Part 1 Question 3: Analyze the exam data

Tasks:

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

1. Read the data from the CSV file into a DataFrame and display the first five rows.

In [ ]:
```
df = pd.read_csv('datasets/exams.csv')
df.head()
```

Out[ ]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writin scor
0	female	group B	bachelor's degree	standard	none	72	72	7
1	female	group C	some college	standard	completed	69	90	8
2	female	group B	master's degree	standard	none	90	95	9
3	male	group A	associate's degree	free/reduced	none	47	57	4
4	male	group C	some college	standard	none	76	78	7

2. Display the basic information for the DataFrame and its columns using the info() method.

In [ ]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 8 columns):
 #   Column                       Non-Null Count  Dtype 
---  ------                       --------------  ----- 
 0   gender                       1000 non-null   object
 1   race/ethnicity               1000 non-null   object
 2   parental level of education  1000 non-null   object
 3   lunch                        1000 non-null   object
 4   test preparation course      1000 non-null   object
 5   math score                   1000 non-null   int64 
 6   reading score                1000 non-null   int64 
 7   writing score                1000 non-null   int64 
dtypes: int64(3), object(5)
memory usage: 62.6+ KB
```

3. Display statistical information for the math score reading score and writing score columns using the describe() method.

In [ ]:
```
df[['math score', 'reading score', 'writing score']].describe()
```

Out[ ]:

	math score	reading score	writing score
count	1000.00000	1000.000000	1000.000000
mean	66.08900	69.169000	68.054000
std	15.16308	14.600192	15.195657
min	0.00000	17.000000	10.000000
25%	57.00000	59.000000	57.750000
50%	66.00000	70.000000	69.000000
75%	77.00000	79.000000	79.000000
max	100.00000	100.000000	100.000000

4. Group the data by the race/ethnicity column and display the mean scores.

In [ ]:
```
df[['race/ethnicity', 'math score', 'reading score', 'writing score']].group
```

Out[ ]:

race/ethnicity	math score	reading score	writing score
group A	61.629213	64.674157	62.674157
group B	63.452632	67.352632	65.600000
group C	64.463950	69.103448	67.827586
group D	67.362595	70.030534	70.145038
group E	73.821429	73.028571	71.407143

5. Display a single column as a DataFrame with bracket notation.

In [ ]:
```
df['gender'].to_frame(name='gender')
```

Out[ ]:

	gender
0	female
1	female
2	female
3	male
4	male
...	...
995	female
996	male
997	female
998	female
999	female

1000 rows × 1 columns

6. Display a single column as a Series with bracket notation.

In [ ]:
```
df['gender']
```

Out[ ]:
```
0      female
1      female
2      female
3        male
4        male
       ...
995    female
996      male
997    female
998    female
999    female
Name: gender, Length: 1000, dtype: object
```

7. Display a single column as a Series with dot notation.

In [ ]:
```
df.gender
```

Out[ ]:
```
0      female
1      female
2      female
3        male
4        male
       ...
995    female
996      male
997    female
998    female
999    female
Name: gender, Length: 1000, dtype: object
```

8. Display only rows for females with a math score greater than or equal to 90.

In [ ]:
```
df[(df['gender'] == 'female')&(df['math score'] > 90)]
```

Out[ ]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	wri s
114	female	group E	bachelor's degree	standard	completed	99	100	
165	female	group C	bachelor's degree	standard	completed	96	100	
179	female	group D	some high school	standard	completed	97	100	
263	female	group E	high school	standard	none	99	93	
451	female	group E	some college	standard	none	100	92	
458	female	group E	bachelor's degree	standard	none	100	100	
501	female	group B	associate's degree	standard	completed	94	87	
503	female	group E	associate's degree	standard	completed	95	89	
521	female	group C	associate's degree	standard	none	91	86	
546	female	group A	some high school	standard	completed	92	100	
566	female	group E	bachelor's degree	free/reduced	completed	92	100	
594	female	group C	bachelor's degree	standard	completed	92	100	
685	female	group E	master's degree	standard	completed	94	99	
712	female	group D	some college	standard	none	98	100	
717	female	group C	associate's degree	standard	completed	96	96	
855	female	group B	bachelor's degree	standard	none	97	97	
886	female	group E	associate's degree	standard	completed	93	100	
903	female	group D	bachelor's degree	free/reduced	completed	93	100	
957	female	group D	master's degree	standard	none	92	100	
962	female	group E	associate's	standard	none	100	100	

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	wri s
			degree					
979	female	group C	associate's degree	standard	none	91	95	

Questions:

1. Does taking a test preparation course improve average scores?

In [ ]:
```
df[['test preparation course', 'math score', 'reading score', 'writing score
```

Out[ ]:

test preparation course	math score	reading score	writing score
completed	69.695531	73.893855	74.418994
none	64.077882	66.534268	64.504673

It can be seen in the table above that the average score for all 3 categories was higher when the test preparation course was taken, so we can say that taking the test improves average scores.

2. Which gender is better on average at math?

In [ ]:
```
df[['gender', 'math score']].groupby('gender').mean()
```

Out[ ]:

gender	math score
female	63.633205
male	68.728216

From the table above we can see that males on average are better at math than females

3. Which gender is better on average at all three subjects? Hint: Start by adding a column to the DataFrame with the total score

In [ ]:
```
df['average score'] = df[['math score', 'reading score', 'writing score']].s
df[['gender', 'average score']].groupby('gender').mean()
```

Out[ ]:

gender	average score
female	69.569498
male	65.837483

From the table above we can see that on average in all 3 subjects, females are better than males.