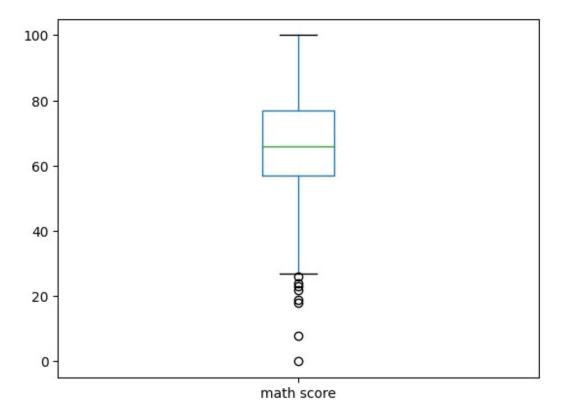
```
#Assignment 2
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read csv('StudentsPerformance.csv')
df
     gender race/ethnicity parental level of education
lunch
     female
                    group B
                                       bachelor's degree
                                                               standard
                                            some college
1
     female
                                                               standard
                    group C
     female
                                         master's degree
                                                               standard
                    group B
                                      associate's degree
                                                           free/reduced
       male
                    group A
       male
                    group C
                                            some college
                                                               standard
     female
995
                                         master's degree
                                                               standard
                    group E
996
       male
                                             high school
                                                           free/reduced
                    group C
997
     female
                                                           free/reduced
                                             high school
                    group C
     female
998
                    group D
                                            some college
                                                               standard
999
     female
                                            some college free/reduced
                    group D
    test preparation course
                              math score
                                           reading score writing score
0
                                       72
                                                       72
                                                                       74
                        none
1
                   completed
                                       69
                                                       90
                                                                       88
2
                                       90
                                                       95
                                                                       93
                        none
3
                                       47
                                                       57
                                                                       44
                        none
                                       76
                                                       78
                                                                       75
                        none
                                                       99
                                                                       95
995
                   completed
                                       88
996
                                       62
                                                       55
                                                                       55
                        none
```

007	1	50	71	65						
997	completed	59	71	65						
998	completed	68	78	77						
999	none	77	86	86						
[1000 rows x 8 columns]										
<pre>df.head()</pre>										
gender rad 0 female 1 female 2 female 3 male 4 male	ce/ethnicity pare group B group C group B group A group C	master's associate's	degree some degree some degree free/	lunch \ tandard tandard tandard reduced tandard						
			•	ng score						
0 1 2 3 4	none completed none none none	72 69 90 47 76	72 90 95 57 78	74 88 93 44 75						
df.tail()										
gender race/ethnicity parental level of education										
	race, etimiterty pe									
lunch \ 995 female	group E	master	's degree	standard						
lunch \ 995 female	group E		-							
lunch \ 995 female 996 male	group E group C	hi	gh school free	e/reduced						
lunch \ 995 female 996 male 997 female	group E group C group C	hiọ hiọ	gh school free	e/reduced e/reduced						
lunch \ 995 female 996 male 997 female 998 female	group E group C group C group D	hi hi some	gh school free gh school free e college	e/reduced e/reduced standard						
lunch \ 995 female 996 male 997 female	group E group C group C	hi hi some	gh school free gh school free e college	e/reduced e/reduced						
lunch \ 995 female 996 male 997 female 998 female 999 female	group E group C group C group D group D	hiq hiq some	gh school free gh school free e college e college free	e/reduced e/reduced standard e/reduced						
lunch \ 995 female 996 male 997 female 998 female 999 female test pre	group E group C group C group D group D	hiq some some math score read:	gh school free gh school free e college e college free ing score wri	e/reduced e/reduced standard e/reduced ting score						
lunch \ 995 female 996 male 997 female 998 female 999 female test prep	group E group C group D group D caration course completed	high some some some math score read:	gh school free gh school free e college e college free ing score wri	e/reduced e/reduced standard e/reduced ting score						
lunch \ 995 female 996 male 997 female 998 female 999 female test prep 995 996	group E group C group C group D group D coaration course completed none	high some some some some 88	gh school free gh school free e college e college free ing score wri	e/reduced e/reduced standard e/reduced ting score 95						
lunch \ 995 female 996 male 997 female 998 female 999 female test prep	group E group C group D group D caration course completed	high some some some math score read:	gh school free gh school free e college e college free ing score wri	e/reduced e/reduced standard e/reduced ting score						

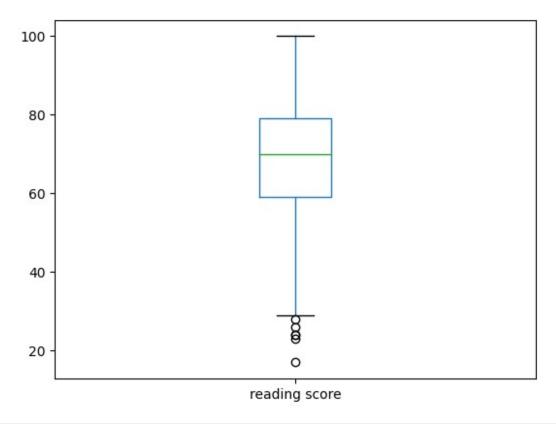
999		none	77	86		86			
<pre>df.describe()</pre>									
count mean std min 25% 50% 75% max df.isr	1000.00000 66.08900 15.16308 0.00000 57.00000 66.00000 77.00000 100.00000	reading score 1000.000000 69.169000 14.600192 17.000000 59.000000 70.000000 79.000000 100.000000	writing score 1000.000000 68.054000 15.195657 10.000000 57.750000 69.000000 79.000000 100.000000						
		thnicity parer False	ntal level of e	education False	False False False				
t	test preparati	on course math	n score readin	ig score	writing	score			
0		False	False	False		False			
1		False	False	False		False			
2		False	False	False		False			
3		False	False	False		False			
4		False	False	False		False			
995		False	False	False		False			
996		False	False	False		False			
997		False	False	False		False			
998		False	False	False		False			

```
999          False     False     False
[1000 rows x 8 columns]

def plot_boxplot(df,ft):
    df.boxplot(column=[ft])
    plt.grid(False)
plt.show()
plot_boxplot(df,'math score')
```

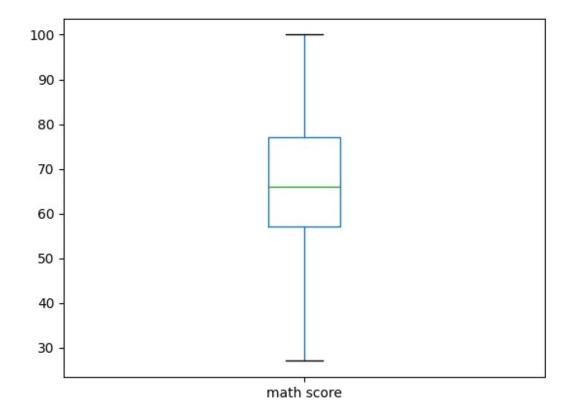


```
def plot_boxplot(df,ft):
    df.boxplot(column=[ft])
    plt.grid(False)
plt.show()
plot_boxplot(df,'reading score')
```



```
def outliers(df,ft):
   Q1=df[ft].quantile(0.25)
   Q3=df[ft].quantile(0.75)
   IQR=Q3-Q1
   lower_bound=Q1-1.5 *IQR
   upper bound=Q3 + 1.5 *IQR
   ls=df.index[(df[ft] < lower_bound) | (df[ft] > upper_bound)]
   return ls
index list=[]
for features in ['math score', 'reading score']:
    index list.extend(outliers(df,features))
index_list
[17, 59, 145, 338, 466, 787, 842, 980, 59, 76, 211, 327, 596, 980]
def remove(df,ls):
    ls=sorted(set(ls))
    df=df.drop(ls)
    return df
df cleaned=remove(df,index list)
df_cleaned.shape
(988, 8)
```

plot_boxplot(df_cleaned,'math score')



plot_boxplot(df_cleaned,'reading score')

