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

In [2]:
```
df=pd.read_csv('F://SWATI ENGG//2021-2022//DS and Big Data//PRACTICALS//student performance.cs
```

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

Out[3]:

	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
5	female	group B	associate's degree	standard	none	71	83	78
6	female	group B	some college	standard	completed	88	95	92
7	male	group B	some college	free/reduced	none	40	43	39
8	male	group D	high school	free/reduced	completed	64	64	67
9	female	group B	high school	free/reduced	none	38	60	50
10	male	group C	associate's degree	standard	none	58	54	52
11	male	group D	associate's degree	standard	none	40	52	43
12	female	group B	high school	standard	none	65	81	73
13	male	group A	some college	standard	completed	78	72	70
14	female	group A	master's degree	standard	none	50	53	58
15	female	group C	some high school	standard	none	69	75	78
16	male	group C	high school	standard	none	88	89	86
17	female	group B	some high school	free/reduced	none	18	32	28
18	male	group C	master's degree	free/reduced	completed	46	42	46
19	female	group C	associate's degree	free/reduced	none	54	58	61
20	male	group D	high school	standard	none	66	69	63
21	female	group B	some college	free/reduced	completed	65	75	70
22	male	group D	some college	standard	none	44	54	53
23	female	group C	some high school	standard	none	69	73	73
24	male	group D	bachelor's degree	free/reduced	completed	74	71	80
25	male	group A	master's degree	free/reduced	none	73	74	72
26	male	group B	some college	standard	none	69	54	55
27	female	group C	bachelor's degree	standard	none	67	69	75
28	male	group C	high school	standard	none	70	70	65
29	female	group D	master's degree	standard	none	62	70	75
...	...	...	...	...	...	...	...	...
970	female	group D	bachelor's degree	standard	none	89	100	100
971	male	group C	some high school	standard	completed	78	72	69
972	female	group A	high school	free/reduced	completed	53	50	60
973	female	group D	some college	free/reduced	none	49	65	61
974	female	group A	some college	standard	none	54	63	67
975	female	group C	some college	standard	completed	64	82	77

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
976	male	group B	some college	free/reduced	completed	60	62	60
977	male	group C	associate's degree	standard	none	62	65	58
978	male	group D	high school	standard	completed	55	41	48
979	female	group C	associate's degree	standard	none	91	95	94
980	female	group B	high school	free/reduced	none	8	24	23
981	male	group D	some high school	standard	none	81	78	78
982	male	group B	some high school	standard	completed	79	85	86
983	female	group A	some college	standard	completed	78	87	91
984	female	group C	some high school	standard	none	74	75	82
985	male	group A	high school	standard	none	57	51	54
986	female	group C	associate's degree	standard	none	40	59	51
987	male	group E	some high school	standard	completed	81	75	76
988	female	group A	some high school	free/reduced	none	44	45	45
989	female	group D	some college	free/reduced	completed	67	86	83
990	male	group E	high school	free/reduced	completed	86	81	75
991	female	group B	some high school	standard	completed	65	82	78
992	female	group D	associate's degree	free/reduced	none	55	76	76
993	female	group D	bachelor's degree	free/reduced	none	62	72	74
994	male	group A	high school	standard	none	63	63	62
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 [10]:
```
df.isnull()
```

Out[10]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False	False
6	False	False	False	False	False	False	False	False
7	False	False	False	False	False	False	False	False
8	False	False	False	False	False	False	False	False
9	False	False	False	False	False	False	False	False
10	False	False	False	False	False	False	False	False
11	False	False	False	False	False	False	False	False
12	False	False	False	False	False	False	False	False
13	False	False	False	False	False	False	False	False
14	False	False	False	False	False	False	False	False
15	False	False	False	False	False	False	False	False
16	False	False	False	False	False	False	False	False
17	False	False	False	False	False	False	False	False
18	False	False	False	False	False	False	False	False
19	False	False	False	False	False	False	False	False
20	False	False	False	False	False	False	False	False
21	False	False	False	False	False	False	False	False
22	False	False	False	False	False	False	False	False
23	False	False	False	False	False	False	False	False
24	False	False	False	False	False	False	False	False
25	False	False	False	False	False	False	False	False
26	False	False	False	False	False	False	False	False
27	False	False	False	False	False	False	False	False
28	False	False	False	False	False	False	False	False
29	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...
970	False	False	False	False	False	False	False	False
971	False	False	False	False	False	False	False	False
972	False	False	False	False	False	False	False	False
973	False	False	False	False	False	False	False	False
974	False	False	False	False	False	False	False	False
975	False	False	False	False	False	False	False	False
976	False	False	False	False	False	False	False	False
977	False	False	False	False	False	False	False	False

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
978	False	False	False	False	False	False	False	False
979	False	False	False	False	False	False	False	False
980	False	False	False	False	False	False	False	False
981	False	False	False	False	False	False	False	False
982	False	False	False	False	False	False	False	False
983	False	False	False	False	False	False	False	False
984	False	False	False	False	False	False	False	False
985	False	False	False	False	False	False	False	False
986	False	False	False	False	False	False	False	False
987	False	False	False	False	False	False	False	False
988	False	False	False	False	False	False	False	False
989	False	False	False	False	False	False	False	False
990	False	False	False	False	False	False	False	False
991	False	False	False	False	False	False	False	False
992	False	False	False	False	False	False	False	False
993	False	False	False	False	False	False	False	False
994	False	False	False	False	False	False	False	False
995	False	False	False	False	False	False	False	False
996	False	False	False	False	False	False	False	False
997	False	False	False	False	False	False	False	False
998	False	False	False	False	False	False	False	False
999	False	False	False	False	False	False	False	False

1000 rows × 8 columns

In [6]:
```
df.isnull().sum()
```

Out[6]:
```
gender                         0
race/ethnicity                 0
parental level of education    0
lunch                          0
test preparation course        0
math score                     0
reading score                  0
writing score                  0
dtype: int64
```

In [11]:
```
df.columns
```

Out[11]:
```
Index(['gender', 'race/ethnicity', 'parental level of education', 'lunch',
       'test preparation course', 'math score', 'reading score',
       'writing score'],
      dtype='object')
```

In [8]:
```
df.isnull().any()
```

Out[8]:
```
gender                         False
race/ethnicity                 False
parental level of education    False
lunch                          False
test preparation course        False
math score                     False
reading score                  False
writing score                  False
dtype: bool
```

In [15]:
```
pd.isnull(df["math score"]) .sum()
```

Out[15]:
```
0
```

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
5	True	True	True	True	True	True	True	True
6	True	True	True	True	True	True	True	True
7	True	True	True	True	True	True	True	True
8	True	True	True	True	True	True	True	True
9	True	True	True	True	True	True	True	True
10	True	True	True	True	True	True	True	True
11	True	True	True	True	True	True	True	True
12	True	True	True	True	True	True	True	True
13	True	True	True	True	True	True	True	True
14	True	True	True	True	True	True	True	True
15	True	True	True	True	True	True	True	True
16	True	True	True	True	True	True	True	True
17	True	True	True	True	True	True	True	True
18	True	True	True	True	True	True	True	True
19	True	True	True	True	True	True	True	True
20	True	True	True	True	True	True	True	True
21	True	True	True	True	True	True	True	True
22	True	True	True	True	True	True	True	True
23	True	True	True	True	True	True	True	True
24	True	True	True	True	True	True	True	True
25	True	True	True	True	True	True	True	True
26	True	True	True	True	True	True	True	True
27	True	True	True	True	True	True	True	True
28	True	True	True	True	True	True	True	True
29	True	True	True	True	True	True	True	True
...	...	...	...	...	...	...	...	...
970	True	True	True	True	True	True	True	True
971	True	True	True	True	True	True	True	True
972	True	True	True	True	True	True	True	True
973	True	True	True	True	True	True	True	True
974	True	True	True	True	True	True	True	True
975	True	True	True	True	True	True	True	True
976	True	True	True	True	True	True	True	True
977	True	True	True	True	True	True	True	True

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
978	True	True	True	True	True	True	True	True
979	True	True	True	True	True	True	True	True
980	True	True	True	True	True	True	True	True
981	True	True	True	True	True	True	True	True
982	True	True	True	True	True	True	True	True
983	True	True	True	True	True	True	True	True
984	True	True	True	True	True	True	True	True
985	True	True	True	True	True	True	True	True
986	True	True	True	True	True	True	True	True
987	True	True	True	True	True	True	True	True
988	True	True	True	True	True	True	True	True
989	True	True	True	True	True	True	True	True
990	True	True	True	True	True	True	True	True
991	True	True	True	True	True	True	True	True
992	True	True	True	True	True	True	True	True
993	True	True	True	True	True	True	True	True
994	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 [17]:
```
df.notnull().sum()
```

Out[17]:
```
gender                         1000
race/ethnicity                 1000
parental level of education    1000
lunch                          1000
test preparation course        1000
math score                     1000
reading score                  1000
writing score                  1000
dtype: int64
```

In [20]:
```
df.head()
df1=df
```

In [21]:
```
df1
```

Out[21]:

	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
5	female	group B	associate's degree	standard	none	71	83	78
6	female	group B	some college	standard	completed	88	95	92
7	male	group B	some college	free/reduced	none	40	43	39
8	male	group D	high school	free/reduced	completed	64	64	67
9	female	group B	high school	free/reduced	none	38	60	50
10	male	group C	associate's degree	standard	none	58	54	52
11	male	group D	associate's degree	standard	none	40	52	43
12	female	group B	high school	standard	none	65	81	73
13	male	group A	some college	standard	completed	78	72	70
14	female	group A	master's degree	standard	none	50	53	58
15	female	group C	some high school	standard	none	69	75	78
16	male	group C	high school	standard	none	88	89	86
17	female	group B	some high school	free/reduced	none	18	32	28
18	male	group C	master's degree	free/reduced	completed	46	42	46
19	female	group C	associate's degree	free/reduced	none	54	58	61
20	male	group D	high school	standard	none	66	69	63
21	female	group B	some college	free/reduced	completed	65	75	70
22	male	group D	some college	standard	none	44	54	53
23	female	group C	some high school	standard	none	69	73	73
24	male	group D	bachelor's degree	free/reduced	completed	74	71	80
25	male	group A	master's degree	free/reduced	none	73	74	72
26	male	group B	some college	standard	none	69	54	55
27	female	group C	bachelor's degree	standard	none	67	69	75
28	male	group C	high school	standard	none	70	70	65
29	female	group D	master's degree	standard	none	62	70	75
...	...	...	...	...	...	...	...	...
970	female	group D	bachelor's degree	standard	none	89	100	100
971	male	group C	some high school	standard	completed	78	72	69
972	female	group A	high school	free/reduced	completed	53	50	60
973	female	group D	some college	free/reduced	none	49	65	61
974	female	group A	some college	standard	none	54	63	67
975	female	group C	some college	standard	completed	64	82	77

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
976	male	group B	some college	free/reduced	completed	60	62	60
977	male	group C	associate's degree	standard	none	62	65	58
978	male	group D	high school	standard	completed	55	41	48
979	female	group C	associate's degree	standard	none	91	95	94
980	female	group B	high school	free/reduced	none	8	24	23
981	male	group D	some high school	standard	none	81	78	78
982	male	group B	some high school	standard	completed	79	85	86
983	female	group A	some college	standard	completed	78	87	91
984	female	group C	some high school	standard	none	74	75	82
985	male	group A	high school	standard	none	57	51	54
986	female	group C	associate's degree	standard	none	40	59	51
987	male	group E	some high school	standard	completed	81	75	76
988	female	group A	some high school	free/reduced	none	44	45	45
989	female	group D	some college	free/reduced	completed	67	86	83
990	male	group E	high school	free/reduced	completed	86	81	75
991	female	group B	some high school	standard	completed	65	82	78
992	female	group D	associate's degree	free/reduced	none	55	76	76
993	female	group D	bachelor's degree	free/reduced	none	62	72	74
994	male	group A	high school	standard	none	63	63	62
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 [19]:
```
from sklearn import preprocessing
le=preprocessing.LabelEncoder()
```

In [22]:
```
df['gender']=le.fit_transform(df['gender'])
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
5	0	group B	associate's degree	standard	none	71	83	78
6	0	group B	some college	standard	completed	88	95	92
7	1	group B	some college	free/reduced	none	40	43	39
8	1	group D	high school	free/reduced	completed	64	64	67
9	0	group B	high school	free/reduced	none	38	60	50
10	1	group C	associate's degree	standard	none	58	54	52
11	1	group D	associate's degree	standard	none	40	52	43
12	0	group B	high school	standard	none	65	81	73
13	1	group A	some college	standard	completed	78	72	70
14	0	group A	master's degree	standard	none	50	53	58
15	0	group C	some high school	standard	none	69	75	78
16	1	group C	high school	standard	none	88	89	86
17	0	group B	some high school	free/reduced	none	18	32	28
18	1	group C	master's degree	free/reduced	completed	46	42	46
19	0	group C	associate's degree	free/reduced	none	54	58	61
20	1	group D	high school	standard	none	66	69	63
21	0	group B	some college	free/reduced	completed	65	75	70
22	1	group D	some college	standard	none	44	54	53
23	0	group C	some high school	standard	none	69	73	73
24	1	group D	bachelor's degree	free/reduced	completed	74	71	80
25	1	group A	master's degree	free/reduced	none	73	74	72
26	1	group B	some college	standard	none	69	54	55
27	0	group C	bachelor's degree	standard	none	67	69	75
28	1	group C	high school	standard	none	70	70	65
29	0	group D	master's degree	standard	none	62	70	75
...	...	...	...	...	...	...	...	...
970	0	group D	bachelor's degree	standard	none	89	100	100
971	1	group C	some high school	standard	completed	78	72	69
972	0	group A	high school	free/reduced	completed	53	50	60
973	0	group D	some college	free/reduced	none	49	65	61
974	0	group A	some college	standard	none	54	63	67
975	0	group C	some college	standard	completed	64	82	77

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
976	1	group B	some college	free/reduced	completed	60	62	60
977	1	group C	associate's degree	standard	none	62	65	58
978	1	group D	high school	standard	completed	55	41	48
979	0	group C	associate's degree	standard	none	91	95	94
980	0	group B	high school	free/reduced	none	8	24	23
981	1	group D	some high school	standard	none	81	78	78
982	1	group B	some high school	standard	completed	79	85	86
983	0	group A	some college	standard	completed	78	87	91
984	0	group C	some high school	standard	none	74	75	82
985	1	group A	high school	standard	none	57	51	54
986	0	group C	associate's degree	standard	none	40	59	51
987	1	group E	some high school	standard	completed	81	75	76
988	0	group A	some high school	free/reduced	none	44	45	45
989	0	group D	some college	free/reduced	completed	67	86	83
990	1	group E	high school	free/reduced	completed	86	81	75
991	0	group B	some high school	standard	completed	65	82	78
992	0	group D	associate's degree	free/reduced	none	55	76	76
993	0	group D	bachelor's degree	free/reduced	none	62	72	74
994	1	group A	high school	standard	none	63	63	62
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 [24]:
```
ohe=preprocessing.OneHotEncoder()
```

In [26]:
```
df['race/ethnicity'].unique()
```

Out[26]:
```
array(['group B', 'group C', 'group A', 'group D', 'group E'],
      dtype=object)
```

In [ ]:
```
enc_df=(enc.fit_transform(df3[['Species']])).toarray()
```

In [31]:
```
racedf=(ohe.fit_transform(df[['race/ethnicity']])).toarray()
```

In [33]:
```
racedf
```

Out[33]:
```
array([[0., 1., 0., 0., 0.],
       [0., 0., 1., 0., 0.],
       [0., 1., 0., 0., 0.],
       ...,
       [0., 0., 1., 0., 0.],
       [0., 0., 0., 1., 0.],
       [0., 0., 0., 1., 0.]])
```

In [36]:
```
racedf=pd.DataFrame(racedf,columns=['racegroup B', 'racegroup C', 'racegroup A', 'racegroup D'
```

In [37]:
```
racedf.head()
```

Out[37]:

	racegroup B	racegroup C	racegroup A	racegroup D	racegroup E
0	0.0	1.0	0.0	0.0	0.0
1	0.0	0.0	1.0	0.0	0.0
2	0.0	1.0	0.0	0.0	0.0
3	1.0	0.0	0.0	0.0	0.0
4	0.0	0.0	1.0	0.0	0.0

In [38]:
```
df.head()
```

Out[38]:

	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

In [40]:
```
df=df.drop([columns=df['race/ethnicity']])
```

```
  File "<ipython-input-40-96683aab9c13>", line 1
    df=df.drop([columns=df['race/ethnicity']])
                        ^
SyntaxError: invalid syntax
```

In [ ]: