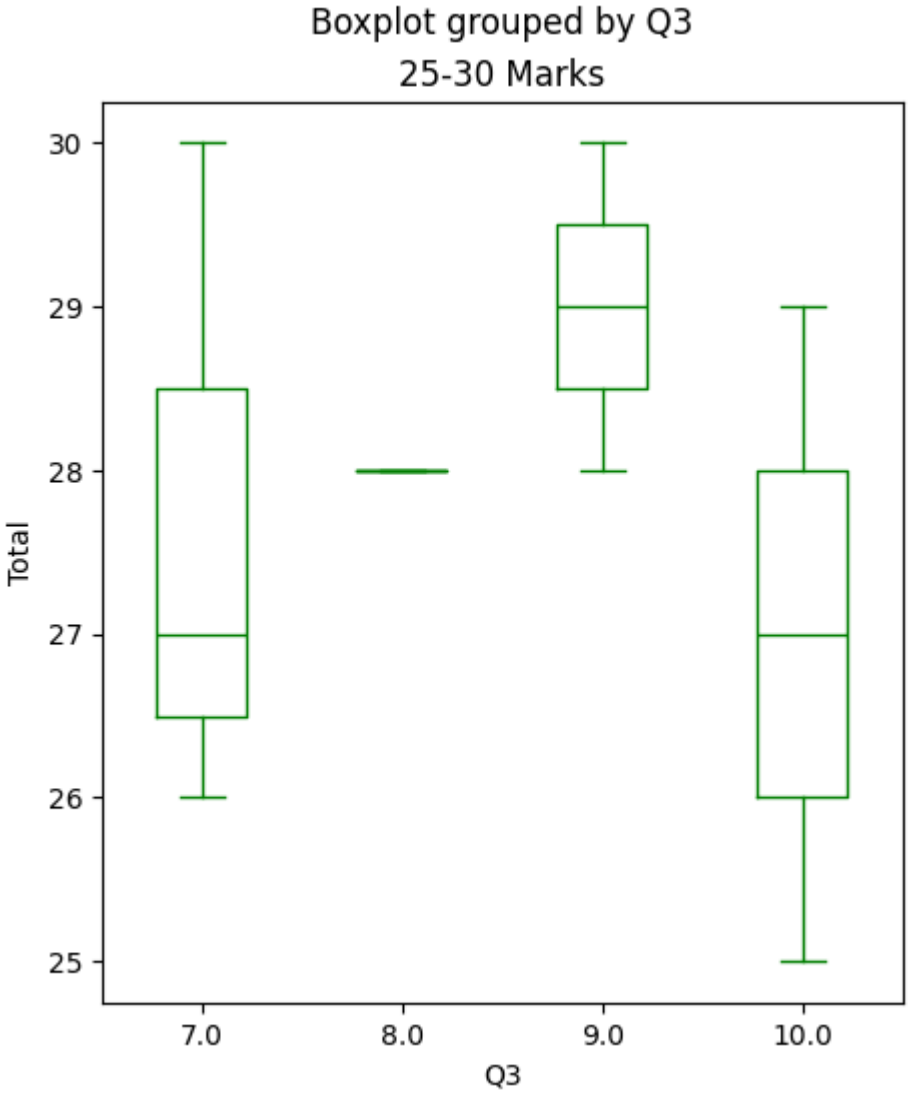


Many of the students in this range got marks between 8and 10, the maximum mark is 10 and minimum mark is0.

In [57]:
```
c.boxplot(by='Q2', column =['Total'], grid = False,color='red',figsize=[5,6])
plt.title("25-30 Marks")
plt.ylabel("Total")
plt.show()
```

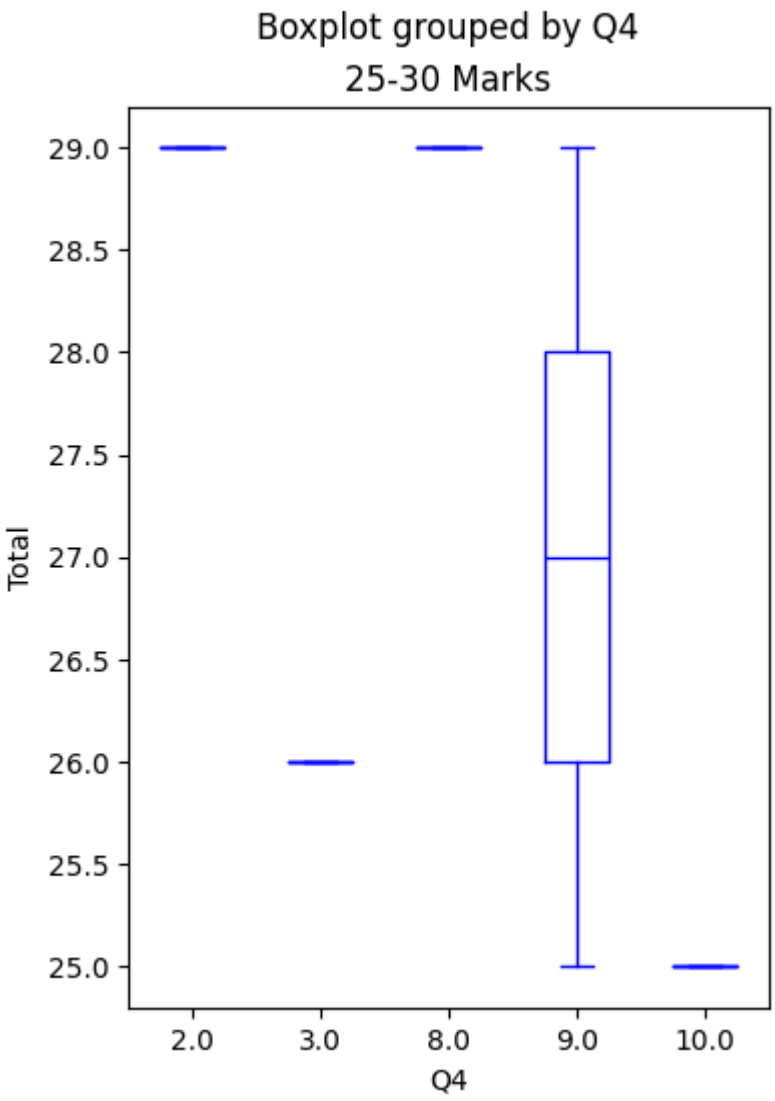


In [59]:
```
c.boxplot(by='Q3', column =['Total'], grid = False,color='green',figsize=[5,6])
plt.title("25-30 Marks")
plt.ylabel("Total")
plt.show()
```



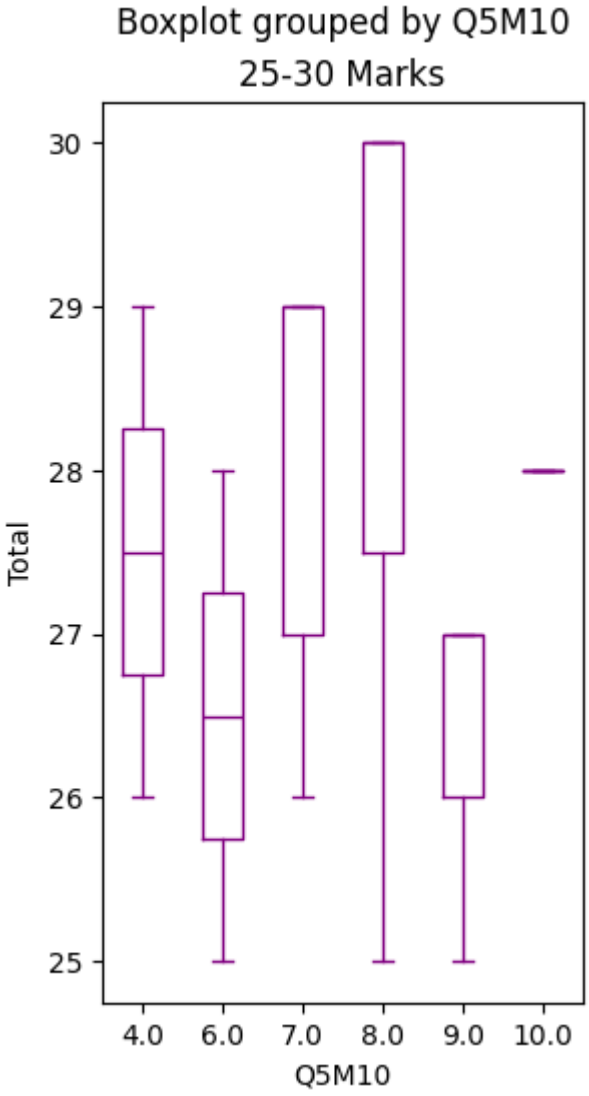
The minimum mark in this question is 0 and maximummark is 10Most of the students have secured marks between 7and 9

In [61]:
```
c.boxplot(by='Q4', column =['Total'], grid = False,color='blue',figsize=[4,6])
plt.title("25-30 Marks")
plt.ylabel("Total")
plt.show()
```



