

u3yi0mdef

February 17, 2025

```
[1]: import pandas as pd
import numpy as np
```

```
[2]: df=pd.read_csv("student.csv")
```

```
[3]: df
```

```
[3]:  maths_score  reading_score  writing_score  placement_score  club_join_year  \
0           78           63           67.0           68           2021
1           67           16           73.0           63           2019
2           62           78           60.0           63           2021
3           73           69           69.0           66           2021
4           67           na           77.0           71           2020
5           68           70           64.0           69           2018
6           79           64           72.0           73           2020
7           78           70           78.0           64           2021
8           NaN           67           75.0           62           2020
9           72           80           79.0           69           2020
10          Na           70           68.0           72           2019
11          65           79           69.0           75           2019
12          62           74           80.0           75           2018
13          66           63           NaN           74           2020
14          76           60           60.0           64           2018
15          70           61           61.0           63           2021
16          72           69           65.0           74           2020
17          80           76           74.0           78           2020
18          65           70           63.0           76           2019
19          71           65           68.0           na           2019
20          68           65           68.0           75           2019
21          77           79           66.0           62           2020
22          68           76           80.0           74           2020
23          80           71           79.0           78           2020
24          62           71           62.0           60           2018
25          62           78           66.0           78           2018
26          66           77           68.0           69           2020
27          60           75           77.0           65           2021
```

	placement_offer_count	Gender
0	1	Female
1	1	Male
2	1	Male
3	1	Male
4	1	Male
5	1	Male
6	1	Male
7	1	Male
8	1	Male
9	1	Male
10	1	Male
11	2	Female
12	2	Female
13	1	Female
14	1	Female
15	1	Female
16	1	Female
17	2	Female
18	2	Female
19	3	Female
20	2	Female
21	1	Female
22	1	Male
23	2	Male
24	1	Male
25	2	Male
26	1	Female
27	1	Male

```
[4]: df.isnull()
```

```
[4]:
```

	maths_score	reading_score	writing_score	placement_score	\
0	False	False	False	False	
1	False	False	False	False	
2	False	False	False	False	
3	False	False	False	False	
4	False	False	False	False	
5	False	False	False	False	
6	False	False	False	False	
7	False	False	False	False	
8	True	False	False	False	
9	False	False	False	False	
10	False	False	False	False	
11	False	False	False	False	
12	False	False	False	False	
13	False	False	True	False	

14	False	False	False	False
15	False	False	False	False
16	False	False	False	False
17	False	False	False	False
18	False	False	False	False
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

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

```
[5]: series = pd.isnull(df["maths_score"])
df[series]
```

```
[5]:  maths_score reading_score writing_score placement_score club_join_year \
      8          NaN          67          75.0          62          2020

      placement_offer_count Gender
      8                   1   Male
```

```
[6]: df.notnull()
```

```
[6]:  maths_score reading_score writing_score placement_score \
0          True          True          True          True
1          True          True          True          True
2          True          True          True          True
3          True          True          True          True
4          True          True          True          True
5          True          True          True          True
6          True          True          True          True
7          True          True          True          True
8         False          True          True          True
9          True          True          True          True
10         True          True          True          True
11         True          True          True          True
12         True          True          True          True
13         True          True          False         True
14         True          True          True          True
15         True          True          True          True
16         True          True          True          True
17         True          True          True          True
18         True          True          True          True
19         True          True          True          True
20         True          True          True          True
21         True          True          True          True
22         True          True          True          True
23         True          True          True          True
24         True          True          True          True
25         True          True          True          True
26         True          True          True          True
27         True          True          True          True

      club_join_year placement_offer_count Gender
0          True          True          True
1          True          True          True
2          True          True          True
3          True          True          True
4          True          True          True
5          True          True          True
6          True          True          True
7          True          True          True
```

8	True	True	True
9	True	True	True
10	True	True	True
11	True	True	True
12	True	True	True
13	True	True	True
14	True	True	True
15	True	True	True
16	True	True	True
17	True	True	True
18	True	True	True
19	True	True	True
20	True	True	True
21	True	True	True
22	True	True	True
23	True	True	True
24	True	True	True
25	True	True	True
26	True	True	True
27	True	True	True

```
[8]: series1 = pd.notnull(df["maths_score"])
df[series1]
```

```
[8]:  maths_score  reading_score  writing_score  placement_score  club_join_year  \
0          78           63          67.0           68           2021
1          67           16          73.0           63           2019
2          62           78          60.0           63           2021
3          73           69          69.0           66           2021
4          67           na          77.0           71           2020
5          68           70          64.0           69           2018
6          79           64          72.0           73           2020
7          78           70          78.0           64           2021
9          72           80          79.0           69           2020
10         Na           70          68.0           72           2019
11         65           79          69.0           75           2019
12         62           74          80.0           75           2018
13         66           63           NaN           74           2020
14         76           60          60.0           64           2018
15         70           61          61.0           63           2021
16         72           69          65.0           74           2020
17         80           76          74.0           78           2020
18         65           70          63.0           76           2019
19         71           65          68.0           na           2019
20         68           65          68.0           75           2019
21         77           79          66.0           62           2020
22         68           76          80.0           74           2020
```

23	80	71	79.0	78	2020
24	62	71	62.0	60	2018
25	62	78	66.0	78	2018
26	66	77	68.0	69	2020
27	60	75	77.0	65	2021

	placement_offer_count	Gender
0	1	Female
1	1	Male
2	1	Male
3	1	Male
4	1	Male
5	1	Male
6	1	Male
7	1	Male
9	1	Male
10	1	Male
11	2	Female
12	2	Female
13	1	Female
14	1	Female
15	1	Female
16	1	Female
17	2	Female
18	2	Female
19	3	Female
20	2	Female
21	1	Female
22	1	Male
23	2	Male
24	1	Male
25	2	Male
26	1	Female
27	1	Male

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

	maths_score	reading_score	writing_score	placement_score	club_join_year	\
0	78	63	67.0	68	2021	
1	67	16	73.0	63	2019	
2	62	78	60.0	63	2021	
3	73	69	69.0	66	2021	
4	67	na	77.0	71	2020	

5	68	70	64.0	69	2018
6	79	64	72.0	73	2020
7	78	70	78.0	64	2021
8	NaN	67	75.0	62	2020
9	72	80	79.0	69	2020
10	Na	70	68.0	72	2019
11	65	79	69.0	75	2019
12	62	74	80.0	75	2018
13	66	63	NaN	74	2020
14	76	60	60.0	64	2018
15	70	61	61.0	63	2021
16	72	69	65.0	74	2020
17	80	76	74.0	78	2020
18	65	70	63.0	76	2019
19	71	65	68.0	na	2019
20	68	65	68.0	75	2019
21	77	79	66.0	62	2020
22	68	76	80.0	74	2020
23	80	71	79.0	78	2020
24	62	71	62.0	60	2018
25	62	78	66.0	78	2018
26	66	77	68.0	69	2020
27	60	75	77.0	65	2021

	placement_offer_count	Gender	gender
0	1	Female	0
1	1	Male	1
2	1	Male	1
3	1	Male	1
4	1	Male	1
5	1	Male	1
6	1	Male	1
7	1	Male	1
8	1	Male	1
9	1	Male	1
10	1	Male	1
11	2	Female	0
12	2	Female	0
13	1	Female	0
14	1	Female	0
15	1	Female	0
16	1	Female	0
17	2	Female	0
18	2	Female	0
19	3	Female	0
20	2	Female	0
21	1	Female	0

22	1	Male	1
23	2	Male	1
24	1	Male	1
25	2	Male	1
26	1	Female	0
27	1	Male	1

```
[11]: missing_values = ["Na", "na"]
df = pd.read_csv("student.csv", na_values = missing_values)
df
```

```
[11]:
```

	maths_score	reading_score	writing_score	placement_score	\
0	78.0	63.0	67.0	68.0	
1	67.0	16.0	73.0	63.0	
2	62.0	78.0	60.0	63.0	
3	73.0	69.0	69.0	66.0	
4	67.0	NaN	77.0	71.0	
5	68.0	70.0	64.0	69.0	
6	79.0	64.0	72.0	73.0	
7	78.0	70.0	78.0	64.0	
8	NaN	67.0	75.0	62.0	
9	72.0	80.0	79.0	69.0	
10	NaN	70.0	68.0	72.0	
11	65.0	79.0	69.0	75.0	
12	62.0	74.0	80.0	75.0	
13	66.0	63.0	NaN	74.0	
14	76.0	60.0	60.0	64.0	
15	70.0	61.0	61.0	63.0	
16	72.0	69.0	65.0	74.0	
17	80.0	76.0	74.0	78.0	
18	65.0	70.0	63.0	76.0	
19	71.0	65.0	68.0	NaN	
20	68.0	65.0	68.0	75.0	
21	77.0	79.0	66.0	62.0	
22	68.0	76.0	80.0	74.0	
23	80.0	71.0	79.0	78.0	
24	62.0	71.0	62.0	60.0	
25	62.0	78.0	66.0	78.0	
26	66.0	77.0	68.0	69.0	
27	60.0	75.0	77.0	65.0	

	club_join_year	placement_offer_count	Gender
0	2021	1	Female
1	2019	1	Male
2	2021	1	Male
3	2021	1	Male
4	2020	1	Male



5	2018	1	Male
6	2020	1	Male
7	2021	1	Male
8	2020	1	Male
9	2020	1	Male
10	2019	1	Male
11	2019	2	Female
12	2018	2	Female
13	2020	1	Female
14	2018	1	Female
15	2021	1	Female
16	2020	1	Female
17	2020	2	Female
18	2019	2	Female
19	2019	3	Female
20	2019	2	Female
21	2020	1	Female
22	2020	1	Male
23	2020	2	Male
24	2018	1	Male
25	2018	2	Male
26	2020	1	Female
27	2021	1	Male

```
[12]: ndf=df
      ndf.fillna(0)
```

```
[12]:
```

	maths_score	reading_score	writing_score	placement_score	\
0	78.0	63.0	67.0	68.0	
1	67.0	16.0	73.0	63.0	
2	62.0	78.0	60.0	63.0	
3	73.0	69.0	69.0	66.0	
4	67.0	0.0	77.0	71.0	
5	68.0	70.0	64.0	69.0	
6	79.0	64.0	72.0	73.0	
7	78.0	70.0	78.0	64.0	
8	0.0	67.0	75.0	62.0	
9	72.0	80.0	79.0	69.0	
10	0.0	70.0	68.0	72.0	
11	65.0	79.0	69.0	75.0	
12	62.0	74.0	80.0	75.0	
13	66.0	63.0	0.0	74.0	
14	76.0	60.0	60.0	64.0	
15	70.0	61.0	61.0	63.0	
16	72.0	69.0	65.0	74.0	
17	80.0	76.0	74.0	78.0	
18	65.0	70.0	63.0	76.0	

19	71.0	65.0	68.0	0.0
20	68.0	65.0	68.0	75.0
21	77.0	79.0	66.0	62.0
22	68.0	76.0	80.0	74.0
23	80.0	71.0	79.0	78.0
24	62.0	71.0	62.0	60.0
25	62.0	78.0	66.0	78.0
26	66.0	77.0	68.0	69.0
27	60.0	75.0	77.0	65.0

	club_join_year	placement_offer_count	Gender
0	2021	1	Female
1	2019	1	Male
2	2021	1	Male
3	2021	1	Male
4	2020	1	Male
5	2018	1	Male
6	2020	1	Male
7	2021	1	Male
8	2020	1	Male
9	2020	1	Male
10	2019	1	Male
11	2019	2	Female
12	2018	2	Female
13	2020	1	Female
14	2018	1	Female
15	2021	1	Female
16	2020	1	Female
17	2020	2	Female
18	2019	2	Female
19	2019	3	Female
20	2019	2	Female
21	2020	1	Female
22	2020	1	Male
23	2020	2	Male
24	2018	1	Male
25	2018	2	Male
26	2020	1	Female
27	2021	1	Male

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

```
[14]:   maths_score  reading_score  writing_score  placement_score  \
0      78.000000          63.0          67.0          68.0
1      67.000000          16.0          73.0          63.0
```

2	62.000000	78.0	60.0	63.0
3	73.000000	69.0	69.0	66.0
4	67.000000	NaN	77.0	71.0
5	68.000000	70.0	64.0	69.0
6	79.000000	64.0	72.0	73.0
7	78.000000	70.0	78.0	64.0
8	69.769231	67.0	75.0	62.0
9	72.000000	80.0	79.0	69.0
10	69.769231	70.0	68.0	72.0
11	65.000000	79.0	69.0	75.0
12	62.000000	74.0	80.0	75.0
13	66.000000	63.0	NaN	74.0
14	76.000000	60.0	60.0	64.0
15	70.000000	61.0	61.0	63.0
16	72.000000	69.0	65.0	74.0
17	80.000000	76.0	74.0	78.0
18	65.000000	70.0	63.0	76.0
19	71.000000	65.0	68.0	NaN
20	68.000000	65.0	68.0	75.0
21	77.000000	79.0	66.0	62.0
22	68.000000	76.0	80.0	74.0
23	80.000000	71.0	79.0	78.0
24	62.000000	71.0	62.0	60.0
25	62.000000	78.0	66.0	78.0
26	66.000000	77.0	68.0	69.0
27	60.000000	75.0	77.0	65.0

	club_join_year	placement_offer_count	Gender
0	2021	1	Female
1	2019	1	Male
2	2021	1	Male
3	2021	1	Male
4	2020	1	Male
5	2018	1	Male
6	2020	1	Male
7	2021	1	Male
8	2020	1	Male
9	2020	1	Male
10	2019	1	Male
11	2019	2	Female
12	2018	2	Female
13	2020	1	Female
14	2018	1	Female
15	2021	1	Female
16	2020	1	Female
17	2020	2	Female
18	2019	2	Female

19	2019	3	Female
20	2019	2	Female
21	2020	1	Female
22	2020	1	Male
23	2020	2	Male
24	2018	1	Male
25	2018	2	Male
26	2020	1	Female
27	2021	1	Male

```
[15]: ndf.replace(to_replace = np.nan, value = -99)
```

```
[15]:
```

	maths_score	reading_score	writing_score	placement_score	\
0	78.000000	63.0	67.0	68.0	
1	67.000000	16.0	73.0	63.0	
2	62.000000	78.0	60.0	63.0	
3	73.000000	69.0	69.0	66.0	
4	67.000000	-99.0	77.0	71.0	
5	68.000000	70.0	64.0	69.0	
6	79.000000	64.0	72.0	73.0	
7	78.000000	70.0	78.0	64.0	
8	69.769231	67.0	75.0	62.0	
9	72.000000	80.0	79.0	69.0	
10	69.769231	70.0	68.0	72.0	
11	65.000000	79.0	69.0	75.0	
12	62.000000	74.0	80.0	75.0	
13	66.000000	63.0	-99.0	74.0	
14	76.000000	60.0	60.0	64.0	
15	70.000000	61.0	61.0	63.0	
16	72.000000	69.0	65.0	74.0	
17	80.000000	76.0	74.0	78.0	
18	65.000000	70.0	63.0	76.0	
19	71.000000	65.0	68.0	-99.0	
20	68.000000	65.0	68.0	75.0	
21	77.000000	79.0	66.0	62.0	
22	68.000000	76.0	80.0	74.0	
23	80.000000	71.0	79.0	78.0	
24	62.000000	71.0	62.0	60.0	
25	62.000000	78.0	66.0	78.0	
26	66.000000	77.0	68.0	69.0	
27	60.000000	75.0	77.0	65.0	

	club_join_year	placement_offer_count	Gender
0	2021	1	Female
1	2019	1	Male
2	2021	1	Male
3	2021	1	Male

4	2020	1	Male
5	2018	1	Male
6	2020	1	Male
7	2021	1	Male
8	2020	1	Male
9	2020	1	Male
10	2019	1	Male
11	2019	2	Female
12	2018	2	Female
13	2020	1	Female
14	2018	1	Female
15	2021	1	Female
16	2020	1	Female
17	2020	2	Female
18	2019	2	Female
19	2019	3	Female
20	2019	2	Female
21	2020	1	Female
22	2020	1	Male
23	2020	2	Male
24	2018	1	Male
25	2018	2	Male
26	2020	1	Female
27	2021	1	Male

```
[16]: ndf.dropna()
```

```
[16]:
```

	maths_score	reading_score	writing_score	placement_score	\
0	78.000000	63.0	67.0	68.0	
1	67.000000	16.0	73.0	63.0	
2	62.000000	78.0	60.0	63.0	
3	73.000000	69.0	69.0	66.0	
5	68.000000	70.0	64.0	69.0	
6	79.000000	64.0	72.0	73.0	
7	78.000000	70.0	78.0	64.0	
8	69.769231	67.0	75.0	62.0	
9	72.000000	80.0	79.0	69.0	
10	69.769231	70.0	68.0	72.0	
11	65.000000	79.0	69.0	75.0	
12	62.000000	74.0	80.0	75.0	
14	76.000000	60.0	60.0	64.0	
15	70.000000	61.0	61.0	63.0	
16	72.000000	69.0	65.0	74.0	
17	80.000000	76.0	74.0	78.0	
18	65.000000	70.0	63.0	76.0	
20	68.000000	65.0	68.0	75.0	
21	77.000000	79.0	66.0	62.0	

22	68.000000	76.0	80.0	74.0
23	80.000000	71.0	79.0	78.0
24	62.000000	71.0	62.0	60.0
25	62.000000	78.0	66.0	78.0
26	66.000000	77.0	68.0	69.0
27	60.000000	75.0	77.0	65.0

	club_join_year	placement_offer_count	Gender
0	2021	1	Female
1	2019	1	Male
2	2021	1	Male
3	2021	1	Male
5	2018	1	Male
6	2020	1	Male
7	2021	1	Male
8	2020	1	Male
9	2020	1	Male
10	2019	1	Male
11	2019	2	Female
12	2018	2	Female
14	2018	1	Female
15	2021	1	Female
16	2020	1	Female
17	2020	2	Female
18	2019	2	Female
20	2019	2	Female
21	2020	1	Female
22	2020	1	Male
23	2020	2	Male
24	2018	1	Male
25	2018	2	Male
26	2020	1	Female
27	2021	1	Male

```
[17]: ndf.dropna(how = 'all')
```

```
[17]:
```

	maths_score	reading_score	writing_score	placement_score	\
0	78.000000	63.0	67.0	68.0	
1	67.000000	16.0	73.0	63.0	
2	62.000000	78.0	60.0	63.0	
3	73.000000	69.0	69.0	66.0	
4	67.000000	NaN	77.0	71.0	
5	68.000000	70.0	64.0	69.0	
6	79.000000	64.0	72.0	73.0	
7	78.000000	70.0	78.0	64.0	
8	69.769231	67.0	75.0	62.0	
9	72.000000	80.0	79.0	69.0	

10	69.769231	70.0	68.0	72.0
11	65.000000	79.0	69.0	75.0
12	62.000000	74.0	80.0	75.0
13	66.000000	63.0	NaN	74.0
14	76.000000	60.0	60.0	64.0
15	70.000000	61.0	61.0	63.0
16	72.000000	69.0	65.0	74.0
17	80.000000	76.0	74.0	78.0
18	65.000000	70.0	63.0	76.0
19	71.000000	65.0	68.0	NaN
20	68.000000	65.0	68.0	75.0
21	77.000000	79.0	66.0	62.0
22	68.000000	76.0	80.0	74.0
23	80.000000	71.0	79.0	78.0
24	62.000000	71.0	62.0	60.0
25	62.000000	78.0	66.0	78.0
26	66.000000	77.0	68.0	69.0
27	60.000000	75.0	77.0	65.0

	club_join_year	placement_offer_count	Gender
0	2021	1	Female
1	2019	1	Male
2	2021	1	Male
3	2021	1	Male
4	2020	1	Male
5	2018	1	Male
6	2020	1	Male
7	2021	1	Male
8	2020	1	Male
9	2020	1	Male
10	2019	1	Male
11	2019	2	Female
12	2018	2	Female
13	2020	1	Female
14	2018	1	Female
15	2021	1	Female
16	2020	1	Female
17	2020	2	Female
18	2019	2	Female
19	2019	3	Female
20	2019	2	Female
21	2020	1	Female
22	2020	1	Male
23	2020	2	Male
24	2018	1	Male
25	2018	2	Male
26	2020	1	Female

27                      2021                      1      Male

```
[18]: ndf.dropna(axis = 1)
```

```
[18]:
```

	maths_score	club_join_year	placement_offer_count	Gender
0	78.000000	2021	1	Female
1	67.000000	2019	1	Male
2	62.000000	2021	1	Male
3	73.000000	2021	1	Male
4	67.000000	2020	1	Male
5	68.000000	2018	1	Male
6	79.000000	2020	1	Male
7	78.000000	2021	1	Male
8	69.769231	2020	1	Male
9	72.000000	2020	1	Male
10	69.769231	2019	1	Male
11	65.000000	2019	2	Female
12	62.000000	2018	2	Female
13	66.000000	2020	1	Female
14	76.000000	2018	1	Female
15	70.000000	2021	1	Female
16	72.000000	2020	1	Female
17	80.000000	2020	2	Female
18	65.000000	2019	2	Female
19	71.000000	2019	3	Female
20	68.000000	2019	2	Female
21	77.000000	2020	1	Female
22	68.000000	2020	1	Male
23	80.000000	2020	2	Male
24	62.000000	2018	1	Male
25	62.000000	2018	2	Male
26	66.000000	2020	1	Female
27	60.000000	2021	1	Male

```
[19]: new_data = ndf.dropna(axis = 0, how = 'any')
      new_data
```

```
[19]:
```

	maths_score	reading_score	writing_score	placement_score	\
0	78.000000	63.0	67.0	68.0	
1	67.000000	16.0	73.0	63.0	
2	62.000000	78.0	60.0	63.0	
3	73.000000	69.0	69.0	66.0	
5	68.000000	70.0	64.0	69.0	
6	79.000000	64.0	72.0	73.0	
7	78.000000	70.0	78.0	64.0	
8	69.769231	67.0	75.0	62.0	
9	72.000000	80.0	79.0	69.0	



10	69.769231	70.0	68.0	72.0
11	65.000000	79.0	69.0	75.0
12	62.000000	74.0	80.0	75.0
14	76.000000	60.0	60.0	64.0
15	70.000000	61.0	61.0	63.0
16	72.000000	69.0	65.0	74.0
17	80.000000	76.0	74.0	78.0
18	65.000000	70.0	63.0	76.0
20	68.000000	65.0	68.0	75.0
21	77.000000	79.0	66.0	62.0
22	68.000000	76.0	80.0	74.0
23	80.000000	71.0	79.0	78.0
24	62.000000	71.0	62.0	60.0
25	62.000000	78.0	66.0	78.0
26	66.000000	77.0	68.0	69.0
27	60.000000	75.0	77.0	65.0

	club_join_year	placement_offer_count	Gender
0	2021	1	Female
1	2019	1	Male
2	2021	1	Male
3	2021	1	Male
5	2018	1	Male
6	2020	1	Male
7	2021	1	Male
8	2020	1	Male
9	2020	1	Male
10	2019	1	Male
11	2019	2	Female
12	2018	2	Female
14	2018	1	Female
15	2021	1	Female
16	2020	1	Female
17	2020	2	Female
18	2019	2	Female
20	2019	2	Female
21	2020	1	Female
22	2020	1	Male
23	2020	2	Male
24	2018	1	Male
25	2018	2	Male
26	2020	1	Female
27	2021	1	Male

[ ]: