Analysis of Student Performance in Math, Reading, and Writing

preprocessing data

1 Dataset information

	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

```
data.shape
```

```
(1000, 8)
```

The dataset about student perfromace that contain 1000 rows with 8 col

preprocessing data

2 Check null

```
data.isnull().sum()
```

	0
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

There are no null in datasets , can not check redundant because the have col that redundemt data such as level of education

• Analysis Distribution of Categories

```
gender_count=data['gender'].value_counts()
```

```
gender_count
```

gender	count
female	518
male	482

dtype: int64

group C	319
group D	262
group B	190
group E	140
group A	89

dtype: int64

```
count_level=data['parental level of e
```

```
count_level
```

parental level of education	count
some college	226
associate's degree	222
high school	196
some high school	179
bachelor's degree	118
master's degree	59

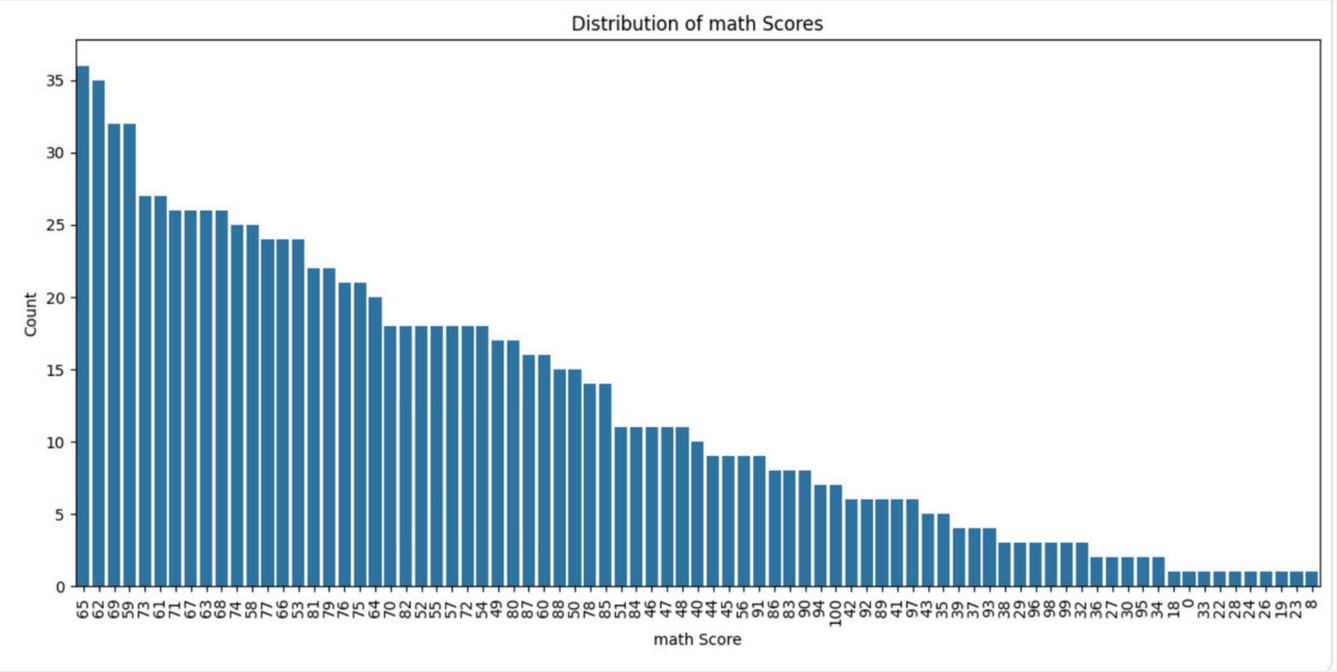
Analysis Distribution of Categories

- **Gender Distribution: 518 females vs 482 males (balanced dataset).**
- **Parental Education: Most parents have Some College (226) or Associate's Degree (222), least have Master's Degree (59).**
- **Groups: Group C is the largest (319), while Group A is the smallest (89).**

Recommendations

Ensure balanced programs for both genders.
Provide support for students from mid-education families.
Promote awareness of higher education opportunities.
Review group distribution to ensure fair resources.

Analysis math score




```
Mean math score: 66.089
Mean reading score: 69.169
```

```
very_good = data[(data['math score'] >= 80) & (data['math score'] < 90)].groupby('race/ethnicit
print("Number of students with math score very good")
print(very_good)
```

```
Number of students with math score very good
  race/ethnicity  count
0        group A      5
1        group B     21
2        group C     33
3        group D     43
4        group E     33
```

```
exllent = data[(data['math score'] >= 90) ].groupby('race/ethnicity').size().reset_index(name='
print("Number of students with math score exllent")
print(exllent)
```

```
Number of students with math score exllent
  race/ethnicity  count
0        group A      4
1        group B      8
2        group C     16
3        group D      8
4        group E     22
```

```
pass_s = data[(data['math score'] >= 60) & (data['math score'] < 70)].groupby
print("Number of students with math score that pass")
print(pass_s)
```

```
Number of students with math score that pass
  race/ethnicity  count
0        group A     22
1        group B     54
2        group C     95
3        group D     72
4        group E     25
```

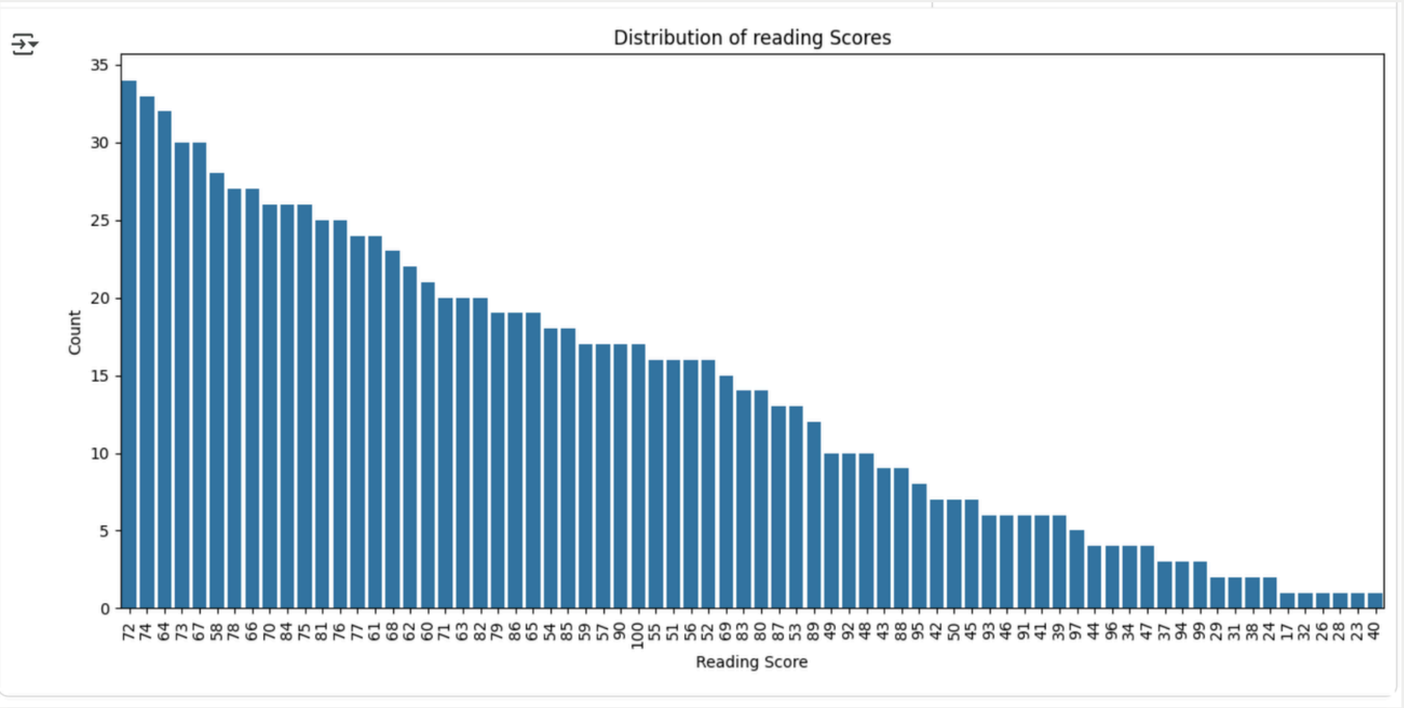
```
good = data[(data['math score'] >= 70) & (data['math score'] < 80)].groupby(
print("Number of students with math score good")
print(good)
```

```
Number of students with math score good
  race/ethnicity  count
0        group A     16
1        group B     36
2        group C     62
3        group D     67
4        group E     35
```

Analysis Math score

- Most students scored between 60–79 (pass and good levels).
- Few students achieved excellent (≥90) scores, with the highest in Group D (18) and Group E (22).
- Group C and Group D have the largest number of students across most score ranges.
- the mean of math greater than degree of pass (60)
- Focus is needed to help students move from pass to higher performance levels.

Analysis reading score



[ ]

```
pass_s_r = data[(data['reading score'] >= 60) & (data['readin
print("Number of students with reading score that pass")
print(pass_s_r)
```

```
Number of students with reading score that pass
  race/ethnicity  count
0        group A     21
1        group B     50
2        group C     81
3        group D     57
4        group E     24
```

[ ]

```
good_r= data[(data['reading score'] >= 70) & (data['reading s
print("Number of students with reading score good")
print(good_r)
```

```
Number of students with reading score good
  race/ethnicity  count
0        group A     15
1        group B     37
2        group C     90
3        group D     80
4        group E     42
```

ommands | + Code | + Text | ▷ Run all

[ ]

```
very_good = data[(data['math score'] >= 80) & (data['math score'] < 90)].groupb
print("Number of students with math score very good")
print(very_good)
```

```
Number of students with math score very good
  race/ethnicity  count
0        group A      5
1        group B     21
2        group C     33
3        group D     43
4        group E     33
```

[ ]

```
exllent = data[(data['math score'] >= 90) ].groupby('race/ethnicity').size().re
print("Number of students with math score exllent")
print(exllent)
```

```
Number of students with math score exllent
  race/ethnicity  count
0        group A      4
1        group B      8
2        group C     16
3        group D      8
4        group E     22
```

Analysis reading score

- **Most students score 60–80.**
- **Best: Group C.**
- **Lowest: Group A.**

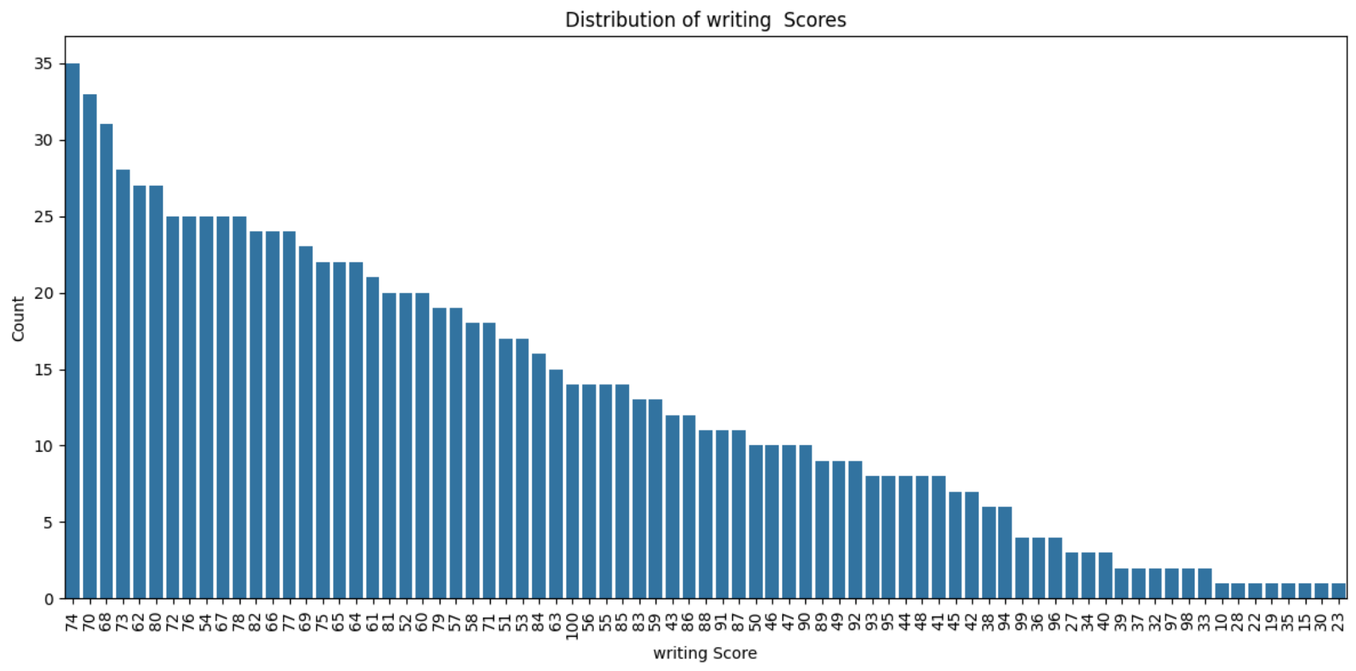
Recommendations

Targeted Support for Group A

Enhance Reading Programs Across Groups

Leverage Group C’s Best Practices
Identify teaching methods or study habits common in Group C and replicate them across other groups

Analysis writing score




```
pass_s_w = data[(data['writing score'] >= 60) & (data['writing score'] < 70)]
print("Number of students with writing score that pass")
print(pass_s_w)
```

```
Number of students with writing score that pass
  race/ethnicity  count
0        group A     24
1        group B     48
2        group C     73
3        group D     60
4        group E     25
```

```
good_w= data[(data['writing score'] >= 70) & (data['writing score'] < 80)]
print("Number of students with writng score good")
print(good_w)
```

```
Number of students with writng score good
  race/ethnicity  count
0        group A     13
1        group B     43
2        group C     77
3        group D     76
4        group E     45
```

```
print("Number of students with writing score very good")
print(very_good_w)
```

```
Number of students with writing score very good
  race/ethnicity  count
0        group A      9
1        group B     26
2        group C     53
3        group D     46
4        group E     22
```

```
exllent_w = data[(data['writing score'] >= 90) ].groupby('race/ethnicity').size().res
print("Number of students with writing score exllent")
print(exllent_w)
```

```
Number of students with writing score exllent
  race/ethnicity  count
0        group A      5
1        group B      9
2        group C     26
3        group D     20
4        group E     18
```

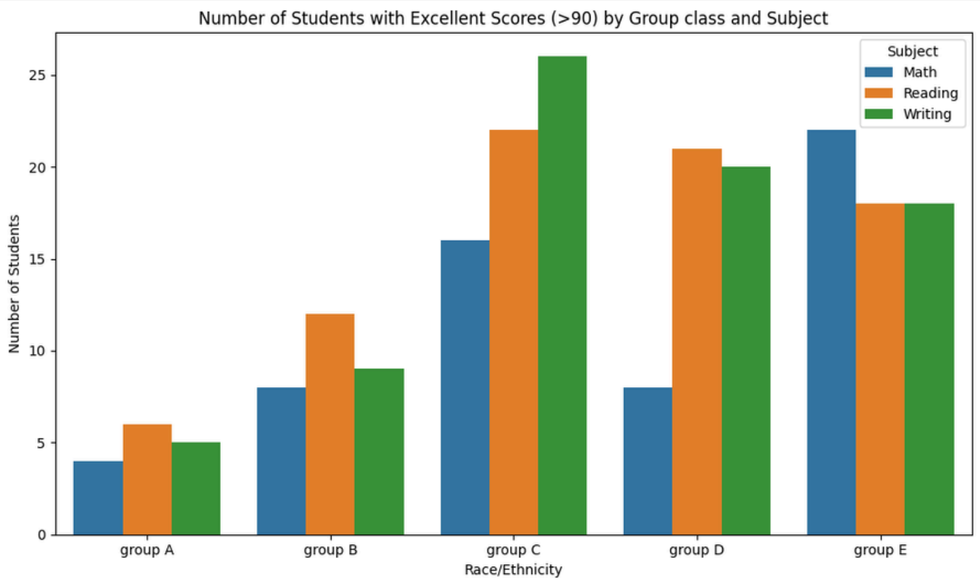
Analysis reading score

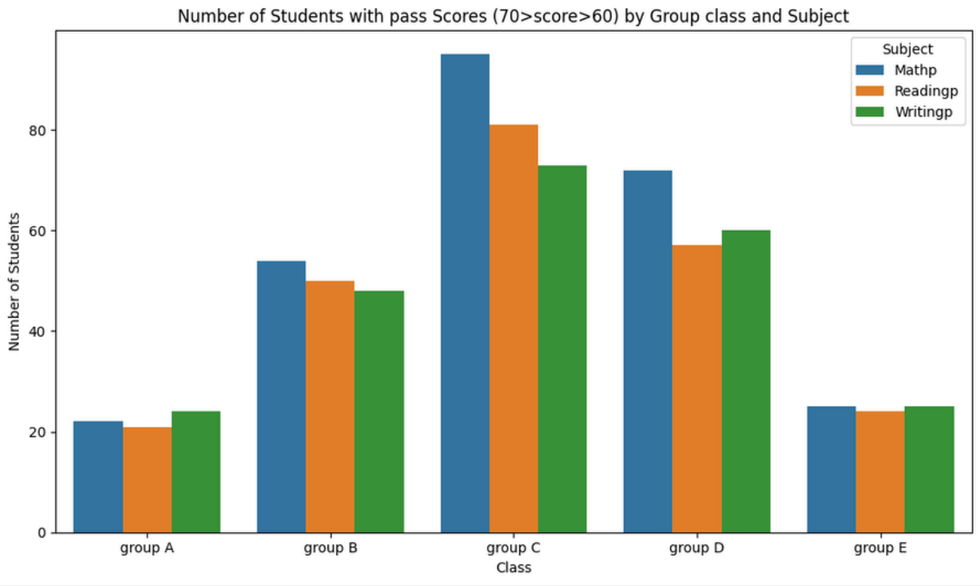
- The majority of students scored between 60 and 80, while fewer achieved very low or very high scores.
- Pass range (60–69): Group D had the largest number of students (60), whereas Group A had the fewest (24).
- Good range (70–79): Group D again performed best with 76 students also C 77 students, while Group A had only 13.
- Very good range (80–89): Group C led with 53 students, in contrast to Group A with just 9.
- Excellent (≥ 90): Group C showed the highest performance (26 students) next D with 20 students , while Group A remained the lowest with only 5.

Recommendations

- **Focus on Group A: weakest performance across all ranges → needs targeted support.**
- **Leverage Group D & C: they dominate in all performance levels, so their study methods/teaching approaches could be replicated.**
- **Encourage Higher-Level Writing: many students reach “good” levels, but fewer move to “excellent”; schools should focus on advanced writing skills.**

Number of Students Achieving Excellent and Passing Scores in Math, Reading, and Writing



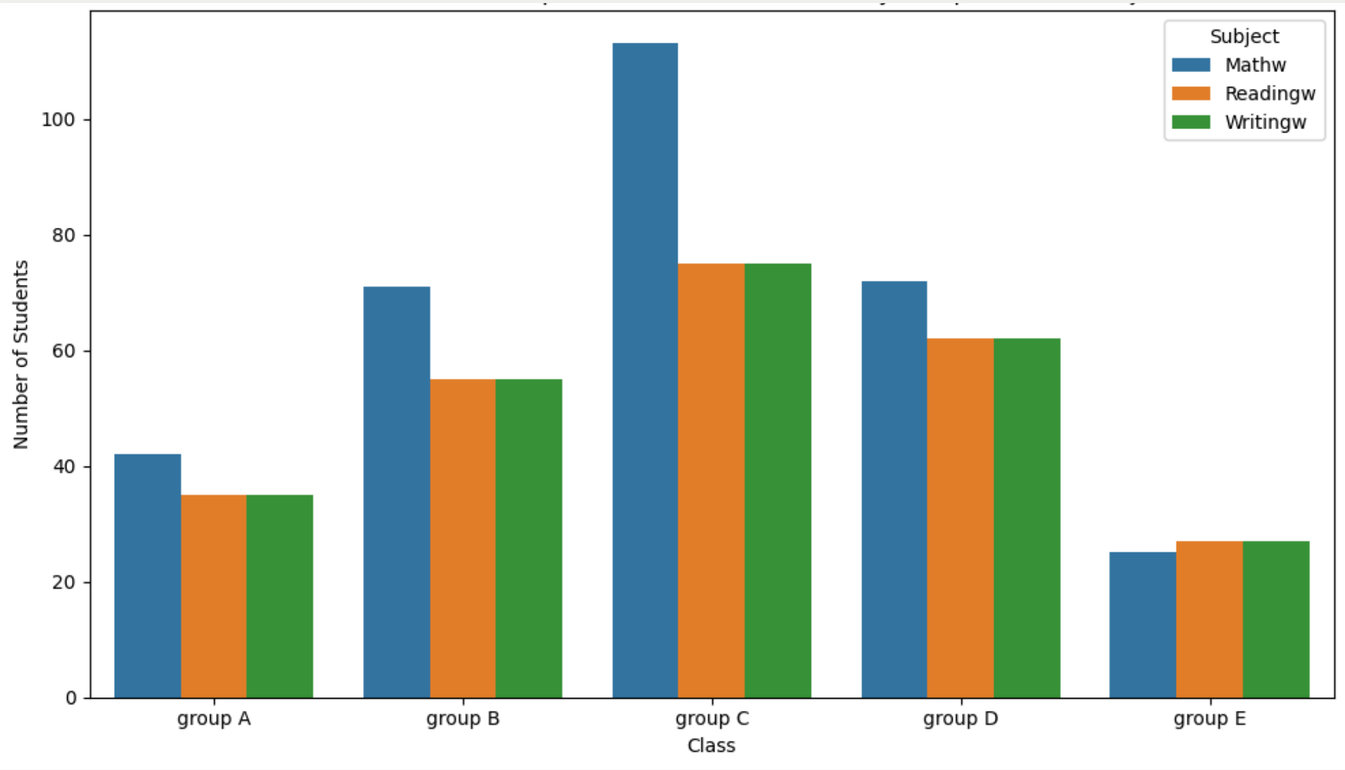


- **Group A shows the weakest performance across all three subjects, with very few students achieving excellent scores.**
- **Group B demonstrates moderate results, with the highest strength in Reading compared to Math and Writing.**
- **Group C stands out as the strongest group overall, especially in Writing (the highest among all groups), followed by Reading, while Math performance is slightly lower.**
- **Group D performs well in Reading and Writing, though their Math results are relatively weaker.**
- **Group E excels in Math (the top performer in this subject), while their Reading and Writing results remain solid but not outstanding.**

Recommendations

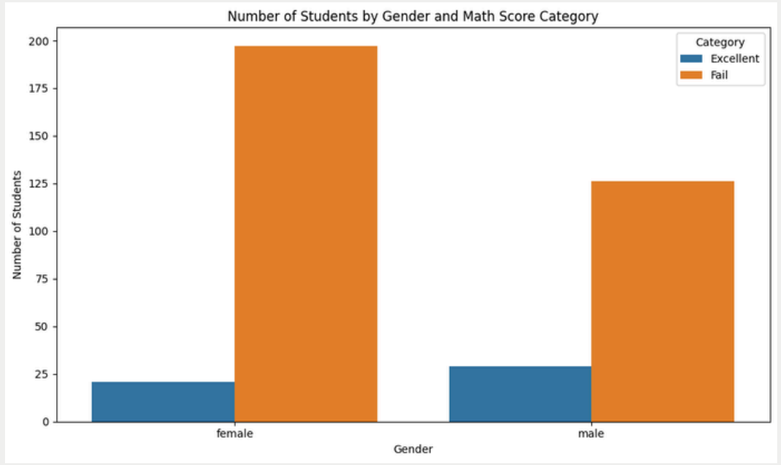
1. **Math Support: All groups except Group E require additional support materials to enhance their math performance. It is also important to investigate the methods or factors that enabled Group E to achieve high accuracy in this subject.**
2. **Reading Focus: Group C has demonstrated strong performance in reading, making it a benchmark for best practices in this area.**
3. **Special Attention to Group A: Group A consistently shows low performance across all subjects, and therefore requires targeted interventions and tailored support programs.**

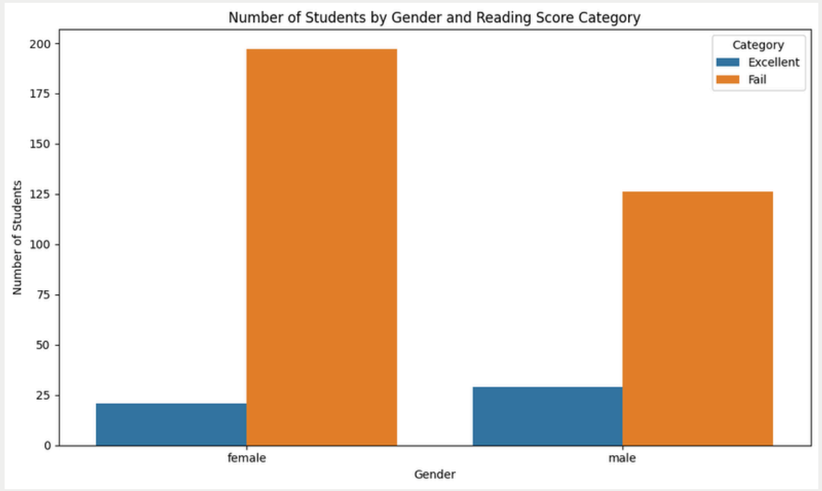
Number of Students Who Failed in Math, Reading, and Writing

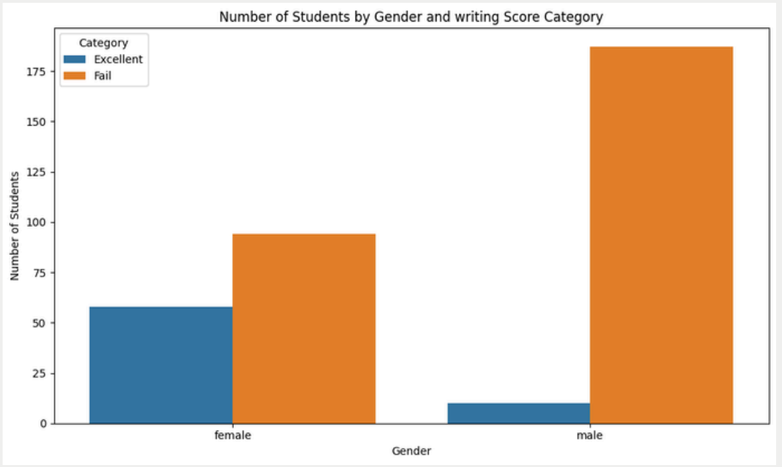


Group C has the highest number of students failing in math, followed by Group B and then Group A.
Across all groups, the number of students failing in Reading and Writing is nearly the same.

Number of Students Who Failed in Math, Reading, and Writing by Gender

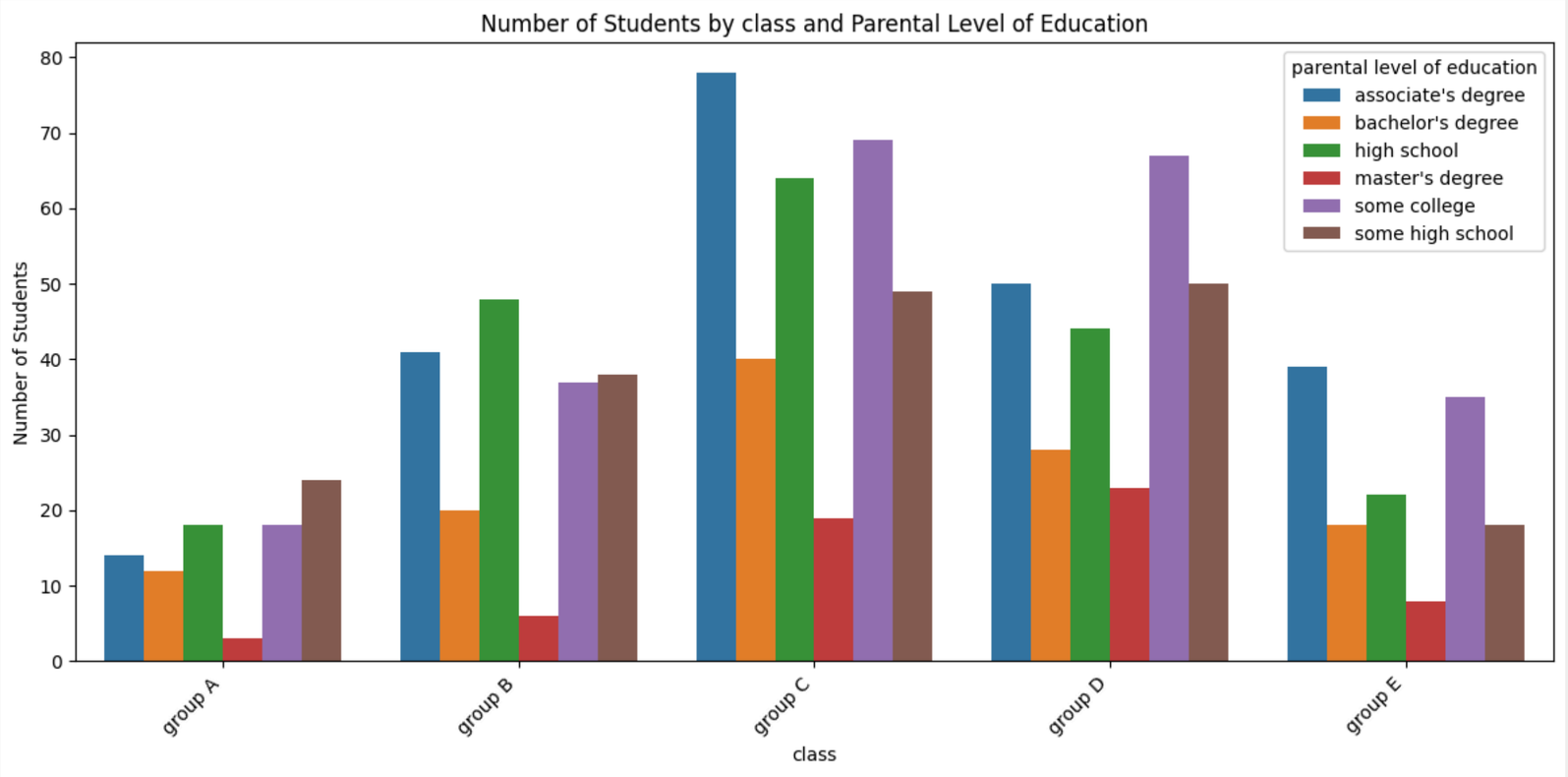






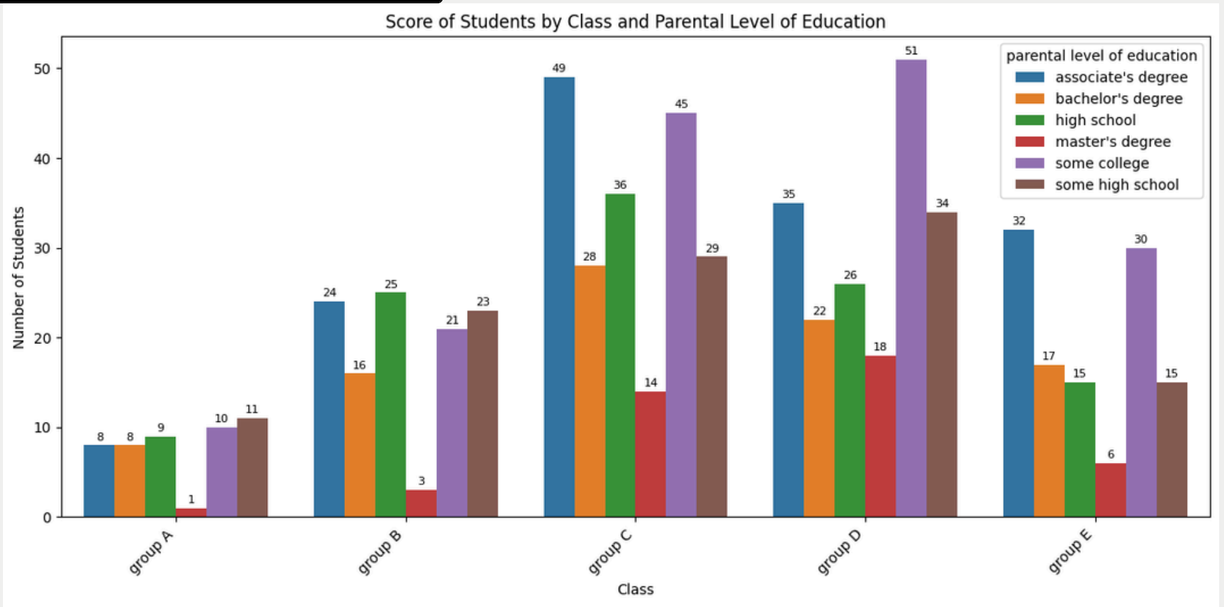
Most femal in math and reading fail but in reading most male fail

Distribution of the number of students in each group by parental level of education



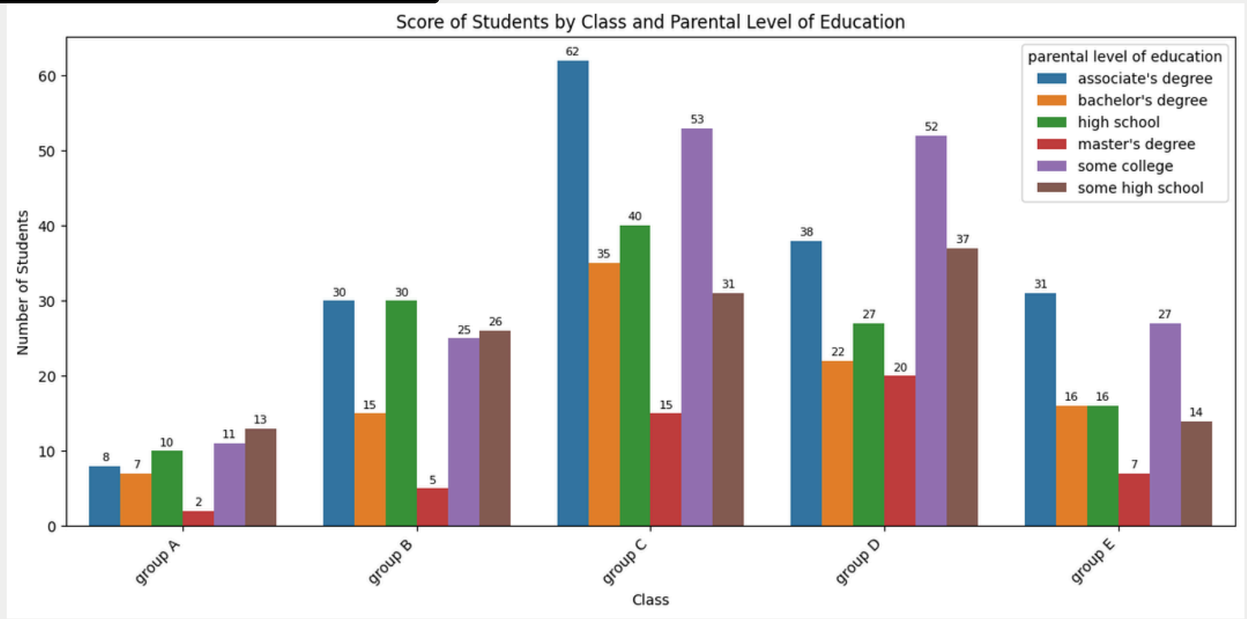
- **Groups C and D contain the largest share of students, strongly linked to parents with associate’s degree or some college education.**
- **Group A has the fewest students, mostly from families with lower educational backgrounds (high school or below).**
- **Group B is heavily influenced by high school–educated parents, while Group E is more balanced between associate’s degree and some college.**
- **Group D have the higer number of student have master degree when comapre to other groups.**

Distribution of Students Who Passed in Math, by Group and Parental Education Level



- Group C
- Shows the largest number of students excelling in mathematics.
- The majority come from families with an Associate’s degree or Some college, followed by High school.
- Group D
- Also demonstrates strong performance, especially among students whose parents have Some college or Some high school education.
- A considerable number come from families with an Associate’s degree.
- Group E
- Medium in size, with parental education most commonly at the Associate’s degree and Some college levels.
- Group B
- Most students in this group come from families with High school or Some high school education.
- Group A
- Has the smallest number of students in mathematics.
- Most parents here have lower education levels (High school or below).

Distribution of Students Who Passed in Reading, by Group and Parental Education Level



Group C

.Has the largest number of students excelling in reading

.Most parents hold an Associate’s degree or Some college, followed by High school

Group D

.Also shows strong performance, especially among students whose parents have Some college education (the highest in this group)

.Followed by Associate’s degree and Some high school

Group E

.Moderate in size, with the majority of parents holding an Associate’s degree or Some college background

Group B

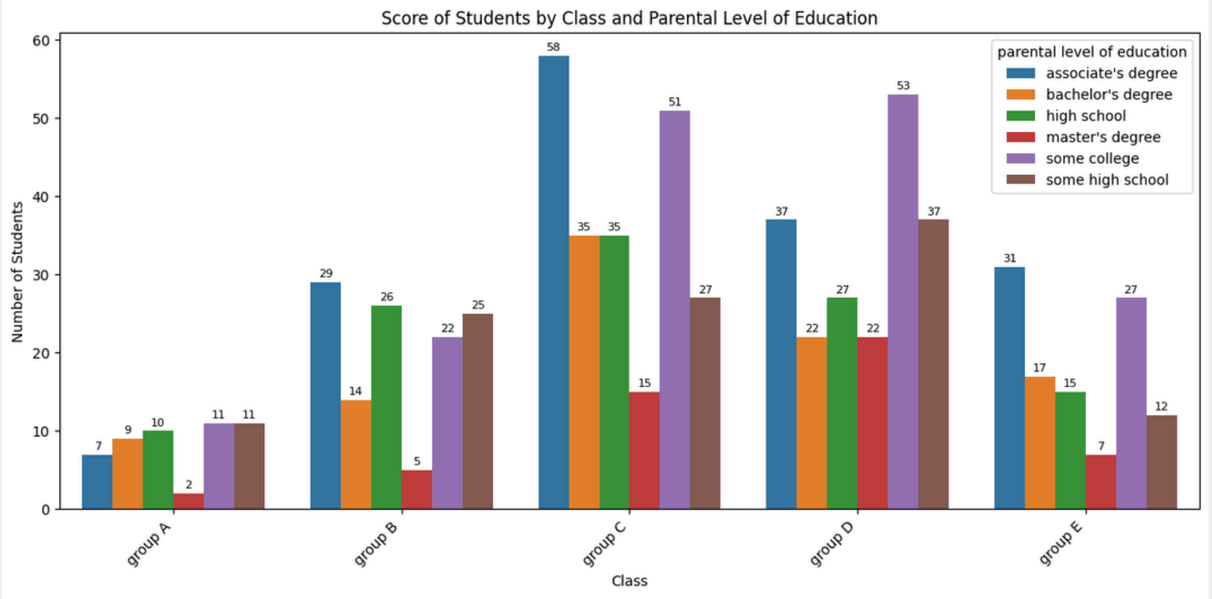
.Medium-level performance, mainly from families with High school or Some high school education

