

In [1]: `import pandas as pd`

In [2]: `import numpy as np`

In [7]: `df = pd.read_csv('/home/ubuntu/Desktop/dataset/StudentsPerformance.csv')`

In [8]: `df`

Out[8]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75
...
995	female	group E	master's degree	standard	completed	88	99	95
996	male	group C	high school	free/reduced	none	62	55	55
997	female	group C	high school	free/reduced	completed	59	71	65
998	female	group D	some college	standard	completed	68	78	77
999	female	group D	some college	free/reduced	none	77	86	86

1000 rows × 8 columns

In [9]: `df.isnull`

Out[9]:

```
<bound method DataFrame.isnull of      gender race/ethnicity parental level of education      lunch \
0    female      group B    bachelor's degree      standard
1    female      group C      some college      standard
2    female      group B    master's degree      standard
3     male      group A    associate's degree  free/reduced
4     male      group C      some college      standard
..     ...      ...      ...      ...
995  female      group E    master's degree      standard
996   male      group C      high school  free/reduced
997  female      group C      high school  free/reduced
998  female      group D      some college      standard
999  female      group D      some college  free/reduced

      test preparation course  math score  reading score  writing score
0                none           72           72           74
1             completed           69           90           88
2                none           90           95           93
3                none           47           57           44
4                none           76           78           75
..                ...           ...           ...           ...
995             completed           88           99           95
996                none           62           55           55
997             completed           59           71           65
998             completed           68           78           77
999                none           77           86           86

[1000 rows x 8 columns]>
```

In [14]: `series = pd.isnull(df["math score"])`

In [15]: `df[series]`

Out[15]:

gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
--------	----------------	-----------------------------	-------	-------------------------	------------	---------------	---------------

In [16]: `df.notnull()`

Out[16]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	True	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True	True
2	True	True	True	True	True	True	True	True
3	True	True	True	True	True	True	True	True
4	True	True	True	True	True	True	True	True
...
995	True	True	True	True	True	True	True	True
996	True	True	True	True	True	True	True	True
997	True	True	True	True	True	True	True	True
998	True	True	True	True	True	True	True	True
999	True	True	True	True	True	True	True	True

1000 rows × 8 columns

In [18]: `from sklearn import preprocessing`

In [19]: `le = preprocessing.LabelEncoder()`

In [20]: `df['gender'] = le.fit_transform(df['gender'])`

In [21]: `newdf = df`

In [22]: `df`

Out[22]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	0	group B	bachelor's degree	standard	none	72	72	74
1	0	group C	some college	standard	completed	69	90	88
2	0	group B	master's degree	standard	none	90	95	93
3	1	group A	associate's degree	free/reduced	none	47	57	44
4	1	group C	some college	standard	none	76	78	75
...
995	0	group E	master's degree	standard	completed	88	99	95
996	1	group C	high school	free/reduced	none	62	55	55
997	0	group C	high school	free/reduced	completed	59	71	65
998	0	group D	some college	standard	completed	68	78	77
999	0	group D	some college	free/reduced	none	77	86	86

1000 rows × 8 columns

In []: