Data Visualization

Topic:
Student's Name: Logan Eichbaum

School of Graduate Professional Studies
DAAN 871 - Data Visualization

Semester: Fall
Year: 2025

DOCUMENT CONTROL

Work carried out by:

Name	Email Address	Task description
Logan Eichbaum	lke5200@psu.edu	

REVISION SHEET

Release No.	Date	Revision Description
1	10/26/2025	Created Research Goal related to Student Performance. Outlined motivation for paper. Presented four different queries to conduct further research into performance factors in high school. Cited source of data and source of claim made.
2	11/02/2025	Added Data Source. Described data. Added features in table that describes what factors were included in the dataset. Identified and explained measurement scales in table.
3	11/09/2025	Drafted purpose of dashboard, including target audience, what data will drive the dashboard, and what questions the dashboard will answer.
4	11/16/2025	Added Design Principles, Visualization Techniques, and Interactivity sections. Included Sample Dashboard for reference. Added Few as citation source.
5	11/30/2025	Complete Report with screenshots of dashboard use cases.

TABLE OF CONTENTS

Contents

PROJECT OVERVIEW

RESEARCH GOAL: What are the most important factors in determining how well a child will do in High School?

MOTIVATION: Children are often told that if they work hard enough, then they can achieve anything they set their mind to. While this should be true, we see time and time again that work ethic is not the only factor in a child's success. For example, according to Eric Hanushek, "schools bring little influence to bear on a child's achievement that is independent of his background and general social context" (Hanushek, 3). This suggests that success is dependent on factors not chosen by students, such as home life, economic status, and more. This project is going to analyze the most important factors in determining success for students in school. I will create a dashboard that allows the user to filter through different variables that determine how much of an influence those variables have over success in high school.

QUERIES:
1. What is the most important factor in determining success in high school children?
2. What demographic factors influence a child's performance (gender, age)?
3. Does time spent studying play any large role independent of other social factors? How much of the child's success are they in control of?

DATA SOURCE:

The majority of my analysis will be done using a dataset about student performance from the UCI public data repository. It is a popular dataset with 30 different features that analyzes student data from two different Portuguese public schools in 2014. Using this dataset, I will create a logistic classification model where I can unravel the most important factors in a successful education in high school. Using this model I can guide users through an interactive dashboard, where they can discover different effects that certain variables have on student performance.

DATA COLLECTION

The student data from the UCI data repository contains 30 different features, with 3 grade columns. The table extracted contains data from 2014 pertaining to Portuguese students that went to two different public schools. It attempts to use a variety of factors to explain the performance of students in three of their periods. This table will be used to model a logistic regression as well as extract meaningful distributions about students during this time.

DATA DESCRIPTION:
Number of Instances: 650
Missing Values: None

Attribute	Data Type	Decription	Example value	Measurement Scale
school	text	School student attended	GP	Nominal
sex	factor	M–Male, F– Female	F	Nominal
age	numeric	Age of student	15	Ratio
address	text	Student's home address type	U	Nominal
famsize	binary	Family size: Less then or equal to 3, or greater than 3	LE3	Nominal
pstatus	binary	Parent's living either together (T) or apart (A)	A	Nominal
medu	factor	Mother's education 0-4 0: none – 4: higher education	3	Ordinal
fedu	factor	Father's eduction 0-4 0: none – 4: higher education	1	Ordinal
mjob	text	Mother's job	reputation	Nominal
fjob	text	Father's job	services	Nominal
reason	text	Reason for choosing school	home	Nominal
guardian	text	Primary guardian	mother	Nominal
traveltime	numeric	Home to school travel time, 1-4 based on length	2	Ordinal
studytime	numeric	Study time per week, 1-4 based on length	3	Ordinal
failures	numeric	Number of classes failed	2	Ratio
schoolsup	binary	Educational support	no	Nominal
famsup	binary	Family educational support	yes	Nominal
paid	binary	Extra paid classes in course subject Math or Portuguese	yes	Nominal
activities	binary	Extra-curricular activities	no	Nominal

nursery	binary	Determines whether student attended nursery school	yes	Nominal
higher	binary	If student wants to go to higher education	yes	Nominal
internet	binary	If student has internet access at home	no	Nominal
romantic	binary	Whether student has romantic relationship	no	Nominal
famrel	factor	Quality of family relationships 1-5 Very bad – very good	3	Ordinal
freetime	factor	Amount of free time student has 1-5	5	Ordinal
goout	factor	How often student goes out with friends 1-5	1	Ordinal
Dalc	factor	Workday alcohol consumption 1-5	2	Ordinal
Walc	factor	Weekend alcohol consumption 1-5	2	Ordinal
health	factor	Health measure for student 1-5	4	Ordinal
absences	numeric	Number of absences	19	Ratio
G1	numeric	First period grade 0-20	15	Ratio
G2	numeric	Second period grade	12	Ratio
G3	numeric	Third period grade	18	Ratio

PURPOSE OF DASHBOARD:

The purpose of this dashboard will be to let the user conduct an exploratory analysis on the UCI dataset for student performance in Peru during 2014. It will allow the user to filter by different features to analyze relationships between different variables and average student grades. By searching for insights in this dashboard the user will be able to generate plans for future education planning, including altering curriculum or adjusting teaching methods to better accommodate all types of students.

USERS:
The intended audience of this visualization is any teacher or education entity that is looking for methods to improve success rates of future students.

WHAT DATA WILL DRIVE EACH VISUALIZATION:

To continue with an analysis solely based on the data, I need to create a calculated field that reflects success in school. Specifically, I will be creating an average of each student's grades over the three periods. I will then conduct a comparative analysis of several different features with average grades to determine any perceived trends in the data.

The first visualization will show a table of factors that have a high impact on grades in school, coming from the logistic regression model. It will drive user navigation and amplify which factors the user should be looking at when thinking about grade performance.

The second visualization will be interactive, with the user being able to switch between different features of the data and compare it with average student grades. This way, the audience can conduct an exploratory analysis on determinants for student performance – with anything from mother's education to time spent traveling.

My last visualization will be a filter that allows the user to add another dimension to the data. For example, if the user wants to look at the higher education factor and its relation to student performance, but also wants to break this down by gender, the user will be able to do this with the filter.

QUESTIONS AND QUERIES TO BE ANSWERED:

- How much of an effect does study time have on grades earned?
- How does parental education affect student performance?
- Do absences take a significant toll on average grades?
- What is the most important factor in influencing average student grades?
- What types of behavioral patterns affect performance?

RATIONALE BEHIND THE CONTENTS OF THE DASHBOARD

DESIGN PRINCIPLES:
The core purpose of visualizations displayed in this paper is to find relationships between factors in a student's life and their grades. In doing so, the user can find different variables to influence when trying to provide better avenues of success with education. As a result, the dashboard must

be very easy to understand, with proper labels and no unnecessary additions.

The design principles that fit these criteria best are Stephen Few's rules for good visualizations. In his principles, Few describes how one should "display neither more nor less than what is relevant," "differences in visual properties should accurately correspond to actual differences they represent," and that one should appeal to short-term memory by combining multiple facts into a single visual [3]. My dashboard will combine these key principles in its display to users – I will only be using a bar chart with few dimensional data, I will scale appropriately from 0, and I will use clean, unaltered data in conjunction with graphic displays for additional analysis.
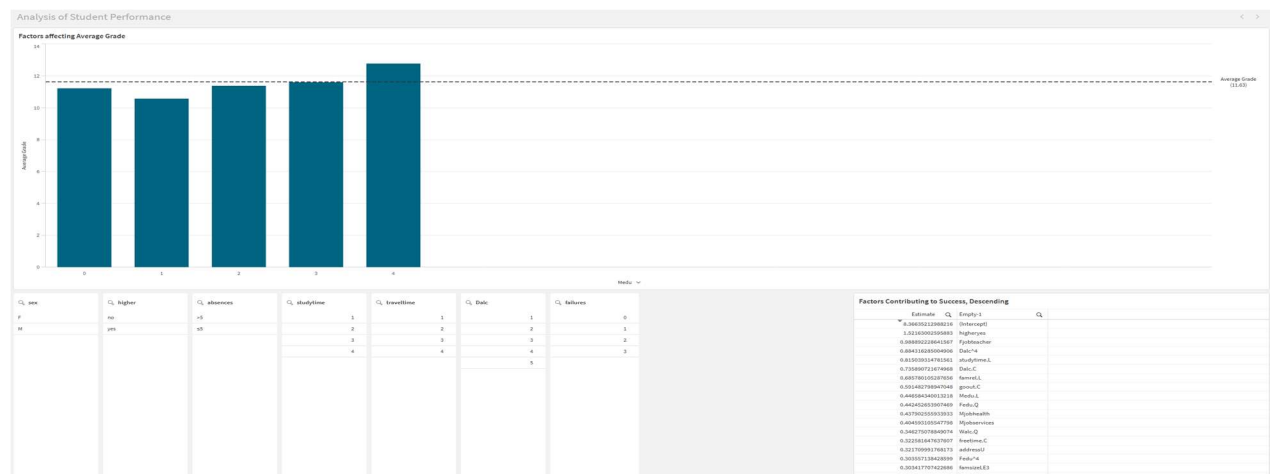
VISUALIZATION TECHNIQUES:
My main goal in the exploratory dashboard is to allow the user to look at relationships between variables and student grades – I will add a table showing how different variables contribute to student grade, but the dashboard will mainly be comparing two variables with the option to add a filter. Therefore, to keep with the theme of simplicity I will only be using an interactive bar chart that makes it very easy to compare features with average grade. The dashboard will include a filter pane as well as a table of the most intuitive search queries, but the main source of exploratory analysis will be the bar chart.
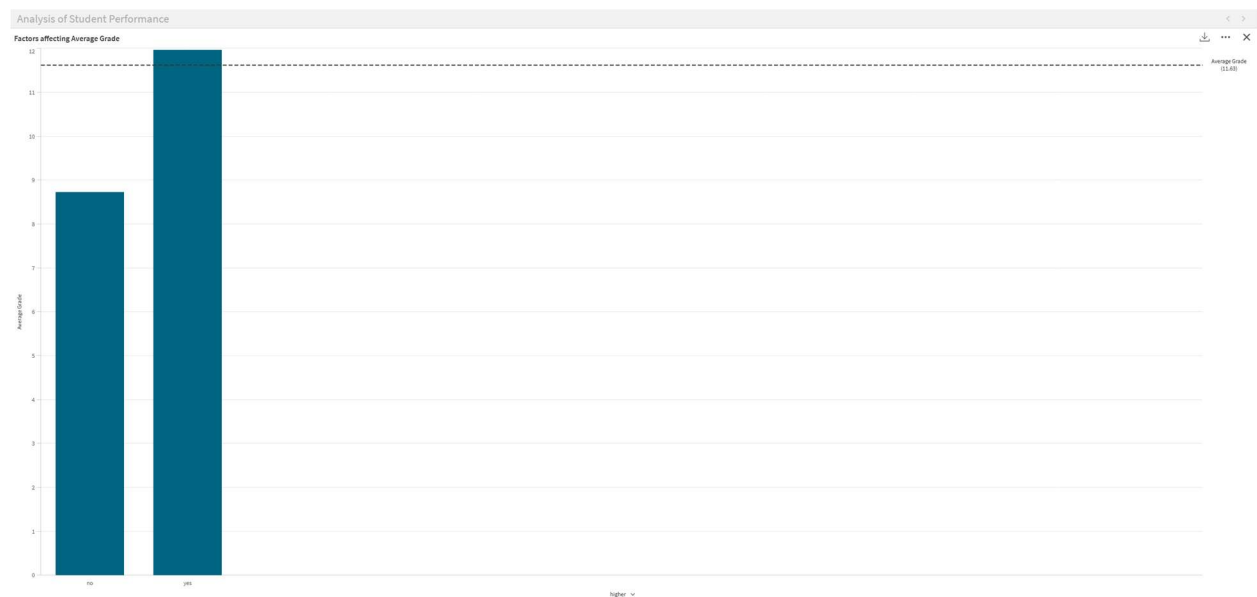
INTERACTIVITY:
The dashboard created will be interactive to allow the user to conduct an exploratory analysis on average grade based on any data available in the feature space of the UCI dataset. Since all of the variables are either binary, factor, or lie in an interval, there are not many dimensions to fit onto an x-axis, and it will be simple to view every possible combination of variable and average grade output. I will also input a filter into the dashboard to allow for a third dimension to be explored when viewing the bar chart for analysis. For example, if the viewer is using the chart to see the relationship between a student's intention to go to higher education and their grade, they can also filter by high travel time to see if that impacts the data at all.
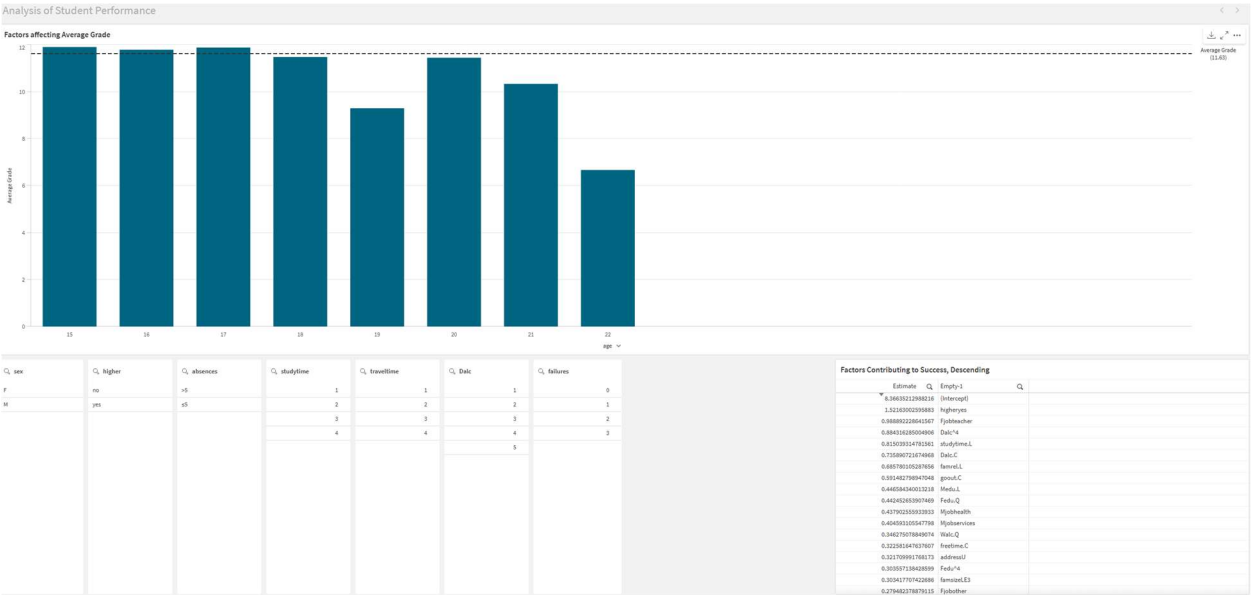
DASHBOARD

My Dashboard is made up of three different sections. The entire top half of the dashboard contains a bar chart that displays the relationship between any variable with the Average Grade of the student. This is meant to have the user conduct an exploratory analysis on the data to determine which variables contribute most to student success. The bar chart contains a constant Average Grade trendline to provide a baseline for determining if factors help increase the average grade, or decrease the average grade. On the bottom left, the user will find several different filters. The filters exist to have the user examine how two different factors can affect student performance at once. Finally, the bottom right of the dashboard is a high level overview of the regression output for the Logistic Regression. It shows in descending order the most important factors that affect Average Grade, which is meant to guide the user in their exploratory analysis.
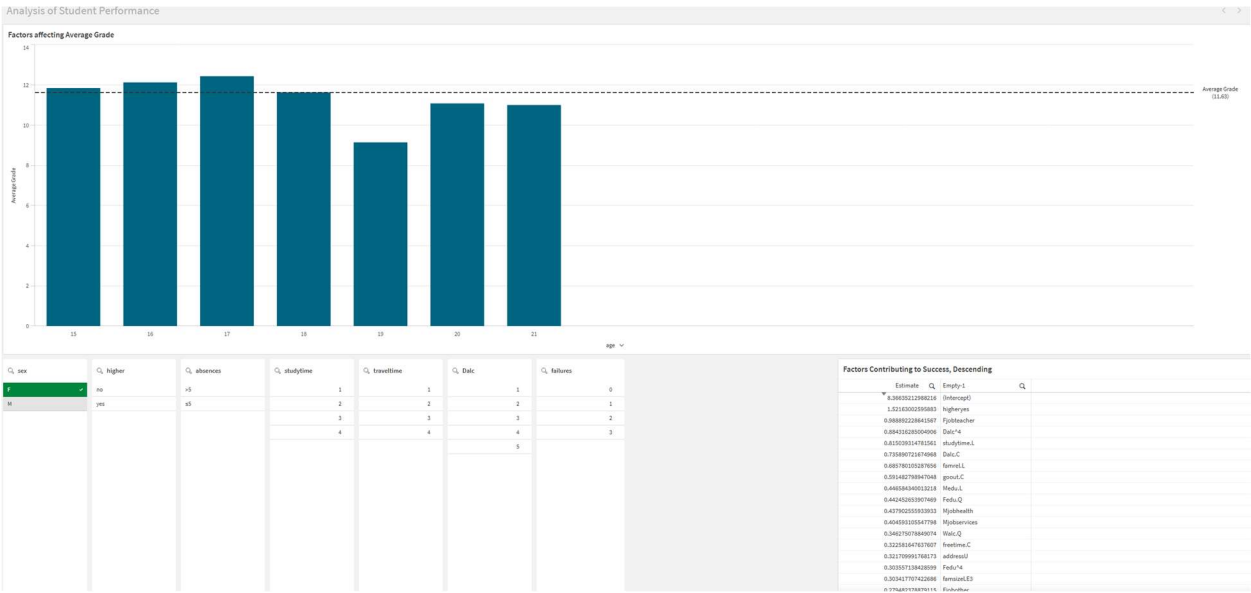

In the full-screen depiction of the bar chart, we can see just how higher education affects student performance. Those who are not seeking to go to college seem to have scores much

below average, whereas those who are tend to score slightly above average. This variable was used because we can see that higher education has the highest effect on average grade, shown on the bottom right in the dashboard.

Here is an initial display of Age vs. Average Grade, which helps us look into the initial question proposed about demographic information and its relation to student performance. We can see that generally, students who are having to retake high school perform worse. This makes sense intuitively, since older students would have had to fail initially in order to retake high school classes. What if we also add a filter for sex on this data?

We can see that the age group 22 disappears, and the average grade increases among every category. If the user then compares sex and average grade, they would see that females
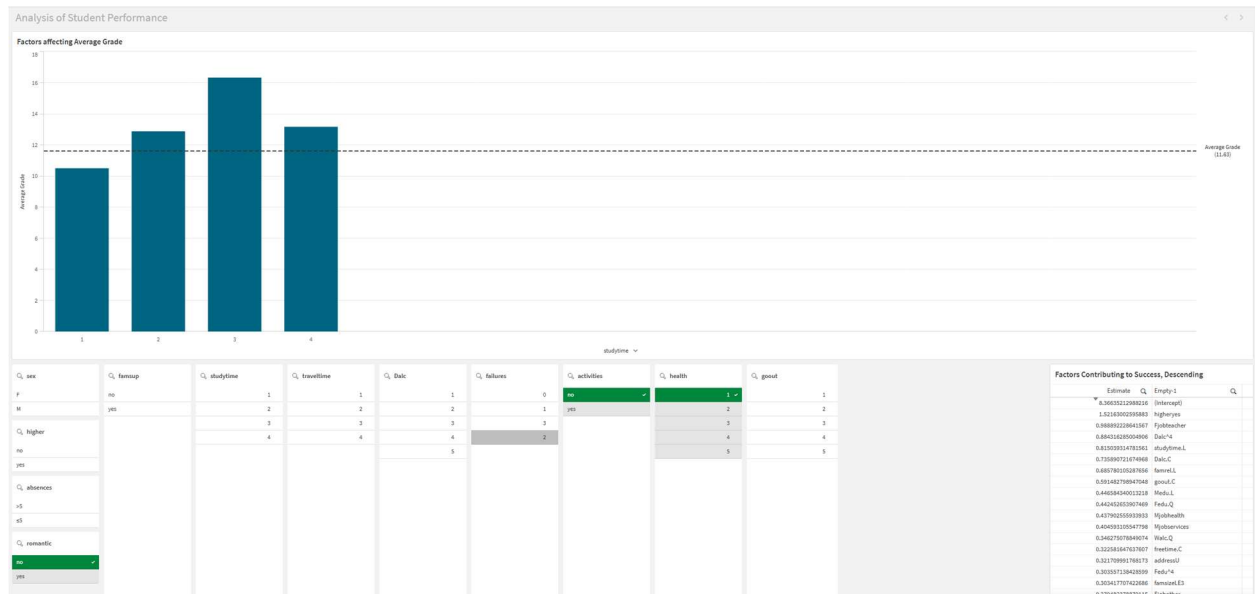
generally score higher than males as well, supporting the evidence to this trend.

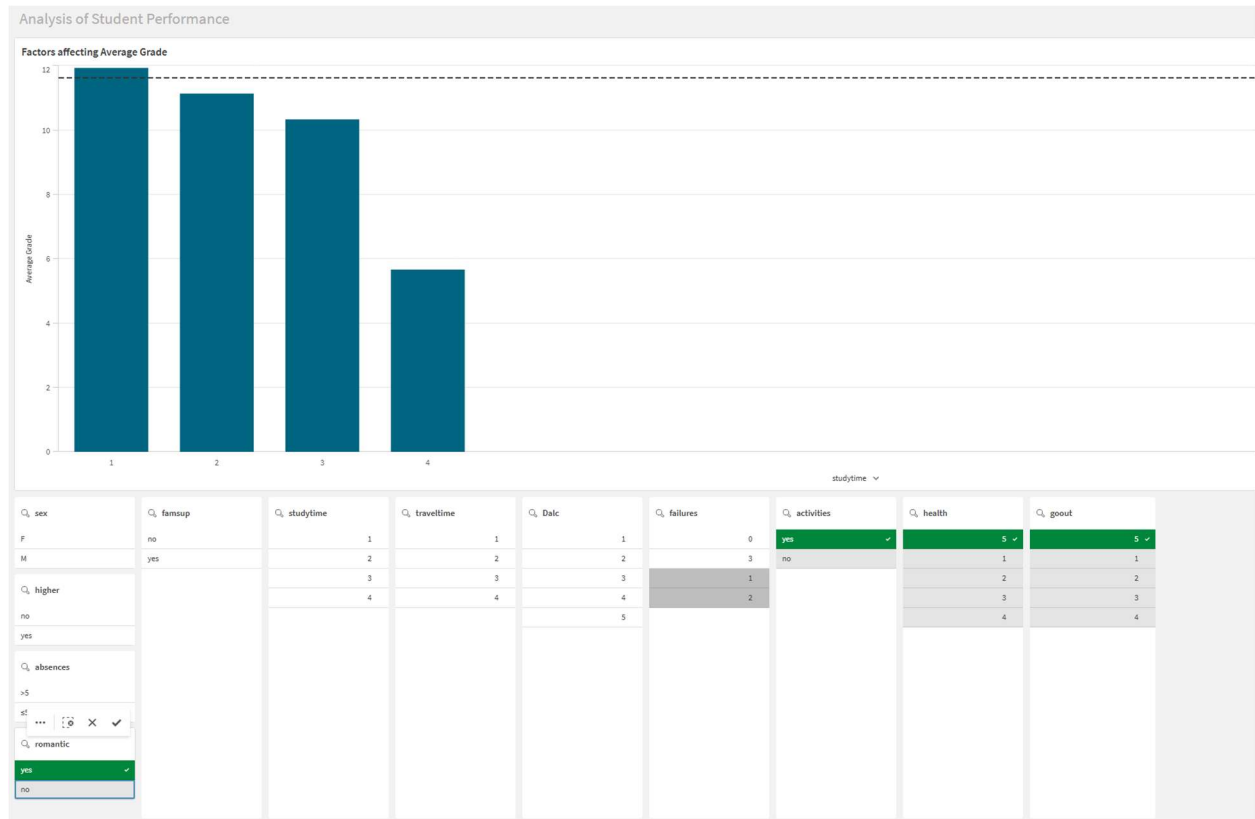## USE CASE: HOW MUCH CAN A STUDENT AFFECT THEIR OWN PERFORMANCE?

Our last initial question for this research project was about how much control a child has over their own success. There are many factors that affect student performance, but a student has little control over external factors in their life. Among things they can control are time spent studying, time spent hanging out with friends, healthy habits, romantic habits, etc. Using this Dashboard, we can explore every aspect that a student can control to determine impact on success. First, shown above, is a simple graph of Time Spent Studying vs. Average Grade. We can see that as hours spend studying per day increases, average grade increases, with marginal benefits after 3 hours.

Let's look at a student who is extremely studious; they don't go outside, they're stressed because of school, and they don't participate in any other extracurricular activities. This would be a student that is extremely dedicated to studying regardless of other external factors. Using the filter pane, we can see how scores change based off of this behavior:

As we can see, students who exhibit this behavior tend to score higher when studying for 2 or more hours.

Now, what about a student who goes out often, takes extracurricular activities, is in a romantic relationship, and cares a lot about their health?

We can that see that, quite unexpectedly, an increase in study time actually leads to a strong decrease in average grade.

REFERENCES

[1]   Hanushek, Eric A. "What matters for student achievement." Education next 16.2 (2016): Accessed October 26, 2025.

[2]   Cortez, Paulo. "Student Performance." UCI Machine Learning Repository, 2008, https://doi.org/10.24432/C5TG7T. Accessed October 26, 2025

[3]   Few, Stephen. 2006. Visual Communication: Core Principles for displaying quantitative information. Perceptual edge.com.