```python
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

```python
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. Count the number of null values in the complete dataset

```python
df.replace("", np.nan, inplace=True)  # Standardizing empty strings as NaN
total_null_values = df.isnull().sum().sum()
print("\nTotal number of null values in dataset:", total_null_values)
```

```
Total number of null values in dataset: 229
```

∨ 4. Drop rows with at least 1 null value

```python
df_no_null_rows = df.dropna()
print("\nDataset after dropping rows with at least one null value:\n", df_no_null_rows.head())
```

```
Dataset after dropping rows with at least one null value:
   roll_no gender race_ethnicity parental_level_of_education  lunch  \
0  std-01   male        group D                some college    1.0
1  std-02   male        group B                 high school    1.0
2  std-03   male        group C             master's degree    1.0
3  std-04   male        group D                some college    1.0
```

```
     4  std-05    male        group C              some college    0.0

        test_preparation_course math_score  reading_score  writing_score  \
     0                     1.0          89           38.0           85.0
     1                     0.0          65          100.0           67.0
     2                     0.0          10           99.0           97.0
     3                     1.0          22           51.0           41.0
     4                     1.0          26           58.0           64.0

        science_score  total_score grade
     0           26.0        238.0     C
     1           96.0        328.0     A
     2           58.0        264.0     B
     3           84.0        198.0     D
     4           65.0        213.0     C
```

✓ 5. Drop rows if all values in that row are missing

```python
df_no_full_null_rows = df.dropna(how='all')
print("\nDataset after dropping rows where all values were null:\n", df_no_full_null_rows.head())
```

```
⇥   Dataset after dropping rows where all values were null:
       roll_no gender race_ethnicity parental_level_of_education  lunch  \
     0  std-01   male        group D              some college    1.0
     1  std-02   male        group B               high school    1.0
     2  std-03   male        group C             master's degree   1.0
     3  std-04   male        group D              some college    1.0
     4  std-05   male        group C              some college    0.0

        test_preparation_course math_score  reading_score  writing_score  \
     0                     1.0          89           38.0           85.0
     1                     0.0          65          100.0           67.0
     2                     0.0          10           99.0           97.0
     3                     1.0          22           51.0           41.0
     4                     1.0          26           58.0           64.0

        science_score  total_score grade
     0           26.0        238.0     C
     1           96.0        328.0     A
     2           58.0        264.0     B
     3           84.0        198.0     D
     4           65.0        213.0     C
```

✓ 6. Dropping columns with at least 1 null value.

```python
df_no_null_columns = df.dropna(axis=1)
print("\nDataset after dropping columns with at least one null value:\n", df_no_null_columns.head())
```

```
⇥   Dataset after dropping columns with at least one null value:
      Empty DataFrame
     Columns: []
     Index: [0, 1, 2, 3, 4]
```

✓ 7. Dropping Rows with at least 1 null value in CSV file

```python
df_cleaned = df.dropna()
df_cleaned.to_csv('cleaned_dataset.csv', index=False)  # Saves the cleaned dataset
print("\nCleaned dataset saved as 'cleaned_dataset.csv'")
```
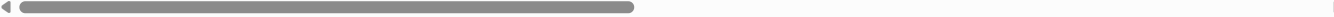
```
⇥   Cleaned dataset saved as 'cleaned_dataset.csv'
```

```python
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)