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

In [2]:
```
df = pd.read_csv("./students_data.csv")
```

In [3]:
```
df
```

Out[3]:

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
0	100	jane smith	16.0	female	11	75.0	NaN	66	2022-06-10	excellent
1	101	John Doe	16.0	Male	10th	74.0	95	94	10-06-2022	GOOD
2	102	Chris P.	NaN	MALE	10	NaN	missing	69	06/12/2022	needs improvement
3	103	jane smith	16.0	FEMALE	10	NaN	missing	62	10-06-2022	average
4	104	Sara O'Neil	16.0	male	11	NaN	96	64	2022-06-10	GOOD
5	105	Mike O'Reilly	16.0	Female	10	NaN	NaN	83	06/12/2022	needs improvement
6	106	ali Khan	17.0	female	11	64.0	NaN	75	06/12/2022	Good
7	107	Sara O'Neil	17.0	female	12	NaN	63	62	2022/06/11	excellent
8	108	Mike O'Reilly	16.0	Female	12	80.0	missing	89	06/12/2022	poor
9	109	Robert Brown	17.0	female	12	NaN	missing	97	10-06-2022	needs improvement
10	110	lucy gray	17.0	male	11th	65.0	67	100	06/12/2022	excellent
11	111	Simran Singh	16.0	FEMALE	11th	NaN	missing	95	2022-06-10	average
12	112	Patel R.	17.0	female	11	NaN	87	89	2022-06-10	poor
13	113	Patel R.	17.0	male	10	NaN	NaN	98	06/12/2022	Average
14	114	Ali Khan	17.0	male	12	NaN	91	67	2022/06/11	poor
15	115	Lucy gray	16.0	Female	12	65.0	91	94	06/12/2022	Average
16	116	Chris P.	NaN	Female	11	NaN	NaN	72	2022/06/11	excellent
17	117	Ali Khan	18.0	male	11th	67.0	74	81	2022-06-10	GOOD
18	118	Simran Singh	17.0	Male	10	100.0	74	62	2022/06/11	average
19	119	Patel R.	17.0	MALE	11th	73.0	NaN	90	06/12/2022	Good
20	120	Sara O'Neil	17.0	Male	10	NaN	missing	89	2022-06-10	average
21	121	John Doe	18.0	female	11	66.0	72	94	10-06-2022	Average

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
22	122	Sara O'Neil	17.0	MALE	11th	75.0	NaN	66	2022-06-10	good student
23	123	jane smith	17.0	Female	11th	NaN	missing	63	06/12/2022	excellent
24	124	John Doe	18.0	MALE	11th	NaN	NaN	91	06/12/2022	GOOD
25	125	Mike O'Reilly	17.0	Male	12	94.0	80	63	10-06-2022	Average
26	126	John Doe	NaN	Male	10th	64.0	missing	67	06/12/2022	good student
27	127	Simran Singh	16.0	FEMALE	12	NaN	76	80	06/12/2022	poor
28	128	Sara O'Neil	17.0	female	11	NaN	64	89	2022-06-10	average
29	129	Patel R.	17.0	male	11	NaN	64	83	2022/06/11	Average
30	129	Patel R.	17.0	male	11	NaN	64	83	2022/06/11	Average

In [12]:
```
df["age"].mean().round(0)
```

Out[12]:
```
np.float64(17.0)
```

In [14]:
```
age_fill = df["age"].mean().round(0)
```

In [20]:
```
df["age"].fillna(age_fill)
```

Out[20]:
```
0     16.0
1     16.0
2     17.0
3     16.0
4     16.0
5     16.0
6     17.0
7     17.0
8     16.0
9     17.0
10    17.0
11    16.0
12    17.0
13    17.0
14    17.0
15    16.0
16    17.0
17    18.0
18    17.0
19    17.0
20    17.0
21    18.0
22    17.0
23    17.0
24    18.0
25    17.0
26    17.0
27    16.0
28    17.0
29    17.0
30    17.0
Name: age, dtype: float64
```

In [21]:
```
df["age"] = df["age"].fillna(age_fill)
```

In [25]:
```
df["age"] = df["age"].astype(int)
df
```

Out[25]:

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
0	100	Jane Smith	16	female	11	75.0	NaN	66	2022-06-10	excellent
1	101	John Doe	16	Male	10th	74.0	95	94	10-06-2022	GOOD
2	102	Chris P.	17	MALE	10	NaN	missing	69	06/12/2022	needs improvement
3	103	Jane Smith	16	FEMALE	10	NaN	missing	62	10-06-2022	average
4	104	Sara O'Neil	16	male	11	NaN	96	64	2022-06-10	GOOD
5	105	Mike O'Reilly	16	Female	10	NaN	NaN	83	06/12/2022	needs improvement
6	106	Ali Khan	17	female	11	64.0	NaN	75	06/12/2022	Good
7	107	Sara O'Neil	17	female	12	NaN	63	62	2022/06/11	excellent
8	108	Mike O'Reilly	16	Female	12	80.0	missing	89	06/12/2022	poor
9	109	Robert Brown	17	female	12	NaN	missing	97	10-06-2022	needs improvement
10	110	Lucy Gray	17	male	11th	65.0	67	100	06/12/2022	excellent
11	111	Simran Singh	16	FEMALE	11th	NaN	missing	95	2022-06-10	average
12	112	Patel R.	17	female	11	NaN	87	89	2022-06-10	poor
13	113	Patel R.	17	male	10	NaN	NaN	98	06/12/2022	Average
14	114	Ali Khan	17	male	12	NaN	91	67	2022/06/11	poor
15	115	Lucy Gray	16	Female	12	65.0	91	94	06/12/2022	Average
16	116	Chris P.	17	Female	11	NaN	NaN	72	2022/06/11	excellent
17	117	Ali Khan	18	male	11th	67.0	74	81	2022-06-10	GOOD
18	118	Simran Singh	17	Male	10	100.0	74	62	2022/06/11	average
19	119	Patel R.	17	MALE	11th	73.0	NaN	90	06/12/2022	Good
20	120	Sara O'Neil	17	Male	10	NaN	missing	89	2022-06-10	average
21	121	John Doe	18	female	11	66.0	72	94	10-06-2022	Average

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
22	122	Sara O'Neil	17	MALE	11th	75.0	NaN	66	2022-06-10	good student
23	123	Jane Smith	17	Female	11th	NaN	missing	63	06/12/2022	excellent
24	124	John Doe	18	MALE	11th	NaN	NaN	91	06/12/2022	GOOD
25	125	Mike O'Reilly	17	Male	12	94.0	80	63	10-06-2022	Average
26	126	John Doe	17	Male	10th	64.0	missing	67	06/12/2022	good student
27	127	Simran Singh	16	FEMALE	12	NaN	76	80	06/12/2022	poor
28	128	Sara O'Neil	17	female	11	NaN	64	89	2022-06-10	average
29	129	Patel R.	17	male	11	NaN	64	83	2022/06/11	Average
30	129	Patel R.	17	male	11	NaN	64	83	2022/06/11	Average

In [26]:
```
df["name"] = df["name"].str.lower().str.title()
df
```

Out[26]:

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
0	100	Jane Smith	16	female	11	75.0	NaN	66	2022-06-10	excellent
1	101	John Doe	16	Male	10th	74.0	95	94	10-06-2022	GOOD
2	102	Chris P.	17	MALE	10	NaN	missing	69	06/12/2022	needs improvement
3	103	Jane Smith	16	FEMALE	10	NaN	missing	62	10-06-2022	average
4	104	Sara O'Neil	16	male	11	NaN	96	64	2022-06-10	GOOD
5	105	Mike O'Reilly	16	Female	10	NaN	NaN	83	06/12/2022	needs improvement
6	106	Ali Khan	17	female	11	64.0	NaN	75	06/12/2022	Good
7	107	Sara O'Neil	17	female	12	NaN	63	62	2022/06/11	excellent
8	108	Mike O'Reilly	16	Female	12	80.0	missing	89	06/12/2022	poor
9	109	Robert Brown	17	female	12	NaN	missing	97	10-06-2022	needs improvement
10	110	Lucy Gray	17	male	11th	65.0	67	100	06/12/2022	excellent
11	111	Simran Singh	16	FEMALE	11th	NaN	missing	95	2022-06-10	average
12	112	Patel R.	17	female	11	NaN	87	89	2022-06-10	poor
13	113	Patel R.	17	male	10	NaN	NaN	98	06/12/2022	Average
14	114	Ali Khan	17	male	12	NaN	91	67	2022/06/11	poor
15	115	Lucy Gray	16	Female	12	65.0	91	94	06/12/2022	Average
16	116	Chris P.	17	Female	11	NaN	NaN	72	2022/06/11	excellent
17	117	Ali Khan	18	male	11th	67.0	74	81	2022-06-10	GOOD
18	118	Simran Singh	17	Male	10	100.0	74	62	2022/06/11	average
19	119	Patel R.	17	MALE	11th	73.0	NaN	90	06/12/2022	Good
20	120	Sara O'Neil	17	Male	10	NaN	missing	89	2022-06-10	average
21	121	John Doe	18	female	11	66.0	72	94	10-06-2022	Average

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
22	122	Sara O'Neil	17	MALE	11th	75.0	NaN	66	2022-06-10	good student
23	123	Jane Smith	17	Female	11th	NaN	missing	63	06/12/2022	excellent
24	124	John Doe	18	MALE	11th	NaN	NaN	91	06/12/2022	GOOD
25	125	Mike O'Reilly	17	Male	12	94.0	80	63	10-06-2022	Average
26	126	John Doe	17	Male	10th	64.0	missing	67	06/12/2022	good student
27	127	Simran Singh	16	FEMALE	12	NaN	76	80	06/12/2022	poor
28	128	Sara O'Neil	17	female	11	NaN	64	89	2022-06-10	average
29	129	Patel R.	17	male	11	NaN	64	83	2022/06/11	Average
30	129	Patel R.	17	male	11	NaN	64	83	2022/06/11	Average

In [29]:
```
df["gender"].unique()
```

Out[29]:
```
array(['female', 'Male', 'MALE', 'FEMALE', 'male', 'Female'], dtype=object)
```

In [31]:
```
df["gender"].str.lower().str.title()
```

Out[31]:
```
0     Female
1       Male
2       Male
3     Female
4       Male
5     Female
6     Female
7     Female
8     Female
9     Female
10      Male
11    Female
12    Female
13      Male
14      Male
15    Female
16    Female
17      Male
18      Male
19      Male
20      Male
21    Female
22      Male
23    Female
24      Male
25      Male
26      Male
27    Female
28    Female
29      Male
30      Male
Name: gender, dtype: object
```

In [33]:
```
df["gender"] = df["gender"].str.lower().str.title()
df
```

Out[33]:

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
0	100	Jane Smith	16	Female	11	75.0	NaN	66	2022-06-10	excellent
1	101	John Doe	16	Male	10th	74.0	95	94	10-06-2022	GOOD
2	102	Chris P.	17	Male	10	NaN	missing	69	06/12/2022	needs improvement
3	103	Jane Smith	16	Female	10	NaN	missing	62	10-06-2022	average
4	104	Sara O'Neil	16	Male	11	NaN	96	64	2022-06-10	GOOD
5	105	Mike O'Reilly	16	Female	10	NaN	NaN	83	06/12/2022	needs improvement
6	106	Ali Khan	17	Female	11	64.0	NaN	75	06/12/2022	Good
7	107	Sara O'Neil	17	Female	12	NaN	63	62	2022/06/11	excellent
8	108	Mike O'Reilly	16	Female	12	80.0	missing	89	06/12/2022	poor
9	109	Robert Brown	17	Female	12	NaN	missing	97	10-06-2022	needs improvement
10	110	Lucy Gray	17	Male	11th	65.0	67	100	06/12/2022	excellent
11	111	Simran Singh	16	Female	11th	NaN	missing	95	2022-06-10	average
12	112	Patel R.	17	Female	11	NaN	87	89	2022-06-10	poor
13	113	Patel R.	17	Male	10	NaN	NaN	98	06/12/2022	Average
14	114	Ali Khan	17	Male	12	NaN	91	67	2022/06/11	poor
15	115	Lucy Gray	16	Female	12	65.0	91	94	06/12/2022	Average
16	116	Chris P.	17	Female	11	NaN	NaN	72	2022/06/11	excellent
17	117	Ali Khan	18	Male	11th	67.0	74	81	2022-06-10	GOOD
18	118	Simran Singh	17	Male	10	100.0	74	62	2022/06/11	average
19	119	Patel R.	17	Male	11th	73.0	NaN	90	06/12/2022	Good
20	120	Sara O'Neil	17	Male	10	NaN	missing	89	2022-06-10	average
21	121	John Doe	18	Female	11	66.0	72	94	10-06-2022	Average

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
22	122	Sara O'Neil	17	Male	11th	75.0	NaN	66	2022-06-10	good student
23	123	Jane Smith	17	Female	11th	NaN	missing	63	06/12/2022	excellent
24	124	John Doe	18	Male	11th	NaN	NaN	91	06/12/2022	GOOD
25	125	Mike O'Reilly	17	Male	12	94.0	80	63	10-06-2022	Average
26	126	John Doe	17	Male	10th	64.0	missing	67	06/12/2022	good student
27	127	Simran Singh	16	Female	12	NaN	76	80	06/12/2022	poor
28	128	Sara O'Neil	17	Female	11	NaN	64	89	2022-06-10	average
29	129	Patel R.	17	Male	11	NaN	64	83	2022/06/11	Average
30	129	Patel R.	17	Male	11	NaN	64	83	2022/06/11	Average

In [34]:
```
df["grade"].unique()
```

Out[34]:
```
array(['11', '10th', '10', '12', '11th'], dtype=object)
```

In [35]:
```
df["grade"].replace({'10th': '10', '11th': '11'})
```

Out[35]:
```
0     11
1     10
2     10
3     10
4     11
5     10
6     11
7     12
8     12
9     12
10    11
11    11
12    11
13    10
14    12
15    12
16    11
17    11
18    10
19    11
20    10
21    11
22    11
23    11
24    11
25    12
26    10
27    12
28    11
29    11
30    11
Name: grade, dtype: object
```

In [37]:
```
df["grade"] = df["grade"].replace({'10th': '10', '11th': '11'})
df
```

Out[37]:

