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 16 17 18 19	61 61 76 75 69	80 63 74 64 74	74.0 70.0 67.0 66.0 67.0	80 71 73 76 72
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 19	Club_Join_Date 2020 2019 2019 2020 2021 2021 2021 2019 2018 2020 2018 2018 2019 2018 2019 2020 2019 2021 2021 2021 2021	Placement_Offer_C	Count 93 75 90 91 88 75 100 95 100 93 89 90 92 89 100 97 95 95 91 94	
df.i		Reading_Score g_Score False		Placement_Score Fa ls 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
                      Placement_Offer_Count
    Club Join Date
      False
              False
0
      False
              False
      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
      False
              False
18
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]
                      Reading_Score
                                                        Placement_Sco
     Math_Score\n
     Writing_Score
                                                                    re
                  66
                                   67
                                                                    77
                                   NaN
                                                                    72
                  70
                                                 75.
                                                                    74
                                   80
```

4 5 6 7 8 9 10 11 11 11 11 11 11 11 11 11 11 11 11	76 60 68 64 80 80 72 63 71 74 76 61 61 75 69 79 80 71 62 75 73 71 64 68	80 70 68 78 74 63 72 78 69 66 65 80 63 74 64 74 76 70 63 71 63 60 70 68 68	69.0 75.0 69.0 65.0 72.0 74.0 75.0 71.0 74.0 70.0 67.0 67.0 67.0 69.0 74.0 69.0 74.0 73.0 65.0 75.0	73 79 70 79 79 74 77 70 76 80 71 73 76 72 71 73 80 71 78 75 72 71
0 1 2 3 4 5 6 7 8 9 10 11 11 11 11 11 11 11 11 11 11 11 11		Placement_Offer_Count 93 75 90 91 88 75 100 95 100 93 89 90 91 92 89 100 97 95 95 91 94 99 77	75.0	

```
22
                2019
                                            89
23
                2021
                                            77
24
                2018
                                            85
25
                2018
                                            84
26
                2018
                                            84
27
                2020
                                            84
28
                                            97
                2018
print(df.columns)
Index(['Math_Score\n ', 'Reading_Score', 'Writing_Score',
'Placement_Score',
        'Club_Join_Date', 'Placement_Offer_Count'],
      dtype='object')
ndf=df
ndf.fillna(0)
                                                         Placement_Sco
                      Reading_Score
     Math_Score\n
     Writing_Score
                                                                     re
                                                   0.
                  66
                                   67
                                                                     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
                                                  69.
                  68
                                                                     77
                                                                     70
                                   78
                                                  65.
                  64
                                                                     76
                                                                     80
                  80
                                   74
                                                  72.
                                                                     71
                                                                     73
                  80
                                   63
                                                  74.
                                                                     76
                  0
                                                                     72
                  72
                                   72
                                                  73.
                                                                     71
                  0
                                                                     73
                  63
                                                  66.
                                   78
                                                                     80
                                                                     71
                                   69
                  71
                                                  75.
                                                                     78
                  0
                                                                     75
                  74
                                   66
                                                  72.
                                                                     72
                                                                     71
```

Club_Join_Date Placement_Offer_Count

```
0
                2020
                                            93
                2019
                                            75
23
                2019
                                            90
                2020
                                            91
                2021
                                            88
4
5
6
7
                2021
                                            75
                2021
                                           100
                2019
                                            95
8 9
                2018
                                           100
                                            93
                2020
10
                2021
                                            89
11
                2019
                                            90
12
                                            92
                2018
13
                                            89
                2018
14
                2019
                                           100
15
                2020
                                            97
16
                                            95
                2019
17
                2021
                                            95
18
                2021
                                            91
19
                2018
                                            94
20
                2020
                                            99
21
                2019
                                            77
22
                2019
                                            89
23
                                            77
                2021
24
                                            85
                2018
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
                                                  a
                                                  5.
                                                  69.
                                                  71.
                                                  69.
                                                  75.
```

69.

15 16 17 18 19 20 21 22 23 24 25 26 27 28	61 76 75 69 79 80 71 62 75 73 71 64 68	80 63 74 64 74 76 70 63 71 63 60 70 68 68	74.0 70.0 67.0 66.0 67.0 72.0 69.0 74.0 69.0 74.0 73.0 65.0 66.0 75.0	80 71 73 76 72 71 73 80 71 78 75 72 71 75
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27	Date Placement_ 2020 2019 2019 2020 2021 2021 2021 2019 2018 2020 2019 2018 2019 2021 2019 2021 2018 2019 2021 2018 2020 2019 2021 2018 2020 2019 2021 2018 2020 2019 2021 2018 2020 2019 2021	93 75 90 91 88 75 100 95 100 93 89 90 92 89 100 97 95 91 94 99 77 85 84 84 84 97		
	Math_Score\r	n Reading_S Writing S	core Placement	_Sco re

234567891113141567891012324526728	78 78 76 60 68 64 80 80 72 63 71 74 76 61 61 75 69 79 80 71 62 75 73 71 64 68	61 74 80 70 68 78 74 63 72 78 69 66 65 80 63 74 64 74 76 70 63 71 63 60 70 68 68	69.0 71.0 69.0 75.0 69.0 72.0 74.0 73.0 66.0 77.0 67.0 67.0 69.0 74.0 73.0 65.0 75.0	74 72 73 73 79 70 79 79 74 77 77 70 76 80 71 73 76 72 71 73 80 71 78 75 72 71 78 75
1 2 3 4 5 6 7 8 9 10 11 2 13 14 15 16 17 18 19 20	Club_Join_Date	Placement_Offer_Count 75 90 91 88 75 100 95 100 95 89 90 91 89 90 91 90 91 91 91 91		

21	2019	77	
22	2019	89	
222324252627	2021	77	
24	2018	85	
25	2018	84	
26	2018	84	
27	2020	84	
28	2018	97	
16 1	na(how = 'all')		

nui.uiopiia(now – all)

Math_Score\n	Reading_Score	Pla	cement_Sco
Writing_Score	11000010 <u> </u>		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	63	7 4	73
80	63	74.	76
0	70	70	72
72	72	73.	71
0	70		73
63	78	66.	80
0	60	75	71
71	69	75.	78
0	6.6	72	75
74 0	66	72.	72
U			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)

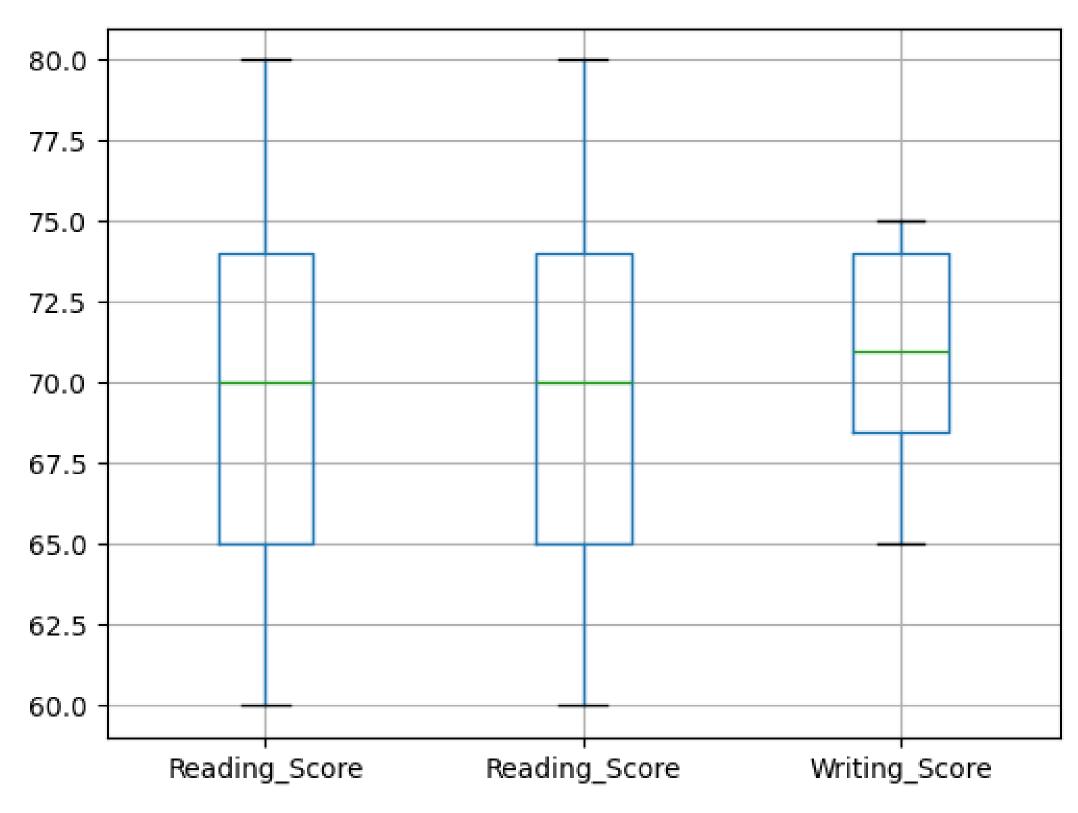
Math Score\n Reading Score Placement Score Club Join Date 20 20 20 19 20 19 20 20 21 20 20 20 19 20 20 20 20 20 20 20 20 20 20 20 20 20
20 21 20 19

24 75 25 73 26 71 27 64 28 68	63 60 70 68 68	78 75 72 71 75	2018 2018 2018 2020 2018
Placement_Offer_0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 new_data =ndf.dropn new_data	93 75 90 91 88 75 100 95 100 93 89 90 92 89 100 97 95 95 91 94 99 77 85 84 84 84 84 97	any')	
Math_Score\n Writing_Score 70	Reading_Score 80	Placemo	ent_Sco re 72
0 78 0	61	69.	74 72 73
78 0 76 0	74 80	71. 69.	73 79 70 79

10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	72 63 71 74 76 61 61 75 69 79 80 71 62 75 73 71 64 68	72 78 69 66 65 80 63 74 64 74 76 70 63 71 63 60 70 68 68	73.0 66.0 75.0 72.0 71.0 74.0 67.0 67.0 69.0 74.0 69.0 74.0 73.0 65.0 66.0 75.0	74 77 77 70 76 80 71 73 76 72 71 73 80 71 78 75 72 71 75
123456789111131456789 1123456789 1123456789 123456789	Club_Join_Date 2019 2019 2020 2021 2021 2021 2019 2018 2020 2021 2019 2018 2019 2019 2020 2011 2019 2021 2019 2021 2018 2018 2018 2018 2018 2020 2019 2019 2019 2019 2019 2019 2019	Placement_Offer_Co	ount 75 90 91 88 75 100 95 100 93 89 90 92 89 100 97 95 95 91 94 99 77 89 77 89 77 85 84 84 84 84 84	

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

<Axes: >



Name - Sushant Jawale Roll no. - 13209