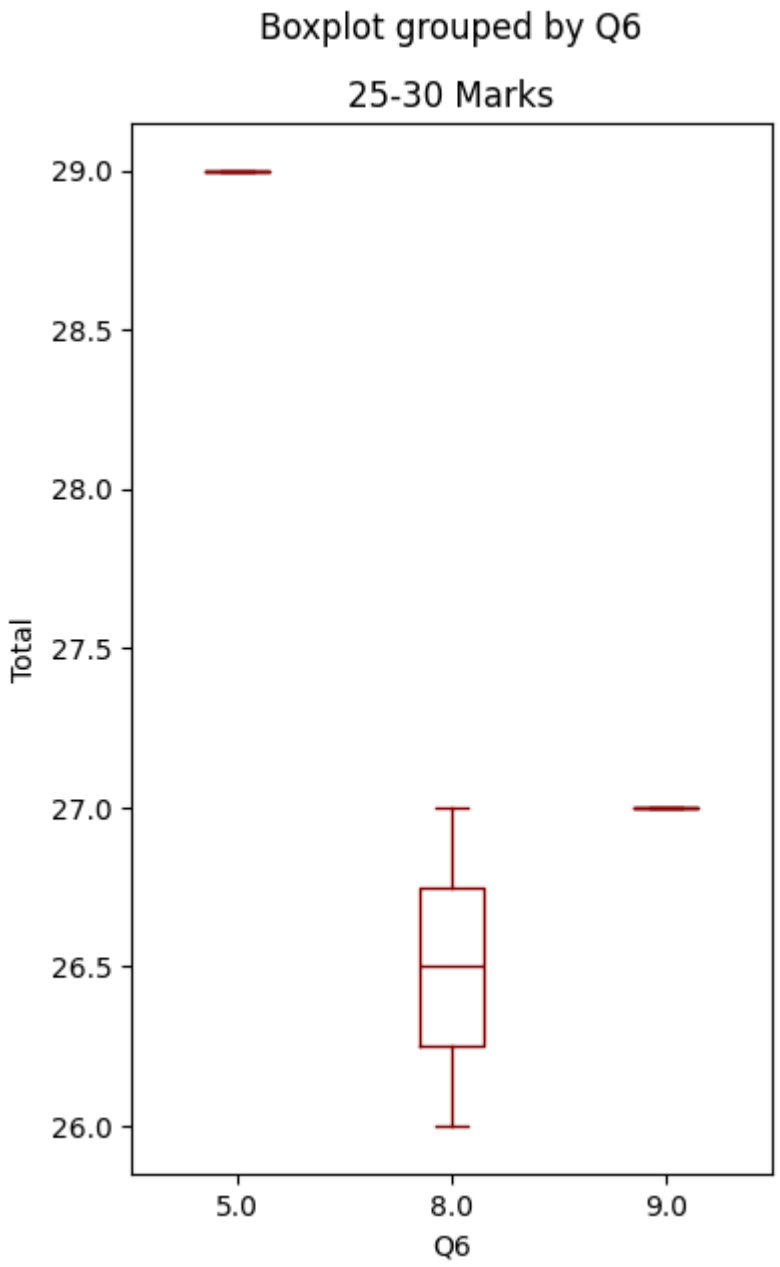
No student in this range secured full mark for thisquestionMinimum mark is 0 and the ones who attemptedsecured low marks.

In [63]:
```
c.boxplot(by='Q5M10', column =['Total'], grid = False,color='purple',figsize=[3
plt.title("25-30 Marks")
plt.ylabel("Total")
plt.show()
```



It can be observed that many students scored marksbetween 4 and 9, the minimum mark remains 0 andnone of the students scored full marks in thisquestion.

In [65]:
```
c.boxplot(by='Q6', column =['Total'], grid = False,color='maroon',figsize=[4,7]
plt.title("25-30 Marks")
plt.ylabel("Total")
plt.show()
```



None of the students who attempted the questionscored full markMany of them didn't attempt this question so minimummark remains 0.