Group A

.The smallest group in reading scores

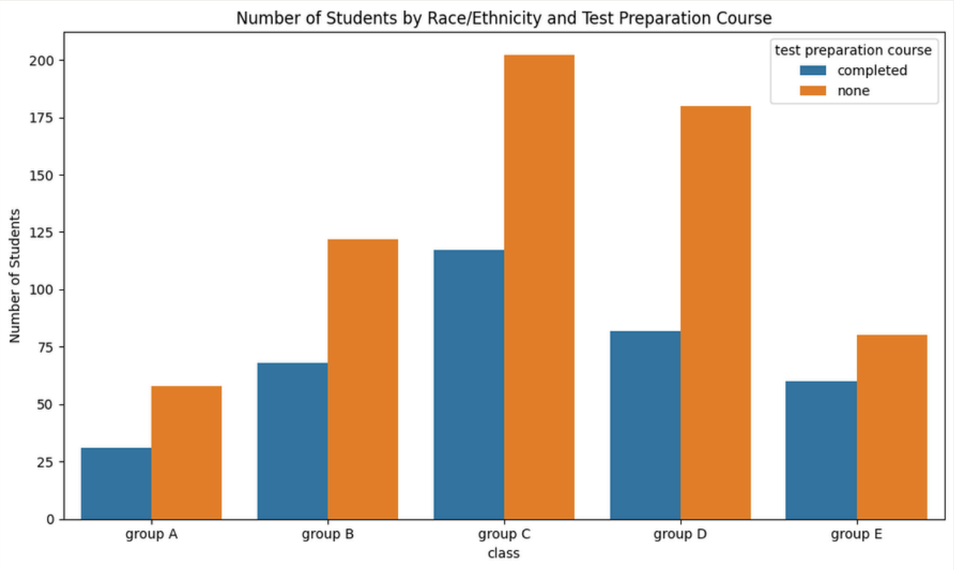
.Most parents here have lower education levels (High school or below)

Distribution of Students Who Passed in Writing, by Group and Parental Education Level



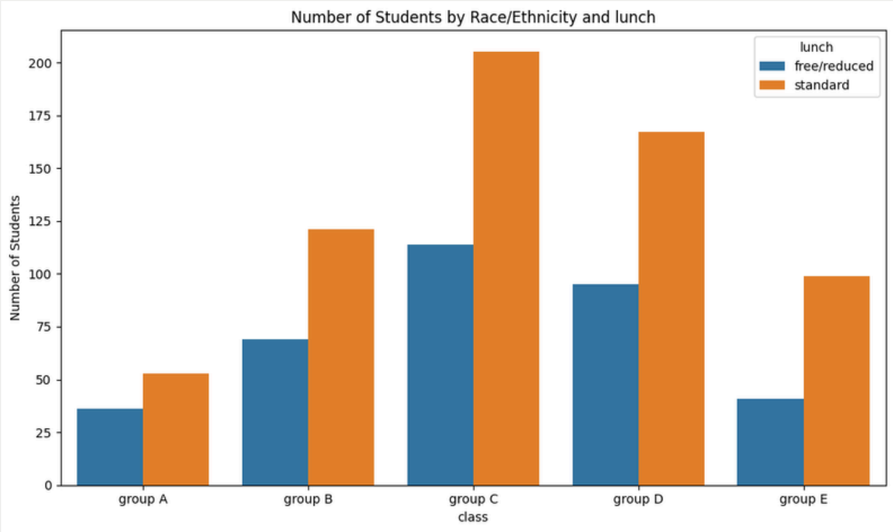
- Group C has the highest performance in writing, mostly from parents with an Associate’s degree and Some college.
- Group D also performs strongly, led by parents with Some college and Associate’s degree.
- Group E shows moderate results, mainly from Associate’s degree and Some college backgrounds.
- Group B is average, dominated by High school and Associate’s degree.
- Group A is the weakest group, with most parents having High school or below education.

Impact of Test Preparation Completion on Student scores



Most students across all groups did not take the test preparation course, which negatively affected their performance in math, reading, and writing scores.

Distribution of Students by Lunch Type



Since most students in all groups had a standard lunch, this factor did not show a significant impact on their scores. Other variables, such as parental education level or test preparation, had a stronger effect on performance.

Recommendations

- Redistribute groups more evenly, while taking into account the parental education level.
- Parental education has a clear impact on student achievement: the higher the parents' education level, the higher the expected performance of their children.
- Provide additional support for students whose parents have lower education levels.
- Revise and enhance the teaching strategy for mathematics, since most groups showed lower performance in this subject.
- Ensure completion of the test preparation course across all groups, as the lack of preparation was observed to negatively affect achievement.
- Develop targeted support plans for Group A students across all subjects, as they consistently demonstrated the lowest performance.