

```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
```

```
df = pd.read_csv("students_dataset.csv")
df
```

	Student ID	Age	Gender	Height	Weight	Blood Type	BMI	Temperature	Heart Rate	Blood Pressure	Cholesterol	Diabetes	Smoking
0	1.0	18.0	Female	161.777924	72.354947	O	27.645835	NaN	95.0	109.0	203.0	No	NaN
1	2.0	NaN	Male	152.069157	47.630941	B	NaN	98.714977	93.0	104.0	163.0	No	No
2	3.0	32.0	Female	182.537664	55.741083	A	16.729017	98.260293	76.0	130.0	216.0	Yes	No
3	NaN	30.0	Male	182.112867	63.332207	B	19.096042	98.839605	99.0	112.0	141.0	No	Yes
4	5.0	23.0	Female	NaN	46.234173	O	NaN	98.480008	95.0	NaN	231.0	No	No
...	...	...	...	...	...	...	...	...	...	...	...	...	...
29210	29211.0	33.0	Female	185.134162	47.351015	B	13.815168	99.008948	83.0	113.0	217.0	No	No
29211	29212.0	19.0	Female	186.542388	45.448234	B	13.060565	98.396925	91.0	101.0	150.0	NaN	No
29212	29213.0	22.0	Female	152.379104	47.568426	B	20.486509	NaN	68.0	NaN	216.0	No	No
29213	29214.0	20.0	Female	196.774227	77.415408	A	19.993599	99.873319	81.0	123.0	129.0	Yes	No
29214	29215.0	32.0	Female	172.979786	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Next steps:

[Generate code with df](#)[View recommended plots](#)[New interactive sheet](#)

## ✓ 1) Check if there are any missing values in the dataset as a whole

```
print("\nAny missing values in the entire dataframe:", df.isnull())
```

```
Any missing values in the entire dataframe:
0      False  False  False  False  False  False  False  False
1      False   True  False  False  False  False  False  True
2      False  False  False  False  False  False  False  False
3       True  False  False  False  False  False  False  False
4      False  False  False   True  False  False  False  True
...      ...   ...   ...   ...   ...   ...   ...   ...
29210   False  False  False  False  False  False  False  False
29211   False  False  False  False  False  False  False  False
29212   False  False  False  False  False  False  False  False
29213   False  False  False  False  False  False  False  False
29214   False  False  False  False  False  True   True   True

      Temperature  Heart Rate  Blood Pressure  Cholesterol  Diabetes  Smoking
0              True        False          False          False        False        True
1              False        False          False          False        False        False
2              False        False          False          False        False        False
3              False        False          False          False        False        False
4              False        False          True           False        False        False
...            ...         ...           ...           ...         ...         ...
29210          False        False          False          False        False        False
29211          False        False          False          False        True        False
29212           True        False          True           False        False        False
29213          False        False          False          False        False        False
29214           True        True           True           True        True        True

[29215 rows x 13 columns]
```

## ✓ 2) Check for missing values across each column

```
print("\nMissing values per column:\n", df.isnull())
```

```
Missing values per column:
Student ID  Age  Gender  Height  Weight  Blood Type  BMI  \
0      False  False  False  False  False  False  False
1      False   True  False  False  False  False  False
2      False  False  False  False  False  False  False
3       True  False  False  False  False  False  False
4      False  False  False   True  False  False  False
...      ...   ...   ...   ...   ...   ...   ...
29210   False  False  False  False  False  False  False
```

```

29211      False False False False False False False
29212      False False False False False False False
29213      False False False False False False False
29214      False False False False True  True  False

      Temperature Heart Rate Blood Pressure Cholesterol Diabetes \
0          True      False      False      False      False      False
1          False     False     False     False     False     False
2          False     False     False     False     False     False
3          False     False     False     False     False     False
4          False     False     True      False     False     False
...      ...      ...      ...      ...      ...      ...
29210      False     False     False     False     False     False
29211      False     False     False     False     False     True
29212      True      False     True      False     False     False
29213      False     False     False     False     False     False
29214      True      True      True      True      True      True

```

```

      Smoking missing_values_per_row
0          True      False
1          False     False
2          False     False
3          False     False
4          False     False
...      ...      ...
29210      False     False
29211      False     False
29212      False     False
29213      False     False
29214      True      False

```

[29215 rows x 14 columns]

### 3) Count missing values per column

```
print("\nCount of missing values per column:\n", df.isnull().sum())
```



```

Count of missing values per column:
Student ID      2902
Age             2847
Gender          2904
Height         2934
Weight         2955
Blood Type     2875
BMI            2924
Temperature    2927
Heart Rate     2807
Blood Pressure 2864
Cholesterol    2909
Diabetes       2982
Smoking        2953
dtype: int64

```

### 4) Count row-wise missing values

```
df['missing_values_per_row'] = df.isnull().sum(axis=1)
print("\nRow-wise missing values:\n", df[['Student ID', 'missing_values_per_row']].head())
```



```

Row-wise missing values:
Student ID missing_values_per_row
0          1.0                    2
1          2.0                    2
2          3.0                    0
3          NaN                    1
4          5.0                    3

```

### 5) Count missing values in the 'Gender' column

```
print("\nMissing values in 'Gender' column:", df['Gender'].isnull().sum())
```



```
Missing values in 'Gender' column: 2904
```

### 6) Group by 'Gender' and count missing values in 'Cholesterol'

```
missing_values_grouped = df.groupby('Gender')['Cholesterol'].apply(lambda x: x.isnull().sum())
print("\nMissing values count grouped by Gender:\n", missing_values_grouped)
```



```
Missing values count grouped by Gender:
Gender
Female    1302
Male      1312
Name: Cholesterol, dtype: int64
```

## 7) Replace missing values in 'BMI' with the column's average value

```
df['BMI'] = df['BMI'].fillna(df['BMI'].mean())
print("Average value of BMI :- \n",df['BMI'].mean())
print("\nMissing values in 'BMI' column replaced with mean value \n",df['BMI'])
```



```
Average value of BMI :-
23.40211153031812
```

```
Missing values in 'BMI' column replaced with mean value
0      27.645835
1      23.402112
2      16.729017
3      19.096042
4      23.402112
...
29210   13.815168
29211   13.060565
29212   20.486509
29213   19.993599
29214   23.402112
Name: BMI, Length: 29215, dtype: float64
```

Start coding or [generate](#) with AI.