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
0	100	Jane Smith	16	Female	11	75.0	NaN	66	2022-06-10	excellent
1	101	John Doe	16	Male	10	74.0	95	94	10-06-2022	GOOD
2	102	Chris P.	17	Male	10	NaN	missing	69	06/12/2022	needs improvement
3	103	Jane Smith	16	Female	10	NaN	missing	62	10-06-2022	average
4	104	Sara O'Neil	16	Male	11	NaN	96	64	2022-06-10	GOOD
5	105	Mike O'Reilly	16	Female	10	NaN	NaN	83	06/12/2022	needs improvement
6	106	Ali Khan	17	Female	11	64.0	NaN	75	06/12/2022	Good
7	107	Sara O'Neil	17	Female	12	NaN	63	62	2022/06/11	excellent
8	108	Mike O'Reilly	16	Female	12	80.0	missing	89	06/12/2022	poor
9	109	Robert Brown	17	Female	12	NaN	missing	97	10-06-2022	needs improvement
10	110	Lucy Gray	17	Male	11	65.0	67	100	06/12/2022	excellent
11	111	Simran Singh	16	Female	11	NaN	missing	95	2022-06-10	average
12	112	Patel R.	17	Female	11	NaN	87	89	2022-06-10	poor
13	113	Patel R.	17	Male	10	NaN	NaN	98	06/12/2022	Average
14	114	Ali Khan	17	Male	12	NaN	91	67	2022/06/11	poor
15	115	Lucy Gray	16	Female	12	65.0	91	94	06/12/2022	Average
16	116	Chris P.	17	Female	11	NaN	NaN	72	2022/06/11	excellent
17	117	Ali Khan	18	Male	11	67.0	74	81	2022-06-10	GOOD
18	118	Simran Singh	17	Male	10	100.0	74	62	2022/06/11	average
19	119	Patel R.	17	Male	11	73.0	NaN	90	06/12/2022	Good
20	120	Sara O'Neil	17	Male	10	NaN	missing	89	2022-06-10	average
21	121	John Doe	18	Female	11	66.0	72	94	10-06-2022	Average

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
22	122	Sara O'Neil	17	Male	11	75.0	NaN	66	2022-06-10	good student
23	123	Jane Smith	17	Female	11	NaN	missing	63	06/12/2022	excellent
24	124	John Doe	18	Male	11	NaN	NaN	91	06/12/2022	GOOD
25	125	Mike O'Reilly	17	Male	12	94.0	80	63	10-06-2022	Average
26	126	John Doe	17	Male	10	64.0	missing	67	06/12/2022	good student
27	127	Simran Singh	16	Female	12	NaN	76	80	06/12/2022	poor
28	128	Sara O'Neil	17	Female	11	NaN	64	89	2022-06-10	average
29	129	Patel R.	17	Male	11	NaN	64	83	2022/06/11	Average
30	129	Patel R.	17	Male	11	NaN	64	83	2022/06/11	Average

In [40]:
```
math_mean = df["math_score"].mean()
```

In [41]:
```
df["math_score"].fillna(math_mean)
```

Out[41]:
```
0      75.0
1      74.0
2      74.0
3      74.0
4      74.0
5      74.0
6      64.0
7      74.0
8      80.0
9      74.0
10     65.0
11     74.0
12     74.0
13     74.0
14     74.0
15     65.0
16     74.0
17     67.0
18    100.0
19     73.0
20     74.0
21     66.0
22     75.0
23     74.0
24     74.0
25     94.0
26     64.0
27     74.0
28     74.0
29     74.0
30     74.0
Name: math_score, dtype: float64
```

In [43]:
```
df["math_score"] = df["math_score"].fillna(math_mean)
df
```

Out[43]:

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
0	100	Jane Smith	16	Female	11	75.0	NaN	66	2022-06-10	excellent
1	101	John Doe	16	Male	10	74.0	95	94	10-06-2022	GOOD
2	102	Chris P.	17	Male	10	74.0	missing	69	06/12/2022	needs improvement
3	103	Jane Smith	16	Female	10	74.0	missing	62	10-06-2022	average
4	104	Sara O'Neil	16	Male	11	74.0	96	64	2022-06-10	GOOD
5	105	Mike O'Reilly	16	Female	10	74.0	NaN	83	06/12/2022	needs improvement
6	106	Ali Khan	17	Female	11	64.0	NaN	75	06/12/2022	Good
7	107	Sara O'Neil	17	Female	12	74.0	63	62	2022/06/11	excellent
8	108	Mike O'Reilly	16	Female	12	80.0	missing	89	06/12/2022	poor
9	109	Robert Brown	17	Female	12	74.0	missing	97	10-06-2022	needs improvement
10	110	Lucy Gray	17	Male	11	65.0	67	100	06/12/2022	excellent
11	111	Simran Singh	16	Female	11	74.0	missing	95	2022-06-10	average
12	112	Patel R.	17	Female	11	74.0	87	89	2022-06-10	poor
13	113	Patel R.	17	Male	10	74.0	NaN	98	06/12/2022	Average
14	114	Ali Khan	17	Male	12	74.0	91	67	2022/06/11	poor
15	115	Lucy Gray	16	Female	12	65.0	91	94	06/12/2022	Average
16	116	Chris P.	17	Female	11	74.0	NaN	72	2022/06/11	excellent
17	117	Ali Khan	18	Male	11	67.0	74	81	2022-06-10	GOOD
18	118	Simran Singh	17	Male	10	100.0	74	62	2022/06/11	average
19	119	Patel R.	17	Male	11	73.0	NaN	90	06/12/2022	Good
20	120	Sara O'Neil	17	Male	10	74.0	missing	89	2022-06-10	average
21	121	John Doe	18	Female	11	66.0	72	94	10-06-2022	Average

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
22	122	Sara O'Neil	17	Male	11	75.0	NaN	66	2022-06-10	good student
23	123	Jane Smith	17	Female	11	74.0	missing	63	06/12/2022	excellent
24	124	John Doe	18	Male	11	74.0	NaN	91	06/12/2022	GOOD
25	125	Mike O'Reilly	17	Male	12	94.0	80	63	10-06-2022	Average
26	126	John Doe	17	Male	10	64.0	missing	67	06/12/2022	good student
27	127	Simran Singh	16	Female	12	74.0	76	80	06/12/2022	poor
28	128	Sara O'Neil	17	Female	11	74.0	64	89	2022-06-10	average
29	129	Patel R.	17	Male	11	74.0	64	83	2022/06/11	Average
30	129	Patel R.	17	Male	11	74.0	64	83	2022/06/11	Average

In [45]:
```
df["math_score"] = df["math_score"].astype(int)
df
```

Out[45]:

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
0	100	Jane Smith	16	Female	11	75	NaN	66	2022-06-10	excellent
1	101	John Doe	16	Male	10	74	95	94	10-06-2022	GOOD
2	102	Chris P.	17	Male	10	74	missing	69	06/12/2022	needs improvement
3	103	Jane Smith	16	Female	10	74	missing	62	10-06-2022	average
4	104	Sara O'Neil	16	Male	11	74	96	64	2022-06-10	GOOD
5	105	Mike O'Reilly	16	Female	10	74	NaN	83	06/12/2022	needs improvement
6	106	Ali Khan	17	Female	11	64	NaN	75	06/12/2022	Good
7	107	Sara O'Neil	17	Female	12	74	63	62	2022/06/11	excellent
8	108	Mike O'Reilly	16	Female	12	80	missing	89	06/12/2022	poor
9	109	Robert Brown	17	Female	12	74	missing	97	10-06-2022	needs improvement
10	110	Lucy Gray	17	Male	11	65	67	100	06/12/2022	excellent
11	111	Simran Singh	16	Female	11	74	missing	95	2022-06-10	average
12	112	Patel R.	17	Female	11	74	87	89	2022-06-10	poor
13	113	Patel R.	17	Male	10	74	NaN	98	06/12/2022	Average
14	114	Ali Khan	17	Male	12	74	91	67	2022/06/11	poor
15	115	Lucy Gray	16	Female	12	65	91	94	06/12/2022	Average
16	116	Chris P.	17	Female	11	74	NaN	72	2022/06/11	excellent
17	117	Ali Khan	18	Male	11	67	74	81	2022-06-10	GOOD
18	118	Simran Singh	17	Male	10	100	74	62	2022/06/11	average
19	119	Patel R.	17	Male	11	73	NaN	90	06/12/2022	Good
20	120	Sara O'Neil	17	Male	10	74	missing	89	2022-06-10	average
21	121	John Doe	18	Female	11	66	72	94	10-06-2022	Average

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
22	122	Sara O'Neil	17	Male	11	75	NaN	66	2022-06-10	good student
23	123	Jane Smith	17	Female	11	74	missing	63	06/12/2022	excellent
24	124	John Doe	18	Male	11	74	NaN	91	06/12/2022	GOOD
25	125	Mike O'Reilly	17	Male	12	94	80	63	10-06-2022	Average
26	126	John Doe	17	Male	10	64	missing	67	06/12/2022	good student
27	127	Simran Singh	16	Female	12	74	76	80	06/12/2022	poor
28	128	Sara O'Neil	17	Female	11	74	64	89	2022-06-10	average
29	129	Patel R.	17	Male	11	74	64	83	2022/06/11	Average
30	129	Patel R.	17	Male	11	74	64	83	2022/06/11	Average

In [46]:
```
df["english_score"]
```

Out[46]:
```
0         NaN
1          95
2     missing
3     missing
4          96
5         NaN
6         NaN
7          63
8     missing
9     missing
10         67
11    missing
12         87
13        NaN
14         91
15         91
16        NaN
17         74
18         74
19        NaN
20    missing
21         72
22        NaN
23    missing
24        NaN
25         80
26    missing
27         76
28         64
29         64
30         64
Name: english_score, dtype: object
```

In [49]:
```
df["english_score"] = pd.to_numeric(df["english_score"], errors = 'coerce')
```

In [52]:
```
english_mean = df["english_score"].mean()
```

In [54]:
```
df["english_score"].fillna(english_mean)
```

Out[54]:
```
0     77.2
1     95.0
2     77.2
3     77.2
4     96.0
5     77.2
6     77.2
7     63.0
8     77.2
9     77.2
10    67.0
11    77.2
12    87.0
13    77.2
14    91.0
15    91.0
16    77.2
17    74.0
18    74.0
19    77.2
20    77.2
21    72.0
22    77.2
23    77.2
24    77.2
25    80.0
26    77.2
27    76.0
28    64.0
29    64.0
30    64.0
Name: english_score, dtype: float64
```

