*Student Performance Regression Analysis Report*

**Introduction**

The purpose of this analysis is to explore how various factors influence students' exam performance. Using a multiple linear regression model, we have examined how a range of personal, social, and academic factors predict the final exam scores of students. This report will present and interpret the results from the regression model, highlighting the most significant predictors of exam scores.

**Data Overview**

[**https://www.kaggle.com/datasets/lainguyn123/student-performance-factors/data**](https://www.kaggle.com/datasets/lainguyn123/student-performance-factors/data)

The dataset used in this analysis contains the following key variables:

Independent Variables (Factors):

* **Peer\_Influence (Neutral, Positive)**
* **Hours\_Studied**
* **Attendance**
* **Parental\_Involvement**
* **Access\_to\_Resources**
* **Extracurricular\_Activities**
* **Sleep\_Hours**
* **Previous\_Scores**
* **Motivation\_Level**
* **Internet\_Access**
* **Tutoring\_Sessions**
* **Family\_Income**
* **Teacher\_Quality**
* **School\_Type**
* **Physical\_Activity**
* **Learning\_Disabilities**
* **Parental\_Education\_Level**
* **Distance\_from\_Home**
* **Gender**

Dependent Variable:

* **Exam Score**

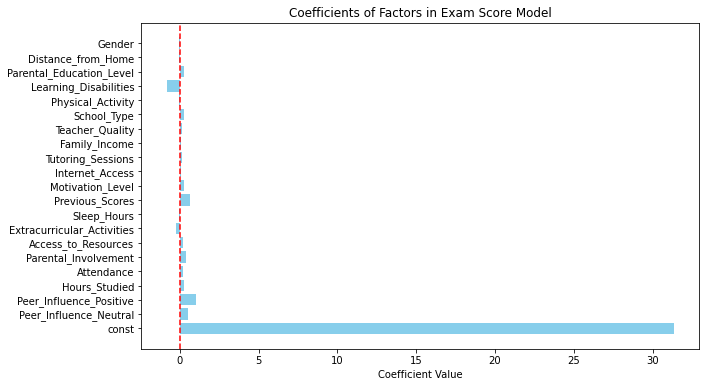
**Regression Results Summary**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Features | Coefficient | Standard Error | t-value | p-value | Confidence Interval Lower | Confidence Interval Upper |
| const | 31.33102861 | 0.358459869 | 87.40456405 | 0 | 30.62833103 | 32.03372618 |
| Peer\_Influence\_Neutral | 0.519644856 | 0.067930196 | 7.649688805 | 2.30265E-14 | 0.386479647 | 0.652810066 |
| Peer\_Influence\_Positive | 1.027918917 | 0.067716016 | 15.17984926 | 3.50694E-51 | 0.895173569 | 1.160664266 |
| Hours\_Studied | 0.294843219 | 0.004183816 | 70.47231988 | 0 | 0.286641583 | 0.303044855 |
| Attendance | 0.19885916 | 0.002173474 | 91.49370794 | 0 | 0.194598447 | 0.203119873 |
| Parental\_Involvement | 0.996539016 | 0.036071963 | 27.62641477 | 4.7097E-159 | 0.925826272 | 1.06725176 |
| Access\_to\_Resources | 1.025704518 | 0.035904125 | 28.56787399 | 2.5227E-169 | 0.955320791 | 1.096088245 |
| Extracurricular\_Activities | 0.555258188 | 0.051090168 | 10.8682004 | 2.78205E-27 | 0.455104892 | 0.655411484 |
| Sleep\_Hours | -0.001639266 | 0.017076109 | -0.095997607 | 0.923525394 | -0.035113976 | 0.031835444 |
| Previous\_Scores | 0.048886313 | 0.001742183 | 28.06038347 | 9.3753E-164 | 0.04547107 | 0.052301556 |
| Motivation\_Level | 0.529696731 | 0.036019519 | 14.70582472 | 3.41556E-48 | 0.459086795 | 0.600306668 |
| Internet\_Access | 0.927612946 | 0.094846865 | 9.780111792 | 1.94718E-22 | 0.741682336 | 1.113543555 |
| Tutoring\_Sessions | 0.497201611 | 0.020364116 | 24.41557572 | 4.1079E-126 | 0.457281341 | 0.537121881 |
| Family\_Income | 0.531875039 | 0.033757478 | 15.75576943 | 6.25639E-55 | 0.465699437 | 0.598050642 |
| Teacher\_Quality | 0.532933327 | 0.042012426 | 12.68513566 | 1.90327E-36 | 0.450575349 | 0.615291305 |
| School\_Type | -0.032199718 | 0.054494223 | -0.590883139 | 0.554619001 | -0.139026064 | 0.074626629 |
| Physical\_Activity | 0.187441142 | 0.024321231 | 7.706893614 | 1.47931E-14 | 0.139763643 | 0.235118642 |
| Learning\_Disabilities | -0.852906283 | 0.081691737 | -10.44054536 | 2.54967E-25 | -1.013048575 | -0.69276399 |
| Parental\_Education\_Level | 0.488105358 | 0.032187804 | 15.16429495 | 4.40944E-51 | 0.425006825 | 0.551203891 |
| Distance\_from\_Home | -0.477405608 | 0.037495633 | -12.73229895 | 1.0565E-36 | -0.550909206 | -0.403902009 |
| Gender | -0.040822626 | 0.050718714 | -0.804882913 | 0.420916304 | -0.140247752 | 0.058602499 |

The following sections detail the relationship between each independent variable and the dependent variable (exam score), based on the coefficients, t-values, p-values, and confidence intervals provided in the regression summary.

Regression Summary Table

|  |  |  |  |
| --- | --- | --- | --- |
| **R-squared** | **Adjusted R-squared** | **F-statistic** | **Prob (F-statistic)** |
| 0.727248898 | 0.726420622 | 878.0278438 | 0 |

****

**Figure 1: Bar Plots for the coefficients**

**A green line with numbers

Description automatically generated**

**Figure 2: Feature Importance for the coefficients**

**Constant (Intercept)**

* **Coefficient: 31.33**.
* This is the intercept term, representing the expected exam score when all other predictors (factors) are zero. It acts as a baseline score.
* Interpretation: When none of the factors have any effect (i.e., all are at zero), the baseline exam score is approximately 31.33.

**Peer Influence (Neutral, Positive)**

**Neutral:**

* **Coefficient: 0.52**.
* This represents the effect of Neutral peer influence compared to the baseline category (probably Negative peer influence). A coefficient of 0.52 means that students with neutral peer influence score 0.52 points higher than those with negative peer influence.
* **Significant**: Yes, the p-value is near zero.
* Insight: Neutral peer influence has a positive but moderate impact on exam scores.

**Positive:**

* **Coefficient: 1.03**.
* This indicates that students with positive peer influence score 1.03 points higher than those with negative peer influence (the baseline).
* **Significant**: Yes.
* Insight: Positive peer influence has a more substantial positive effect on exam scores compared to neutral or negative influence.

**Hours\_Studied:**

* Coefficient: 0.29.
* For every additional hour studied per week, the exam score is expected to increase by 0.29 points.
* Significant: Yes.
* Insight: Hours studied have a significant positive impact on performance, reinforcing that more study time leads to better results.

**Attendance:**

* Coefficient: 0.198.
* Every 1% increase in attendance correlates with a 0.198-point increase in exam score.
* Significant: Yes.
* Insight: Higher attendance is associated with better academic performance.

**Parental\_Involvement:**

* Coefficient: 0.99.
* Higher levels of parental involvement positively impact student exam scores. Each increase in the level of parental involvement leads to a 0.99-point increase.
* Significant: Yes.
* Insight: Parental involvement is a meaningful contributor to better exam performance.

**Access\_to\_Resources:**

* Coefficient: 1.03.
* Greater access to educational resources increases exam scores by 1.03 points.
* Significant: Yes.
* Insight: Students with better access to resources tend to perform better in exams, although the effect is modest.

**Extracurricular\_Activities:**

* Coefficient: 0.56.
* Participation in extracurricular activities is associated with an increase in exam scores (by 0.56 points).
* Significant: Yes.
* Insight: This suggest that students who engage in extracurricular activities are exposed to activities that helps them become better students. Possible reasons includes sofskilss development, social benefits, cognitive benefits, academic reinforcements etc.

**Sleep\_Hours:**

* Coefficient: -0.002.
* The effect of additional sleep hours is negative. Each additional hour of sleep leads to a -0.002-point reduction in exam score.
* Significant: No.
* Insight: The small negative effect suggests that more sleep might slightly reduce performance, but the result is not statistically significant. This could mean there’s an optimal range for sleep, beyond which it might not help.

**Previous\_Scores:**

* Coefficient: 0.05.
* For every 1-point increase in previous scores, the current exam score is expected to rise by 0.05 points.
* Significant: Yes.
* Insight: Students with higher past exam performance tend to do well in current exams.

**Motivation\_Level:**

* Coefficient: 0.53.
* Higher motivation levels contribute to a 0.52-point increase in exam scores.
* Significant: Yes.
* Insight: Motivation positively impacts student performance.

**Internet\_Access:**

* Coefficient: 0.93.
* Internet access has a high positive effect on exam scores.
* Significant: Yes.
* Insight: Access to the internet is a critical factor for improving academic performance, as it helps the students to gain access to educational materials on the internet that can help the students perform better.

**Tutoring\_Sessions:**

* Coefficient: 0.49.
* Each additional tutoring session attended increases the exam score by 0.49 points.
* Significant: Yes.
* Insight: Tutoring helps improve academic performance.

**Family\_Income:**

* Coefficient: 0.53.
* Higher family income leads to an increase in exam scores (by 0.53 points).
* Significant: Yes
* Insight: While family income has a positive effect, it’s not a strong predictor of exam success in this dataset.

**Teacher\_Quality:**

* Coefficient: 0.53.
* Better teacher quality leads to a 0.53-point increase in exam scores.
* Significant: Yes.
* Insight: Teacher quality is a very important factor in exam performance.

**Physical\_Activity:**

* Coefficient: 0.19.
* Physical activity has a very small and positive effect on exam scores.
* Significant: Yes.
* Insight: This suggests that physical activity has a significant direct impact on academic performance.

**Learning\_Disabilities:**

* Coefficient: -0.85.
* Having a learning disability is associated with an 0.83-point decrease in exam scores.
* Significant: Yes.
* Insight: Learning disabilities negatively impact exam performance significantly, highlighting the need for additional support for these students.

**Parental\_Education\_Level:**

* Coefficient: 0.48.
* Higher parental education levels increase student exam scores by 0.48 points.
* Significant: Yes.
* Insight: Students whose parents have higher education levels tend to perform better, indicating parental education influences academic success.

**Distance\_from\_Home:**

* Coefficient: -0.58.
* A greater distance from home to school reduces exam scores.
* Significant: Yes.
* Insight: The impact of distance from home is not negligible in predicting exam scores, as this could also affect the motivation of the student, and if the the parents income are low, this could in turn affect the attendance of the students, as they might not be able to make it to school on time, and sometimes they might not make it to school at all during harder times.

**Gender (Male):**

* Coefficient: -0.04.
* Male students score slightly lower than female students (the baseline), but the difference is very small and not statistically significant.
* Significant: No.
* Insight: Gender does not seem to be a major factor in predicting exam scores.

**Conclusion and Insights**

From this analysis, several key factors have a significant impact on students' exam scores:

* **Peer\_Influence\_Positive, Access\_to\_Resources, Parental\_Involvement, Internet\_Access** are the strongest positive predictors of exam success.
* **Learning Disabilities** and **Distance\_from\_Home** negatively affect student performance, suggesting areas where additional support may be necessary.
* **Family Income**, **Access\_to\_Resources**, and **School Type** indicate socioeconomic factors that contribute to disparities in exam performance.

These results provide actionable insights into areas that can be targeted for improvement, particularly for students at risk of lower academic performance.