GROUP 2

PREDICTING & ANALYSING STUDENT PERFORMANCE AT EASTMINSTER USING REGRESSION MODELS

SHAHLY FAYEEK

BASILU MUNAWEERA

SAADIQ ASHFAAQ

PURPOSE

- Understand and analyse the various factors influencing students' exam scores.

- Provide actionable insights for students, educators, and policymakers to enhance academic outcomes.

PROBLEM STATEMENT AND BACKGROUND

Exam performance varies widely among students, influenced by multiple factors. Lack of clarity on which factors are most significant hinders targeted interventions.

Significance:
Improving academic performance is critical for personal growth, career prospects, and societal advancement. Identifying these factors can help optimize learning environments and resource allocation.

Background:
Previous studies suggest links between study habits, socioeconomic background, and exam results. However, gaps remain in understanding how these factors interact and their relative importance.

GAP ANALYSIS

Existing Solutions

- General strategies like improving teaching methods, increasing study time, and offering additional resources.
- Broad recommendations without detailed personalisation.

Limitations of Current Understanding

- Limited focus on how personal, environmental, and socioeconomic factors interact.
- Insufficient data-driven models to quantify their relative impact.

Proposed Solution

- Develop a comprehensive analysis using statistical and machine learning models.
- Identify and rank factors such as study habits, parental involvement, and resource access.

OBJECTIVES AND RESEARCH QUESTIONS

Main Objective

To identify and analyze the factors that significantly impact students' exam scores.

Specific Objectives:

- Investigate the influence of personal habits, such as study hours and sleep patterns.
- Examine environmental factors, including access to resources and parental involvement.
- Analyse socioeconomic and demographic factors like income, school type, and gender.

Research Questions

- What factors impact students’ exam scores, and how significant is their influence?
- How does access to resources (e.g., internet, tutoring) affect exam scores?
- Are there notable differences in exam performance based on demographic factors, such as gender or school type?
- How does the socioeconomic status of students’ families impact their academic results?

METHODOLOGY

Data Source: Acquired from Kaggle: Student Performance Factors Dataset. It's a secondary dataset which has already recorded data

Data Collection: Contains variables such as study habits, sleep hours, attendance, access to resources, parental involvement, and more.

Data Preprocessing: Cleaned the dataset by removing columns with missing values.
Split the data into train model (70%) and test model (30%) sets for analysis.

Techniques and Tools: Utilised statistical and machine learning models, including stepwise regression and linear modelling.

Model Building: Examined multiple models (forward selection, backward elimination, stepwise regression). Final model identifies significant predictors of exam scores.

ANALYSIS & FINDINGS

MODEL PERFORMANCE

```
Residual standard error: 2.261 on 4613 degrees of freedom
Multiple R-squared:  0.6662,    Adjusted R-squared:  0.6654
F-statistic: 837.1 on 11 and 4613 DF,  p-value: < 2.2e-16
```

- **Residual standard error:** Measures the average deviation of observed values from predicted values
- **Multiple R-squared:** Measures the proportion of variance in the dependent variable explained by the independent variables
- **F-statistic:** Measures the overall significance of the regression model

ANALYSIS & FINDINGS

Insights

Significant Predictors

- Positive Impacts:
  - Attendance: Each unit increase adds 0.20 points to the score.
  - Hours Studied: Each additional hour adds 0.29 points.
  - High Parental Involvement: Adds 2.02 points to the score.
  - High Access to Resources: Adds 2.10 points.
  - Tutoring Sessions: Adds 0.53 points per session.

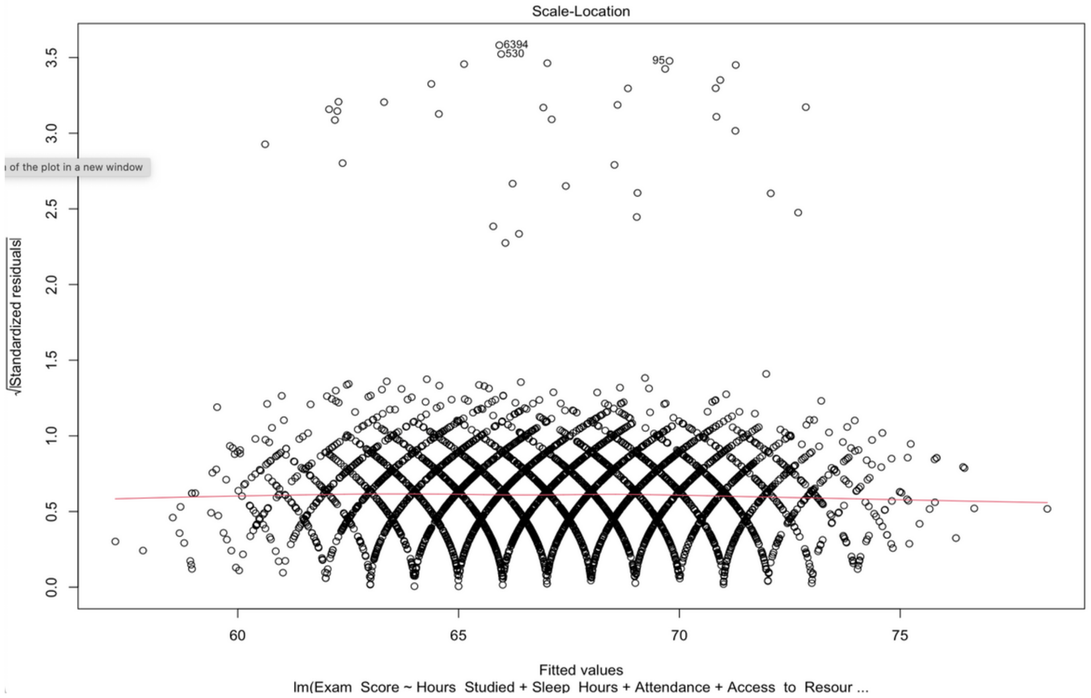
- Negative Impact:
  - Lack of Extracurricular Activities: Decreases score by 0.57 points.

Coefficients:

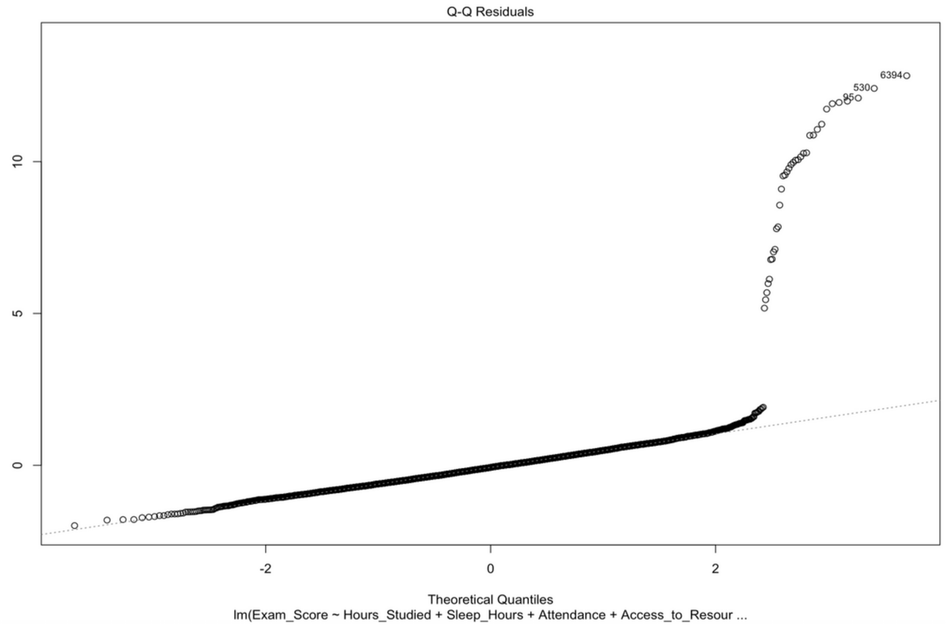
	Estimate
(Intercept)	37.774897
Attendance	0.200184
Hours_Studied	0.291041
Previous_Scores	0.047210
Tutoring_Sessions	0.526581
Parental_InvolvementHigh	2.024660
Access_to_ResourcesHigh	2.098006
Access_to_ResourcesMedium	1.132335
Parental_InvolvementMedium	0.880272
Extracurricular_ActivitiesNo	-0.574145
Internet_AccessYes	0.936959
Physical_Activity	0.152363

ANALYSIS & FINDINGS

Visualisations



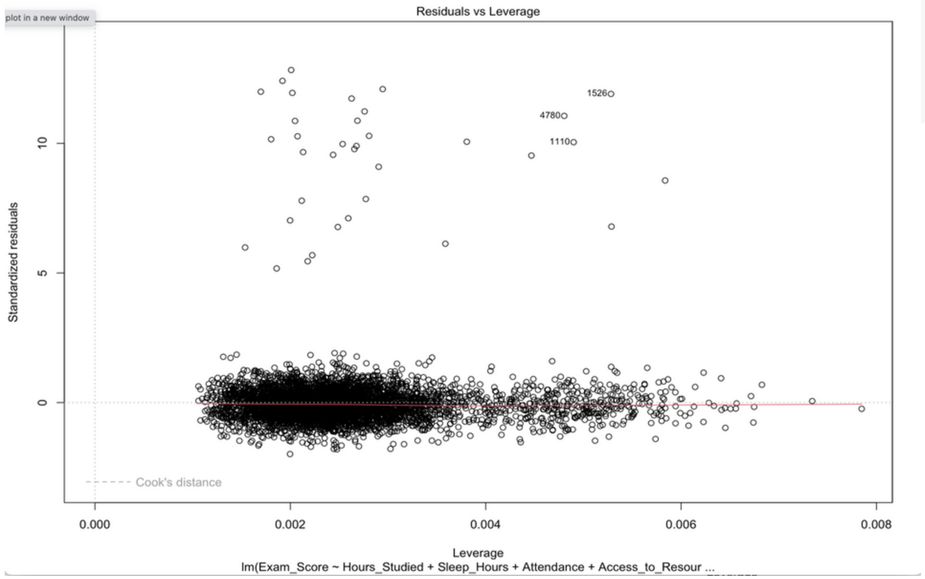
- This plot examines the homoscedasticity (constant variance) assumption. Residuals should spread evenly along the horizontal line. A cone-like shape indicates heteroscedasticity.

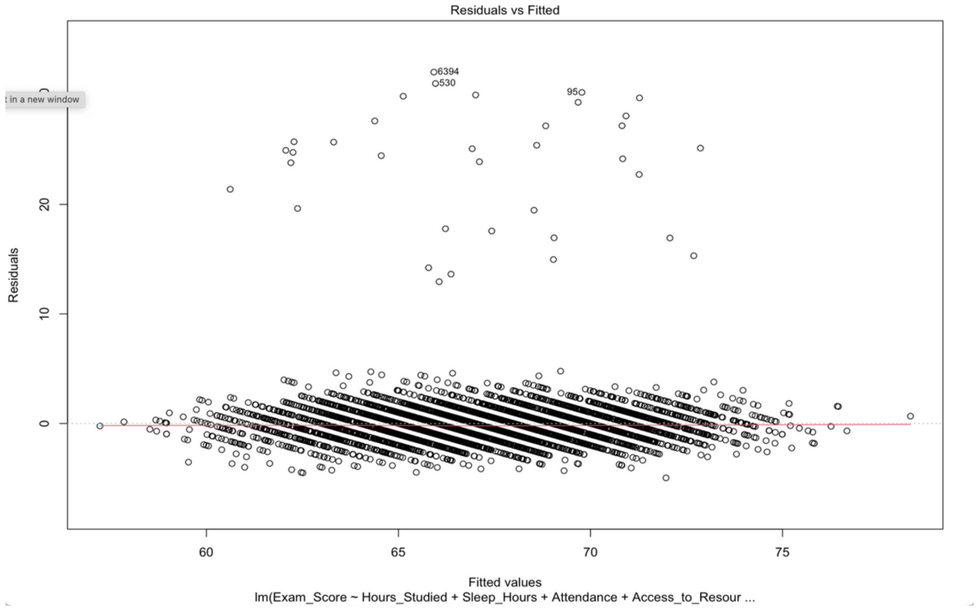


- This plot assesses the normality of residuals. The residuals should lie close to the diagonal line. Deviations suggest non-normality, which may affect statistical inference.

ANALYSIS & FINDINGS

Visualisations





- This plot identifies influential observations. Points with high leverage and large residuals may disproportionately affect the model. Look for Cook's distance to identify outliers.

- This plot checks the assumption of linearity. It shows the residuals (errors) against the fitted values. Ideally, the points should be randomly scattered around the horizontal line, indicating a good fit.

RECOMMENDATION

Personal Level

- Encourage effective time management and consistent study schedules.
- Promote healthy sleep habits to optimise cognitive function.

Parental Involvement

- Increase awareness of the positive impact of active parental engagement.
- Provide resources for parents to support their children's education effectively.

Educational Institutions

- Enhance access to resources like tutoring and internet facilities.
- Focus on improving teacher quality and personalised learning strategies.
- Reduce the allocated time for extracurricular activities.

Policy Recommendations

- Develop targeted interventions for underperforming groups based on socioeconomic factors.
- Advocate for public funding to reduce resource disparities between public and private schools.

CHALLENGES & LIMITATIONS

- **Data Collection Issues:** Incomplete or inconsistent data can affect the accuracy of the analysis.
- **Model Limitations:** Some models may not perform well with certain types of data.
- **Resource Constraints:** Limited access to high-quality resources can impact the study.
- **External Factors:** Uncontrollable variables such as socio-economic conditions can influence results.

IMPACT & FUTURE WORK

Impact

- **Educational Insights:** Provides valuable insights into factors affecting student performance.
- **Policy Recommendations:** Can inform educational policies and resource allocation.
- **Improved Interventions:** Helps in designing targeted interventions for students.

Future Work

- **Expanded Data Collection:** Collecting more diverse and comprehensive data.
- **Advanced Modelling Techniques:** Exploring more sophisticated models to improve accuracy.
- **Longitudinal Studies:** Conducting studies over a longer period to observe trends and changes.

CONCLUSION

Summary:

- We explored factors influencing student performance at Eastminster.
- Identified key predictors impacting academic outcomes.
- Developed regression models to analyze and predict exam performance.

GROUP 2

THANK YOU.

SHAHLY FAYEEK

BASILU MUNAWEERA

SAADIQ ASHFAAQ