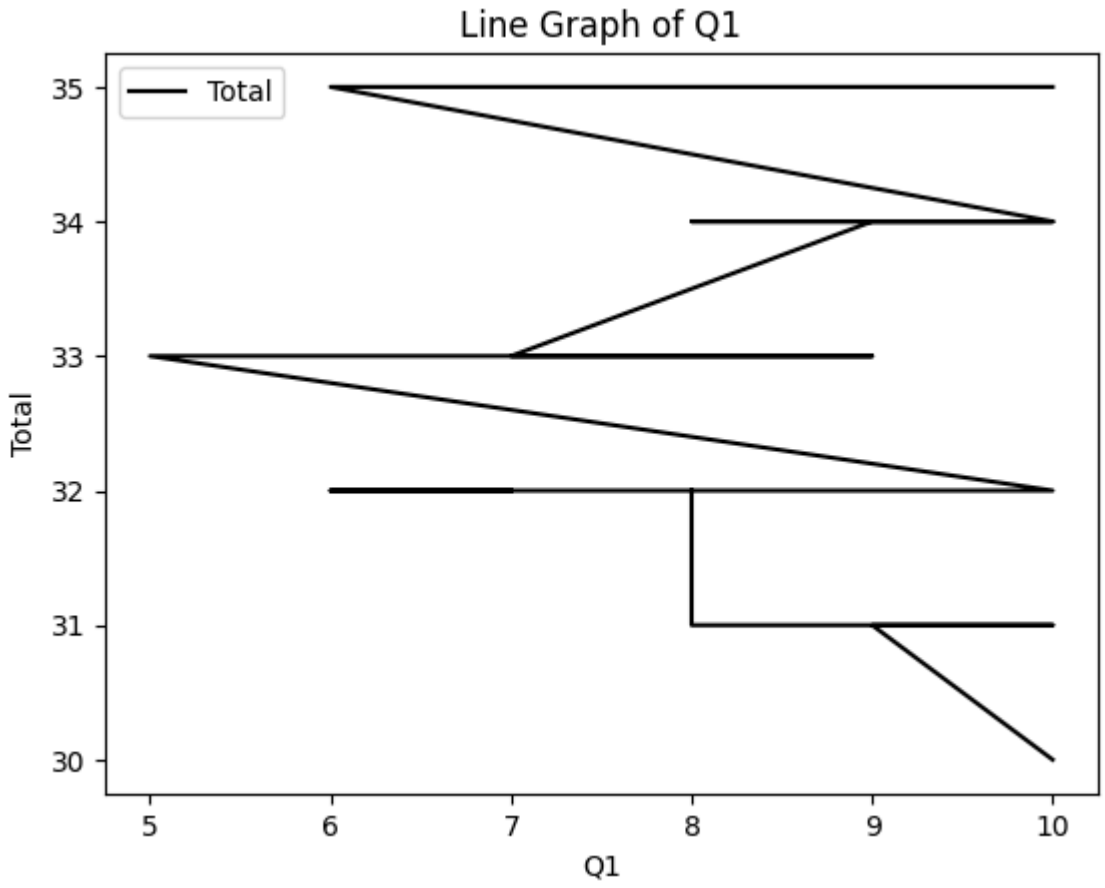
In [67]:
```
d=DF.loc[(DF['Total']>=30) & (DF['Total']<=35)]
d=d.reset_index()
d
```

Out[67]:

	index	Total	Q1aM4	Q1bM6	Q2aM6	Q2bM4	Q3aM5	Q3bM5	Q4aM3	Q4bM7
0	35	30	4.0	6.0	6.0	4.0	NaN	1.0	NaN	NaN
1	16	30	4.0	NaN	6.0	4.0	5.0	2.0	NaN	NaN
2	14	30	4.0	6.0	6.0	2.0	4.0	5.0	3.0	NaN
3	66	31	4.0	5.0	5.0	2.0	5.0	3.0	1.0	5.0
4	80	31	4.0	6.0	6.0	2.0	2.0	5.0	NaN	NaN
5	37	31	4.0	4.0	6.0	4.0	NaN	NaN	NaN	NaN
6	27	32	2.0	6.0	6.0	1.0	5.0	5.0	3.0	3.0
7	15	32	3.0	NaN	2.0	1.0	5.0	5.0	3.0	7.0
8	1	32	4.0	3.0	4.0	3.0	NaN	NaN	3.0	6.0
9	13	32	3.0	3.0	6.0	4.0	3.0	5.0	NaN	NaN
10	81	32	3.0	6.0	3.0	4.0	5.0	3.0	NaN	NaN
11	32	32	4.0	6.0	6.0	4.0	2.0	NaN	NaN	NaN
12	78	33	2.0	3.0	6.0	4.0	5.0	5.0	NaN	NaN
13	43	33	4.0	5.0	NaN	NaN	NaN	NaN	3.0	4.0
14	2	33	4.0	5.0	5.0	1.0	5.0	5.0	NaN	NaN
15	24	33	1.0	6.0	6.0	3.0	5.0	5.0	3.0	3.0
16	58	34	4.0	5.0	6.0	3.0	NaN	NaN	3.0	NaN
17	7	34	4.0	6.0	6.0	4.0	NaN	NaN	2.0	NaN
18	12	34	4.0	4.0	5.0	3.0	2.0	2.0	2.0	1.0
19	45	34	2.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN
20	21	34	4.0	6.0	5.0	3.0	5.0	5.0	NaN	3.0
21	9	35	2.0	4.0	5.0	4.0	5.0	5.0	NaN	NaN
22	56	35	2.0	6.0	NaN	NaN	NaN	NaN	3.0	7.0
23	70	35	4.0	6.0	6.0	4.0	5.0	5.0	NaN	NaN

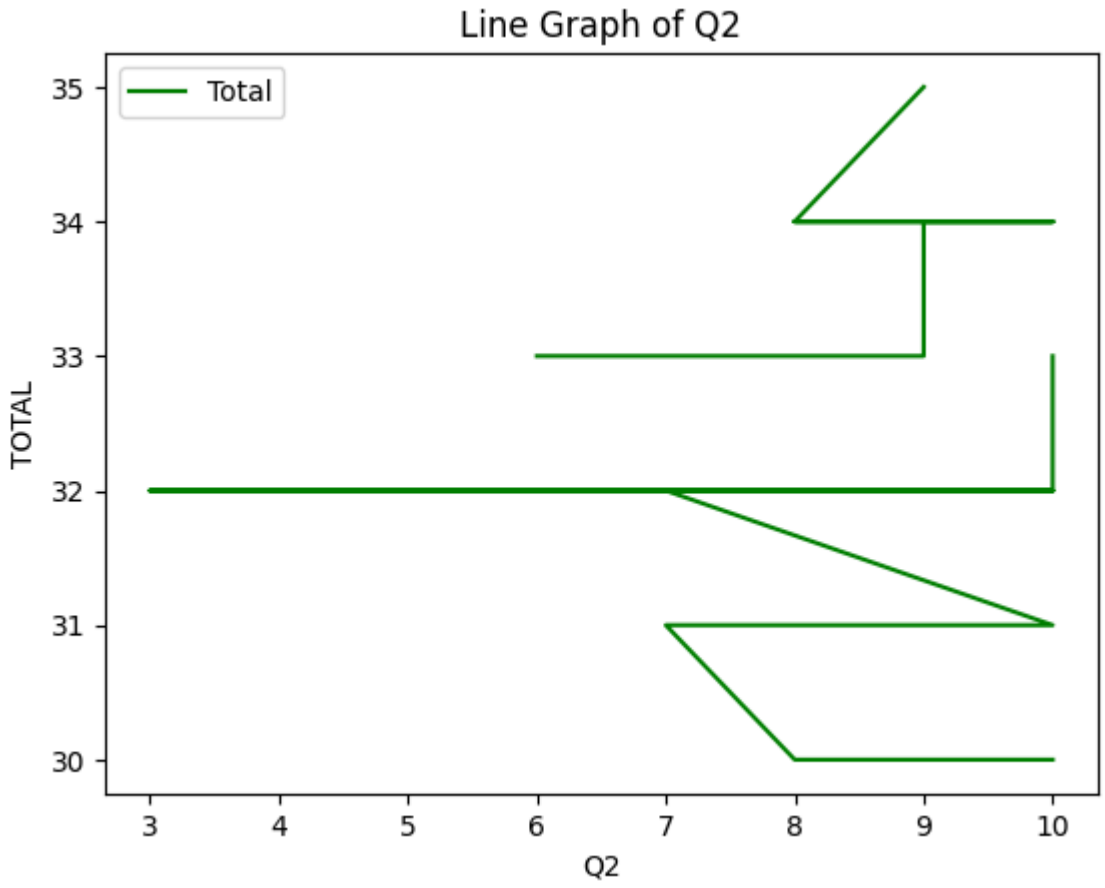
Total marks 30-35 is filtered from the data set

In [73]:
```
d.plot.line(x='Q1',y='Total',color='black')
plt.title("Line Graph of Q1")
plt.ylabel("Total")
plt.show()
```



