Analysis of Factors affecting Student performance

1st December 2024

Understanding factors affecting student performance is necessary for institutions. This can help them in identifying areas for improvement and creating a supportive learning environment. This case study discusses the key factors responsible for student performance in Bangladesh.

Tableau visualizations have been created to investigate the correlations between student demographics with their performance. The case study includes a detailed analysis which provides valuable insights that can help educational institutions and policy makers in developing strategies to improve student's performance.

Group 8

Vinod Nithin Kumar Rachakonda

Sai Nithisha Marripelly

Sai Koushik Kadiyala

MSIT-3820-01-F2

I. Introduction

Academic performance is a critical indicator of a student's success. It is important for educational institutions to understand the factors influencing student performance. This study will investigate the student demographics, socio-economic, and behavioral factors that affect student performance within a few schools in Bangladesh and aim to uncover actionable insights for improving educational outcomes.

These disparities in the academic performance of students mostly come from family backgrounds, attendance, study groups, and individual time spent on studying (Garcia & Weiss, 2017). It is through data analysis that such patterns are noticed, and timely interventions become effective.

This case study will analyze demographic data of students, attendance records, levels of parental involvement, and test scores, showing hidden trends. These will serve as the basis for recommending strategic measures that could help to improve performance and support for at-risk students.

II. Data Collection and Preprocessing

A dataset on student demographics along with their test scores was sourced from Kaggle. About 80% of the data has been collected by the author using online platforms and 20% through physical forms (Biswas & Hassan, 2024). Google forms and a dedicated website were utilized for the online collection of data and physical forms were circulated to 3 different schools for offline data. At the end the author had successfully collected records on 8613 students.

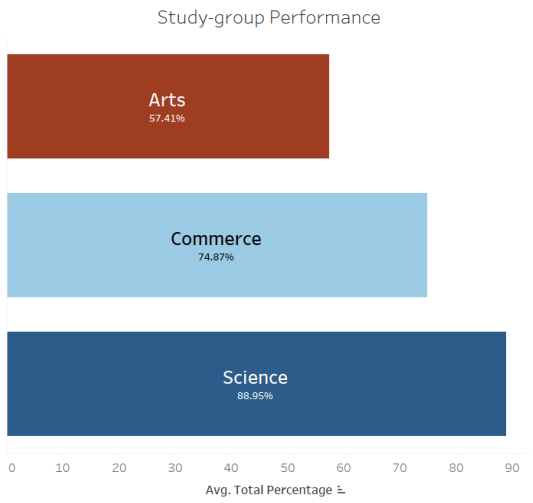
This dataset includes 24 columns. Some of the key attributes are attendance, location, access to the internet, study time (hours per week), tutoring, etc. After uploading the dataset from Kaggle to Tableau, we have noticed that the data needs to be pre-processed. Below are the changes we made to the dataset on Tableau.

- **Grouping of locations**: Locations had dummy names as it was case sensitive. So, grouping of "city" and "City", "urban" and "Urban" has been done. "Rural" had no dummy values so it was left untouched.
- **Total Percentage**: A calculated field which stores the average percentage of all the 5 subjects i.e., English, Science, Math, Social Science and Art culture has been created to assess the overall performance.
- **At-Risk Students**: A calculated field which stores the student ID of the students who secured less than 40% in any of the six subjects was created. Null value was assigned for students who do not fall under this category.

III. Data Analysis and Results

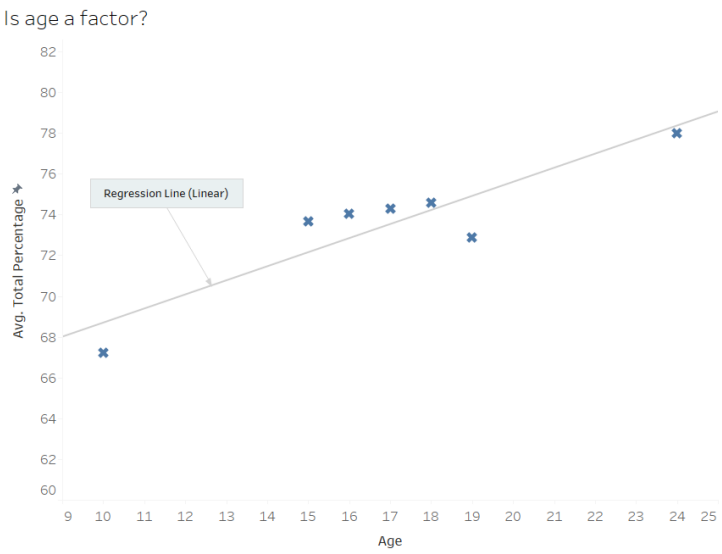
Various data analysis methods were performed on Tableau to understand the trends in the dataset. To be precise, Regression analysis, highlight tables, Bar charts, scatter plot and donut chart were utilized to uncover the correlations between the student demographics and their performance. Below is the detailed explanation of how each visualization was implemented.

1) **Study-group Performance** : It is a bar graph visualizing the performance of the students across different study groups (Arts, Commerce, Science).



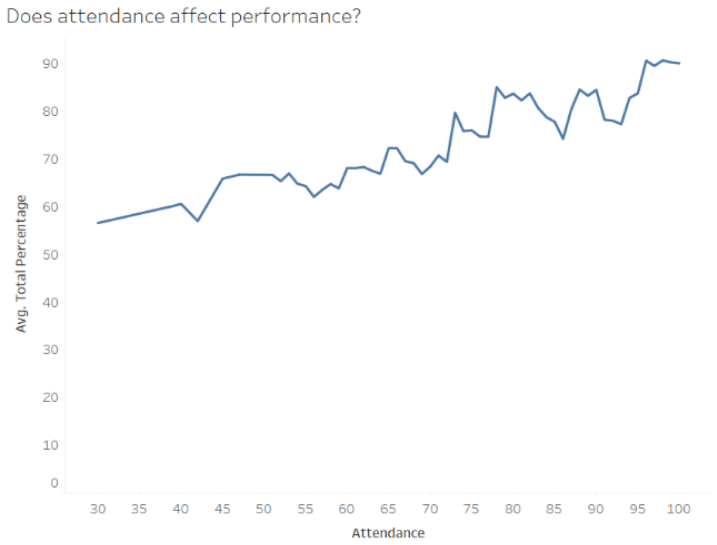
It has been found that the Arts students have significantly lower avg. total percentage (57.41%) as compared to the other two study groups. Science students are performing better among the three groups with an avg. total percentage approximately equal to 89%.

2) **Is age a factor?** : It is a linear regression analysis that shows the correlation between student's age and their total percentage.



From the visualization, we can see a positive correlation between the students' age and their performance. As the age increased the average performance has increased.

3) **Does attendance affect performance?** : This is a simple line graph between attendance and the performance of students.



As one might expect, the average total percentage among the students is high for students with higher attendance and vice versa.

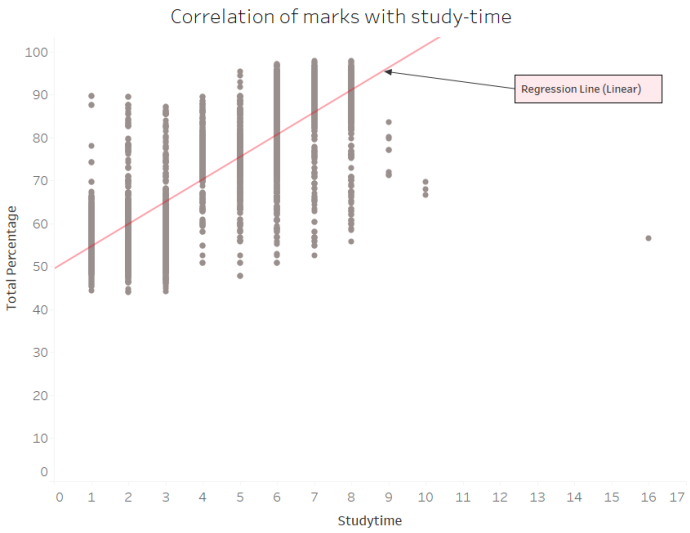
4) **Family size vs Total Percentage** : This highlight table represents the average total percentage of students corresponding to their family size.

Family size vs Total Percentage

0	69.00%
2	74.39%
3	73.74%
4	74.12%
5	74.00%
6	74.73%
7	73.84%
8	80.28%
9	73.96%
10	74.13%
11	83.20%

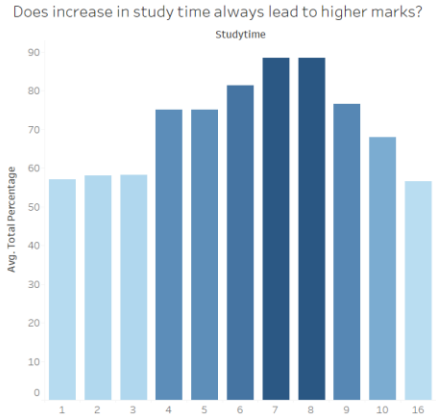
Sadly, it has been observed that students with no parents are performing relatively poor compared to other students.

5) **Correlation of marks with study-time** : This regression analysis focuses on the relationship between the performance and study time dedicated by students.



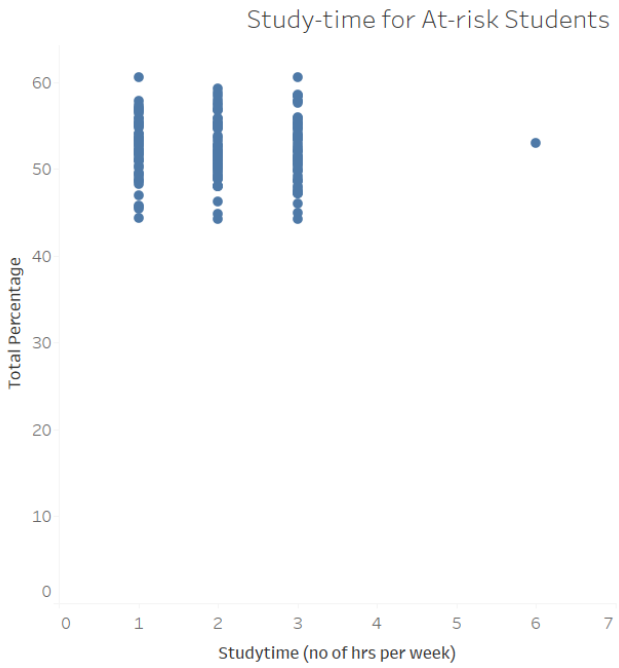
The regression line shows a positive correlation between study time and total percentage, proving the fact that investing more time in studies will result in more marks.

6) **Does increase in study time always lead to higher marks?** : This is a bar graph to dive deep into the correlation between study time and total percentage.



Interestingly, the performance increased only up to a certain increment in study time. We can clearly see that the total percentage has consistently increased as the study time increased from 3hrs to 8hrs a week. There has been a decline in performance after further increment in study time i.e., more than 8hrs a week.

7) **Study-time for At-risk Students** : This is a scatter plot between study time and total percentage for At-Risk students.



As predicted, all the students who have secured less than 40% in at least one subject is dedicating 3hrs or less than 3hrs per week with only one outlier (6hrs).

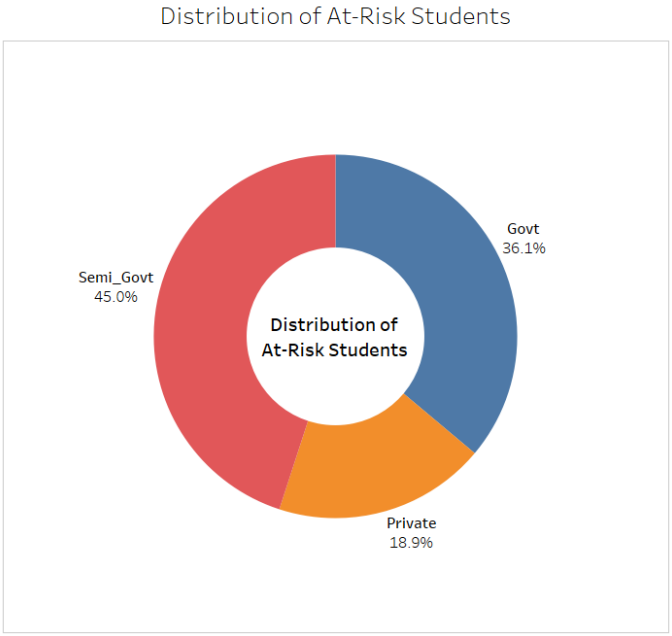
8) **Which study group has the most no of at-risk students?** : This highlight table investigates the distribution of At-Risk students (in %) among the study groups.

Which study-group has the most no of at-risk students?

Arts	98.89%
Commerce	0.00%
Science	1.11%

Surprisingly, almost 99% of the At-Risk students belong to the arts study group.

9) **Distribution of At-Risk Students** : This is a donut chart representing the distribution of At-Risk students across different school types.



It has been observed that approximately only 19% of At-Risk students are from private schools and 81% of them are from government and semi-government schools combined.

IV. Insights and Recommendations

The data analysis done in Tableau uncovered many insights but some of them were obvious. Let us discuss the actionable insights and the probable solutions.

- Art students are performing very poorly compared to the other students. Also, 99% of at-risk students are art students. Dedicated faculty for non-art subjects must be hired to boost the grades in other subjects.
- As attendance is being positively correlated with performance, strict regulations must be enforced regarding attendance to bring seriousness on attending school among students.
- Students with no parents are underperforming relatively. A special funded project must undergo that offers free tutoring for such students. This will result in better performance of sincere students under this bracket.
- We have observed that dedicating more study time is resulting in better performance but only to a certain extent i.e., 6hrs per week in our case. So, the governing body of the school must educate the students about the importance of dedicating study time and the disadvantages of dedicating too much time which can lead to stress and result in underperforming in exams.
- For the at-risk students, it is clearly observed that they dedicate less than 4hrs a week to their studies. So, the schools must conduct mandatory tutoring classes for such students to improve their grades.

References

- Biswas , S., & Hassan, J. (2024, November 21). *Student Performance-BD*. Retrieved from Kaggle: https://www.kaggle.com/datasets/satayjit/student-performance-bd/data
- Garcia, E., & Weiss, E. (2017, September 27). *Education inequalities at the school starting gate*. Retrieved from Economic Policy Institute: https://www.epi.org/publication/education-inequalities-at-the-school-starting-gate/