dsbda-practical-2

February 9, 2024

```
[2]: import pandas as pd
```

```
[3]: import numpy as num
```

```
[5]: df = pd.read_csv(r"E:\BE\6th sem\Data Science and Big Data analytics\Practical␣
      ↪Outputs\customer_shopping_data.csv")
```

```
[6]: df
```

```
[6]:       invoice_no customer_id  gender  age          category  quantity    price  \
     0        I138884     C241288  Female   28          Clothing         5  1500.40
     1        I317333     C111565    Male   21             Shoes         3  1800.51
     2        I127801     C266599    Male   20          Clothing         1   300.08
     3        I173702     C988172  Female   66             Shoes         5  3000.85
     4        I337046     C189076  Female   53             Books         4    60.60
     …            …           …       …   …                 …       …       …
     99452    I219422     C441542  Female   45          Souvenir         5    58.65
     99453    I325143     C569580    Male   27  Food & Beverage         2    10.46
     99454    I824010     C103292    Male   63  Food & Beverage         2    10.46
     99455    I702964     C800631    Male   56        Technology         4  4200.00
     99456    I232867     C273973  Female   36          Souvenir         3    35.19

           payment_method invoice_date     shopping_mall
     0        Credit Card     5/8/2022            Kanyon
     1         Debit Card   12/12/2021    Forum Istanbul
     2               Cash    9/11/2021         Metrocity
     3        Credit Card   16/05/2021      Metropol AVM
     4               Cash   24/10/2021            Kanyon
     …                …           …                 …
     99452    Credit Card   21/09/2022            Kanyon
     99453           Cash   22/09/2021    Forum Istanbul
     99454     Debit Card   28/03/2021         Metrocity
     99455           Cash   16/03/2021      Istinye Park
     99456    Credit Card   15/10/2022  Mall of Istanbul

     [99457 rows x 10 columns]
```

[8]:

```
df.isnull()
```

[8]:

```
       invoice_no  customer_id  gender    age  category  quantity  price  \
0           False        False   False  False     False     False  False
1           False        False   False  False     False     False  False
2           False        False   False  False     False     False  False
3           False        False   False  False     False     False  False
4           False        False   False  False     False     False  False
…           …          …     …    …       …       …    …
99452       False        False   False  False     False     False  False
99453       False        False   False  False     False     False  False
99454       False        False   False  False     False     False  False
99455       False        False   False  False     False     False  False
99456       False        False   False  False     False     False  False

       payment_method  invoice_date  shopping_mall
0               False         False          False
1               False         False          False
2               False         False          False
3               False         False          False
4               False         False          False
…               …           …            …
99452           False         False          False
99453           False         False          False
99454           False         False          False
99455           False         False          False
99456           False         False          False

[99457 rows x 10 columns]
```

[12]:

```
series = pd.isnull(df["invoice_no"])
df[series]
```

[12]:

```
Empty DataFrame
Columns: [invoice_no, customer_id, gender, age, category, quantity, price,
payment_method, invoice_date, shopping_mall]
Index: []
```