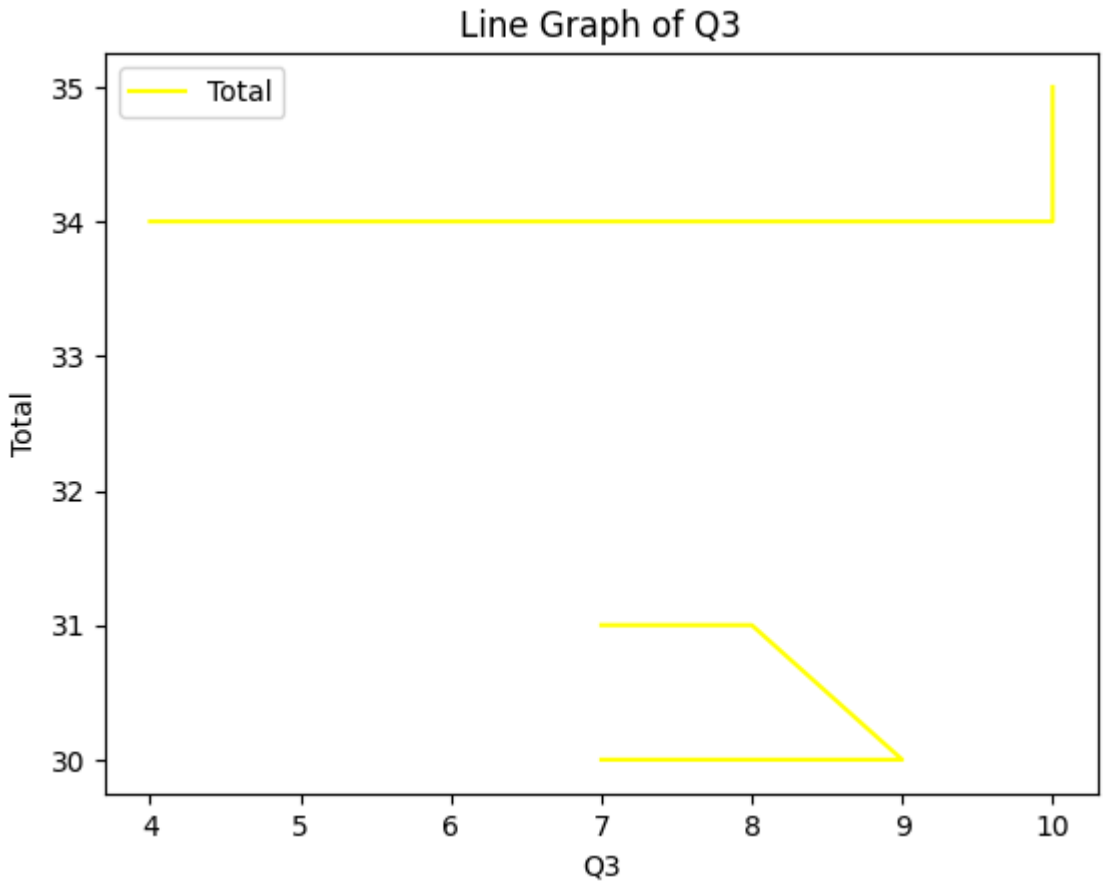
It can be deduced from the above line graph thatmajority of the students in this range scored marksbetween 7 and 10.

In [77]:
```
d.plot.line(x='Q2',y='Total',color='green')
plt.title("Line Graph of Q2")
plt.ylabel("TOTAL")
plt.show()
```



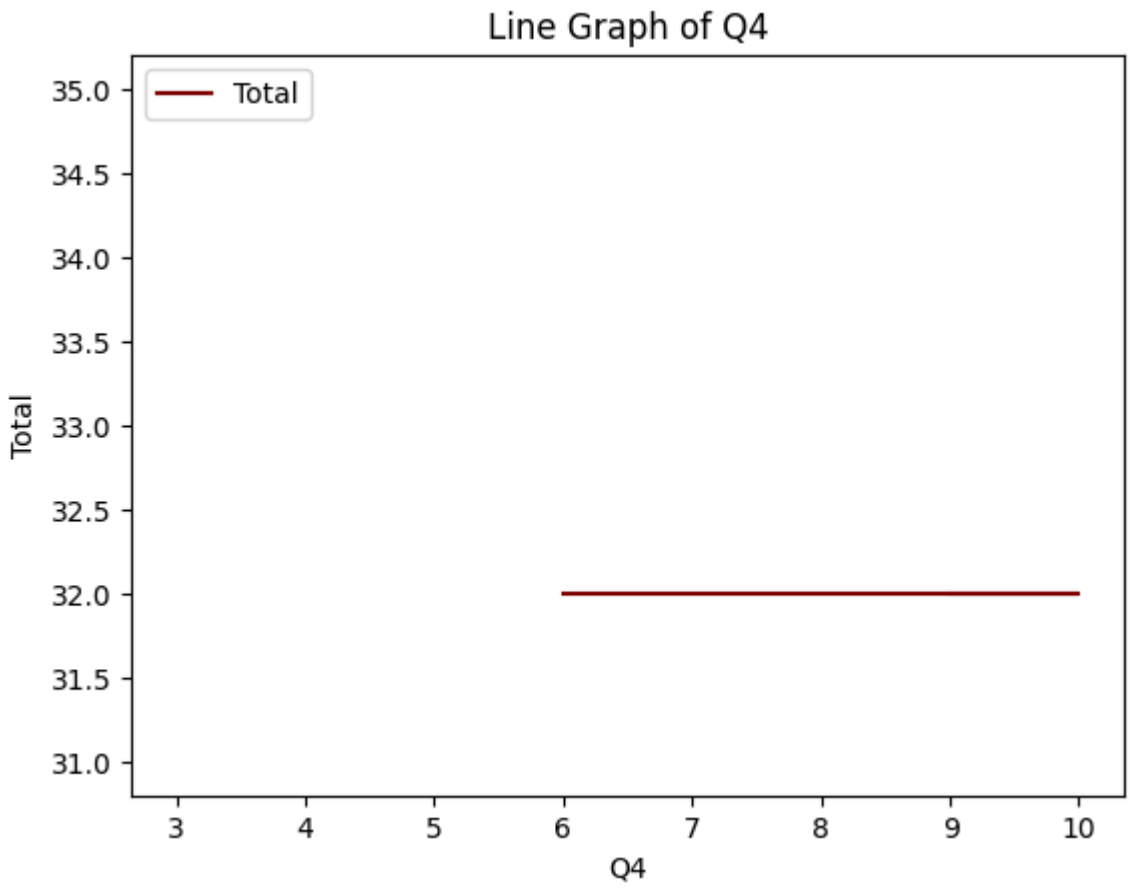
The students who attempted this question scoredmarks between 8 and 10 overall, also the minimummark is zero whereas the maximum mark is 10.

