Student Result Analysis

1. Project Overview

- This project aims to analyze a large student performance dataset to understand how demographic, social, and family-related factors influence academic achievement across Mathematics, Reading, and Writing.
- The focus is on identifying patterns, evaluating correlations, and understanding how factors like gender, parental education, marital status, and ethnic background impact student outcomes.
- The project also involves performing data cleaning, handling missing values, removing unnecessary fields, and creating visualizations to derive meaningful insights.

2. Dataset Summary

The dataset contains 30,641 student records with 14 variables, including demographic details, family background, and academic scores.

	Gender	EthnicGroup	ParentEduc	LunchType	TestPrep	ParentMaritalStatus	PracticeSport	IsFirstChild	NrSiblings
0	female	NaN	bachelor's degree	standard	none	married	regularly	yes	3.0
1	female	group C	some college	standard	NaN	married	sometimes	yes	0.0
2	female	group B	master's degree	standard	none	single	sometimes	yes	4.0
3	male	group A	associate's degree	free/reduced	none	married	never	no	1.0
4	male	group C	some college	standard	none	married	sometimes	yes	0.0

TransportMeans	WklyStudyHours	MathScore	ReadingScore	WritingScore
school_bus	< 5	71	71	74
NaN	5 - 10	69	90	88
school_bus	< 5	87	93	91
NaN	5 - 10	45	56	42
school_bus	5 - 10	76	78	75

2.1 Key Attributes

- Target variables: MathScore, ReadingScore, WritingScore
- Categorical variables: Gender, EthnicGroup, ParentEduc, LunchType, etc.
- Numeric variables: NrSiblings, scores

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30641 entries, 0 to 30640
Data columns (total 14 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Gender               30641 non-null  object
 1   EthnicGroup          28801 non-null  object
 2   ParentEduc           28796 non-null  object
 3   LunchType            30641 non-null  object
 4   TestPrep             28811 non-null  object
 5   ParentMaritalStatus  29451 non-null  object
 6   PracticeSport        30010 non-null  object
 7   IsFirstChild         29737 non-null  object
 8   NrSiblings           29069 non-null  float64
 9   TransportMeans       27507 non-null  object
 10  WklyStudyHours       29686 non-null  object
 11  MathScore            30641 non-null  int64
 12  ReadingScore         30641 non-null  int64
 13  WritingScore         30641 non-null  int64
dtypes: float64(1), int64(3), object(10)
memory usage: 3.3+ MB
```

- Missing values:
    - Highest missing: TransportMeans (3134), EthnicGroup (1840), ParentEduc (1845)
    - No missing values in scoring columns, so no imputation required for analysis.

```
Unnamed: 0                0
Gender                    0
EthnicGroup            1840
ParentEduc             1845
LunchType                 0
TestPrep               1830
ParentMaritalStatus    1190
PracticeSport           631
IsFirstChild            904
NrSiblings             1572
TransportMeans         3134
WklyStudyHours          955
MathScore                 0
ReadingScore              0
WritingScore              0
dtype: int64
```

2.2 Statistical Highlights

- Average scores:
    - Math: 66.56
    - Reading: 69.37
    - Writing: 68.41

	Unnamed: 0	NrSiblings	MathScore	ReadingScore	WritingScore
count	30641.000000	29069.000000	30641.000000	30641.000000	30641.000000
mean	499.556607	2.145894	66.558402	69.377533	68.418622
std	288.747894	1.458242	15.361616	14.758952	15.443525
min	0.000000	0.000000	0.000000	10.000000	4.000000
25%	249.000000	1.000000	56.000000	59.000000	58.000000
50%	500.000000	2.000000	67.000000	70.000000	69.000000
75%	750.000000	3.000000	78.000000	80.000000	79.000000
max	999.000000	7.000000	100.000000	100.000000	100.000000

- Score ranges are valid: 0–100, indicating no data entry anomalies.

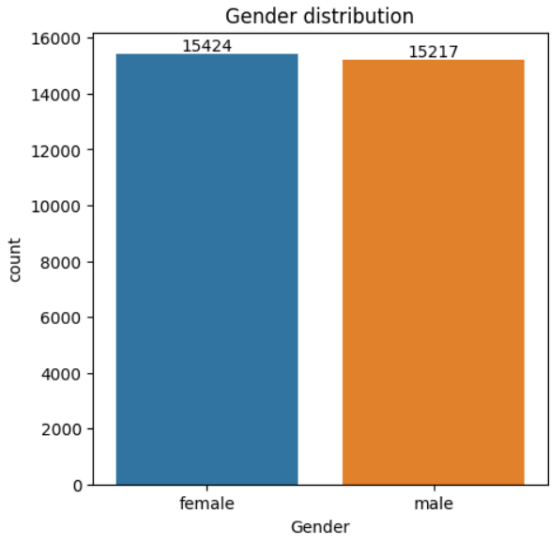
3. Data Transformation

- Removed the Unnamed: 0 column, which served no analytical purpose and only added redundancy to the dataset.
- Verified that the dataset contains zero duplicate rows, indicating high integrity and consistency of the provided data.
- Reviewed missing values and confirmed that while certain demographic fields contain gaps, the score-related fields are completely intact, enabling accurate evaluation of academic patterns.
- Prepared the dataset for visualization and further exploration by ensuring all key fields were clean, structured, and ready for analysis.

4. Data Visualization

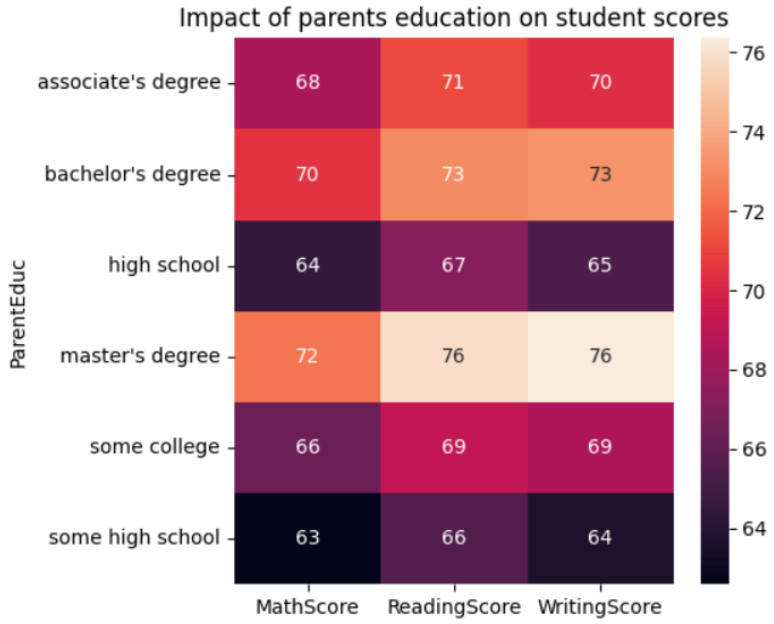
4.1 Gender Distribution

- The visual analysis showed that the dataset contains female students are slightly more than male students.
- This balanced distribution ensures that gender-based insights are naturally unbiased and reflective of the full dataset.



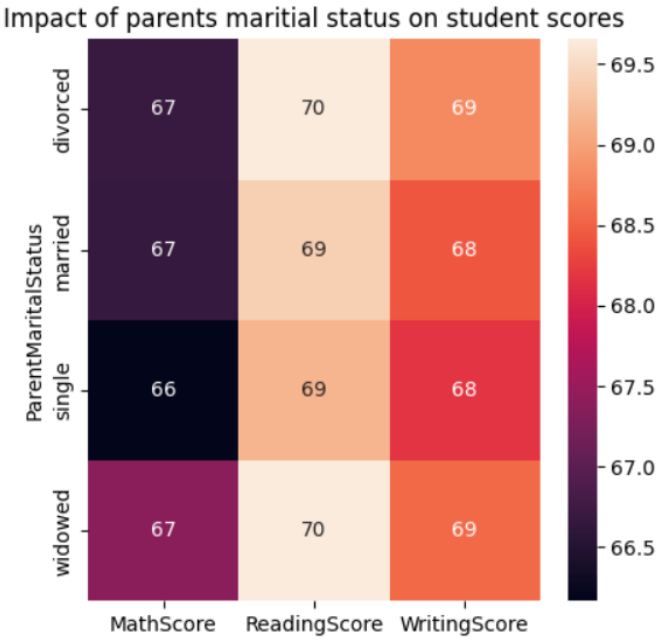
4.2 Impact of Parental Education on Student Scores

- Students with parents who hold Master's degrees consistently achieve the highest average scores across all three subjects.
- Those whose parents have only completed some high school education scored the lowest on average, demonstrating a strong correlation between parental education and student academic success.
- The trend clearly indicates that the educational background of parents significantly contributes to shaping the academic readiness and support available to their children.



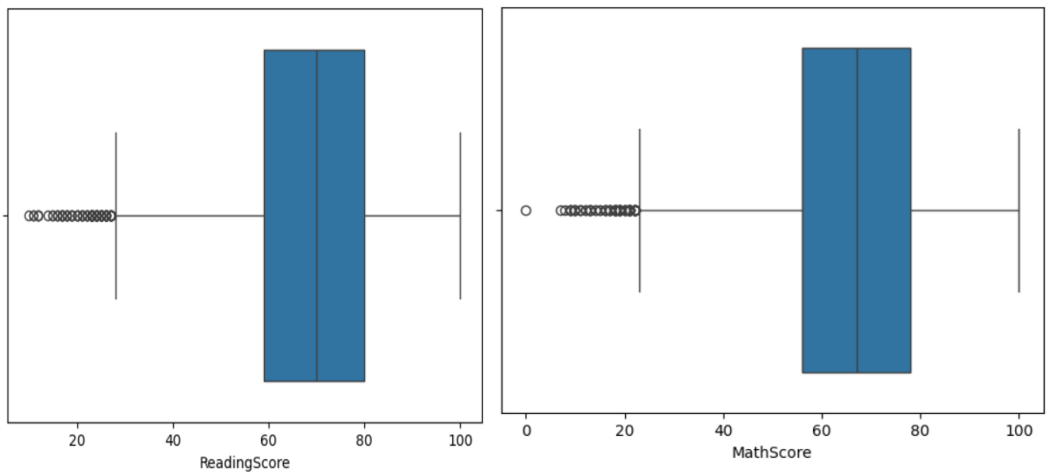
4.3 Impact of Parent Marital Status

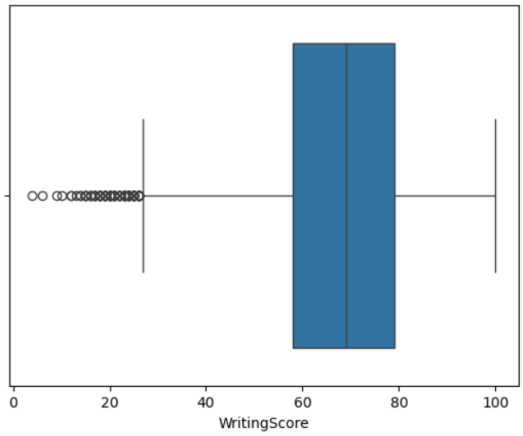
- An analysis of marital status shows minimal score differences between students of married, divorced, single, or widowed parents.
- This suggests that marital structure alone does not significantly affect academic performance, and other environmental or educational factors play a much larger role.



4.4 Outlier Detection

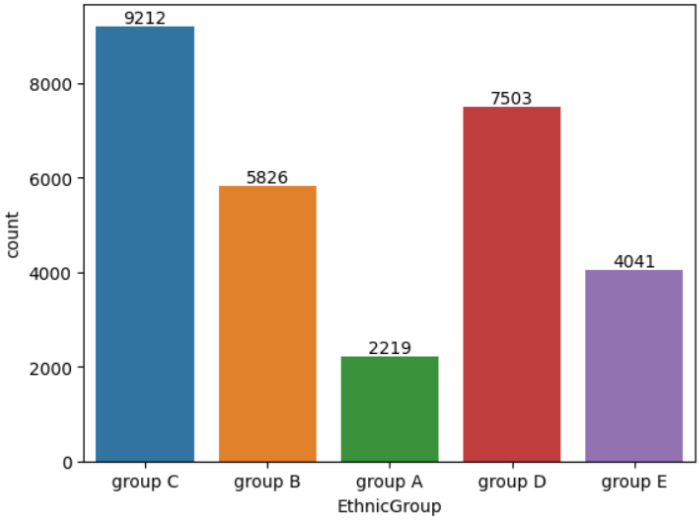
- Boxplots for Math, Reading, and Writing scores reveal no abnormal outliers, as all values fall within the expected 0–100 range.
- The absence of extreme or irregular data points reinforces the dataset's reliability and supports the validity of the conclusions drawn from it.
- Students are relatively week in Maths than Reading or Writing.

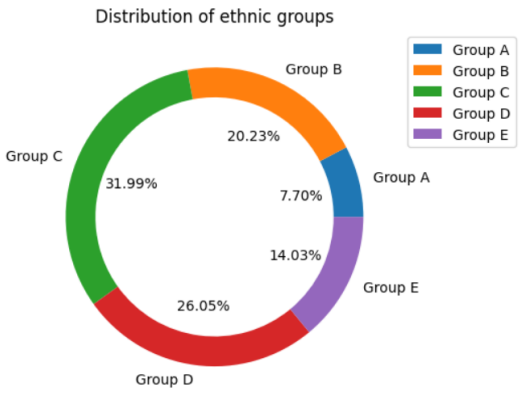




4.5 Ethnic Group Distribution

- Ethnic Group C is the largest represented demographic among all ethnic groups in the dataset, accounting for nearly one-third of all students.
- Other groups show varying but reasonable representation, making it possible to compare academic differences across ethnic categories if needed.





5. Conclusion

- The dataset is clean, large, and free from critical missing values in the academic score fields, making it reliable for analysis.
- Parental education emerges as the most influential factor affecting student performance, with higher education levels strongly associated with higher scores.
- Mathematics stands out as the comparatively weaker subject for most students when compared with Reading and Writing.
- Parent marital status has minimal impact, suggesting students' academic performance depends more on educational environment and parental involvement than family structure.
- Ethnic group distribution is diverse, with Group C having the highest representation.
- Overall, the dataset provides rich insights into factors shaping student performance and helps guide targeted improvement strategies.

6. Recommendations (Schools & Educators)

- Strengthen math support with remedial classes, concept-based teaching, and engaging digital tools to improve problem-solving skills.
- Support parents with lower education through simple workshops, easy study guides, and regular communication to help them assist their children effectively.
- Implement consistent test preparation using mock exams, revision periods, and guided practice sessions to boost exam readiness.
- Create supportive study spaces such as quiet study hours, supervised homework zones, and resource rooms for extra help.

- Use continuous assessments to track progress, identify weak areas early, and personalize learning support.
- Promote literacy activities like reading clubs and writing workshops to strengthen Reading and Writing skills.
- Develop teacher mentorship programs and small-group learning sessions for students needing focused guidance.
- Teach effective study habits including time management, organized note-taking, and regular review routines.
- Improve parent–teacher communication to keep families informed and involved in student progress.
- Integrate digital learning tools to make subjects interactive, engaging, and personalized for students.