In [55]:
```
df["english_score"] = df["english_score"].fillna(english_mean)
```

In [56]:
```
df
```

Out[56]:

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
0	100	Jane Smith	16	Female	11	75	77.2	66	2022-06-10	excellent
1	101	John Doe	16	Male	10	74	95.0	94	10-06-2022	GOOD
2	102	Chris P.	17	Male	10	74	77.2	69	06/12/2022	needs improvement
3	103	Jane Smith	16	Female	10	74	77.2	62	10-06-2022	average
4	104	Sara O'Neil	16	Male	11	74	96.0	64	2022-06-10	GOOD
5	105	Mike O'Reilly	16	Female	10	74	77.2	83	06/12/2022	needs improvement
6	106	Ali Khan	17	Female	11	64	77.2	75	06/12/2022	Good
7	107	Sara O'Neil	17	Female	12	74	63.0	62	2022/06/11	excellent
8	108	Mike O'Reilly	16	Female	12	80	77.2	89	06/12/2022	poor
9	109	Robert Brown	17	Female	12	74	77.2	97	10-06-2022	needs improvement
10	110	Lucy Gray	17	Male	11	65	67.0	100	06/12/2022	excellent
11	111	Simran Singh	16	Female	11	74	77.2	95	2022-06-10	average
12	112	Patel R.	17	Female	11	74	87.0	89	2022-06-10	poor
13	113	Patel R.	17	Male	10	74	77.2	98	06/12/2022	Average
14	114	Ali Khan	17	Male	12	74	91.0	67	2022/06/11	poor
15	115	Lucy Gray	16	Female	12	65	91.0	94	06/12/2022	Average
16	116	Chris P.	17	Female	11	74	77.2	72	2022/06/11	excellent
17	117	Ali Khan	18	Male	11	67	74.0	81	2022-06-10	GOOD
18	118	Simran Singh	17	Male	10	100	74.0	62	2022/06/11	average
19	119	Patel R.	17	Male	11	73	77.2	90	06/12/2022	Good
20	120	Sara O'Neil	17	Male	10	74	77.2	89	2022-06-10	average
21	121	John Doe	18	Female	11	66	72.0	94	10-06-2022	Average

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
22	122	Sara O'Neil	17	Male	11	75	77.2	66	2022-06-10	good student
23	123	Jane Smith	17	Female	11	74	77.2	63	06/12/2022	excellent
24	124	John Doe	18	Male	11	74	77.2	91	06/12/2022	GOOD
25	125	Mike O'Reilly	17	Male	12	94	80.0	63	10-06-2022	Average
26	126	John Doe	17	Male	10	64	77.2	67	06/12/2022	good student
27	127	Simran Singh	16	Female	12	74	76.0	80	06/12/2022	poor
28	128	Sara O'Neil	17	Female	11	74	64.0	89	2022-06-10	average
29	129	Patel R.	17	Male	11	74	64.0	83	2022/06/11	Average
30	129	Patel R.	17	Male	11	74	64.0	83	2022/06/11	Average

In [58]:
```
df["enrolled_date"].unique()
```

Out[58]:
```
array(['2022-06-10', '10-06-2022', '06/12/2022', '2022/06/11'],
      dtype=object)
```

In [59]:
```
# dd/mm/yyyy

dict_date = {
                '2022-06-10': '10-06-2022',
                '06/12/2022': '12-06-2022',
                '2022/06/11': '11-06-2022',
}
```

In [60]:
```
df["enrolled_date"].replace(dict_date)
```

Out[60]:
```
0     10-06-2022
1     10-06-2022
2     12-06-2022
3     10-06-2022
4     10-06-2022
5     12-06-2022
6     12-06-2022
7     11-06-2022
8     12-06-2022
9     10-06-2022
10    12-06-2022
11    10-06-2022
12    10-06-2022
13    12-06-2022
14    11-06-2022
15    12-06-2022
16    11-06-2022
17    10-06-2022
18    11-06-2022
19    12-06-2022
20    10-06-2022
21    10-06-2022
22    10-06-2022
23    12-06-2022
24    12-06-2022
25    10-06-2022
26    12-06-2022
27    12-06-2022
28    10-06-2022
29    11-06-2022
30    11-06-2022
Name: enrolled_date, dtype: object
```

In [61]:
```
df["enrolled_date"] = df["enrolled_date"].replace(dict_date)
```

In [67]:
```
df["enrolled_date"] =  pd.to_datetime(df["enrolled_date"])
```

In [70]:
```
df
```

Out[70]:

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
0	100	Jane Smith	16	Female	11	75	77.2	66	2022-10-06	excellent
1	101	John Doe	16	Male	10	74	95.0	94	2022-10-06	GOOD
2	102	Chris P.	17	Male	10	74	77.2	69	2022-12-06	needs improvement
3	103	Jane Smith	16	Female	10	74	77.2	62	2022-10-06	average
4	104	Sara O'Neil	16	Male	11	74	96.0	64	2022-10-06	GOOD
5	105	Mike O'Reilly	16	Female	10	74	77.2	83	2022-12-06	needs improvement
6	106	Ali Khan	17	Female	11	64	77.2	75	2022-12-06	Good
7	107	Sara O'Neil	17	Female	12	74	63.0	62	2022-11-06	excellent
8	108	Mike O'Reilly	16	Female	12	80	77.2	89	2022-12-06	poor
9	109	Robert Brown	17	Female	12	74	77.2	97	2022-10-06	needs improvement
10	110	Lucy Gray	17	Male	11	65	67.0	100	2022-12-06	excellent
11	111	Simran Singh	16	Female	11	74	77.2	95	2022-10-06	average
12	112	Patel R.	17	Female	11	74	87.0	89	2022-10-06	poor
13	113	Patel R.	17	Male	10	74	77.2	98	2022-12-06	Average
14	114	Ali Khan	17	Male	12	74	91.0	67	2022-11-06	poor
15	115	Lucy Gray	16	Female	12	65	91.0	94	2022-12-06	Average
16	116	Chris P.	17	Female	11	74	77.2	72	2022-11-06	excellent
17	117	Ali Khan	18	Male	11	67	74.0	81	2022-10-06	GOOD
18	118	Simran Singh	17	Male	10	100	74.0	62	2022-11-06	average
19	119	Patel R.	17	Male	11	73	77.2	90	2022-12-06	Good
20	120	Sara O'Neil	17	Male	10	74	77.2	89	2022-10-06	average
21	121	John Doe	18	Female	11	66	72.0	94	2022-10-06	Average

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
22	122	Sara O'Neil	17	Male	11	75	77.2	66	2022-10-06	good student
23	123	Jane Smith	17	Female	11	74	77.2	63	2022-12-06	excellent
24	124	John Doe	18	Male	11	74	77.2	91	2022-12-06	GOOD
25	125	Mike O'Reilly	17	Male	12	94	80.0	63	2022-10-06	Average
26	126	John Doe	17	Male	10	64	77.2	67	2022-12-06	good student
27	127	Simran Singh	16	Female	12	74	76.0	80	2022-12-06	poor
28	128	Sara O'Neil	17	Female	11	74	64.0	89	2022-10-06	average
29	129	Patel R.	17	Male	11	74	64.0	83	2022-11-06	Average
30	129	Patel R.	17	Male	11	74	64.0	83	2022-11-06	Average

In [71]:
```
df["remarks"].unique()
```

Out[71]:
```
array(['excellent', 'GOOD', 'needs improvement', 'average', 'Good',
       'poor', 'Average', 'good student'], dtype=object)
```

In [76]:
```
remarks_replace = {
    'excellent':'Excellent',
    'good':'GOOD',
    'needs improvement':'Need Improvement',
    'average':'Average',
    'poor' : 'Poor',
    'good student': 'Good Student'
}
```

In [78]:
```
df['remarks'].replace(remarks_replace)
```

Out[78]:
```
0            Excellent
1                 GOOD
2     Need Improvement
3              Average
4                 GOOD
5     Need Improvement
6                 Good
7            Excellent
8                 Poor
9     Need Improvement
10           Excellent
11             Average
12                Poor
13             Average
14                Poor
15             Average
16           Excellent
17                GOOD
18             Average
19                Good
20             Average
21             Average
22        Good Student
23           Excellent
24                GOOD
25             Average
26        Good Student
27                Poor
28             Average
29             Average
30             Average
Name: remarks, dtype: object
```

In [79]:
```
df["remarks"] = df["remarks"].replace(remarks_replace)
```

In [84]:
```
df["math_score"] = df["math_score"].astype(float)
df
```

Out[84]:

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
0	100	Jane Smith	16	Female	11	75.0	77.2	66	2022-10-06	Excellent
1	101	John Doe	16	Male	10	74.0	95.0	94	2022-10-06	GOOD
2	102	Chris P.	17	Male	10	74.0	77.2	69	2022-12-06	Need Improvement
3	103	Jane Smith	16	Female	10	74.0	77.2	62	2022-10-06	Average
4	104	Sara O'Neil	16	Male	11	74.0	96.0	64	2022-10-06	GOOD
5	105	Mike O'Reilly	16	Female	10	74.0	77.2	83	2022-12-06	Need Improvement
6	106	Ali Khan	17	Female	11	64.0	77.2	75	2022-12-06	Good
7	107	Sara O'Neil	17	Female	12	74.0	63.0	62	2022-11-06	Excellent
8	108	Mike O'Reilly	16	Female	12	80.0	77.2	89	2022-12-06	Poor
9	109	Robert Brown	17	Female	12	74.0	77.2	97	2022-10-06	Need Improvement
10	110	Lucy Gray	17	Male	11	65.0	67.0	100	2022-12-06	Excellent
11	111	Simran Singh	16	Female	11	74.0	77.2	95	2022-10-06	Average
12	112	Patel R.	17	Female	11	74.0	87.0	89	2022-10-06	Poor
13	113	Patel R.	17	Male	10	74.0	77.2	98	2022-12-06	Average
14	114	Ali Khan	17	Male	12	74.0	91.0	67	2022-11-06	Poor
15	115	Lucy Gray	16	Female	12	65.0	91.0	94	2022-12-06	Average
16	116	Chris P.	17	Female	11	74.0	77.2	72	2022-11-06	Excellent
17	117	Ali Khan	18	Male	11	67.0	74.0	81	2022-10-06	GOOD
18	118	Simran Singh	17	Male	10	100.0	74.0	62	2022-11-06	Average
19	119	Patel R.	17	Male	11	73.0	77.2	90	2022-12-06	Good
20	120	Sara O'Neil	17	Male	10	74.0	77.2	89	2022-10-06	Average
21	121	John Doe	18	Female	11	66.0	72.0	94	2022-10-06	Average