[13]:

```
series1 = pd.notnull(df["customer_id"])
df[series1]
```

[13]:

```
      invoice_no customer_id  gender  age         category  quantity    price  \
0        I138884     C241288  Female   28         Clothing         5  1500.40
1        I317333     C111565    Male   21            Shoes         3  1800.51
2        I127801     C266599    Male   20         Clothing         1   300.08
3        I173702     C988172  Female   66            Shoes         5  3000.85
4        I337046     C189076  Female   53            Books         4    60.60
```

```
…          …         …     …  …              …       …      …
99452    I219422     C441542  Female   45          Souvenir         5    58.65
99453    I325143     C569580    Male   27  Food & Beverage          2    10.46
99454    I824010     C103292    Male   63  Food & Beverage          2    10.46
99455    I702964     C800631    Male   56       Technology          4  4200.00
99456    I232867     C273973  Female   36          Souvenir         3    35.19

      payment_method invoice_date      shopping_mall
0        Credit Card     5/8/2022             Kanyon
1         Debit Card   12/12/2021     Forum Istanbul
2               Cash    9/11/2021          Metrocity
3        Credit Card   16/05/2021       Metropol AVM
4               Cash   24/10/2021             Kanyon
…              …          …                …
99452    Credit Card   21/09/2022             Kanyon
99453           Cash   22/09/2021     Forum Istanbul
99454     Debit Card   28/03/2021          Metrocity
99455           Cash   16/03/2021       Istinye Park
99456    Credit Card   15/10/2022  Mall of Istanbul

[99457 rows x 10 columns]
```

```
[22]: df = pd.read_csv(r"E:\BE\6th sem\Data Science and Big Data analytics\Practical␣
       ↪Outputs\StudentsPerformance.csv")
```

```
[15]: df
```

```
[15]:     Math Score  Reading Score  Writing Score  Placement Score  Club Join Year  \
      0         95.0           81.0           93.0             85.0              23
      1         83.0           79.0           64.0             97.0              23
      2        100.0           96.0           70.0             86.0              23
      3         88.0           83.0          100.0             81.0              23
      4         73.0           72.0           96.0             62.0              23
      5         63.0           76.0            NaN            100.0              23
      6         72.0           87.0           99.0             94.0              23
      7         90.0           75.0           91.0             66.0              23
      8         97.0           79.0           74.0             77.0              23
      9          NaN           60.0           77.0             70.0              23
      10        77.0           65.0           61.0             90.0              23
      11        75.0           94.0           88.0             91.0              23
      12        85.0           79.0           63.0             93.0              23
      13        80.0           79.0          100.0              NaN              23
      14        81.0           81.0           88.0             61.0              23
      15        80.0           64.0           98.0            100.0              23
      16        66.0           85.0           94.0             84.0              23
      17        99.0           80.0           75.0             93.0              23
      18        85.0            NaN           81.0             99.0              23
```

```
19        76.0          99.0          89.0            84.0             23
20        34.0          23.0          45.0            21.0             23
21        43.0          23.0          34.0            56.0             23
22        45.0          65.0          67.0            43.0             23
23        12.0          24.0           NaN            10.0             23
24        23.0          21.0          43.0            54.0             23
25        11.0          12.0          13.0            11.0             23
26        13.0          14.0          54.0            44.0             23
27        43.0          54.0          65.0            76.0             23
28        43.0          44.0          54.0            65.0             23

    Placement  Offer Count  gender
0        98.0           99  female
1        66.0           97  female
2        84.0           69  female
3        79.0           79  female
4        64.0           64  female
5        83.0           74  female
6        72.0           68  female
7        88.0           86  female
8        90.0           91  female
9        81.0           69    Male
10       60.0           67    Male
11       90.0           71    Male
12       99.0           80    Male
13       76.0           95    Male
14       91.0           84    Male
15       69.0           96    Male
16       60.0           62    Male
17       72.0           76    Male
18       63.0           73    Male
19       97.0           87    Male
20       45.0           65  female
21       56.0           56  female
22       45.0           76  female
23       56.0           87  female
24       54.0           23  female
25        NaN           16    Male
26       77.0           65    Male
27       43.0           43    Male
28       11.0           12    Male
```

[23]:

```python
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['gender'] = le.fit_transform(df['gender'])
newdf=df
df
```

```
[23]:     Math Score  Reading Score  Writing Score  Placement Score  Club Join Year  \
      0         95.0           81.0           93.0             85.0              23
      1         83.0           79.0           64.0             97.0              23
      2        100.0           96.0           70.0             86.0              23
      3         88.0           83.0          100.0             81.0              23
      4         73.0           72.0           96.0             62.0              23
      5         63.0           76.0            NaN            100.0              23
      6         72.0           87.0           99.0             94.0              23
      7         90.0           75.0           91.0             66.0              23
      8         97.0           79.0           74.0             77.0              23
      9          NaN           60.0           77.0             70.0              23
      10        77.0           65.0           61.0             90.0              23
      11        75.0           94.0           88.0             91.0              23
      12        85.0           79.0           63.0             93.0              23
      13        80.0           79.0          100.0              NaN              23
      14        81.0           81.0           88.0             61.0              23
      15        80.0           64.0           98.0            100.0              23
      16        66.0           85.0           94.0             84.0              23
      17        99.0           80.0           75.0             93.0              23
      18        85.0            NaN           81.0             99.0              23
      19        76.0           99.0           89.0             84.0              23
      20        34.0           23.0           45.0             21.0              23
      21        43.0           23.0           34.0             56.0              23
      22        45.0           65.0           67.0             43.0              23
      23        12.0           24.0            NaN             10.0              23
      24        23.0           21.0           43.0             54.0              23
      25        11.0           12.0           13.0             11.0              23
      26        13.0           14.0           54.0             44.0              23
      27        43.0           54.0           65.0             76.0              23
      28        43.0           44.0           54.0             65.0              23

          Placement  Offer Count  gender
      0        98.0           99       1
      1        66.0           97       1
      2        84.0           69       1
      3        79.0           79       1
      4        64.0           64       1
      5        83.0           74       1
      6        72.0           68       1
      7        88.0           86       1
      8        90.0           91       1
      9        81.0           69       0
      10       60.0           67       0
      11       90.0           71       0
      12       99.0           80       0
      13       76.0           95       0
      14       91.0           84       0
```

```
15        69.0             96        0
16        60.0             62        0
17        72.0             76        0
18        63.0             73        0
19        97.0             87        0
20        45.0             65        1
21        56.0             56        1
22        45.0             76        1
23        56.0             87        1
24        54.0             23        1
25         NaN             16        0
26        77.0             65        0
27        43.0             43        0
28        11.0             12        0
```

[24]:

```
m_v=df['Math Score'].mean()
df['Math Score'].fillna(value=m_v, inplace=True)
df
```

```
C:\Users\Rohan\AppData\Local\Temp\ipykernel_2356\464449955.py:2: FutureWarning:
A value is trying to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.


  df['Math Score'].fillna(value=m_v, inplace=True)
```

[24]:

```
    Math Score  Reading Score  Writing Score  Placement Score  Club Join Year  \
0    95.000000           81.0           93.0             85.0              23
1    83.000000           79.0           64.0             97.0              23
2   100.000000           96.0           70.0             86.0              23
3    88.000000           83.0          100.0             81.0              23
4    73.000000           72.0           96.0             62.0              23
5    63.000000           76.0            NaN            100.0              23
6    72.000000           87.0           99.0             94.0              23
7    90.000000           75.0           91.0             66.0              23
8    97.000000           79.0           74.0             77.0              23
9    65.428571           60.0           77.0             70.0              23
10   77.000000           65.0           61.0             90.0              23
11   75.000000           94.0           88.0             91.0              23
12   85.000000           79.0           63.0             93.0              23
```

```
13   80.000000           79.0          100.0              NaN              23
14   81.000000           81.0           88.0             61.0              23
15   80.000000           64.0           98.0            100.0              23
16   66.000000           85.0           94.0             84.0              23
17   99.000000           80.0           75.0             93.0              23
18   85.000000            NaN           81.0             99.0              23
19   76.000000           99.0           89.0             84.0              23
20   34.000000           23.0           45.0             21.0              23
21   43.000000           23.0           34.0             56.0              23
22   45.000000           65.0           67.0             43.0              23
23   12.000000           24.0            NaN             10.0              23
24   23.000000           21.0           43.0             54.0              23
25   11.000000           12.0           13.0             11.0              23
26   13.000000           14.0           54.0             44.0              23
27   43.000000           54.0           65.0             76.0              23
28   43.000000           44.0           54.0             65.0              23

    Placement  Offer Count  gender
0        98.0           99       1
1        66.0           97       1
2        84.0           69       1
3        79.0           79       1
4        64.0           64       1
5        83.0           74       1
6        72.0           68       1
7        88.0           86       1
8        90.0           91       1
9        81.0           69       0
10       60.0           67       0
11       90.0           71       0
12       99.0           80       0
13       76.0           95       0
14       91.0           84       0
15       69.0           96       0
16       60.0           62       0
17       72.0           76       0
18       63.0           73       0
19       97.0           87       0
20       45.0           65       1
21       56.0           56       1
22       45.0           76       1
23       56.0           87       1
24       54.0           23       1
25        NaN           16       0
26       77.0           65       0
27       43.0           43       0
28       11.0           12       0
```

[29]:
```
import numpy as np
df.replace(to_replace=np.nan, value=-99, inplace=True)
```

[30]:
```
df
```

[30]:
```
    Math Score  Reading Score  Writing Score  Placement Score  Club Join Year  \
0    95.000000           81.0           93.0             85.0              23
1    83.000000           79.0           64.0             97.0              23
2   100.000000           96.0           70.0             86.0              23
3    88.000000           83.0          100.0             81.0              23
4    73.000000           72.0           96.0             62.0              23
5    63.000000           76.0          -99.0            100.0              23
6    72.000000           87.0           99.0             94.0              23
7    90.000000           75.0           91.0             66.0              23
8    97.000000           79.0           74.0             77.0              23
9    65.428571           60.0           77.0             70.0              23
10   77.000000           65.0           61.0             90.0              23
11   75.000000           94.0           88.0             91.0              23
12   85.000000           79.0           63.0             93.0              23
13   80.000000           79.0          100.0            -99.0              23
14   81.000000           81.0           88.0             61.0              23
15   80.000000           64.0           98.0            100.0              23
16   66.000000           85.0           94.0             84.0              23
17   99.000000           80.0           75.0             93.0              23
18   85.000000          -99.0           81.0             99.0              23
19   76.000000           99.0           89.0             84.0              23
20   34.000000           23.0           45.0             21.0              23
21   43.000000           23.0           34.0             56.0              23
22   45.000000           65.0           67.0             43.0              23
23   12.000000           24.0          -99.0             10.0              23
24   23.000000           21.0           43.0             54.0              23
25   11.000000           12.0           13.0             11.0              23
26   13.000000           14.0           54.0             44.0              23
27   43.000000           54.0           65.0             76.0              23
28   43.000000           44.0           54.0             65.0              23

    Placement Offer Count  gender
0        98.0          99       1
1        66.0          97       1
2        84.0          69       1
3        79.0          79       1
4        64.0          64       1
5        83.0          74       1
6        72.0          68       1
7        88.0          86       1
8        90.0          91       1
9        81.0          69       0
```

```
10         60.0             67        0
11         90.0             71        0
12         99.0             80        0
13         76.0             95        0
14         91.0             84        0
15         69.0             96        0
16         60.0             62        0
17         72.0             76        0
18         63.0             73        0
19         97.0             87        0
20         45.0             65        1
21         56.0             56        1
22         45.0             76        1
23         56.0             87        1
24         54.0             23        1
25        -99.0             16        0
26         77.0             65        0
27         43.0             43        0
28         11.0             12        0
```

[31]:
```
df.dropna()
```

[31]:
```
    Math Score  Reading Score  Writing Score  Placement Score  Club Join Year  \
0    95.000000           81.0           93.0             85.0              23
1    83.000000           79.0           64.0             97.0              23
2   100.000000           96.0           70.0             86.0              23
3    88.000000           83.0          100.0             81.0              23
4    73.000000           72.0           96.0             62.0              23
5    63.000000           76.0          -99.0            100.0              23
6    72.000000           87.0           99.0             94.0              23
7    90.000000           75.0           91.0             66.0              23
8    97.000000           79.0           74.0             77.0              23
9    65.428571           60.0           77.0             70.0              23
10   77.000000           65.0           61.0             90.0              23
11   75.000000           94.0           88.0             91.0              23
12   85.000000           79.0           63.0             93.0              23
13   80.000000           79.0          100.0            -99.0              23
14   81.000000           81.0           88.0             61.0              23
15   80.000000           64.0           98.0            100.0              23
16   66.000000           85.0           94.0             84.0              23
17   99.000000           80.0           75.0             93.0              23
18   85.000000          -99.0           81.0             99.0              23
19   76.000000           99.0           89.0             84.0              23
20   34.000000           23.0           45.0             21.0              23
21   43.000000           23.0           34.0             56.0              23
22   45.000000           65.0           67.0             43.0              23
23   12.000000           24.0          -99.0             10.0              23
```

```
24   23.000000         21.0         43.0          54.0            23
25   11.000000         12.0         13.0          11.0            23
26   13.000000         14.0         54.0          44.0            23
27   43.000000         54.0         65.0          76.0            23
28   43.000000         44.0         54.0          65.0            23

    Placement  Offer Count  gender
0        98.0           99       1
1        66.0           97       1
2        84.0           69       1
3        79.0           79       1
4        64.0           64       1
5        83.0           74       1
6        72.0           68       1
7        88.0           86       1
8        90.0           91       1
9        81.0           69       0
10       60.0           67       0
11       90.0           71       0
12       99.0           80       0
13       76.0           95       0
14       91.0           84       0
15       69.0           96       0
16       60.0           62       0
17       72.0           76       0
18       63.0           73       0
19       97.0           87       0
20       45.0           65       1
21       56.0           56       1
22       45.0           76       1
23       56.0           87       1
24       54.0           23       1
25      -99.0           16       0
26       77.0           65       0
27       43.0           43       0
28       11.0           12       0
```

[33]:
```
df.dropna(how = 'all')
```

[33]:
```
    Math Score  Reading Score  Writing Score  Placement Score  Club Join Year  \
0    95.000000           81.0           93.0             85.0              23
1    83.000000           79.0           64.0             97.0              23
2   100.000000           96.0           70.0             86.0              23
3    88.000000           83.0          100.0             81.0              23
4    73.000000           72.0           96.0             62.0              23
5    63.000000           76.0          -99.0            100.0              23
6    72.000000           87.0           99.0             94.0              23
```

```
7    90.000000           75.0           91.0             66.0              23
8    97.000000           79.0           74.0             77.0              23
9    65.428571           60.0           77.0             70.0              23
10   77.000000           65.0           61.0             90.0              23
11   75.000000           94.0           88.0             91.0              23
12   85.000000           79.0           63.0             93.0              23
13   80.000000           79.0          100.0            -99.0              23
14   81.000000           81.0           88.0             61.0              23
15   80.000000           64.0           98.0            100.0              23
16   66.000000           85.0           94.0             84.0              23
17   99.000000           80.0           75.0             93.0              23
18   85.000000          -99.0           81.0             99.0              23
19   76.000000           99.0           89.0             84.0              23
20   34.000000           23.0           45.0             21.0              23
21   43.000000           23.0           34.0             56.0              23
22   45.000000           65.0           67.0             43.0              23
23   12.000000           24.0          -99.0             10.0              23
24   23.000000           21.0           43.0             54.0              23
25   11.000000           12.0           13.0             11.0              23
26   13.000000           14.0           54.0             44.0              23
27   43.000000           54.0           65.0             76.0              23
28   43.000000           44.0           54.0             65.0              23

    Placement  Offer Count  gender
0        98.0           99       1
1        66.0           97       1
2        84.0           69       1
3        79.0           79       1
4        64.0           64       1
5        83.0           74       1
6        72.0           68       1
7        88.0           86       1
8        90.0           91       1
9        81.0           69       0
10       60.0           67       0
11       90.0           71       0
12       99.0           80       0
13       76.0           95       0
14       91.0           84       0
15       69.0           96       0
16       60.0           62       0
17       72.0           76       0
18       63.0           73       0
19       97.0           87       0
20       45.0           65       1
21       56.0           56       1
22       45.0           76       1
```

```
23       56.0           87       1
24       54.0           23       1
25      -99.0           16       0
26       77.0           65       0
27       43.0           43       0
28       11.0           12       0
```

[34]:
```
df.dropna(axis = 1)
```

[34]:
```
    Math Score  Reading Score  Writing Score  Placement Score  Club Join Year  \
0    95.000000           81.0           93.0             85.0              23
1    83.000000           79.0           64.0             97.0              23
2   100.000000           96.0           70.0             86.0              23
3    88.000000           83.0          100.0             81.0              23
4    73.000000           72.0           96.0             62.0              23
5    63.000000           76.0          -99.0            100.0              23
6    72.000000           87.0           99.0             94.0              23
7    90.000000           75.0           91.0             66.0              23
8    97.000000           79.0           74.0             77.0              23
9    65.428571           60.0           77.0             70.0              23
10   77.000000           65.0           61.0             90.0              23
11   75.000000           94.0           88.0             91.0              23
12   85.000000           79.0           63.0             93.0              23
13   80.000000           79.0          100.0            -99.0              23
14   81.000000           81.0           88.0             61.0              23
15   80.000000           64.0           98.0            100.0              23
16   66.000000           85.0           94.0             84.0              23
17   99.000000           80.0           75.0             93.0              23
18   85.000000          -99.0           81.0             99.0              23
19   76.000000           99.0           89.0             84.0              23
20   34.000000           23.0           45.0             21.0              23
21   43.000000           23.0           34.0             56.0              23
22   45.000000           65.0           67.0             43.0              23
23   12.000000           24.0          -99.0             10.0              23
24   23.000000           21.0           43.0             54.0              23
25   11.000000           12.0           13.0             11.0              23
26   13.000000           14.0           54.0             44.0              23
27   43.000000           54.0           65.0             76.0              23
28   43.000000           44.0           54.0             65.0              23

    Placement  Offer Count  gender
0        98.0           99       1
1        66.0           97       1
2        84.0           69       1
3        79.0           79       1
4        64.0           64       1
5        83.0           74       1
```

```
6        72.0           68       1
7        88.0           86       1
8        90.0           91       1
9        81.0           69       0
10       60.0           67       0
11       90.0           71       0
12       99.0           80       0
13       76.0           95       0
14       91.0           84       0
15       69.0           96       0
16       60.0           62       0
17       72.0           76       0
18       63.0           73       0
19       97.0           87       0
20       45.0           65       1
21       56.0           56       1
22       45.0           76       1
23       56.0           87       1
24       54.0           23       1
25      -99.0           16       0
26       77.0           65       0
27       43.0           43       0
28       11.0           12       0
```

[35]:

```
new_data = df.dropna(axis = 0, how ='any')
new_data
```

[35]:

```
    Math Score  Reading Score  Writing Score  Placement Score  Club Join Year  \
0    95.000000           81.0           93.0             85.0              23
1    83.000000           79.0           64.0             97.0              23
2   100.000000           96.0           70.0             86.0              23
3    88.000000           83.0          100.0             81.0              23
4    73.000000           72.0           96.0             62.0              23
5    63.000000           76.0          -99.0            100.0              23
6    72.000000           87.0           99.0             94.0              23
7    90.000000           75.0           91.0             66.0              23
8    97.000000           79.0           74.0             77.0              23
9    65.428571           60.0           77.0             70.0              23
10   77.000000           65.0           61.0             90.0              23
11   75.000000           94.0           88.0             91.0              23
12   85.000000           79.0           63.0             93.0              23
13   80.000000           79.0          100.0            -99.0              23
14   81.000000           81.0           88.0             61.0              23
15   80.000000           64.0           98.0            100.0              23
16   66.000000           85.0           94.0             84.0              23
17   99.000000           80.0           75.0             93.0              23
18   85.000000          -99.0           81.0             99.0              23
```

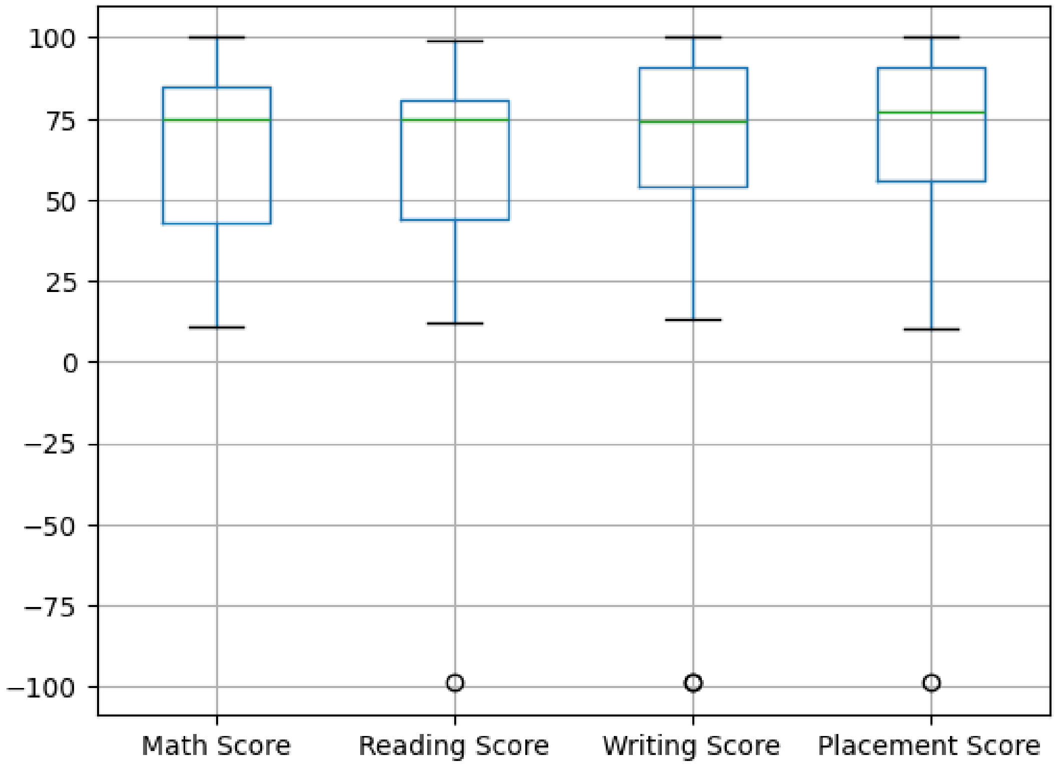
```
19   76.000000           99.0          89.0           84.0          23
20   34.000000           23.0          45.0           21.0          23
21   43.000000           23.0          34.0           56.0          23
22   45.000000           65.0          67.0           43.0          23
23   12.000000           24.0         -99.0           10.0          23
24   23.000000           21.0          43.0           54.0          23
25   11.000000           12.0          13.0           11.0          23
26   13.000000           14.0          54.0           44.0          23
27   43.000000           54.0          65.0           76.0          23
28   43.000000           44.0          54.0           65.0          23

    Placement  Offer Count  gender
0        98.0           99       1
1        66.0           97       1
2        84.0           69       1
3        79.0           79       1
4        64.0           64       1
5        83.0           74       1
6        72.0           68       1
7        88.0           86       1
8        90.0           91       1
9        81.0           69       0
10       60.0           67       0
11       90.0           71       0
12       99.0           80       0
13       76.0           95       0
14       91.0           84       0
15       69.0           96       0
16       60.0           62       0
17       72.0           76       0
18       63.0           73       0
19       97.0           87       0
20       45.0           65       1
21       56.0           56       1
22       45.0           76       1
23       56.0           87       1
24       54.0           23       1
25      -99.0           16       0
26       77.0           65       0
27       43.0           43       0
28       11.0           12       0
```

[36]:

```
col = ['Math Score', 'Reading Score', 'Writing Score','Placement Score']
df.boxplot(col)
```

[36]:

```
<Axes: >
```



[37]:
```
print(np.where(df['Math Score']>90))
print(np.where(df['Reading Score']<25))
print(np.where(df['Writing Score']<30))
```

```
(array([ 0,  2,  8, 17], dtype=int64),)
(array([18, 20, 21, 23, 24, 25, 26], dtype=int64),)
(array([ 5, 23, 25], dtype=int64),)
```

[42]:
```
df=pd.read_csv(r"E:\BE\6th sem\Data Science and Big Data analytics\Practical␣
 ↪Outputs\StudentsPerformance.csv")
df
```

[42]:
```
    Math Score  Reading Score  Writing Score  Placement Score  Club Join Year  \
0         95.0           81.0           93.0             85.0              23
1         83.0           79.0           64.0             97.0              23
2        100.0           96.0           70.0             86.0              23
3         88.0           83.0          100.0             81.0              23
4         73.0           72.0           96.0             62.0              23
5         63.0           76.0            NaN            100.0              23
6         72.0           87.0           99.0             94.0              23
7         90.0           75.0           91.0             66.0              23
8         97.0           79.0           74.0             77.0              23
```

```
9          NaN           60.0           77.0             70.0              23
10        77.0           65.0           61.0             90.0              23
11        75.0           94.0           88.0             91.0              23
12        85.0           79.0           63.0             93.0              23
13        80.0           79.0          100.0              NaN              23
14        81.0           81.0           88.0             61.0              23
15        80.0           64.0           98.0            100.0              23
16        66.0           85.0           94.0             84.0              23
17        99.0           80.0           75.0             93.0              23
18        85.0            NaN           81.0             99.0              23
19        76.0           99.0           89.0             84.0              23
20        34.0           23.0           45.0             21.0              23
21        43.0           23.0           34.0             56.0              23
22        45.0           65.0           67.0             43.0              23
23        12.0           24.0            NaN             10.0              23
24        23.0           21.0           43.0             54.0              23
25        11.0           12.0           13.0             11.0              23
26        13.0           14.0           54.0             44.0              23
27        43.0           54.0           65.0             76.0              23
28        43.0           44.0           54.0             65.0              23

    Placement  Offer Count  gender
0        98.0           99  female
1        66.0           97  female
2        84.0           69  female
3        79.0           79  female
4        64.0           64  female
5        83.0           74  female
6        72.0           68  female
7        88.0           86  female
8        90.0           91  female
9        81.0           69    Male
10       60.0           67    Male
11       90.0           71    Male
12       99.0           80    Male
13       76.0           95    Male
14       91.0           84    Male
15       69.0           96    Male
16       60.0           62    Male
17       72.0           76    Male
18       63.0           73    Male
19       97.0           87    Male
20       45.0           65  female
21       56.0           56  female
22       45.0           76  female
23       56.0           87  female
24       54.0           23  female
```

```
25        NaN           16    Male
26       77.0           65    Male
27       43.0           43    Male
28       11.0           12    Male
```

[45]:

```
import  matplotlib.pyplot as plt
import pandas as pd
import numpy as num
import sklearn as k
```

[46]:

```
df=pd.read_csv(r"E:\BE\6th sem\Data Science and Big Data analytics\Practical␣
 ↪Outputs\StudentsPerformance.csv")
```

[47]:

```
df
```

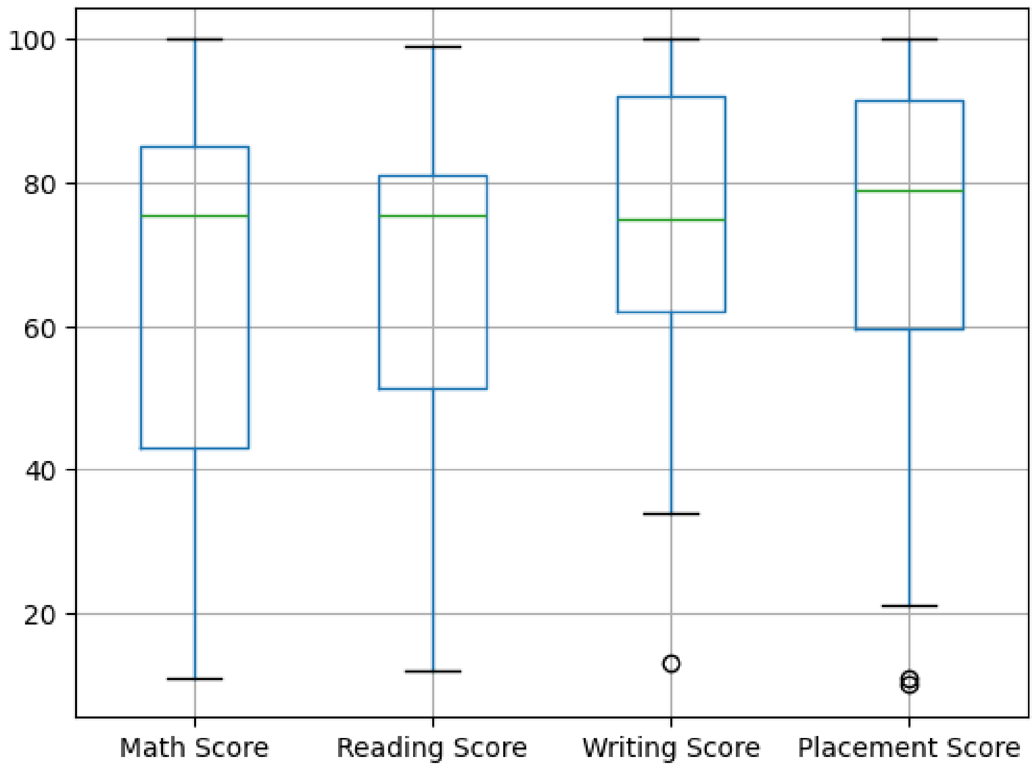
[47]:

```
    Math Score  Reading Score  Writing Score  Placement Score  Club Join Year  \
0         95.0           81.0           93.0             85.0              23
1         83.0           79.0           64.0             97.0              23
2        100.0           96.0           70.0             86.0              23
3         88.0           83.0          100.0             81.0              23
4         73.0           72.0           96.0             62.0              23
5         63.0           76.0            NaN            100.0              23
6         72.0           87.0           99.0             94.0              23
7         90.0           75.0           91.0             66.0              23
8         97.0           79.0           74.0             77.0              23
9          NaN           60.0           77.0             70.0              23
10        77.0           65.0           61.0             90.0              23
11        75.0           94.0           88.0             91.0              23
12        85.0           79.0           63.0             93.0              23
13        80.0           79.0          100.0              NaN              23
14        81.0           81.0           88.0             61.0              23
15        80.0           64.0           98.0            100.0              23
16        66.0           85.0           94.0             84.0              23
17        99.0           80.0           75.0             93.0              23
18        85.0            NaN           81.0             99.0              23
19        76.0           99.0           89.0             84.0              23
20        34.0           23.0           45.0             21.0              23
21        43.0           23.0           34.0             56.0              23
22        45.0           65.0           67.0             43.0              23
23        12.0           24.0            NaN             10.0              23
24        23.0           21.0           43.0             54.0              23
25        11.0           12.0           13.0             11.0              23
26        13.0           14.0           54.0             44.0              23
27        43.0           54.0           65.0             76.0              23
28        43.0           44.0           54.0             65.0              23

    Placement  Offer Count  gender
```

```
0        98.0          99  female
1        66.0          97  female
2        84.0          69  female
3        79.0          79  female
4        64.0          64  female
5        83.0          74  female
6        72.0          68  female
7        88.0          86  female
8        90.0          91  female
9        81.0          69    Male
10       60.0          67    Male
11       90.0          71    Male
12       99.0          80    Male
13       76.0          95    Male
14       91.0          84    Male
15       69.0          96    Male
16       60.0          62    Male
17       72.0          76    Male
18       63.0          73    Male
19       97.0          87    Male
20       45.0          65  female
21       56.0          56  female
22       45.0          76  female
23       56.0          87  female
24       54.0          23  female
25        NaN          16    Male
26       77.0          65    Male
27       43.0          43    Male
28       11.0          12    Male
```

[48]:

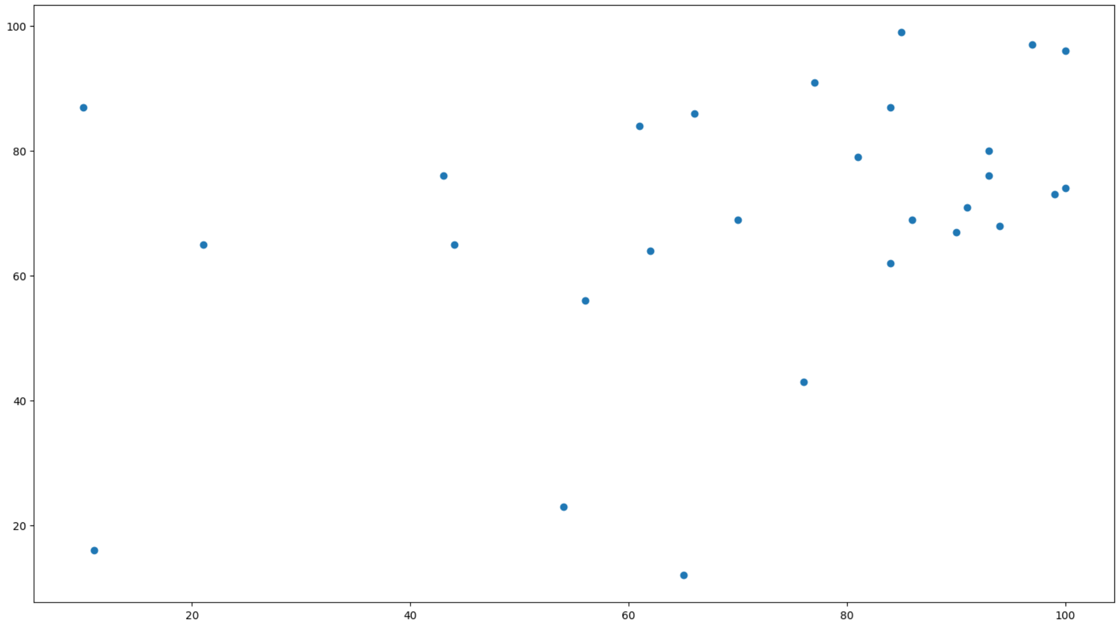
```
col = ['Math Score', 'Reading Score' , 'Writing Score','Placement Score']
df.boxplot(col)
plt.show()
```



[49]:
```
print(np.where(df['Math Score']>90))
print(np.where(df['Reading Score']<25))
print(np.where(df['Writing Score']<30))
```

```
(array([ 0,  2,  8, 17], dtype=int64),)
(array([20, 21, 23, 24, 25, 26], dtype=int64),)
(array([25], dtype=int64),)
```

[50]:
```
fig, ax = plt.subplots(figsize = (18,10))
ax.scatter(df['Placement Score'], df['Offer Count'])
plt.show()
```



[53]:

```
name = "Rohan Dhadke"
roll = 13136
```

[55]:

```
print(name)
print(roll)
```

```
Rohan Dhadke
13136
```

[ ]:

```

```