In [79]:
```
d.plot.line(x='Q3',y='Total',color='yellow')
plt.title("Line Graph of Q3")
plt.ylabel("Total")
plt.show()
```



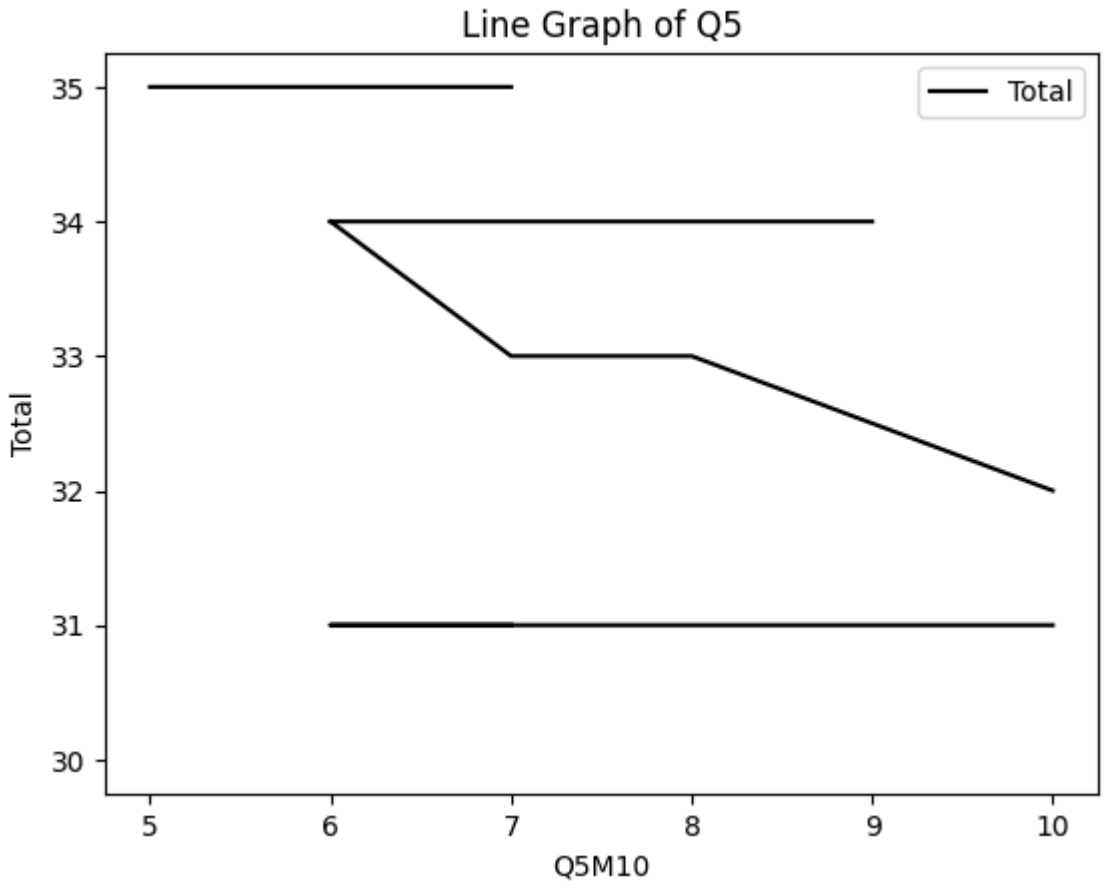
The students in this range scored 6-10 marks on anaverage

In [81]:
```
d.plot.line(x='Q4',y='Total',color='maroon')
plt.title("Line Graph of Q4")
plt.ylabel("Total")
plt.show()
```



The minimum mark for the question is 0 whereashighest mark scored is 10

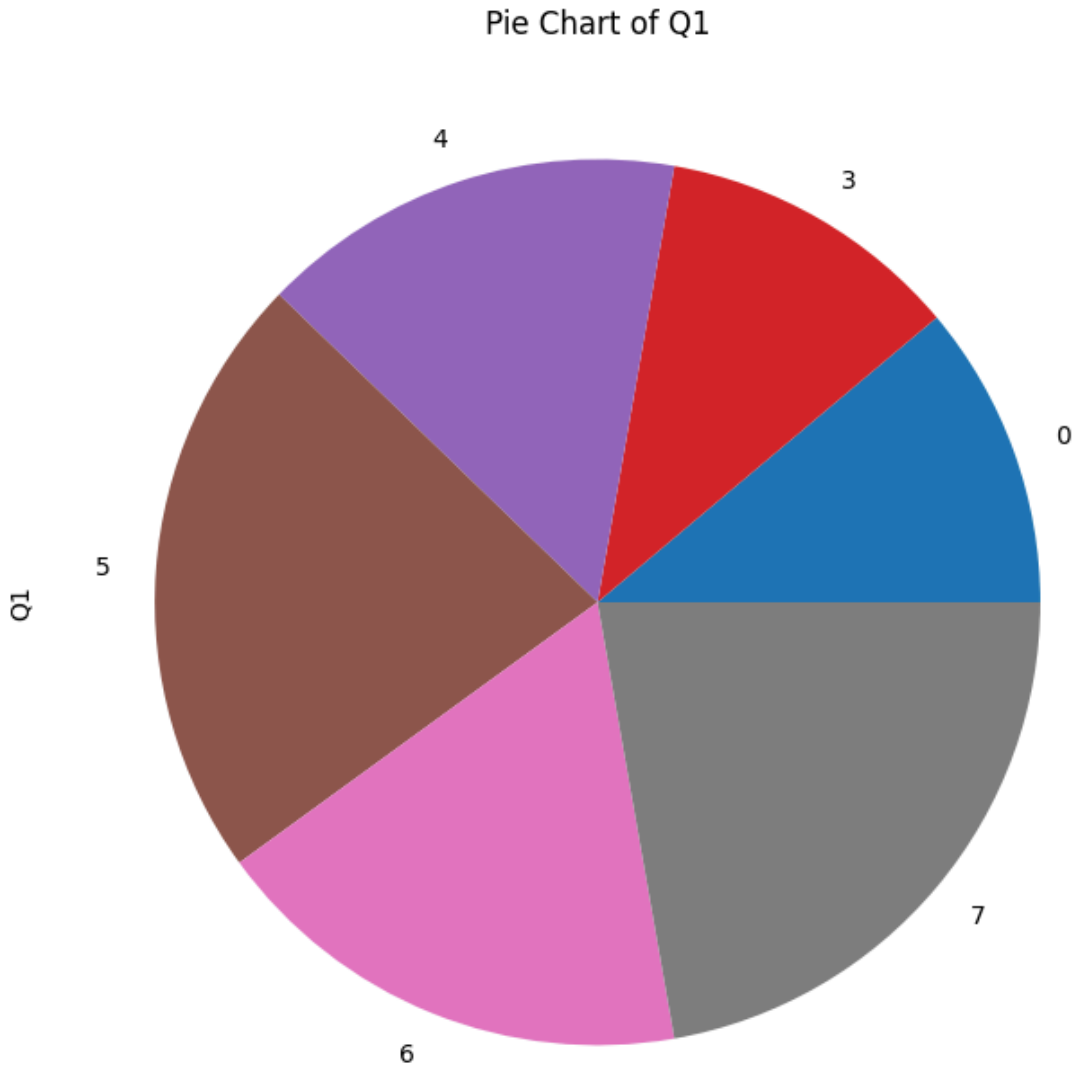
In [83]:
```
d.plot.line(x='Q5M10',y='Total',color='black')
plt.title("Line Graph of Q5")
plt.ylabel("Total")
plt.show()
```



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

In [92]:
```
a['Q1'].plot(kind='pie',subplots=True,figsize=(8,8))
plt.title("Pie Chart of Q1")
```

Out[92]:
```
Text(0.5, 1.0, 'Pie Chart of Q1')
```



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

AMITH YADAV

2211CS010360

In [ ]: