In [1]:

```
import pandas as pd
import numpy as np
```

In [2]:

```
dict1={"Maths_Score":[60,62,10,78,77,300,78,80,200,67],
       "Reading_Score":[70,288,90,54,89,65,80,65,89,20],
       "Writing_Score":[70,43,80,56,95,56,99,45,90,100],
      "Placement_Count":[500,1,8,5,9,255,6,9,5,3],
      "Region":["Mumbai","Buldhana","Mumbai","Baner","Nagpur","Dadar","Indore","Wardha","
      "Gender":["Male","Female","Female","Male","Male","Male","Female","Male","Female","M
      "Placement_Count_Year":[2020,2022,2018,2017,2023,2015,2012,2020,2010,2005]}
```

In [3]:

```
df=pd.DataFrame(dict1)
df
```

Out[3]:

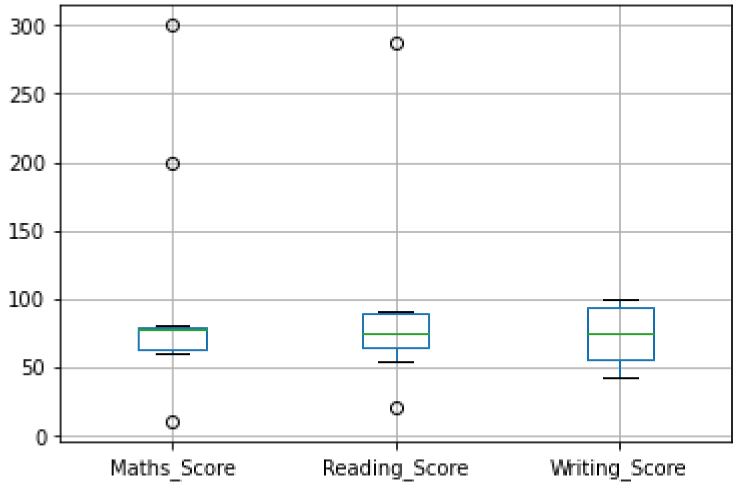
	Maths_Score	Reading_Score	Writing_Score	Placement_Count	Region	Gender	Placeme
0	60	70	70	500	Mumbai	Male	
1	62	288	43	1	Buldhana	Female	
2	10	90	80	8	Mumbai	Female	
3	78	54	56	5	Baner	Male	
4	77	89	95	9	Nagpur	Male	
5	300	65	56	255	Dadar	Male	
6	78	80	99	6	Indore	Female	
7	80	65	45	9	Wardha	Male	
8	200	89	90	5	Indore	Female	
9	67	20	100	3	Mumbai	Male	

In [4]:

```
col1=['Maths_Score','Reading_Score','Writing_Score']
df.boxplot(col1)
```

Out[4]:

```
<AxesSubplot:>
```



In [5]:

```
rscore=df['Reading_Score']
q1=np.percentile(rscore,25)
q3=np.percentile(rscore,75)
print(q1,q3)
```

```
65.0 89.0
```

In [6]:

```
iqr=q3-q1
print(iqr)
```

```
24.0
```

In [7]:

```
l_bound=q1-1.5*iqr
u_bound=q3+1.5*iqr
print(l_bound,u_bound)
```

```
29.0 125.0
```

In [8]:

```
r_outlier=[]
for i in rscore:
    if(i<l_bound or i>u_bound):
        r_outlier.append(i)
print(r_outlier)
```

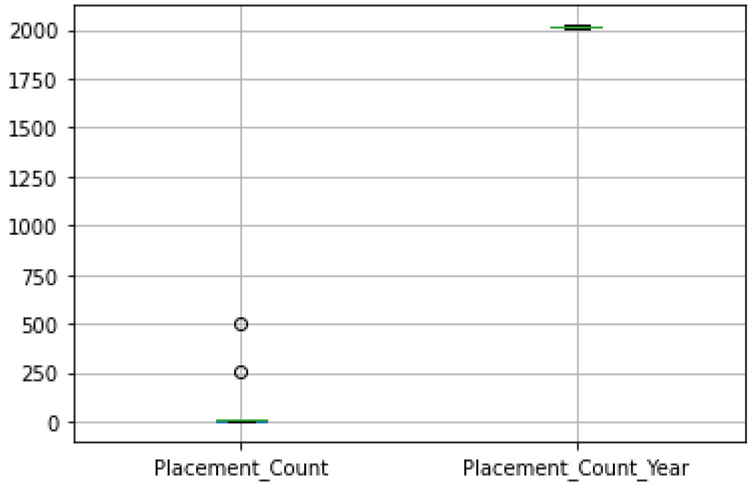
```
[288, 20]
```

In [9]:

```
col2=["Placement_Count","Placement_Count_Year"]
df.boxplot(col2)
```

Out[9]:

```
<AxesSubplot:>
```



In [10]:

```
pscore=df["Placement_Count"]
Q1=np.percentile(pscore,25)
Q3=np.percentile(pscore,75)
print(Q1,Q3)
```

```
5.0 9.0
```

In [11]:

```
IQR=Q3-Q1
print(IQR)
```

```
4.0
```

In [12]:

```
L_BOUND=Q1-1.5*IQR
U_BOUND=Q3+1.5*IQR
print(L_BOUND,U_BOUND)
```

```
-1.0 15.0
```

In [13]:

```
outlier=[]
for i in pscore:
    if(i<L_BOUND or i>U_BOUND):
        outlier.append(i)
print(outlier)
```

```
[500, 255]
```

In [14]:

```
median=np.median(rscore)
median
```

Out[14]:

```
75.0
```

In [15]:

```
df['Reading_Score']=np.where(df['Reading_Score']>u_bound,median,df['Reading_Score'])
```

In [16]:

```
df['Reading_Score']
```

Out[16]:

```
0    70.0
1    75.0
2    90.0
3    54.0
4    89.0
5    65.0
6    80.0
7    65.0
8    89.0
9    20.0
Name: Reading_Score, dtype: float64
```

In [17]:

```
median=np.median(rscore)
df['Reading_Score']=np.where(df['Reading_Score']<l_bound,median,df['Reading_Score'])
df['Reading_Score']
```

Out[17]:

```
0    70.0
1    75.0
2    90.0
3    54.0
4    89.0
5    65.0
6    80.0
7    65.0
8    89.0
9    75.0
Name: Reading_Score, dtype: float64
```

In [18]:

```
MEDIAN=np.median(pscore)
df['Placement_Count']=np.where(df['Placement_Count']<L_BOUND,MEDIAN,df['Placement_Count'
df['Placement_Count']
```

Out[18]:

```
0    500.0
1      1.0
2      8.0
3      5.0
4      9.0
5    255.0
6      6.0
7      9.0
8      5.0
9      3.0
Name: Placement_Count, dtype: float64
```

In [19]:

```
MEDIAN=np.median(pscore)
df['Placement_Count']=np.where(df['Placement_Count']>U_BOUND,MEDIAN,df['Placement_Count'
df['Placement_Count']
```

Out[19]:

```
0    7.0
1    1.0
2    8.0
3    5.0
4    9.0
5    7.0
6    6.0
7    9.0
8    5.0
9    3.0
Name: Placement_Count, dtype: float64
```

In [20]:

```
mscore=df['Maths_Score']
a1=np.percentile(mscore,25)
a3=np.percentile(mscore,75)
print(a1,a3)
```

```
63.25 79.5
```

In [21]:

```
IQR=a3-a1
print(IQR)
```

```
16.25
```

In [22]:

```
L_BOUND=a1-1.5*IQR
U_BOUND=a3+1.5*IQR
print(L_BOUND,U_BOUND)
```

```
38.875 103.875
```

In [23]:

```
outlier=[]
for i in mscore:
    if(i<L_BOUND or i>U_BOUND):
        outlier.append(i)
print(outlier)
```

```
[10, 300, 200]
```

In [24]:

```
MEDIAN=np.median(mscore)
df['Maths_Score']=np.where(df['Maths_Score']>U_BOUND,MEDIAN,df['Maths_Score'])
df['Maths_Score']
```

Out[24]:

```
0    60.0
1    62.0
2    10.0
3    78.0
4    77.0
5    77.5
6    78.0
7    80.0
8    77.5
9    67.0
Name: Maths_Score, dtype: float64
```

In [25]:

```
MEDIAN=np.median(mscore)
df['Maths_Score']=np.where(df['Maths_Score']<L_BOUND,MEDIAN,df['Maths_Score'])
df['Maths_Score']
```

Out[25]:

```
0    60.0
1    62.0
2    77.5
3    78.0
4    77.0
5    77.5
6    78.0
7    80.0
8    77.5
9    67.0
Name: Maths_Score, dtype: float64
```

In [26]:

```
df=pd.get_dummies(df["Region"])
df
```

Out[26]:

	Baner	Buldhana	Dadar	Indore	Mumbai	Nagpur	Wardha
0	0	0	0	0	1	0	0
1	0	1	0	0	0	0	0
2	0	0	0	0	1	0	0
3	1	0	0	0	0	0	0
4	0	0	0	0	0	1	0
5	0	0	1	0	0	0	0
6	0	0	0	1	0	0	0
7	0	0	0	0	0	0	1
8	0	0	0	1	0	0	0
9	0	0	0	0	1	0	0

In [27]:

```
#from sklearn.preprocessing import LabelEncoder
#label_encoder = LabelEncoder()
#df['Gender']=label_encoder.fit_transform(df['Gender'])
#df
```

In [32]:

```
#from sklearn.preprocessing import OneHotEncoder
#ohe_encoder = OneHotEncoder()
#df['Region']=ohe_encoder.fit_transform(df[['Region']]).toarray()
#df
```