STUDENTS PERFORMANCE ASSESSMENT

SQL Fundamentals

By Ankit Prajapati

PROJECT DESCRIPTION:

THE PURPOSE OF THIS PROJECT IS TO ANALYZE THE PERFORMANCE OF STUDENTS BASED ON VARIOUS FACTORS SUCH AS GENDER, RACE/ETHNICITY, PARENTAL EDUCATION LEVEL, LUNCH TYPE, AND TEST PREPARATION COURSE COMPLETION. BY UNDERSTANDING THESE FACTORS, WE AIM TO DRAW MEANINGFUL INSIGHTS THAT CAN HELP IMPROVE EDUCATIONAL OUTCOMES AND PROVIDE TARGETED RECOMMENDATIONS FOR STUDENTS.

PROJECT AIM:

THIS PROJECT AIMS TO DEVELOP A DATA-DRIVEN APPROACH FOR STUDENTS PERFORMANCE AND ITS COMPLICATIONS AT AN EARLY STAGE USING SQL.

TECH STACK USED:

Used Software While Making The Project:

1. MYSQL Work Bench 8.0 (For Working, analyzing, and reporting Insights)
2. Microsoft PowerPoint (For Presenting the detailed analysis)
3. Power Bi (For Visualize Insights)

STUDENTS PERFORMANCE: CALCULATE THE AVERAGE SCORE OF STUDENTS IN ALL SUBJECTS?

```
select avg(math_score) as average_math,
       avg(reading_score) as average_reading,
       avg(writing_score) as average_writing
from students;
```

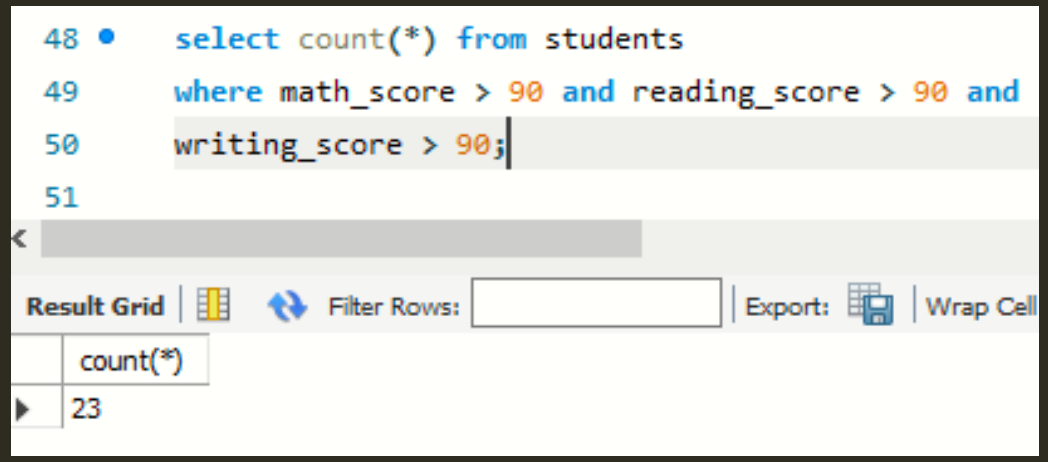
Result Grid | Filter Rows:

average_math	average_reading	average_writing
66.0890	69.1690	68.0540

NOTE: OBSERVING THE RESULTS FOR EACH SUBJECT, IT IS CLEAR THAT MATHEMATICS IS THE SUBJECT IN WHICH STUDENTS FACE THE GREATEST DIFFICULTY.

THE SUBJECT IN WHICH STUDENTS FIND IT EASIEST IS READING.

FIND NUMBER OF STUDENTS WHO SCORE ABOVE 90 IN ALL SUBJECTS.

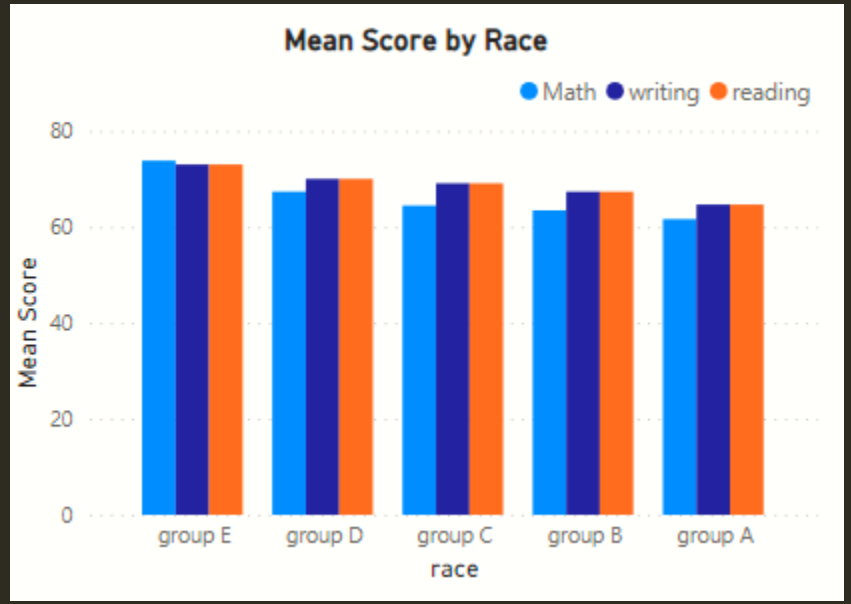


PERFORMANCE BY RACE/ETHNICITY: FIND AVERAGE SCORES STUDENTS IN DIFFERENT RACE GROUP

```
54 •  select race,
55        avg(math_score) as avg_math_score,
56        avg(reading_score) as avg_reading_score,
57        avg(writing_score) as avg_writing_score
58    from students
59    group by race;
```

Result Grid | Filter Rows: | Export: | Wrap Ce

race	avg_math_score	avg_reading_score	avg_writing_score
group B	63.4526	67.3526	65.6000
group C	64.4639	69.1034	67.8276
group A	61.6292	64.6742	62.6742
group D	67.3626	70.0305	70.1450
group E	73.8214	73.0286	71.4071



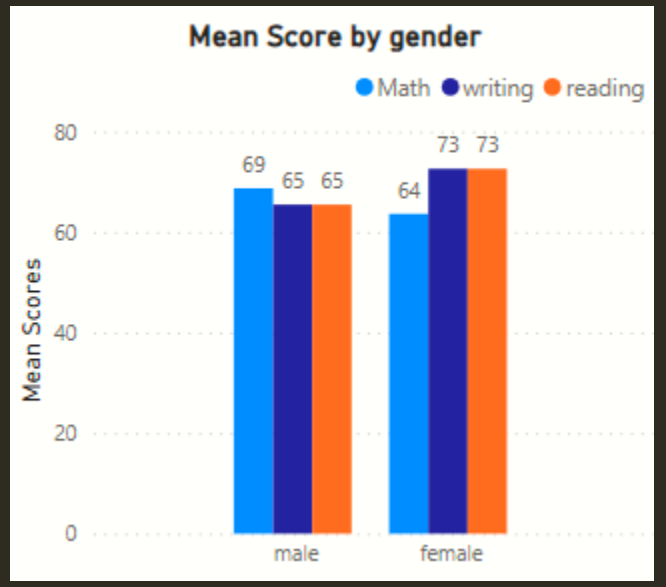
NOTE: AS SEEN GROUP E HAS THE HIGHEST AVERAGE SCORES FOLLOWED BY D,C,B,A

PERFORMANCE ANALYSIS BY GENDER: CALCULATE THE AVERAGE SCORE OF STUDENTS IN ALL SUBJECTS BY THEIR GENDER

```
47 •  select gender, avg(math_score) as average_math,
48          avg(reading_score) as average_reading,
49          avg(writing_score) as average_writing
50    from students
51    group by gender;
```

Result Grid | Filter Rows: | Export: | Wrap Cel

gender	average_math	average_reading	average_writing
female	63.6332	72.6081	72.4672
male	68.7282	65.4730	63.3112



NOTE: IT CAN BE SEEN THAT FEMALES HAVE HIGHER AVERAGE SCORES IN READING AND WRITING WHILE MALES HAVE HIGHER AVERAGE IN MATH.

FIND THE TOP 10 STUDENTS WITH THE HIGHEST OVERALL SCORE (AVERAGE OF MATH, READING, WRITING)

```sql
SELECT gender, race,
  Round((math_score + reading_score + writing_score) / 3, 2) AS overall_score
FROM students
ORDER BY overall_score DESC
limit 10;
```

Result Grid | Filter Rows:

gender	race	overall_score
female	group E	100.00
female	group E	100.00
male	group E	100.00
female	group E	99.67
female	group D	99.00
female	group D	99.00
male	group D	98.67
female	group C	98.67
female	group D	97.67
male	group E	97.67

FIND RACE GROUPS WHERE THE AVERAGE MATH SCORE IS GREATER THAN 70

```
83 •  select race, avg(math_score) as avg_math
84    from students
85    group by race
86    having avg_math > 70;
87
```

Result Grid | Filter Rows: | Export:

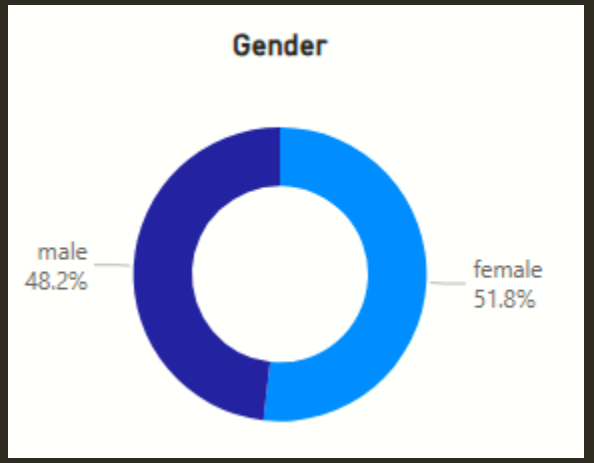
race	avg_math
group E	73.8214

HOW MANY MALE AND FEMALE THERE ARE IN NUMBERS AND PERCENTAGE?

```
90 •  select *,
91        Round(frequency/sum(frequency) over() * 100,2) as percentage
92    from(
93    select gender, count(*) as frequency
94    from students
95    group by gender
96    order by frequency desc) tab1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

gender	frequency	percentage
female	518	51.80
male	482	48.20



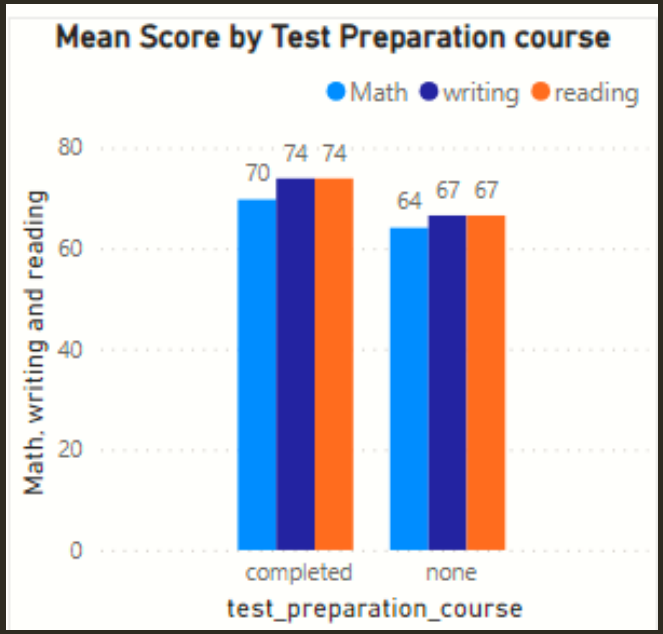
NOTE: THERE ARE 518 FEMALES AND 482 MALES, RESULTING IN A POPULATION THAT IS 52% FEMALE AND 48% MALE.

WRITE AN SQL QUERY TO CALCULATE THE AVERAGE SCORE FOR STUDENTS BASED ON WHETHER THEY COMPLETED A TEST PREPARATION COURSE?

```
120 •   select test_preparation_course,
121             avg(math_score) as avg_math_score,
122             avg(reading_score) as avg_reading_score,
123             avg(writing_score) as avg_writing_score
124     from students
125     group by test_preparation_course;
126
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

test_preparation_course	avg_math_score	avg_reading_score	avg_writing_score
none	64.0779	66.5343	64.5047
completed	69.6955	73.8939	74.4190



NOTE: STUDENTS WHO COMPLETED TEST PREPARATION COURSE HAVE HIGHER AVERAGE SCORES COMPARED TO THOSE WHO DID NOT COMPLETE THE COURSE.

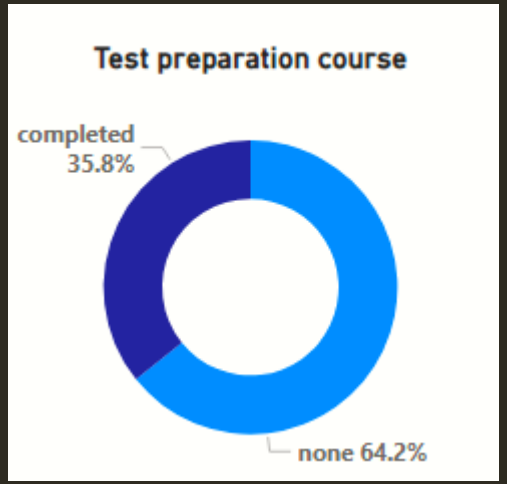
IT INDICATE THE COURSE LIKELY HELP IMPROVE THEIR PERFORMANCE

PARTICIPATION IN PREPARATION COURSE: FIND NUMBER OF STUDENTS WHO COMPLETED TEST PREPARATION COURSE AND THOSE WHO DID NOT?

```
115 •   select *,
116         Round(frequency/sum(frequency) over() * 100,2) as percentage
117     from(
118     select test_preparation_course,
119           count(*) as frequency
120     from students
121     group by test_preparation_course) per;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

test_preparation_course	frequency	percentage
none	642	64.20
completed	358	35.80



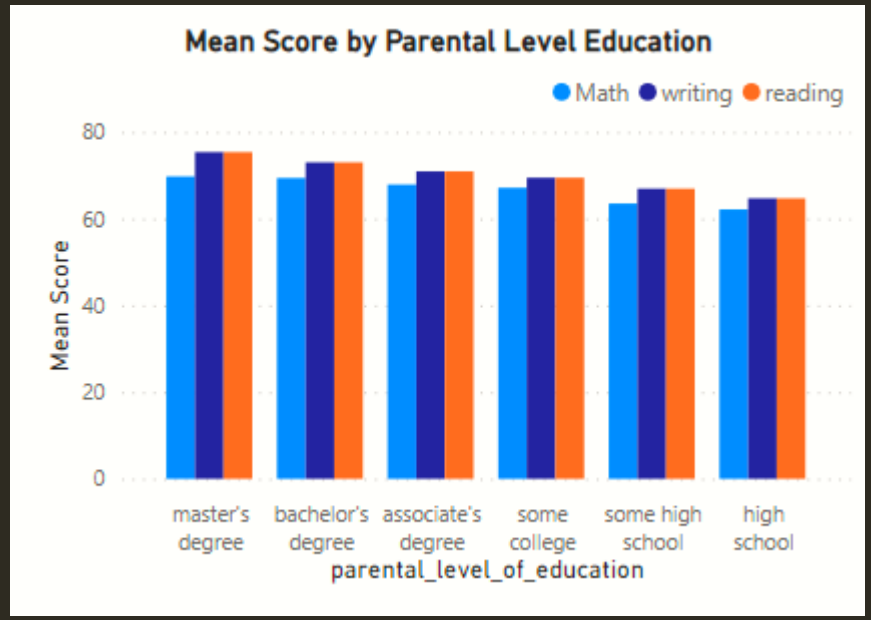
NOTE: THE MAJORITY OF GROUP CONSISTS OF STUDENTS WHO DO NOT PARTICIPATE IN PREPARATORY COURSE WITH 64% OF GROUP.

SCORE DISTRIBUTION BY PARENTAL LEVEL OF EDUCATION: CALCULATE AVERAGE MATH SCORE BY PARENTAL LEVEL EDUCATION?

```
55 •    select parental_level_of_education,
56          avg(math_score) as avg_score
57      from students
58      group by parental_level_of_education;
```

Result Grid | Filter Rows: | Export:

parental_level_of_education	avg_score
bachelor's degree	69.3898
some college	67.1283
master's degree	69.7458
associate's degree	67.8829
high school	62.1378
some high school	63.4972



NOTE: PARENT'S WITH MASTER'S HAVE ACHIEVE HIGHER SCORES WHILE PARENTS WITH ONLY HIGH SCHOOL HAVE LOWEST SCORES.

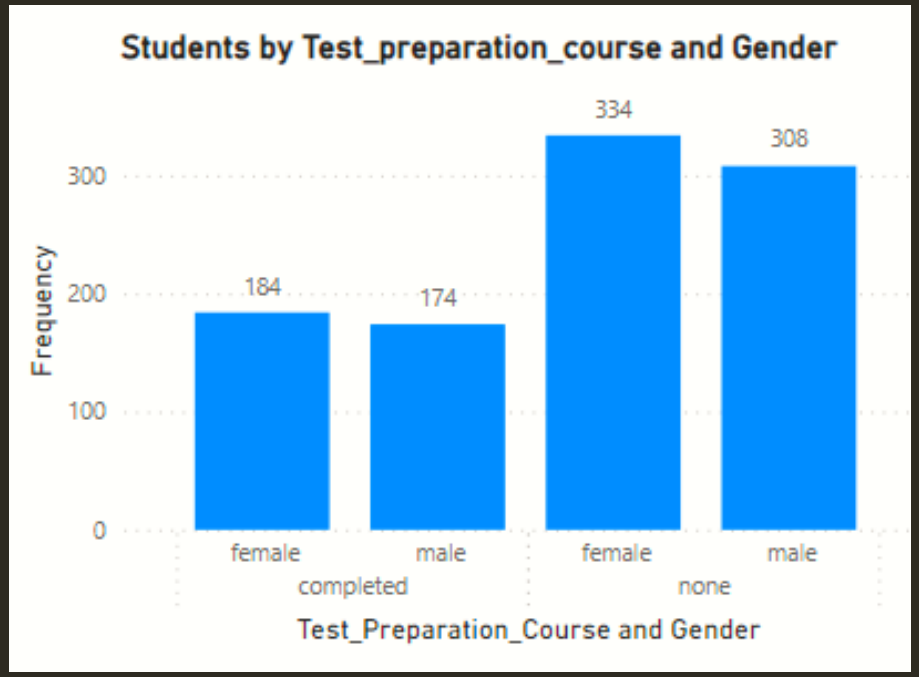
SCORE DISTRIBUTION: FREQUENCY DISTRIBUTION FOR MATH SCORES

