Bresa McClatchy

March 2, 2025

DSC-530

Summary of Analysis

The question of this analysis was: what factors are most associated with student dropout rates and academic success? Specifically, we examined the impact of scholarship status, 1st semester grades, age at enrollment, 2nd semester grades, and target outcomes (e.g., dropout, graduate, or enrolled).

To understand the exploratory data analysis, we should first understand the outcomes of the Target variable. While over 2000 of the students from the dataset graduated, nearly 1500 dropped out, which is a significant amount. When we look at the age of enrollment, a significant portion of them are under the age of 30, with most of the students aged 18. This also leads to a weak negative correlation observed at the age at enrollment and 1st semester grades, suggesting that older students have slightly lower grades.

It was also found that there is a strong correlation between scholarship status and 2nd semester grades. Those who held a scholarship had higher 2nd semester grades than those who didn't, which indicates that students who tend to perform well in the first semester also performed well in the second semester.

Variables that could have enhanced the analysis would be data on study habits, which could prove more insight into academic performance, and course difficulty, which would help contextualize some of the grade distributions. Not a variable, but I might add that deeper analysis

could be done, specifically with the Target variable against other variables, such as age at enrollment or semester averages.

The challenge most faced was the skewness of some of the data. Specifically, the grade averages of 0 for the 1st and 2nd semester grades. I think it's difficult to say whether or not that many students failed with complete 0s, or if there is simply data that is missing. When you look at the dropout rate, it could make sense, but there are no averages between 0 and 10 on the chart, so I'm wondering if there's data missing that should be there. If there is, then it would change the look of the charts and some of the outcomes. Another challenge was the lack of variables that could better be used for a deeper analysis of the statistical question.