```
import pandas as pd
import numpy as np


df = pd.read_csv('/content/Student_performance_10k.csv')
df
```

	roll_no	gender	race_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writi
0	std-01	male	group D	some college	1.0	1.0	89	38.0	
1	std-02	male	group B	high school	1.0	0.0	65	100.0	
2	std-03	male	group C	master's degree	1.0	0.0	10	99.0	
3	std-04	male	group D	some college	1.0	1.0	22	51.0	
4	std-05	male	group C	some college	0.0	1.0	26	58.0	
...	...	...	...	...	...	...	...	...	
9995	std-9996	female	group C	some college	1.0	0.0	78	60.0	
9996	std-9997	female	group C	bachelor's degree	1.0	0.0	37	92.0	
9997	std-9998	female	group B	associate's degree	1.0	1.0	70	71.0	
9998	std-9999	female	group B	some high school	1.0	0.0	87	60.0	

Next steps: Generate code with df | View recommended plots | New interactive sheet

2. Check for null values in the dataset

```
print("\nChecking for null values:\n", df.isnull().sum())
```

```
Checking for null values:
 roll_no                        1
gender                         18
race_ethnicity                 23
parental_level_of_education    22
lunch                          24
test_preparation_course        23
math_score                     24
reading_score                  25
writing_score                  24
science_score                  23
total_score                    19
grade                           3
dtype: int64
```

3. Replace missing values (nulls) with standard NaN value

```
df.replace("", np.nan, inplace=True)  # Replaces empty strings with NaN
print("\nAfter replacing missing values with NaN:\n", df.isnull().sum())
df
```

```
After replacing missing values with NaN:
 roll_no                        1
gender                         18
race_ethnicity                 23
parental_level_of_education    22
lunch                          24
test_preparation_course        23
math_score                     24
reading_score                  25
writing_score                  24
science_score                  23
total_score                    19
grade                           3
dtype: int64
None
```

Like what you see? Visit the data table notebook to learn more about interactive tables.

1 to 25 of 10000 entries Filter

index	roll_no	gender	race_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score ▲	reading_score	writing_score	science
99	std-100	male	group C	some high school	0.0	0.0	0	37.0	62.0	
160	std-161	male	group C	master's degree	NaN	0.0	0	93.0	66.0	
270	std-271	male	group A	high school	1.0	1.0	0	85.0	99.0	
574	std-575	male	group D	some college	0.0	1.0	0	48.0	86.0	
1180	std-1181	male	group A	bachelor's degree	1.0	0.0	0	72.0	45.0	
1304	std-1305	male	group B	high school	1.0	0.0	0	50.0	64.0	
1460	std-1461	male	group C	associate's degree	1.0	0.0	0	82.0	84.0	
1728	std-1729	male	group C	some college	0.0	0.0	0	71.0	91.0	
2017	std-2018	male	group D	associate's degree	1.0	1.0	0	53.0	74.0	
2110	std-2111	male	group E	bachelor's degree	1.0	1.0	0	94.0	52.0	
2164	std-2165	male	group A	associate's degree	1.0	0.0	0	73.0	53.0	
2216	std-2217	male	group D	some college	1.0	1.0	0	43.0	33.0	
2484	std-2485	male	group D	some college	0.0	0.0	0	63.0	36.0	
2603	std-2604	male	group C	master's degree	1.0	0.0	0	66.0	48.0	
2709	std-2710	male	group B	some high school	1.0	1.0	0	100.0	51.0	
2749	std-2750	male	group E	some high school	0.0	0.0	0	73.0	55.0	
2782	std-2783	male	group B	high school	1.0	1.0	0	86.0	68.0	
2936	std-2937	male	group E	master's degree	0.0	0.0	0	49.0	45.0	

Next steps: Generate code with df | View recommended plots | New interactive sheet

4. Replace missing values in 'Math Score' with mean value

```
df['math_score'] = df['math_score'].fillna(df['math_score'].mean())
df['math_score']
```

	math_score
0	89.0
1	65.0
2	10.0
3	22.0
4	26.0
...	...
9995	78.0
9996	37.0
9997	70.0
9998	87.0
9999	31.0

10000 rows × 1 columns

dtype: float64

5. Replace missing values in 'Reading Score' with standard deviation value

```
df['reading_score'].fillna(df['reading_score'].std(), inplace=True)
df['reading_score']
```

```
<ipython-input-15-9cfea9a912e5>:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained ass
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col
```

```
df['reading_score'].fillna(df['reading_score'].std(), inplace=True)
```

	reading_score
0	38.0
1	100.0
2	99.0
3	51.0
4	58.0
...	...
9995	60.0
9996	92.0
9997	71.0
9998	60.0
9999	89.0

10000 rows × 1 columns

dtype: float64

6. Replace missing values in 'Place' with common value "Nashik"

```
df['parental_level_of_education'].fillna("Nashik", inplace=True)
df['parental_level_of_education']
```