```
Lab Assignment 2

AIM:-Create an “Academic performance” dataset of students and perform
the following operations using
Python.
1. Scan all variables for missing values and inconsistencies. If there
are missing values and/or
inconsistencies, use any of the suitable techniques to deal with them.
2. Scan all numeric variables for outliers. If there are outliers, use
any of the suitable
techniques to deal with them.
3. Apply data transformations on at least one of the variables. The
purpose of this
transformation should be one of the following reasons: to change the
scale for better
understanding of the variable, to convert a non-linear relation into a
linear one, or to
decrease the skewness and convert the distribution into a normal
distribution.
Reason and document your approach properly.

Name: Tanuja Mahajan
B2_13227
Practical No:- 2

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
from scipy import stats
file_path=r"C:\Users\System21\Desktop\studentdata.csv"
df=pd.read_csv(file_path)
df.head(20)
```

```
     Math Score\n   Reading Score  Writing Score  Placement Score
                                                    N
                                                   a
                                                   N
                                                   7
                                                   5.
                                                   0
                                                   69.
                                                   0
                                                   71.
                                                   0
                                                   69.
                                                   0
                                                   75.
                                                   0
                                                   69.
```

```
15             61            80           74.0              80
16             61            63           70.0              71
17             76            74           67.0              73
18             75            64           66.0              76
19             69            74           67.0              72
```

```
    Club_Join_Date  Placement_Offer_Count
0             2020                     93
1             2019                     75
2             2019                     90
3             2020                     91
4             2021                     88
5             2021                     75
6             2021                    100
7             2019                     95
8             2018                    100
9             2020                     93
10            2021                     89
11            2019                     90
12            2018                     92
13            2018                     89
14            2019                    100
15            2020                     97
16            2019                     95
17            2021                     95
18            2021                     91
19            2018                     94
```

```
df.isnull()
```

```
     Math_Score\n   Reading_Score                  Placement_Sco
          Writing_Score False                      re
                 False   True                                Fa
          False          False                               ls
                         False                               e
          False          False                               Fa
                         False                               ls
          False          False                               e
                         False                               Fa
          False          False                               ls
                         False                               e
          False          False                               Fa
                         False                               ls
          False          False                               e
                         False                               Fa
          False          False                               ls
                         False                               e
          False          False                               Fa
                         False                               ls
          False          False                               e
```

```
19          False          False          False            False
20          False          False          False            False
21          False          False          False            False
22          False          False          False            False
23          False          False          False            False
24          False          False          False            False
25          False          False          False            False
26          False          False          False            False
27          False          False          False            False
28          False          False          False            False
```

```
    Club_Join_Date  Placement_Offer_Count
0    False   False
1    False   False
2    False   False
3    False   False
4    False   False
5    False   False
6    False   False
7    False   False
8    False   False
9    False   False
10   False   False
11   False   False
12   False   False
13   False   False
14   False   False
15   False   False
16   False   False
17   False   False
18   False   False
19   False   False
20   False   False
21   False   False
22   False   False
23   False   False
24   False   False
25   False   False
26   False   False
27   False   False
28   False   False
```

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

```
     Math_Score\n   Reading_Score                  Placement_Sco
     Writing_Score                                            re
                66             67                             77
                               NaN                            72
                70             80             75.             74
```

```
4           76          80         69.0              73
5           60          70         75.0              73
6           68          68         69.0              79
7           64          78         65.0              70
8           80          74         72.0              79
9           80          63         74.0              79
10          72          72         73.0              74
11          63          78         66.0              77
12          71          69         75.0              77
13          74          66         72.0              70
14          76          65         71.0              76
15          61          80         74.0              80
16          61          63         70.0              71
17          76          74         67.0              73
18          75          64         66.0              76
19          69          74         67.0              72
20          79          76         72.0              71
21          80          70         69.0              73
22          71          63         74.0              80
23          62          71         69.0              71
24          75          63         74.0              78
25          73          60         73.0              75
26          71          70         65.0              72
27          64          68         66.0              71
28          68          68         75.0              75
```

```
    Club_Join_Date  Placement_Offer_Count
0             2020                     93
1             2019                     75
2             2019                     90
3             2020                     91
4             2021                     88
5             2021                     75
6             2021                    100
7             2019                     95
8             2018                    100
9             2020                     93
10            2021                     89
11            2019                     90
12            2018                     92
13            2018                     89
14            2019                    100
15            2020                     97
16            2019                     95
17            2021                     95
18            2021                     91
19            2018                     94
20            2020                     99
21            2019                     77
```

```
22            2019                        89
23            2021                        77
24            2018                        85
25            2018                        84
26            2018                        84
27            2018                        84
28            2018                        97
```

```python
print(df.columns)
```

```
Index(['Math_Score\n ', 'Reading_Score', 'Writing_Score',
'Placement_Score',
       'Club_Join_Date', 'Placement_Offer_Count'],
      dtype='object')
```

```python
ndf=df
ndf.fillna(0)
```

```
     Math_Score\n   Reading_Score                 Placement_Sco
     Writing_Score                                           re
                66             67             0.             77
                0                                            72
                70             80            75.             74
                0                                            72
                78             61            69.             73
                0                                            73
                78             74            71.             79
                0                                            70
                76             80            69.             79
                0                                            79
                60             70            75.             74
                0                                            77
                68             68            69.             77
                0                                            70
                64             78            65.             76
                0                                            80
                80             74            72.             71
                0                                            73
                80             63            74.             76
                0                                            72
                72             72            73.             71
                0                                            73
                63             78            66.             80
                0                                            71
                71             69            75.             78
                0                                            75
                74             66            72.             72
                0                                            71
```

```
    Club_Join_Date  Placement_Offer_Count
```

```
0           2020                      93
1           2019                      75
2           2019                      90
3           2020                      91
4           2021                      88
5           2021                      75
6           2021                     100
7           2019                      95
8           2018                     100
9           2020                      93
10          2021                      89
11          2019                      90
12          2018                      92
13          2018                      89
14          2019                     100
15          2020                      97
16          2019                      95
17          2021                      95
18          2021                      91
19          2018                      94
20          2020                      99
21          2019                      77
22          2019                      89
23          2021                      77
24          2018                      85
25          2018                      84
26          2018                      84
27          2020                      84
28          2018                      97
```

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

```
     Math Score\n    Reading Score  Writing Score  Placement Score
                                                N
                                               a
                                               N
                                               7
                                               5.
                                               0
                                               69.
                                               0
                                               71.
                                               0
                                               69.
                                               0
                                               75.
                                               0
                                               69.
```

```
15         61         80         74.0              80
16         61         63         70.0              71
17         76         74         67.0              73
18         75         64         66.0              76
19         69         74         67.0              72
20         79         76         72.0              71
21         80         70         69.0              73
22         71         63         74.0              80
23         62         71         69.0              71
24         75         63         74.0              78
25         73         60         73.0              75
26         71         70         65.0              72
27         64         68         66.0              71
28         68         68         75.0              75

    Club_Join_Date  Placement_Offer_Count
0             2020                     93
1             2019                     75
2             2019                     90
3             2020                     91
4             2021                     88
5             2021                     75
6             2021                    100
7             2019                     95
8             2018                    100
9             2020                     93
10            2021                     89
11            2019                     90
12            2018                     92
13            2018                     89
14            2019                    100
15            2020                     97
16            2019                     95
17            2021                     95
18            2021                     91
19            2018                     94
20            2020                     99
21            2019                     77
22            2019                     89
23            2021                     77
24            2018                     85
25            2018                     84
26            2018                     84
27            2020                     84
28            2018                     97
```

```
ndf.dropna()
```

```
                  Math_Score\n   Reading_Score    Placement_Sco
                                 Writing_Score               re
```

```
2               78            61          69.0                74
3               78            74          71.0                72
4               76            80          69.0                73
5               60            70          75.0                73
6               68            68          69.0                79
7               64            78          65.0                70
8               80            74          72.0                79
9               80            63          74.0                79
10              72            72          73.0                74
11              63            78          66.0                77
12              71            69          75.0                77
13              74            66          72.0                70
14              76            65          71.0                76
15              61            80          74.0                80
16              61            63          70.0                71
17              76            74          67.0                73
18              75            64          66.0                76
19              69            74          67.0                72
20              79            76          72.0                71
21              80            70          69.0                73
22              71            63          74.0                80
23              62            71          69.0                71
24              75            63          74.0                78
25              73            60          73.0                75
26              71            70          65.0                72
27              64            68          66.0                71
28              68            68          75.0                75
```

```
    Club_Join_Date  Placement_Offer_Count
1             2019                     75
2             2019                     90
3             2020                     91
4             2021                     88
5             2021                     75
6             2021                    100
7             2019                     95
8             2018                    100
9             2020                     93
10            2021                     89
11            2019                     90
12            2018                     92
13            2018                     89
14            2019                    100
15            2020                     97
16            2019                     95
17            2021                     95
18            2021                     91
19            2018                     94
20            2020                     99
```

```
21            2019                         77
22            2019                         89
23            2021                         77
24            2018                         85
25            2018                         84
26            2018                         84
27            2020                         84
28            2018                         97
```

```
ndf.dropna(how = 'all')
```

```
     Math_Score\n   Reading_Score                  Placement_Sco
     Writing_Score                                            re
                66              67                            77
                                NaN                           72
                70              80           75.              74
                0                                             72
                78              61           69.              73
                0                                             73
                78              74           71.              79
                0                                             70
                76              80           69.              79
                0                                             79
                60              70           75.              74
                0                                             77
                68              68           69.              77
                0                                             70
                64              78           65.              76
                0                                             80
                80              74           72.              71
                0                                             73
                80              63           74.              76
                0                                             72
                72              72           73.              71
                0                                             73
                63              78           66.              80
                0                                             71
                71              69           75.              78
                0                                             75
                74              66           72.              72
                0                                             71
```

```
    Club_Join_Date  Placement_Offer_Count
0             2020                     93
1             2019                     75
2             2019                     90
3             2020                     91
4             2021                     88
5             2021                     75
6             2021                    100
```

```
7            2019                      95
8            2018                     100
9            2020                      93
10           2021                      89
11           2019                      90
12           2018                      92
13           2018                      89
14           2019                     100
15           2020                      97
16           2019                      95
17           2021                      95
18           2021                      91
19           2018                      94
20           2020                      99
21           2019                      77
22           2019                      89
23           2021                      77
24           2018                      85
25           2018                      84
26           2018                      84
27           2020                      84
28           2018                      97
```

```
ndf.dropna(axis = 1)
```

```
     Math Score\n   Reading Score  Placement Score  Club Join Date
                                                               20
                                                               20
                                                               20
                                                               19
                                                               20
                                                               19
                                                               20
                                                               20
                                                               20
                                                               21
                                                               20
                                                               21
                                                               20
                                                               21
                                                               20
                                                               19
                                                               20
                                                               18
                                                               20
                                                               20
                                                               20
                                                               21
                                                               20
                                                               19
```

```
24           75            63              78             2018
25           73            60              75             2018
26           71            70              72             2018
27           64            68              71             2020
28           68            68              75             2018

    Placement_Offer_Count
0                      93
1                      75
2                      90
3                      91
4                      88
5                      75
6                     100
7                      95
8                     100
9                      93
10                     89
11                     90
12                     92
13                     89
14                    100
15                     97
16                     95
17                     95
18                     91
19                     94
20                     99
21                     77
22                     89
23                     77
24                     85
25                     84
26                     84
27                     84
28                     97
```

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

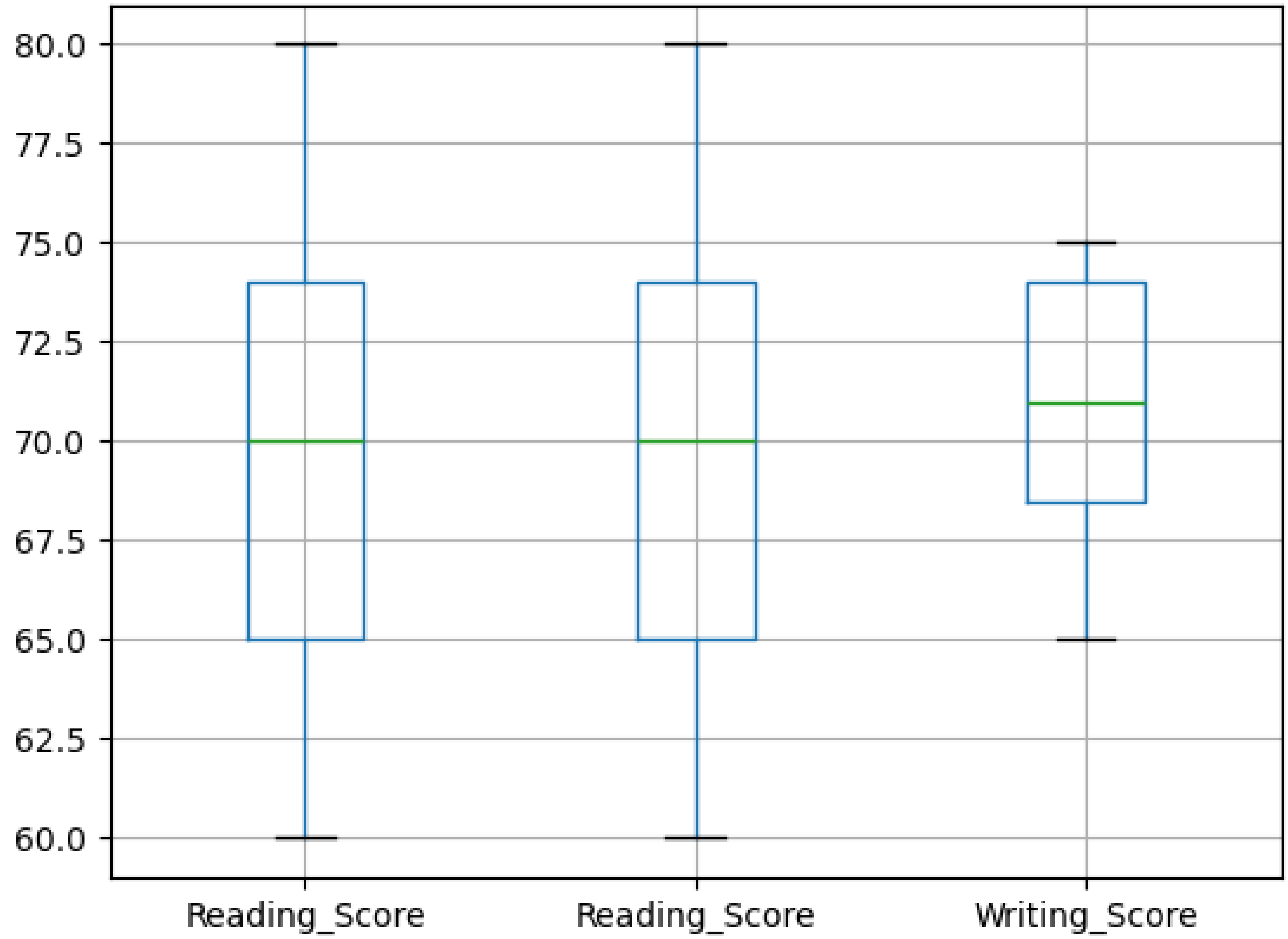
```
     Math_Score\n   Reading_Score              Placement_Sco
     Writing_Score                                        re
              70             80          75.              72
              0                                           74
              78             61          69.              72
              0                                           73
              78             74          71.              73
              0                                           79
              76             80          69.              70
              0                                           79
```

```
10          72          72        73.0          74
11          63          78        66.0          77
12          71          69        75.0          77
13          74          66        72.0          70
14          76          65        71.0          76
15          61          80        74.0          80
16          61          63        70.0          71
17          76          74        67.0          73
18          75          64        66.0          76
19          69          74        67.0          72
20          79          76        72.0          71
21          80          70        69.0          73
22          71          63        74.0          80
23          62          71        69.0          71
24          75          63        74.0          78
25          73          60        73.0          75
26          71          70        65.0          72
27          64          68        66.0          71
28          68          68        75.0          75

    Club_Join_Date  Placement_Offer_Count
1             2019                     75
2             2019                     90
3             2020                     91
4             2021                     88
5             2021                     75
6             2021                    100
7             2019                     95
8             2018                    100
9             2020                     93
10            2021                     89
11            2019                     90
12            2018                     92
13            2018                     89
14            2019                    100
15            2020                     97
16            2019                     95
17            2021                     95
18            2021                     91
19            2018                     94
20            2020                     99
21            2019                     77
22            2019                     89
23            2021                     77
24            2018                     85
25            2018                     84
26            2018                     84
27            2020                     84
28            2018                     97
```

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

```
<Axes: >
```



```
Name - Sushant Jawale
Roll no. - 13209
```