Stat 512 Project Cover Page

1. Project Topic: Factors Impacting Student Performance: An Applied Regression Analysis

2. Group number: 15

3. List of group members: Anish Tiwari, Isabel Turner, Eshita Vani, Shreya Vasant

4. Project YouTube link: https://www.youtube.com/watch?v=z_9qPAQxA7Q

5. Project background introduction (why this is an important question, what has been done on the question, what are your major research questions in the project, etc.):

There are many different factors that impact student's performance at school. Performance at school is typically defined as the test scores or the grade accumulated. Some domains that play a role impacting performance include sleep habits, study habits, extracurricular activities, and previous scores. Prior studies have found individual relationships between these factors and performance, but their combined effects are less understood. This project aims to understand those relationships of combined effects. Our major research questions are as follows:

1. Research question 1: How does time spent outside of studying impact a student's exam performance?
2. Research question 2: does prior knowledge impact future exam performance the same amount as current effort?
3. Research question 3: Does sleep or study hours have the same effect on performance, or not (i.e which healthy habit is valuable)
4. Research question 4: Are extra-curriculars impactful in improving student performance?

6. Project result highlight (what are the major findings of your project, what do you consider the most contribution of this project):

Major findings of this project are that all the factors selected (sleep habits, study habits, extracurricular activities, and previous scores) contribute significantly to a student's performance. We found out that as people often have a common notion that students' involvement in extra-curriculars could either boost their performance or degrade it, our analysis showed that the interaction of study hours and extra-curriculars have no impact on the student performance. The most valuable contribution of this project would be understanding how individual factors, and their interaction effects can work together to offer valuable insights into academic outcomes and give rise to a holistic predictive model which many education institutes lack today.

7. Project data introduction (the exact data resource, a table summarizes variable notation and definition, such as the one on the first page in the homework).

This study was an exploratory observational study. The selected dataset, entitled "Student Performance Dataset" is a set of data including 10,000 student records meant to examine the impact of several different parameters on student test performance extracted from Kaggle.

https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression

Table 1. Variables utilized for analysis of student performance

Variable Identifier	Category	Name	Description
Y	Response Variable	Performance Index	Student academic performance expressed as a range between 10 and 100 with higher numbers indicating better performance
X1	Continuous Explanatory Variable	Hours Studied	Total hours studied by a student in preparation for test
X2	Continuous Explanatory Variable	Previous Scores	Average score by student on previous tests
X3	Categorical Explanatory Variable	Extracurricular Activities	Whether or not a student participates in extracurricular activities (Yes or No)
X4	Continuous Explanatory Variable	Sleep Hours	Average number of hours of sleep per day for a student
X5	Continuous Explanatory Variable	Sample Question Papers Practiced	Number of sample question papers practiced in preparation for test

8. Briefly describe what you learn from the project and what is the most challenging part.

We learnt how to apply multiple linear regression techniques to real world data, validating model assumption using diagnostic checks, interpreting effect of interaction terms and individual variable in the model and using best subset procedures in model selection to improve the model fit. The most challenging part of this project was how to interpret interaction effect results and translate statistical conclusions into real world conclusions offering valuable insights.

9. In one sentence, what is your advice for the future student to deliver a high-quality project in the course.

Build a well-validated model early and spend more time exploring interactions, which will help your data tell a great story.