

✓ Analysis of Student Exam Performance

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv("StudentsPerformance.csv")
df.head()
```

 Show hidden output

```
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
```

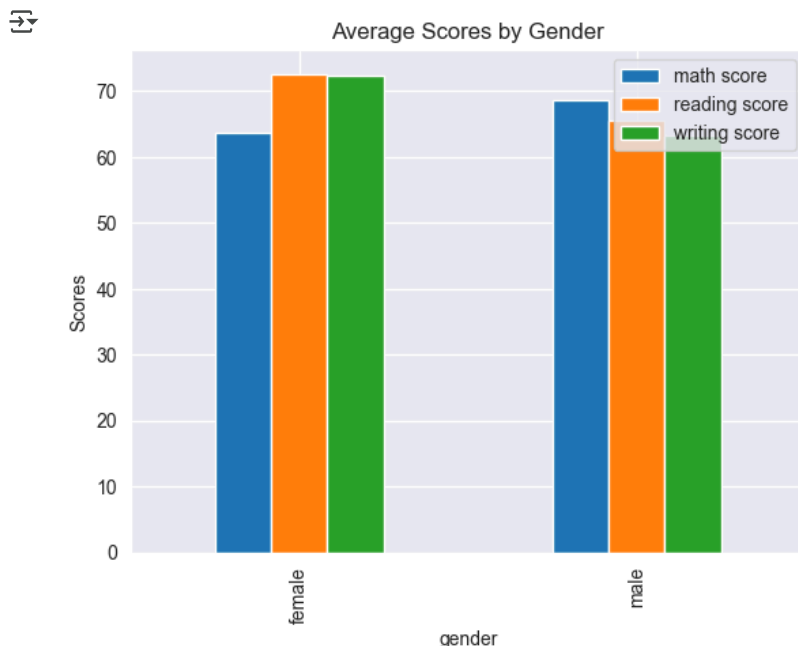
```
df.isnull().sum()
```

```
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
```

✓ Question 1:

How do male and female students differ in their average scores for math, reading, and writing?

```
df.groupby('gender')[['math score', 'reading score', 'writing score']].mean().plot(kind='bar')
plt.title("Average Scores by Gender")
plt.ylabel("Scores")
plt.show()
```



Question 2:

What is the impact of completing a test preparation course on student performance in all three subjects?

```
df.groupby('test preparation course')[['math score', 'reading score', 'writing score']].mean().plot(kind='bar')  
plt.title("Test Prep Course vs Performance")  
plt.ylabel("Average Score")  
plt.show()
```



Question 3:

Does the type of lunch (standard vs. free/reduced) have any effect on student scores?

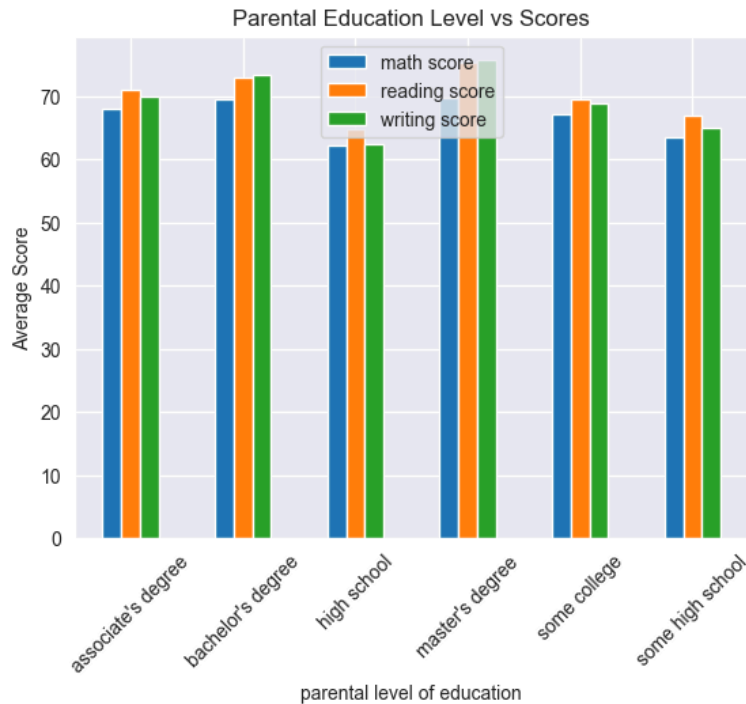
```
df.groupby('lunch')[['math score', 'reading score', 'writing score']].mean().plot(kind='bar')  
plt.title("Lunch Type vs Scores")  
plt.ylabel("Average Score")  
plt.show()
```



Question 4:

How does the parental level of education affect students' performance in exams?

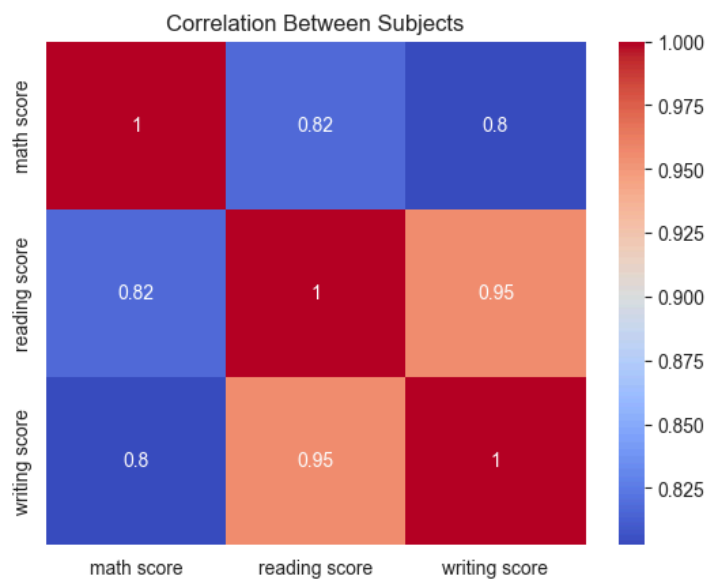
```
df.groupby('parental level of education')[['math score', 'reading score', 'writing score']].mean().plot(kind='bar')
plt.title("Parental Education Level vs Scores")
plt.xticks(rotation=45)
plt.ylabel("Average Score")
plt.show()
```



Question 5:

What is the correlation between math, reading, and writing scores?

```
sns.heatmap(df[['math score', 'reading score', 'writing score']].corr(), annot=True, cmap="coolwarm")
plt.title("Correlation Between Subjects")
plt.show()
```



Final Insights:

1. Gender:

- Female students generally perform better in reading and writing.
- Male students perform slightly better in math.

2. Test Preparation:

- Students who completed the test preparation course scored higher in all three subjects.

3. Lunch Type:

- Students with standard lunch performed better than those with free/reduced lunch.

4. Parental Education Level:

- Higher parental education levels are positively associated with higher student performance.

5. Subject Correlation:

- There is a strong positive correlation between math, reading, and writing scores, meaning students who perform well in one subject tend to do well in others too.