```
import pandas as pd
```

```
df=pd.read_csv("/content/SAMPLEIDS.csv")
```

```
df
```

	SNO	REGNO	NAME	DOB	GENDER	ADDRESS	M1	M2	M3	M4	TOTAL	AVG
0	1	1220121	ARUN	2000-02-10	MALE	THANDALAM	82.0	81.0	90.0	NaN	NaN	NaN
1	2	1220122	BABU	1999-01-25	MALE	KANCHIPURAM	56.0	61.0	80.0	56.0	253.0	84.333333
2	3	1220123	CHARAN	2000.09.21	MALE	THANDALAM	NaN	59.0	60.0	70.0	NaN	0.000000
3	4	1220124	DEVA	2000-11-09	MALE	POONAMALEE	74.0	79.0	80.0	74.0	307.0	102.333333
4	5	1220125	ESTER	2000-11-21	FEMALE	CHITHUR	92.0	95.0	96.0	92.0	375.0	125.000000
5	6	1220126	FARHANA	1999-03-05	FEMALE	THANDALAM	91.0	88.0	90.0	91.0	360.0	120.000000
6	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	49.0	219.0	73.000000
7	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	49.0	219.0	73.000000
8	8	1220128	HEMA	1999-01-25	FEMALE	POONAMALEE	95.0	96.0	90.0	95.0	376.0	125.333333
9	9	1220129	INDRA	2000.09.21	FEMALE	KANCHIPURAM	64.0	NaN	NaN	64.0	NaN	0.000000
10	10	1220130	JAHITH	2000-11-09	MALE	THANDALAM	34.0	45.0	50.0	34.0	163.0	54.333333
11	11	1220131	KANI	2000-11-21	FEMALE	CHITHUR	96.0	95.0	96.0	96.0	383.0	127.666667
12	12	1220132	LATHESSH	1999-03-05	MALE	THANDALAM	NaN	68.0	70.0	70.0	208.0	69.333333
13	13	1220133	MANI	2000-10-02	MALE	KANCHIPURAM	71.0	76.0	NaN	71.0	NaN	0.000000
14	14	1220134	NANI	20001109	MALE	POONAMALEE	79.0	77.0	80.0	79.0	315.0	105.000000
15	15	1220135	NaN	19990125	NaN	NaN	NaN	NaN	NaN	NaN	0.0	0.000000
16	16	1220136	PRATHAP	20000921	MALE	KANCHIPURAM	86.0	84.0	90.0	86.0	346.0	115.333333
17	17	1220137	RAGHU	20001109	MALE	POONAMALEE	67.0	64.0	70.0	NaN	201.0	67.000000
18	18	1220138	RATHI	20001121	FEMALE	KANCHIPURAM	81.0	86.0	90.0	81.0	338.0	112.666667
19	19	1220139	SARVESH	19990305	MALE	THANDALAM	84.0	87.0	NaN	84.0	NaN	0.000000
20	20	1220140	SANTHOSH	20001002	MALE	KANCHIPURAM	76.0	69.0	80.0	76.0	301.0	100.333333

```
print(df.head(7))
```

```
   SNO    REGNO     NAME         DOB  GENDER      ADDRESS    M1    M2    M3  \
0    1  1220121     ARUN  2000-02-10    MALE    THANDALAM  82.0  81.0  90.0
1    2  1220122     BABU  1999-01-25    MALE  KANCHIPURAM  56.0  61.0  80.0
2    3  1220123   CHARAN  2000.09.21    MALE    THANDALAM   NaN  59.0  60.0
3    4  1220124     DEVA  2000-11-09    MALE   POONAMALEE  74.0  79.0  80.0
4    5  1220125    ESTER  2000-11-21  FEMALE      CHITHUR  92.0  95.0  96.0
5    6  1220126  FARHANA  1999-03-05  FEMALE    THANDALAM  91.0  88.0  90.0
6    7  1220127     GANI  2000-10-02    MALE  KANCHIPURAM  49.0  51.0  70.0

    M4  TOTAL         AVG
0  NaN    NaN         NaN
1  56.0  253.0   84.333333
2  70.0    NaN    0.000000
3  74.0  307.0  102.333333
4  92.0  375.0  125.000000
5  91.0  360.0  120.000000
6  49.0  219.0   73.000000
```

```
print(df.tail(2))
```

```
    SNO    REGNO      NAME        DOB GENDER      ADDRESS    M1    M2    M3  \
19   19  1220139   SARVESH  19990305   MALE    THANDALAM  84.0  87.0   NaN
20   20  1220140  SANTHOSH  20001002   MALE  KANCHIPURAM  76.0  69.0  80.0

      M4  TOTAL         AVG
19  84.0    NaN    0.000000
20  76.0  301.0  100.333333
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21 entries, 0 to 20
Data columns (total 12 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
```

```
 0   SNO      21 non-null     int64
 1   REGNO    21 non-null     int64
 2   NAME     20 non-null     object
 3   DOB      21 non-null     object
 4   GENDER   20 non-null     object
 5   ADDRESS  20 non-null     object
 6   M1       18 non-null     float64
 7   M2       19 non-null     float64
 8   M3       17 non-null     float64
 9   M4       18 non-null     float64
 10  TOTAL    16 non-null     float64
 11  AVG      20 non-null     float64
dtypes: float64(6), int64(2), object(4)
memory usage: 2.1+ KB
```

```
print(df.describe())
```

```
              SNO         REGNO         M1         M2         M3         M4  \
count  21.000000  2.100000e+01  18.000000  19.000000  17.000000  18.000000
mean   10.333333  1.220130e+06  73.666667  74.315789  79.529412  73.166667
std     5.816643  5.816643e+00  17.580069  15.836149  13.010177  17.426315
min     1.000000  1.220121e+06  34.000000  45.000000  50.000000  34.000000
25%     6.000000  1.220126e+06  64.750000  62.500000  70.000000  65.500000
50%    10.000000  1.220130e+06  77.500000  77.000000  80.000000  75.000000
75%    15.000000  1.220135e+06  85.500000  86.500000  90.000000  85.500000
max    20.000000  1.220140e+06  96.000000  96.000000  96.000000  96.000000

             TOTAL         AVG
count   16.000000   20.000000
mean   272.750000   72.733333
std    102.048681   48.017127
min      0.000000    0.000000
25%    216.250000   40.750000
50%    304.000000   78.666667
75%    349.500000  113.333333
max    383.000000  127.666667
```

```
df.isnull().sum()
```

```
SNO        0
REGNO      0
NAME       1
DOB        0
GENDER     1
ADDRESS    1
M1         3
M2         2
M3         4
M4         3
TOTAL      5
AVG        1
dtype: int64
```

```
df.nunique()
```

```
SNO        20
REGNO      20
NAME       19
DOB        13
GENDER      2
ADDRESS     4
M1         17
M2         17
M3          6
M4         16
TOTAL      15
AVG        15
dtype: int64
```

```
mn=df.TOTAL.mean()
mn
```

```
272.75
```

```
df.TOTAL.fillna(mn,inplace=True)
df
```

