

# Midterm on Final Grades

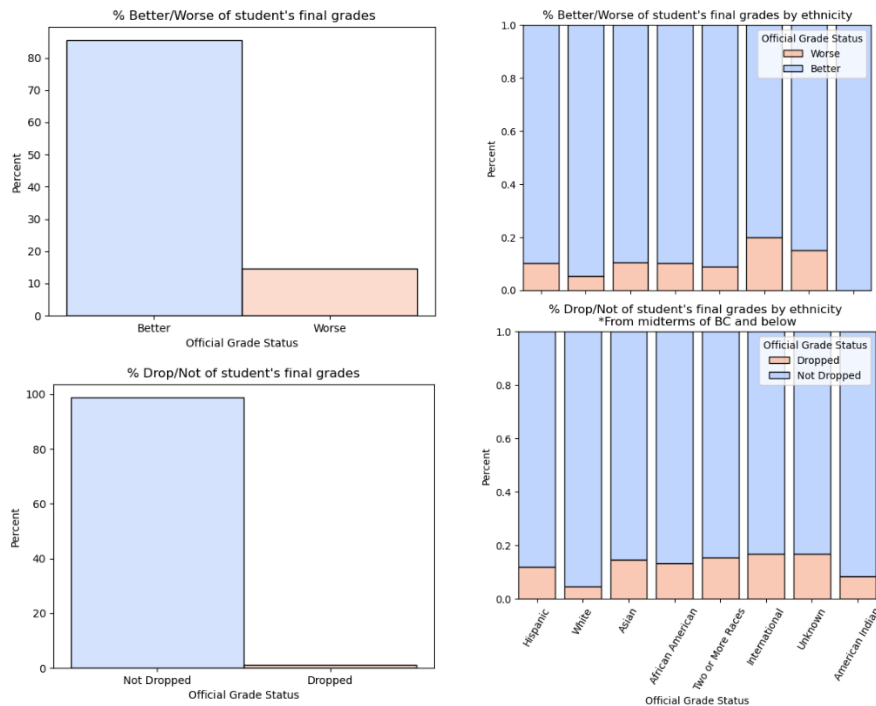
## Overview:

This analysis aims to explain the connection between, and predict final grades, based on students' midterm grades. This analysis was performed on 18,204 observations of first year business students course grades going back to the 2014-2015 academic year. Other variables considered were a student's ethnicity, their semester credit total, total credits entering the university, standardized test scores, and course level. There were two ways final grade performance was measured, first whether the student dropped or not and second if their final grade was better or worse than their midterm grade.

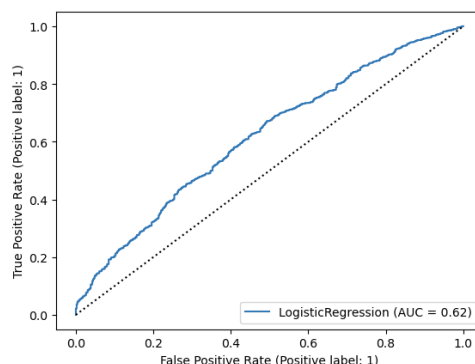
## Key insights:

- As expected, there is a high positive correlation between student's midterm grade and their final grade with a coefficient of 0.719 grade points.
- For the logistic regression models predicting drop/not and doing better/worse in a course, the independent variables are correlated with students' final grades, but they do not explain much of the variability in what leads a student to drop or perform worse.
- The only other notable correlations amongst other variables are with students' math standardized test score (0.21 coefficient) and number of credits they have when entering the university (0.22 coefficient), both positive.
- Of students with midterm grades lower than a BC, 26.5% of them end with a C, while only 0.037% students end with an A.

The data was heavily imbalanced in each case as seen below. There are small numbers of students who drop and perform worse which lead to the binary logistic regression models having bias towards the highly populated category.



The low accuracy due to the unbalanced data can also be seen in the following ROC curve for predicting Better or Worse final grades. The dotted line signifies classifying randomly, while the blue line is how well the model predicts true positives and true negatives.



When running a logistic model to predict whether a student's final grade is better or worse than their midterm grade using all parameters except semester total credits, it overly predicts a student to do better. This is because of how many observations are of students doing better in our dataset. The cross table of this particular model's predictions:

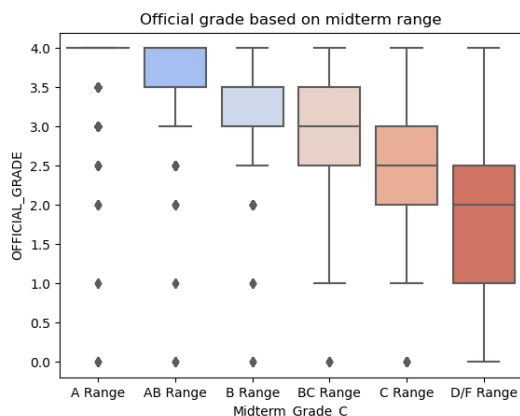
	Predicted Worse	Predicted Better
True Worse	0	1876
True Better	0	10,989

We ran many logistic model variations, with different parameter combinations and methods to balance the data but there was not enough explainability from the data to create an accurate model to predict whether a student drops/stays in a course or if they receive a better vs. worse final grade. While the models did not explain the data very well, midterm grades and incoming test credits were still statistically significant in predicting student performance.

When running a linear regression model on all observations predicting final grades, we saw higher variance explained than the logistic models. The table below provides the regression coefficients for the variables effect on the students predicted final grade.

Variable	Coefficient
Midterm Grade	0.63
Math Test Score	0.03
Semester Credit Total	0.007
Incoming Test Credits	0.0045

This chart shows, based on the grade range student's midterm grades were in, the range that their final grades (numeric) lies in. For students with midterm grades in the D/F range, we see there is a lot of variety in where their final grade may end. As opposed to students who received an A midterm, their final grade distribution range stays at an A.



The final grade distributions based on midterm grade ranges can further be broken down by ethnicity groups. The chart below excludes Native Hawaiians since there is not enough observations to make inferences. An interesting take away was that for white students with D/F midterm grades they're final grades generally end up higher than all other ethnicities. Another, African Americans who have a midterm grade in the A range, have a wider distribution of final grades that lay lower than all other groups.

