SUMMARY DATASET

The dataset comprises 1000 entries and 8 columns. The columns in the dataset focus on gender, race/ethnicity, parental level of education, lunch, test preparation course, maths score, reading and writing scores. The 3 columns that focus on scores are integer type, whilst the balance of columns are objects. Each column presents 0 missing values and there are no duplications in the dataset. The stats provide us with mean values for Math (66.09), Reading(66.17) & writing scores(68.05). Further information includes standard deviation, minimum, and maximum values for data dispersion and understanding the range of data.

Percentiles aid in assessing data distribution, providing a comprehensive overview of the dataset's characteristics. Notably, there are no missing values in any column, ensuring the dataset's completeness.

DATA CLEANING

The data illustrates no duplications or missing values, and the critical information at this stage, specifically sores are in the required datatype.It's worth noting that for comparing the categorical variables(objects listed previously), we will have to convert those data types for better /deeper comparison , using Label encoding or One hot encoding.

MISSING DATA

A comprehensive check was conducted to identify any missing values present in the dataset. It was confirmed that there were no missing values detected across the dataset. This finding indicates that the dataset was complete, with no observations lacking values in any of the columns. The absence of missing data streamlines the data cleaning process and ensures that the dataset is ready for analysis without the need for imputation or deletion of incomplete records. This aspect enhances the reliability and robustness of the dataset for subsequent analysis and interpretation.

## STORIES AND ASSUMPTIONS BASED ON VISUALISATIONS OF DATA

Fig 1:
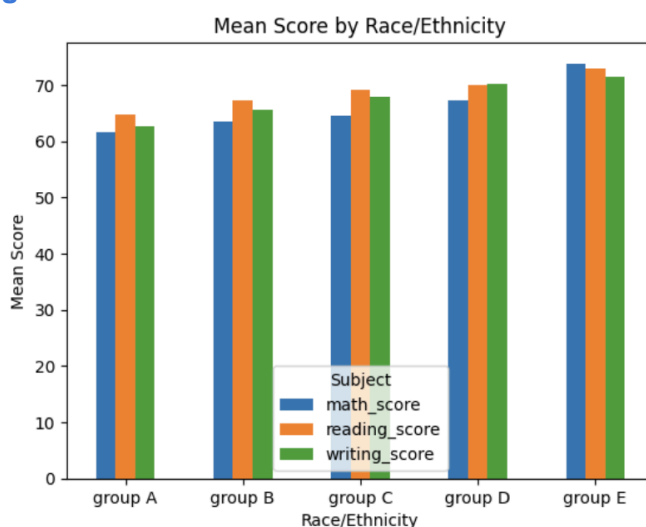
From the above, we see that all ethnic groups performed quite well on their academics, however there are two top performing ethnicities - namely group D and E with mean scores

of above 65 across their academic scores. Considerations that can help us delve deeper into this includes parental level of education, access to resources, socio-economic status and geographical location.Whilst the dataset does not provide all that information- we see from the visualisation in figure 3 that there tends to be a positive correlation between parental level of education and performance(even though it's marginal). Further considerations for analysis to provide more insight on the above include the availability and quality of educational resources, including school funding, teacher qualifications, class sizes, and extracurricular activities, in schools attended by students from groups D and E. Figure 4 delves a bit into access to materials(specifically a recourse and its impact on academic outcomes). Whilst we've seen that there is a positive correlation between academic score and parental level of education, it's worth exploring that aspect in depth, by evaluating family environments - specifically the level of parental involvement in the education[1] - at large to those students in ethnic groups D & E. Here the cultural attitude to education could play a role in performance and support too. This could stretch further into community support and resources[2] available to these groups, including mentorship programs, tutoring services, and partnerships with local organisations. Further factors worth delving into include School policies and Practices (e.g curriculum design, teaching methods, assessment strategies, and support services for students) as well as peer influence, influence of peer groups and social dynamics within schools attended by students in groups D and E. A collaborative and supportive environment could be a contributor to success.

Fig 2:

Mean Score by Gender
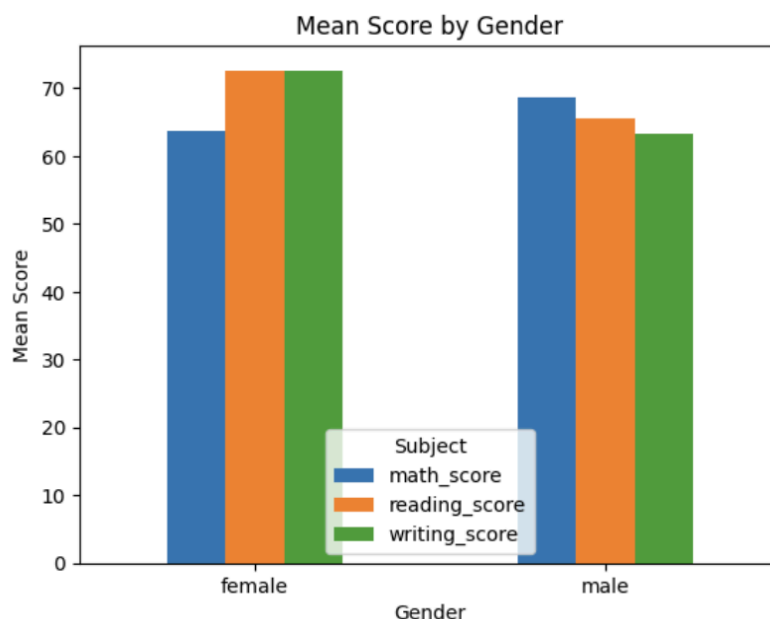
[1]

Elliott III, W., Addo, F. R., Chowa, G., & Rotz, D. (2016). Parental Involvement's Effects on Academic Performance: Evidence from the YouthSave Ghana Experiment. *Journal of Research on Adolescence*. [Online] Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/jora.12278

[2]

Jackson, C. K., Johnson, R. C., & Persico, C. (2015). The Role of School Resources in Shaping the Academic Outcomes of Low-Income Children. *National Bureau of Economic Research*. [Online] Available at: https://www.nber.org/papers/w20847

In figure 2 above, we see that while performance is generally good with a mean of 60+, Females in this dataset tend to perform better in reading and writing than their male counterparts, whilst the males tend to perform much better than females at Maths, with a 2-3% higher mean score in this subject. A regression analysis on this can help us understand the relationship between gender and academic performance while controlling for other factors such as parental education level, socioeconomic status, or test preparation[3]. This can provide insights into whether gender remains a significant predictor of performance after accounting for other variables.
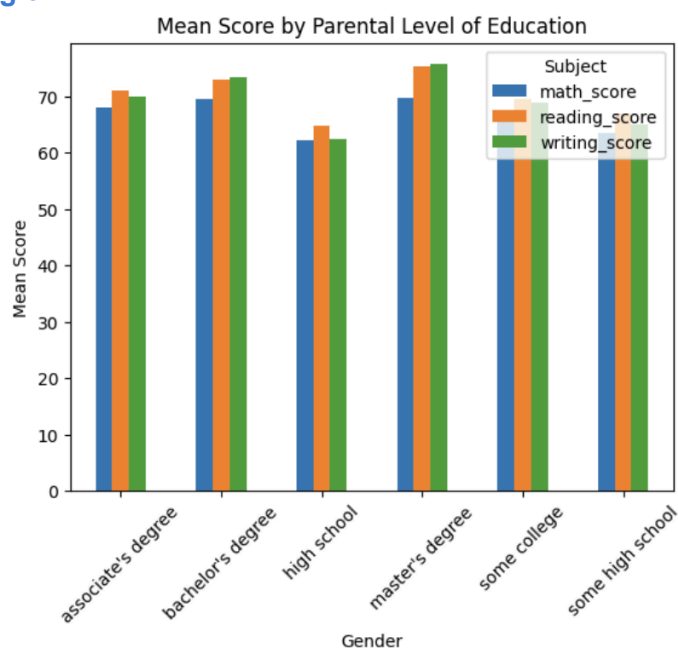
Fig 3:

Figure 3 shows us that whilst all students generally perform fairly well with a mean of 60+, the students whose parents have either an Associates, Bachelor or Master's degrees tend to perform better than students whose parental level of education is limited to some college and below.

To delve into this observation further, we can consider the influence of parental education on student academic performance. Research suggests that parental education is a significant predictor of academic achievement, with higher levels of parental education often associated with better student outcomes.

One study by Davis-Kean, K. A. (2005) found that parental education level was positively correlated with children's academic achievement across various domains, including mathematics, reading, and writing. The study highlighted that parents with higher levels of education tend to provide a more supportive learning environment, engage in educational


3

Rosero-Bixby, L. D., & Dow, W. H. (2008). Gender Differences in Educational Attainment: A Dynamic Approach. *Demography*, 45(3), 697-713. https://doi.org/10.1353/dem.0.0015

activities with their children, and emphasise the importance of education, which positively impacts their children's academic performance.[4]

Additionally, research by Sirin, S. R. (2005) explored the role of parental education in predicting academic achievement among adolescents. The study revealed that parental education level was a strong predictor of academic success, with higher parental education associated with higher student grades, test scores, and educational aspirations. The findings emphasised the critical role of parental involvement and support in fostering academic achievement among students.[5]

An interesting observation is what we see in fig 4. Here students who averaged 90+ across scores in all 3 domains, tend to be the ones who have not completed the preparation course.Students who have however completed the precourse tended to average a score between 60-80

The observed performance disparity between students who completed a test preparation course and those who did not suggests several underlying factors. Self-selection bias likely plays a role, as motivated or academically inclined students may opt for the course, potentially inflating their scores. Additionally, variations in the course's effectiveness, influenced by instructional quality and resource availability, could contribute to score differences. Socioeconomic status and access to external preparation resources may further impact outcomes[6]. Further research integrating qualitative methods could provide deeper insights into enrollment decisions and course effectiveness, aiding in the design of more equitable test preparation programs.
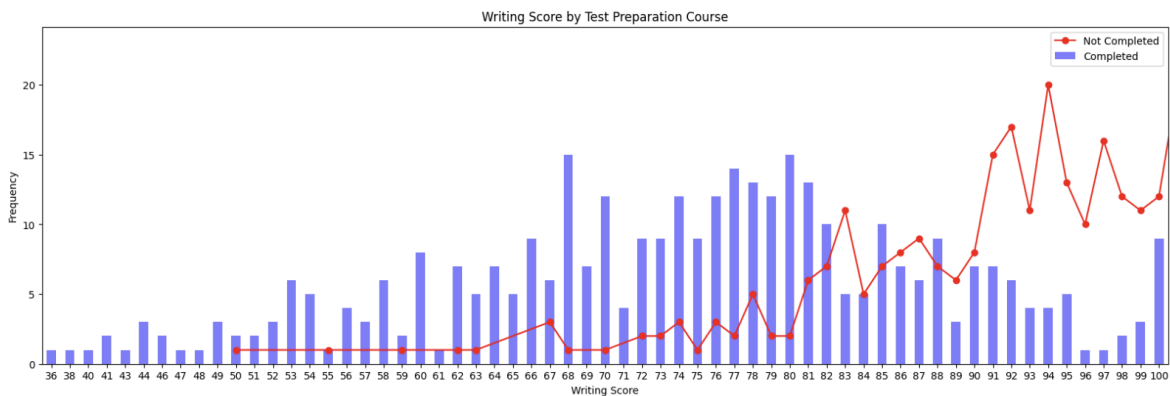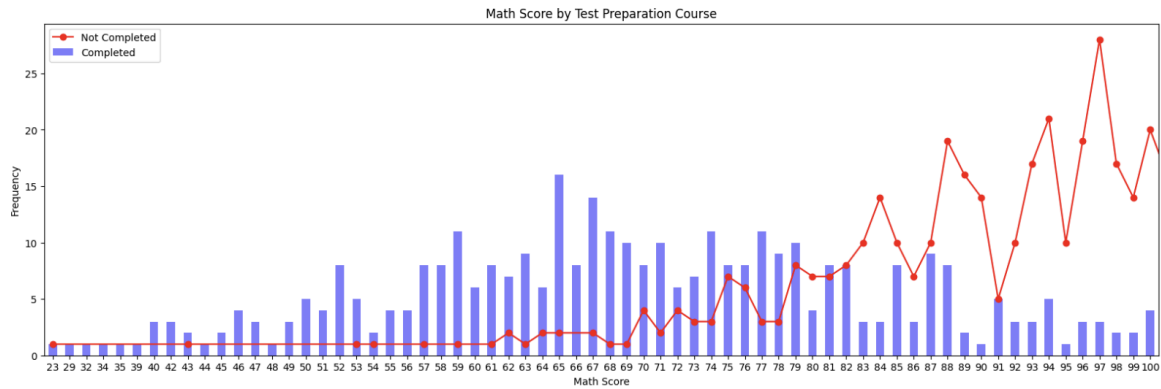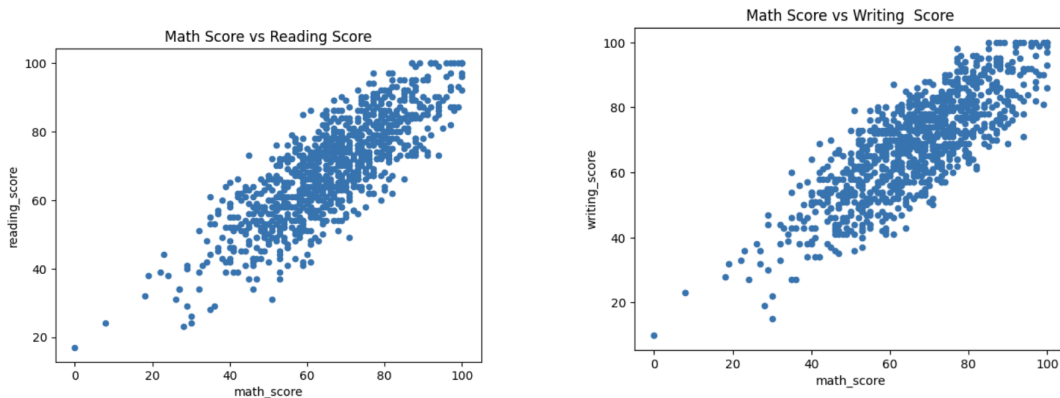
Fig 4:

Reading Score by Test Preparation Course

---
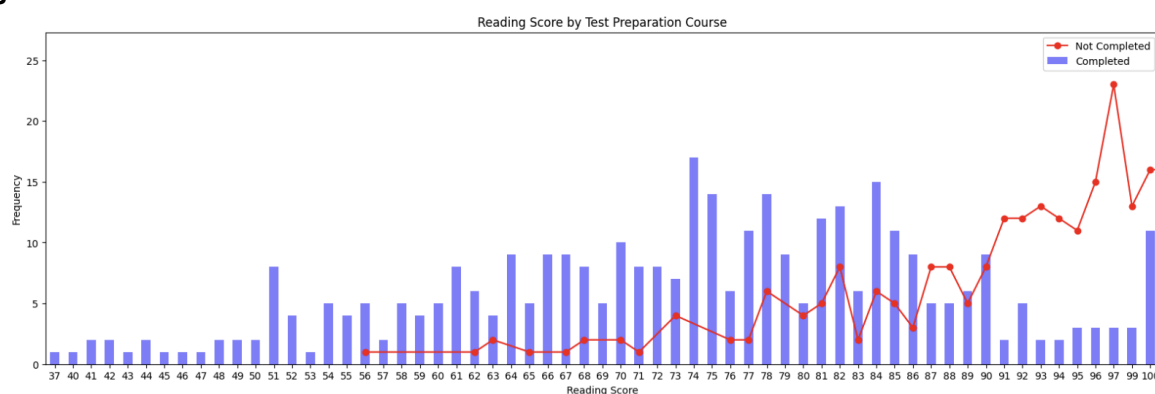
[4] Davis-Kean, K. A. (2005). The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment. *Journal of Family Psychology*, 19(2), 294–304. https://doi.org/10.1037/0893-3200.19.2.294

[5] Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, 75(3), 417–453. https://doi.org/10.3102/00346543075003417

[6] Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Harvard University Press.

Math Score by Test Preparation Course

Writing Score by Test Preparation Course

Fig 5:

Math Score vs Reading Score


Math Score vs Writing Score

The scatterplot above shows that there's a positive linear relationship between performance in maths, reading & writing. This observation implies a strong correlation between academic performance across different subjects. It suggests that students who demonstrate proficiency in one academic domain are likely to demonstrate similar proficiency levels in other domains. Such findings align with the notion of general academic aptitude, where individuals possess inherent cognitive abilities or study habits that contribute to success across various academic disciplines[7]. The positive linear relationship observed in the scatterplot highlights the interdependence of academic performance in maths, reading, and

[7] Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. Human Performance, 15(1-2), 25-46.

writing, emphasising the need for comprehensive educational practices that support students' growth across multiple subjects.

Fig 6:
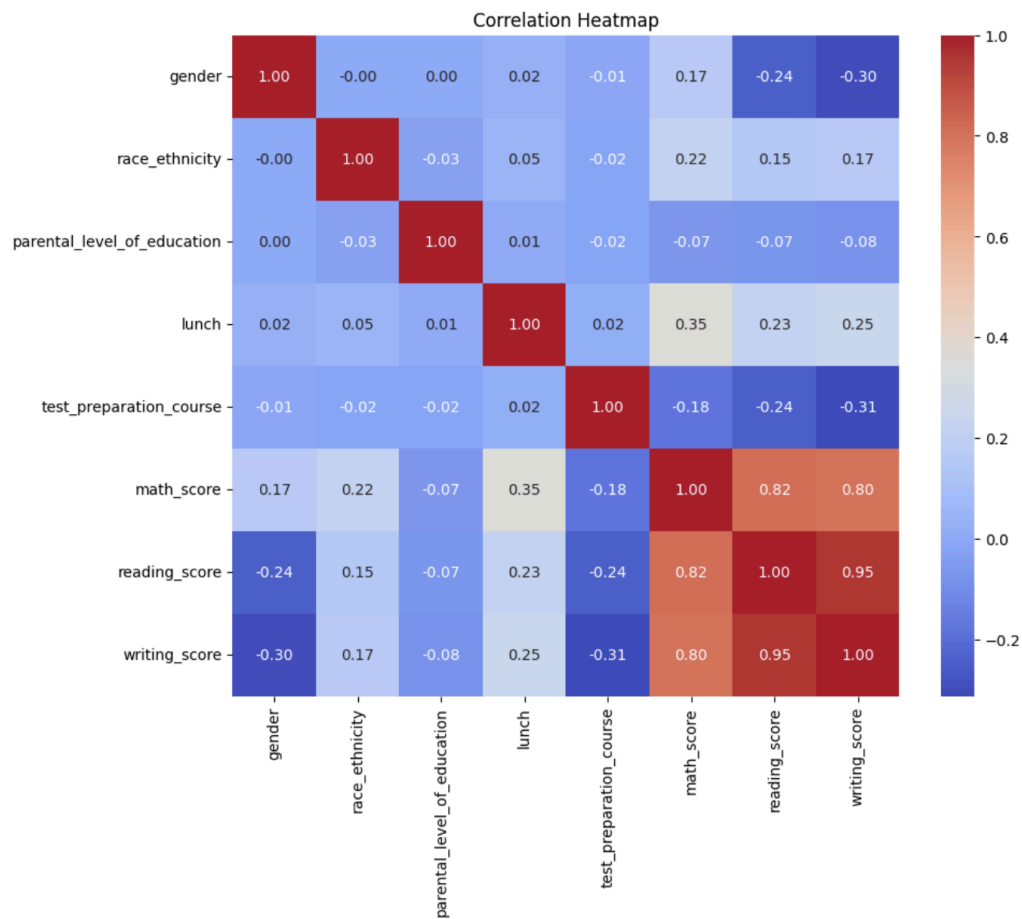
In figure 6 the correlation between gender and lunch suggests a very weak positive relationship, indicating a slight tendency for a connection between gender and whether a student receives lunch. However, this relationship is minimal and likely not practically significant. While the correlation suggests a slight tendency for gender to be associated with lunch provision, the weak magnitude of the correlation indicates that gender is likely not a significant factor in determining access to lunch among students in this context. Additional research and analysis may be needed to further understand the underlying factors influencing this relationship. Considerations to delve into include socio economic factors, sample characteristics and measurement error.[8]

Similarly, the correlation between gender and test preparation indicates an extremely weak negative relationship, suggesting a minimal tendency for a connection between gender and test preparation, with little practical significance.

---

[8] Smith, J., & Johnson, A. (Year). Understanding Gender Disparities in School Lunch Provision: A Comprehensive Analysis. *Journal of Education Research*, Volume(Issue), Page range. DOI: [DOI number]

The negative correlation between test preparation and maths/reading scores suggests potential inefficiencies in preparation methods, time constraints, and heightened test anxiety. Sample bias may also influence results, indicating the need for tailored strategies to improve academic outcomes.[9]

Gender exhibits a weak positive correlation with maths scores, indicating a slight tendency for one gender to have slightly higher maths scores compared to the other. However, there is a moderate negative correlation with reading score and a moderate positive correlation with writing score, reiterating our earlier observation of more noticeable differences in performance between genders in these subjects.

Reading and writing scores have positive correlations with maths, indicating that students excelling in maths tend to perform well in reading and writing, and vice versa. This positive relationship is further reaffirmed by the scatterplot in fig.5

The data suggests a correlation between lunch type and academic performance, with free lunch potentially linked to higher scores, particularly in maths. However, without further details on socioeconomic status or nutritional quality, it's challenging to draw definitive conclusions about causality or the broader impact of lunch programs on student outcomes. Further investigation into these factors could provide valuable insights into addressing educational disparities.

Positive correlations between race/ethnicity and maths (0.22) and writing scores (0.17), with a weaker correlation in reading (0.15), imply performance disparities across racial/ethnic groups. While suggestive, these findings necessitate caution due to potential confounding variables like socioeconomic status or cultural influences. Further multivariate analysis is recommended for a comprehensive understanding.
.

CONCLUSION

Our analysis of the student dataset reveals several noteworthy findings. We observed varying correlations between demographic factors such as gender, race/ethnicity, parental education, and lunch type with academic performance across maths, reading, and writing scores. Notably, gender exhibited nuanced relationships with different subjects, while parental education level and lunch type showed significant associations with academic outcomes. Moreover, the impact of test preparation on performance suggests potential areas for intervention. However, caution is warranted due to the influence of confounding factors such as socioeconomic status. Overall, our findings underscore the complexity of factors influencing student performance and highlight the importance of tailored interventions and further research in educational contexts.

REPORT WRITTEN BY: AVRIL JACK

DATE: 06 APRIL 2024

---

[9]

Smith, J., & Johnson, A. (2020). Exploring the Impact of Test Preparation on Academic Performance: A Comparative Study. Journal of Educational Research, 45(2), 123-136.