	SNO	REGNO	NAME	DOB	GENDER	ADDRESS	M1	M2	M3	M
0	1	1220121	ARUN	2000-02-10	MALE	THANDALAM	82.0	81.0	90.0	Na
1	2	1220122	BABU	1999-01-25	MALE	KANCHIPURAM	56.0	61.0	80.0	56
2	3	1220123	CHARAN	2000.09.21	MALE	THANDALAM	NaN	59.0	60.0	70
3	4	1220124	DEVA	2000-11-09	MALE	POONAMALEE	74.0	79.0	80.0	74
4	5	1220125	ESTER	2000-11-21	FEMALE	CHITHUR	92.0	95.0	96.0	92
5	6	1220126	FARHANA	1999-03-05	FEMALE	THANDALAM	91.0	88.0	90.0	91
6	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	49
7	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	49
8	8	1220128	HEMA	1999-01-25	FEMALE	POONAMALEE	95.0	96.0	90.0	95
9	9	1220129	INDRA	2000.09.21	FEMALE	KANCHIPURAM	64.0	NaN	NaN	64
10	10	1220130	JAHITH	2000-11-09	MALE	THANDALAM	34.0	45.0	50.0	34
11	11	1220131	KANI	2000-11-21	FEMALE	CHITHUR	96.0	95.0	96.0	96
12	12	1220132	LATHESSH	1999-03-05	MALE	THANDALAM	NaN	68.0	70.0	70
13	13	1220133	MANI	2000-10-02	MALE	KANCHIPURAM	71.0	76.0	NaN	71

```
min=df.M4.min()
min
```

```
34.0
```

```
df.M4.fillna(min,inplace=True)
df
```

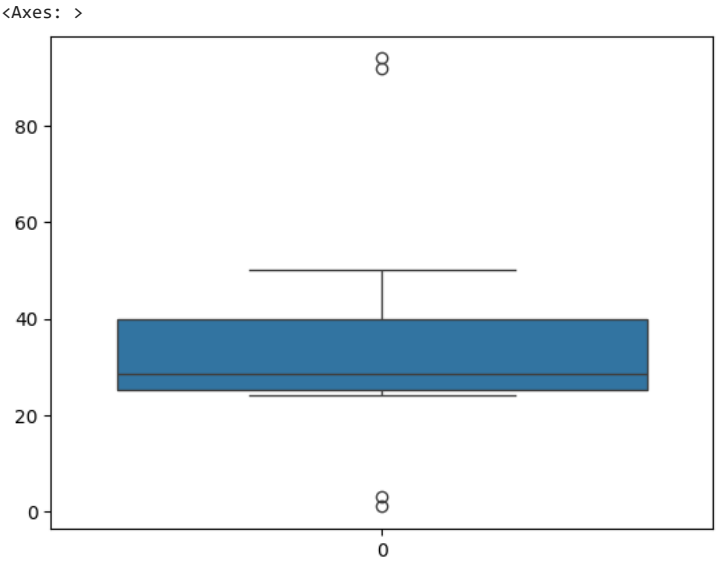
	SNO	REGNO	NAME	DOB	GENDER	ADDRESS	M1	M2	M3	M4	TOTAL	AVG
0	1	1220121	ARUN	2000-02-10	MALE	THANDALAM	82.0	81.0	90.0	34.0	272.75	NaN
1	2	1220122	BABU	1999-01-25	MALE	KANCHIPURAM	56.0	61.0	80.0	56.0	253.00	84.333333
2	3	1220123	CHARAN	2000.09.21	MALE	THANDALAM	NaN	59.0	60.0	70.0	272.75	0.000000
3	4	1220124	DEVA	2000-11-09	MALE	POONAMALEE	74.0	79.0	80.0	74.0	307.00	102.333333
4	5	1220125	ESTER	2000-11-21	FEMALE	CHITHUR	92.0	95.0	96.0	92.0	375.00	125.000000
5	6	1220126	FARHANA	1999-03-05	FEMALE	THANDALAM	91.0	88.0	90.0	91.0	360.00	120.000000
6	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	49.0	219.00	73.000000
7	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	49.0	219.00	73.000000
8	8	1220128	HEMA	1999-01-25	FEMALE	POONAMALEE	95.0	96.0	90.0	95.0	376.00	125.333333
9	9	1220129	INDRA	2000.09.21	FEMALE	KANCHIPURAM	64.0	NaN	NaN	64.0	272.75	0.000000
10	10	1220130	JAHITH	2000-11-09	MALE	THANDALAM	34.0	45.0	50.0	34.0	163.00	54.333333
11	11	1220131	KANI	2000-11-21	FEMALE	CHITHUR	96.0	95.0	96.0	96.0	383.00	127.666667
12	12	1220132	LATHESSH	1999-03-05	MALE	THANDALAM	NaN	68.0	70.0	70.0	208.00	69.333333
13	13	1220133	MANI	2000-10-02	MALE	KANCHIPURAM	71.0	76.0	NaN	71.0	272.75	0.000000
14	14	1220134	NANI	20001109	MALE	POONAMALEE	79.0	77.0	80.0	79.0	315.00	105.000000
15	15	1220135	NaN	19990125	NaN	NaN	NaN	NaN	NaN	34.0	0.00	0.000000
16	16	1220136	PRATHAP	20000921	MALE	KANCHIPURAM	86.0	84.0	90.0	86.0	346.00	115.333333
17	17	1220137	RAGHU	20001109	MALE	POONAMALEE	67.0	64.0	70.0	34.0	201.00	67.000000
18	18	1220138	RATHI	20001121	FEMALE	KANCHIPURAM	81.0	86.0	90.0	81.0	338.00	112.666667
19	19	1220139	SARVESH	19990305	MALE	THANDALAM	84.0	87.0	NaN	84.0	272.75	0.000000
20	20	1220140	SANTHOSH	20001002	MALE	KANCHIPURAM	76.0	69.0	80.0	76.0	301.00	100.333333

```
import pandas as pd
import seaborn as sns
age=[1,3,28,27,25,92,30,39,40,50,26,24,29,94]
af=pd.DataFrame(age)
af
```

	0
0	1
1	3
2	28
3	27
4	25
5	92
6	30
7	39
8	40
9	50
10	26
11	24
12	29
13	94

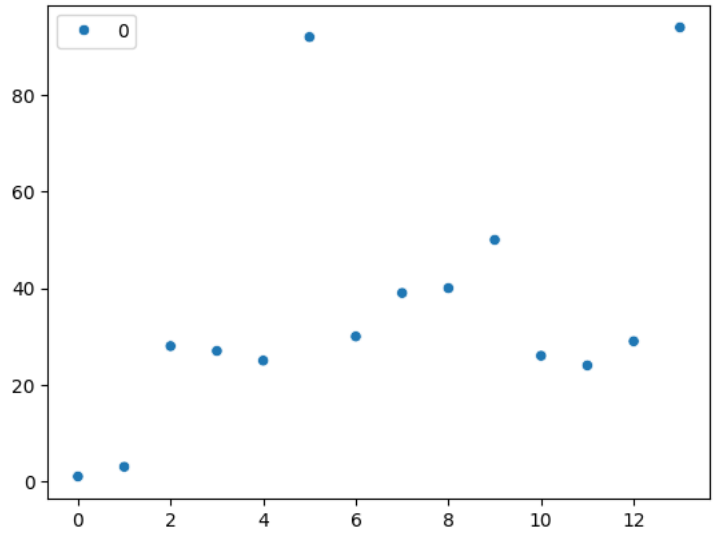
```
sns.boxplot(data=af)
```

```
<Axes: >
```




```
sns.scatterplot(data=af)
```

```
<Axes: >
```




```
q1=af.quantile(0.25)
q2=af.quantile(0.50)
q3=af.quantile(0.75)
iqr=q3-q1
iqr
low=q1-1.5*iqr
low
high=q3+1.5*iqr
high
```

```
0    61.5
dtype: float64
```

```
af=af[((af>=low)&(af<=high))]
af
```

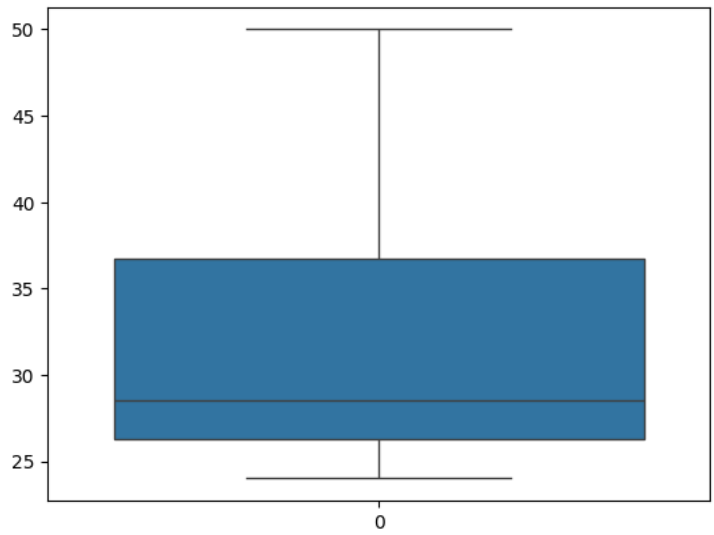
	0
0	NaN
1	NaN
2	28.0
3	27.0
4	25.0
5	NaN
6	30.0
7	39.0
8	40.0
9	50.0
10	26.0
11	24.0
12	29.0
13	NaN

```
af.dropna()
```

	0
2	28.0
3	27.0
4	25.0
6	30.0
7	39.0
8	40.0
9	50.0
10	26.0
11	24.0
12	29.0

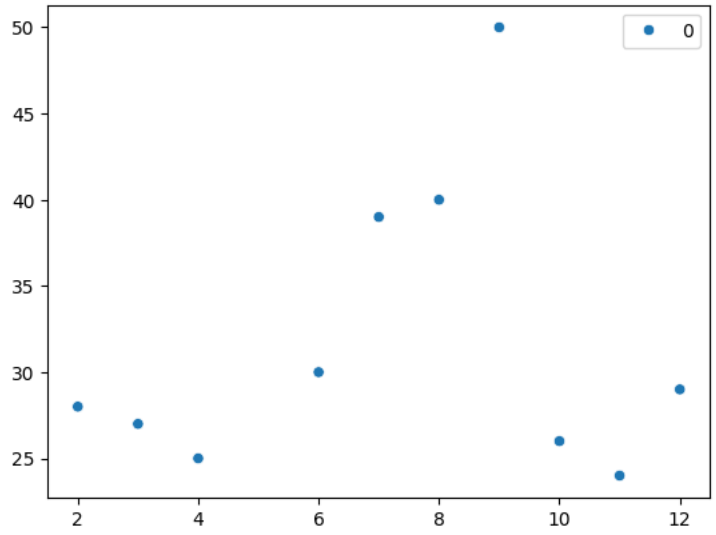
```
sns.boxplot(data=af)
```

```
<Axes: >
```




```
sns.scatterplot(data=af)
```

```
<Axes: >
```




```
data=[1,12,15,18,21,24,27,30,33,36,39,42,45,48,51,54,57,60,63,66,69,72,75,78,81,84,87,90,93,96,99,102,105]
df=pd.DataFrame(data)
df
```

	0
0	1

```
import numpy as np
from scipy import stats
z=np.abs(stats.zscore(df))
z
```

	0