```sql
SELECT
    CASE
        WHEN math_score BETWEEN 0 AND 10 THEN '0-10'
        WHEN math_score BETWEEN 11 AND 20 THEN '11-20'
        WHEN math_score BETWEEN 21 AND 30 THEN '21-30'
        WHEN math_score BETWEEN 31 AND 40 THEN '31-40'
        WHEN math_score BETWEEN 41 AND 50 THEN '41-50'
        WHEN math_score BETWEEN 51 AND 60 THEN '51-60'
        WHEN math_score BETWEEN 61 AND 70 THEN '61-70'
        WHEN math_score BETWEEN 71 AND 80 THEN '71-80'
        WHEN math_score BETWEEN 81 AND 90 THEN '81-90'
        WHEN math_score BETWEEN 91 AND 100 THEN '91-100'
    END AS score_range,
    COUNT(*) AS count
FROM students
GROUP BY score_range
ORDER BY score_range;
```

Result Grid | Filter R

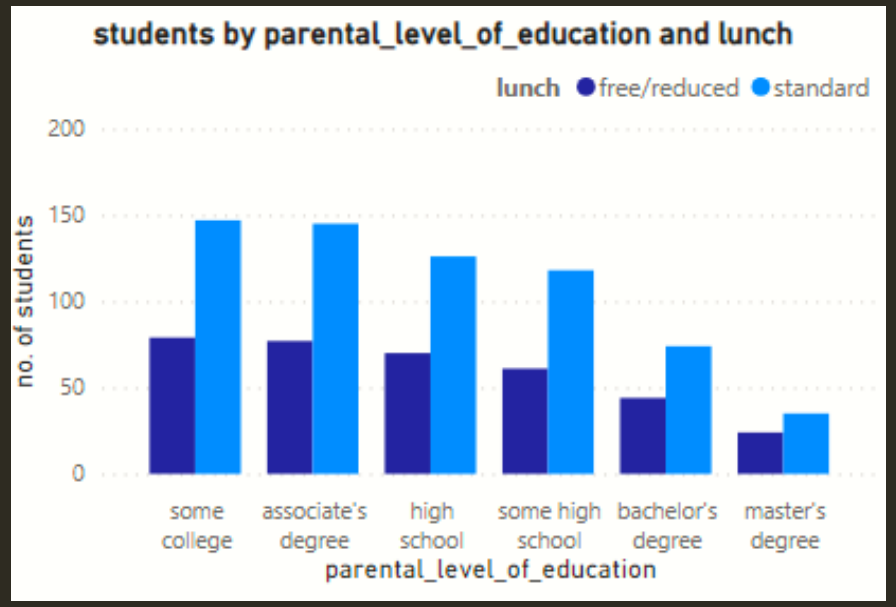
score_range	count
0-10	2
11-20	2
21-30	12
31-40	34
41-50	100
51-60	189
61-70	270
71-80	215
81-90	126
91-100	50

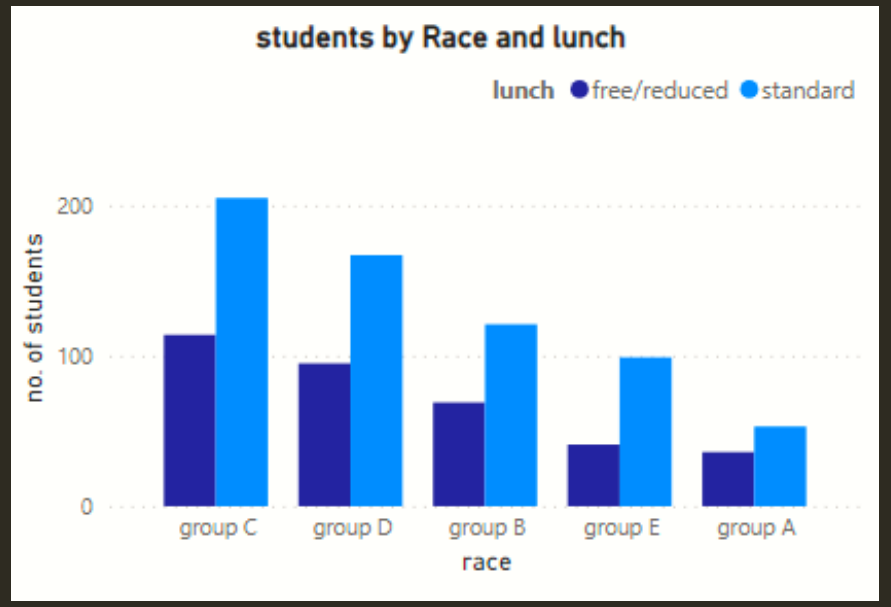
IN POWER BI, CREATE CHART THAT EXPLORE THE NUMBER OF MALE/FEMALE WHO COMPLETED THE TEST PREPARATION COURSE IN THE DATASET?



NOTE: IT IS OBSERVED THAT MORE THAN HALF DID NOT COMPLETE THE PREPARATION COURSE

DOES RACE OR PARENTAL LEVEL OF EDUCATION AFFECT THE LUNCH TAKEN BY STUDENTS?





NOTE: BOTH RACE AND PARENT'S LEVEL OF EDUCATION SEEM TO HAVE NO EFFECT ON LUNCH TAKEN BY STUDENTS

Analysis and Findings:

- Subject Difficulty and Performance: Mathematics is the most challenging subject for students, Reading is the subject in which students perform the best
- Performance by Race: Group E scores highest, followed by Groups D, C, B, and A.
- Gender Based Performance: Females outperform male in reading and writing. Male have a slight edge over female in mathematics. This indicates a gender-based difference in subject-specific performance
- Impact of test course: Students who complete the course score higher. A majority (64%) of the students did not participate in the test preparation course.
- High Achievers: Few students score above 90 in all subjects.
- Parental Education Level: Students with parents who have a master's degree tend to achieve higher scores and parents who completed high school have the lowest score. Higher parental education may have positive influene.

Recommendations for Students:

Focus on Challenging Subjects: Students should allocated more study time and seek additional help to improve their math skills.

Parental Involvement: Encouraging parents to be involved in their children's education, especially those with lower educational

Female students should work on mathematics. Male students should work on their reading and writing.

Test Preparation Course: Students are encouraged to participate in test preparation courses, it indicates these courses are beneficial in boosting academic performance.

Targeted Support: Schools and educators should provide targeted support and resources for racial group and for students who are struggling, based on analysis.

Balanced Academic Students should aim for a well-rounded academic approach, trying to do well in all subjects rather than focusing on just one. Utilize available support for difficult subjects.


Thank You