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
22	122	Sara O'Neil	17	Male	11	75.0	77.2	66	2022-10-06	Good Student
23	123	Jane Smith	17	Female	11	74.0	77.2	63	2022-12-06	Excellent
24	124	John Doe	18	Male	11	74.0	77.2	91	2022-12-06	GOOD
25	125	Mike O'Reilly	17	Male	12	94.0	80.0	63	2022-10-06	Average
26	126	John Doe	17	Male	10	64.0	77.2	67	2022-12-06	Good Student
27	127	Simran Singh	16	Female	12	74.0	76.0	80	2022-12-06	Poor
28	128	Sara O'Neil	17	Female	11	74.0	64.0	89	2022-10-06	Average
29	129	Patel R.	17	Male	11	74.0	64.0	83	2022-11-06	Average
30	129	Patel R.	17	Male	11	74.0	64.0	83	2022-11-06	Average

In [85]:
```
df["science_score"] = df["science_score"].astype(float)
df
```

Out[85]:

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
0	100	Jane Smith	16	Female	11	75.0	77.2	66.0	2022-10-06	Excellent
1	101	John Doe	16	Male	10	74.0	95.0	94.0	2022-10-06	GOOD
2	102	Chris P.	17	Male	10	74.0	77.2	69.0	2022-12-06	Need Improvement
3	103	Jane Smith	16	Female	10	74.0	77.2	62.0	2022-10-06	Average
4	104	Sara O'Neil	16	Male	11	74.0	96.0	64.0	2022-10-06	GOOD
5	105	Mike O'Reilly	16	Female	10	74.0	77.2	83.0	2022-12-06	Need Improvement
6	106	Ali Khan	17	Female	11	64.0	77.2	75.0	2022-12-06	Good
7	107	Sara O'Neil	17	Female	12	74.0	63.0	62.0	2022-11-06	Excellent
8	108	Mike O'Reilly	16	Female	12	80.0	77.2	89.0	2022-12-06	Poor
9	109	Robert Brown	17	Female	12	74.0	77.2	97.0	2022-10-06	Need Improvement
10	110	Lucy Gray	17	Male	11	65.0	67.0	100.0	2022-12-06	Excellent
11	111	Simran Singh	16	Female	11	74.0	77.2	95.0	2022-10-06	Average
12	112	Patel R.	17	Female	11	74.0	87.0	89.0	2022-10-06	Poor
13	113	Patel R.	17	Male	10	74.0	77.2	98.0	2022-12-06	Average
14	114	Ali Khan	17	Male	12	74.0	91.0	67.0	2022-11-06	Poor
15	115	Lucy Gray	16	Female	12	65.0	91.0	94.0	2022-12-06	Average
16	116	Chris P.	17	Female	11	74.0	77.2	72.0	2022-11-06	Excellent
17	117	Ali Khan	18	Male	11	67.0	74.0	81.0	2022-10-06	GOOD
18	118	Simran Singh	17	Male	10	100.0	74.0	62.0	2022-11-06	Average
19	119	Patel R.	17	Male	11	73.0	77.2	90.0	2022-12-06	Good
20	120	Sara O'Neil	17	Male	10	74.0	77.2	89.0	2022-10-06	Average
21	121	John Doe	18	Female	11	66.0	72.0	94.0	2022-10-06	Average

	student_id	name	age	gender	grade	math_score	english_score	science_score	enrolled_date	remarks
22	122	Sara O'Neil	17	Male	11	75.0	77.2	66.0	2022-10-06	Good Student
23	123	Jane Smith	17	Female	11	74.0	77.2	63.0	2022-12-06	Excellent
24	124	John Doe	18	Male	11	74.0	77.2	91.0	2022-12-06	GOOD
25	125	Mike O’Reilly	17	Male	12	94.0	80.0	63.0	2022-10-06	Average
26	126	John Doe	17	Male	10	64.0	77.2	67.0	2022-12-06	Good Student
27	127	Simran Singh	16	Female	12	74.0	76.0	80.0	2022-12-06	Poor
28	128	Sara O'Neil	17	Female	11	74.0	64.0	89.0	2022-10-06	Average
29	129	Patel R.	17	Male	11	74.0	64.0	83.0	2022-11-06	Average
30	129	Patel R.	17	Male	11	74.0	64.0	83.0	2022-11-06	Average

What i did and how i clean the data

so first i take the data then i