REPORT

# Modeling Student Outcomes from Admission Data: Dropout to Graduate

KAYLIE LE

**May 31, 2025**

**Table of Content**

*Introduction*.....3

*PART 1*.....4

*PART 2*.....10

*Summary*.....16

*References*.....17

## Introduction

This analysis examines a dataset of 4,424 undergraduate students across multiple academic programs, aimed at predicting student outcomes: Dropout, Enrolled, or Graduate. The data includes demographic, academic, and socio-economic variables collected at the time of enrollment, along with performance metrics after the first and second semesters. The project was supported by the SATDAP program in Portugal, with the goal of reducing dropout rates in higher education.

The classification problem is threefold and imbalanced, with most students concentrated in one category. Key indicators suggest clear distinctions among student outcomes:

- Dropout students tend to have lower admission grades, older age profiles, and greater financial hardship.
- Enrolled students represent a transitional group with moderate academic progress and mixed financial status.
- Graduates show strong academic performance, timely tuition payment, and a higher rate of scholarship support.

These patterns suggest that early academic performance and financial readiness are critical factors in predicting long-term student success. Identifying at-risk students at the point of enrollment could enable institutions to implement targeted interventions early in the academic journey.

## Part 1

*Produce several descriptive statistics tables for the entire sample*

| | n | mean | sd | min | max |
|---|---|---|---|---|---|
| Marital.status | 4424 | 1.178571e+00 | 0.6057469 | 1.00 | 6.00000 |
| Application.mode | 4424 | 1.866908e+01 | 17.4846823 | 1.00 | 57.00000 |
| Application.order | 4424 | 1.727848e+00 | 1.3137931 | 0.00 | 9.00000 |
| Course | 4424 | 8.856643e+03 | 2063.5664162 | 33.00 | 9991.00000 |
| Daytime.evening.attendance. | 4424 | 8.908228e-01 | 0.3118967 | 0.00 | 1.00000 |
| Previous.qualification | 4424 | 4.577758e+00 | 10.2165923 | 1.00 | 43.00000 |
| Previous.qualification..grade. | 4424 | 1.326133e+02 | 13.1883317 | 95.00 | 190.00000 |
| Nacionality | 4424 | 1.873192e+00 | 6.9145140 | 1.00 | 109.00000 |
| Mother.s.qualification | 4424 | 1.956193e+01 | 15.6031863 | 1.00 | 44.00000 |
| Father.s.qualification | 4424 | 2.227532e+01 | 15.3431078 | 1.00 | 44.00000 |
| Mother.s.occupation | 4424 | 1.096090e+01 | 26.4182529 | 0.00 | 194.00000 |
| Father.s.occupation | 4424 | 1.103232e+01 | 25.2630402 | 0.00 | 195.00000 |
| Admission.grade | 4424 | 1.269781e+02 | 14.4820008 | 95.00 | 190.00000 |
| Displaced | 4424 | 5.483725e-01 | 0.4977109 | 0.00 | 1.00000 |
| Educational.special.needs | 4424 | 1.152803e-02 | 0.1067601 | 0.00 | 1.00000 |
| Debtor | 4424 | 1.136980e-01 | 0.3174800 | 0.00 | 1.00000 |
| Tuition.fees.up.to.date | 4424 | 8.806510e-01 | 0.3242354 | 0.00 | 1.00000 |
| Gender | 4424 | 3.517179e-01 | 0.4775604 | 0.00 | 1.00000 |
| Scholarship.holder | 4424 | 2.484177e-01 | 0.4321442 | 0.00 | 1.00000 |
| Age.at.enrollment | 4424 | 2.326514e+01 | 7.5878156 | 17.00 | 70.00000 |
| International | 4424 | 2.486438e-02 | 0.1557293 | 0.00 | 1.00000 |
| Curricular.units.1st.sem..credited. | 4424 | 7.099910e-01 | 2.3605066 | 0.00 | 20.00000 |
| Curricular.units.1st.sem..enrolled. | 4424 | 6.270570e+00 | 2.4801782 | 0.00 | 26.00000 |
| Curricular.units.1st.sem..evaluations. | 4424 | 8.299051e+00 | 4.1791056 | 0.00 | 45.00000 |
| Curricular.units.1st.sem..approved. | 4424 | 4.706600e+00 | 3.0942380 | 0.00 | 26.00000 |
| Curricular.units.1st.sem..grade. | 4424 | 1.064082e+01 | 4.8436634 | 0.00 | 18.87500 |
| Curricular.units.1st.sem..without.evaluations. | 4424 | 1.376582e-01 | 0.6908802 | 0.00 | 12.00000 |
| Curricular.units.2nd.sem..credited. | 4424 | 5.418174e-01 | 1.9185461 | 0.00 | 19.00000 |
| Curricular.units.2nd.sem..enrolled. | 4424 | 6.232143e+00 | 2.1959508 | 0.00 | 23.00000 |
| Curricular.units.2nd.sem..evaluations. | 4424 | 8.063291e+00 | 3.9479509 | 0.00 | 33.00000 |
| Curricular.units.2nd.sem..approved. | 4424 | 4.435805e+00 | 3.0147639 | 0.00 | 20.00000 |
| Curricular.units.2nd.sem..grade. | 4424 | 1.023021e+01 | 5.2108080 | 0.00 | 18.57143 |
| Curricular.units.2nd.sem..without.evaluations. | 4424 | 1.503165e-01 | 0.7537741 | 0.00 | 12.00000 |
| Unemployment.rate | 4424 | 1.156614e+01 | 2.6638505 | 7.60 | 16.20000 |
| Inflation.rate | 4424 | 1.228029e+00 | 1.3827107 | -0.80 | 3.70000 |
| GDP | 4424 | 1.968807e-03 | 2.2699354 | -4.06 | 3.51000 |

**Table 1. Descriptive Statistics for the Entire Sample**

The overall dataset includes 4,424 undergraduate students, representing a diverse student population across various degree programs. The average admission grade was approximately 127, with a standard deviation of about 14.5, suggesting moderate variability in academic readiness at the point of enrollment.

On average, students enrolled in 6.2 curricular units in the first semester and 6.2 units in the second semester. However, the mean number of approved units was lower: about 4.1 in the first semester and 4.4 in the second, indicating that not all students passed all enrolled courses.

The average age at enrollment was around 23.3 years, but the range extended from 17 to 70, showing the presence of both traditional and non-traditional students.

Financial and social factors may also play a role. For example, the mean value for "Tuition fees up to date" is 0.88 (on a 0–1 scale), suggesting that the majority of students paid on time, but a notable portion faced financial delays.

In terms of academic engagement, the relatively high standard deviations in variables like evaluations and grades (around 4 to 5) show that student performance varied widely, possibly indicating different levels of preparation or support needs.

*Key finding:*

- There is significant diversity and variability among students in terms of academic readiness, demographic background, and performance, which suggests that a one-size-fits-all approach to student success is ineffective.
- Early interventions should be targeted — especially for students with low admission grades, low approval rates, financial struggles, or non-traditional enrollment patterns — to reduce dropout rates and improve academic outcomes.

*Produce several descriptive statistics tables for group Target group (Dropout, Enrolled, Graduate)*

*Droupout Group*

| | n | mean | sd | min | max |
|---|---|---|---|---|---|
| Marital.status | 1421 | 1.26108374 | 0.7219584 | 1.00 | 6.00000 |
| Application.mode | 1421 | 23.71287825 | 17.1505934 | 1.00 | 53.00000 |
| Application.order | 1421 | 1.59324419 | 1.2157694 | 1.00 | 6.00000 |
| Course | 1421 | 8755.62983814 | 2278.7794774 | 33.00 | 9991.00000 |
| Daytime.evening.attendance. | 1421 | 0.85432794 | 0.3529013 | 0.00 | 1.00000 |
| Previous.qualification | 1421 | 5.31104856 | 10.3099336 | 1.00 | 43.00000 |
| Previous.qualification..grade. | 1421 | 131.11407460 | 12.8733538 | 95.00 | 190.00000 |
| Nacionality | 1421 | 1.95707248 | 7.8389773 | 1.00 | 109.00000 |
| Mother.s.qualification | 1421 | 21.03518649 | 15.5238172 | 1.00 | 44.00000 |
| Father.s.qualification | 1421 | 22.68332160 | 15.3570990 | 1.00 | 44.00000 |
| Mother.s.occupation | 1421 | 10.11611541 | 20.6171309 | 0.00 | 194.00000 |
| Father.s.occupation | 1421 | 10.14144968 | 20.0652557 | 0.00 | 194.00000 |
| Admission.grade | 1421 | 124.96136524 | 15.1251593 | 95.00 | 190.00000 |
| Displaced | 1421 | 0.47079521 | 0.4993221 | 0.00 | 1.00000 |
| Educational.special.needs | 1421 | 0.01196341 | 0.1087594 | 0.00 | 1.00000 |
| Debtor | 1421 | 0.21956369 | 0.4140968 | 0.00 | 1.00000 |
| Tuition.fees.up.to.date | 1421 | 0.67839550 | 0.4672566 | 0.00 | 1.00000 |
| Gender | 1421 | 0.49331457 | 0.5001313 | 0.00 | 1.00000 |
| Scholarship.holder | 1421 | 0.09429979 | 0.2923482 | 0.00 | 1.00000 |
| Age.at.enrollment | 1421 | 26.06896552 | 8.7040237 | 18.00 | 70.00000 |
| International | 1421 | 0.02251935 | 0.1484174 | 0.00 | 1.00000 |
| Curricular.units.1st.sem..credited. | 1421 | 0.60942998 | 2.1046952 | 0.00 | 18.00000 |
| Curricular.units.1st.sem..enrolled. | 1421 | 5.82125264 | 2.3263037 | 0.00 | 21.00000 |
| Curricular.units.1st.sem..evaluations. | 1421 | 7.75158339 | 4.9215900 | 0.00 | 31.00000 |
| Curricular.units.1st.sem..approved. | 1421 | 2.55172414 | 2.8575879 | 0.00 | 21.00000 |
| Curricular.units.1st.sem..grade. | 1421 | 7.25665553 | 6.0311201 | 0.00 | 18.00000 |
| Curricular.units.1st.sem..enrolled. | 1421 | 5.82125264 | 2.3263037 | 0.00 | 21.00000 |
| Curricular.units.1st.sem..evaluations. | 1421 | 7.75158339 | 4.9215900 | 0.00 | 31.00000 |
| Curricular.units.1st.sem..approved. | 1421 | 2.55172414 | 2.8575879 | 0.00 | 21.00000 |
| Curricular.units.1st.sem..grade. | 1421 | 7.25665553 | 6.0311201 | 0.00 | 18.00000 |
| Curricular.units.1st.sem..without.evaluations. | 1421 | 0.19211823 | 0.7945907 | 0.00 | 8.00000 |
| Curricular.units.2nd.sem..credited. | 1421 | 0.44968332 | 1.6795481 | 0.00 | 16.00000 |
| Curricular.units.2nd.sem..enrolled. | 1421 | 5.78043631 | 2.1082494 | 0.00 | 18.00000 |
| Curricular.units.2nd.sem..evaluations. | 1421 | 7.17382125 | 4.8173736 | 0.00 | 25.00000 |
| Curricular.units.2nd.sem..approved. | 1421 | 1.94018297 | 2.5736908 | 0.00 | 16.00000 |
| Curricular.units.2nd.sem..grade. | 1421 | 5.89933885 | 6.1187399 | 0.00 | 17.71429 |
| Curricular.units.2nd.sem..without.evaluations. | 1421 | 0.23786066 | 0.9942097 | 0.00 | 12.00000 |
| Unemployment.rate | 1421 | 11.61639690 | 2.7682592 | 7.60 | 16.20000 |
| Inflation.rate | 1421 | 1.28395496 | 1.4049696 | -0.80 | 3.70000 |
| GDP | 1421 | -0.15085855 | 2.2517682 | -4.06 | 3.51000 |

**Table 2. Descriptive Statistics – Dropout Group**

Students in the Dropout group had an average admission grade of about 124.96, which is slightly lower than the overall sample average. This may suggest that lower academic readiness at entry could be associated with a higher risk of dropping out.

On average, dropout students enrolled in around 5.8 to 7.5 courses per semester, but only passed approximately 2.5 to 2.9 of them, indicating a considerable gap between enrollment and successful completion.

The standard deviations for course evaluations and grades are relatively high (around 4.9 to 6.1), reflecting inconsistent academic performance within this group.

The mean age at enrollment was about 26, and the age range went up to 70, which might indicate the presence of non-traditional or returning students—some of whom may have faced challenges balancing education with other responsibilities.

Additionally, the low average on the variable "Tuition fees up to date" (mean = 0.68) suggests that financial stress could have been a factor in student dropout.

*Key finding:*

Students in the Dropout group show early signs of academic and financial vulnerability, including lower admission grades, fewer passed courses, high variability in performance, and indicators of life complexity (such as age and financial stress). These factors suggest a clear profile of students at risk, calling for early detection and targeted support.

*Enrolled Group*

| | n | mean | sd | min | max |
|---|---|---|---|---|---|
| Marital.status | 794 | 1.152393e+00 | 0.5710110 | 1.00 | 6.00 |
| Application.mode | 794 | 1.979975e+01 | 18.0178964 | 1.00 | 53.00 |
| Application.order | 794 | 1.625945e+00 | 1.2135768 | 1.00 | 9.00 |
| Course | 794 | 8.868811e+03 | 2040.9016403 | 33.00 | 9991.00 |
| Daytime.evening.attendance. | 794 | 9.055416e-01 | 0.2926498 | 0.00 | 1.00 |
| Previous.qualification | 794 | 4.785894e+00 | 11.0582319 | 1.00 | 42.00 |
| Previous.qualification..grade. | 794 | 1.312084e+02 | 12.8721109 | 96.00 | 190.00 |
| Nacionality | 794 | 2.079345e+00 | 7.9476536 | 1.00 | 108.00 |
| Mother.s.qualification | 794 | 1.762217e+01 | 15.5476995 | 1.00 | 43.00 |
| Father.s.qualification | 794 | 2.092065e+01 | 15.5867989 | 1.00 | 41.00 |
| Mother.s.occupation | 794 | 1.472040e+01 | 37.2517073 | 0.00 | 194.00 |
| Father.s.occupation | 794 | 1.445844e+01 | 35.3360578 | 0.00 | 193.00 |
| Admission.grade | 794 | 1.255343e+02 | 13.7925592 | 95.00 | 190.00 |
| Displaced | 794 | 5.453401e-01 | 0.4982539 | 0.00 | 1.00 |
| Educational.special.needs | 794 | 1.385390e-02 | 0.1169581 | 0.00 | 1.00 |
| Debtor | 794 | 1.133501e-01 | 0.3172201 | 0.00 | 1.00 |
| Tuition.fees.up.to.date | 794 | 9.471033e-01 | 0.2239684 | 0.00 | 1.00 |
| Gender | 794 | 3.866499e-01 | 0.4872892 | 0.00 | 1.00 |
| Scholarship.holder | 794 | 1.637280e-01 | 0.3702618 | 0.00 | 1.00 |
| Age.at.enrollment | 794 | 2.236902e+01 | 6.3026724 | 17.00 | 54.00 |
| International | 794 | 3.022670e-02 | 0.1713185 | 0.00 | 1.00 |
| Curricular.units.1st.sem..credited. | 794 | 5.075567e-01 | 1.7157584 | 0.00 | 14.00 |
| Curricular.units.1st.sem..enrolled. | 794 | 5.964736e+00 | 1.9876709 | 0.00 | 17.00 |
| Curricular.units.1st.sem..evaluations. | 794 | 9.341310e+00 | 3.4634654 | 0.00 | 24.00 |
| Curricular.units.1st.sem..approved. | 794 | 4.318640e+00 | 2.2893977 | 0.00 | 15.00 |
| Curricular.units.1st.sem..grade. | 794 | 1.112526e+01 | 3.6750839 | 0.00 | 17.00 |
| Curricular.units.1st.sem..without.evaluations. | 794 | 1.775819e-01 | 0.7414591 | 0.00 | 8.00 |
| Curricular.units.2nd.sem..credited. | 794 | 3.589421e-01 | 1.3290343 | 0.00 | 12.00 |
| Curricular.units.2nd.sem..enrolled. | 794 | 5.938287e+00 | 1.8314804 | 0.00 | 17.00 |
| Curricular.units.2nd.sem..evaluations. | 794 | 9.435768e+00 | 3.5665576 | 0.00 | 28.00 |
| Curricular.units.2nd.sem..approved. | 794 | 4.057935e+00 | 2.1801972 | 0.00 | 12.00 |
| Curricular.units.2nd.sem..grade. | 794 | 1.111736e+01 | 3.6012405 | 0.00 | 17.60 |
| Curricular.units.2nd.sem..without.evaluations. | 794 | 1.876574e-01 | 0.7804670 | 0.00 | 8.00 |
| Unemployment.rate | 794 | 1.127254e+01 | 2.6281389 | 7.60 | 16.20 |
| Inflation.rate | 794 | 1.211713e+00 | 1.3733553 | -0.80 | 3.70 |
| GDP | 794 | 5.328715e-02 | 2.3156312 | -4.06 | 3.51 |

**Table 3. Descriptive Statistics – Enrolled Group**

The Enrolled group consists of 794 students who have not yet dropped out or graduated, making them an important focus for early intervention strategies. Their average admission grade is approximately 125.5, slightly higher than the Dropout group but still lower than the Graduate group.

On average, these students enrolled in around 5.9 curricular units per semester, but only passed 4.3 units in the first semester and 4.1 in the second. The gap between enrollment and approval suggests room for improvement in academic performance or support.

The standard deviation in grades and evaluations is moderate (around 3.5 to 3.6), indicating some performance variation, but not as wide as in the Dropout group. This may reflect a mix of students who are either improving or at risk.

Interestingly, this group has a relatively high percentage of students up to date with tuition fees (mean = 0.94) and fewer financial delays, which may positively influence retention.

With an average age at enrollment of 22.4 years, these students are younger than the Dropout group, possibly representing more traditional learners still in the middle of their academic journey.

*Key finding:*

- The Enrolled group is a transitional and diverse segment, offering high potential for recovery or success.
- Strategic interventions at this stage can be pivotal in nudging students toward graduation rather than dropout.

*Graduate Group*

| | n | mean | sd | min | max |
|---|---|---|---|---|---|
| Marital.status | 2209 | 1.134903e+00 | 0.5255255 | 1.00 | 6.00000 |
| Application.mode | 2209 | 1.501811e+01 | 16.6272715 | 1.00 | 57.00000 |
| Application.order | 2209 | 1.851064e+00 | 1.3955156 | 0.00 | 6.00000 |
| Course | 2209 | 8.917248e+03 | 1919.4448367 | 33.00 | 9991.00000 |
| Daytime.evening.attendance. | 2209 | 9.090086e-01 | 0.2876620 | 0.00 | 1.00000 |
| Previous.qualification | 2209 | 4.031236e+00 | 9.8056024 | 1.00 | 43.00000 |
| Previous.qualification..grade. | 2209 | 1.340827e+02 | 13.3421206 | 97.00 | 184.40000 |
| Nacionality | 2209 | 1.745134e+00 | 5.7906374 | 1.00 | 103.00000 |
| Mother.s.qualification | 2209 | 1.931145e+01 | 15.5909226 | 1.00 | 43.00000 |
| Father.s.qualification | 2209 | 2.249977e+01 | 15.2257506 | 1.00 | 43.00000 |
| Mother.s.occupation | 2209 | 1.015301e+01 | 24.9020571 | 0.00 | 194.00000 |
| Father.s.occupation | 2209 | 1.037392e+01 | 23.7892043 | 0.00 | 195.00000 |
| Admission.grade | 2209 | 1.287944e+02 | 14.0714398 | 95.00 | 190.00000 |
| Displaced | 2209 | 5.993662e-01 | 0.4901378 | 0.00 | 1.00000 |
| Educational.special.needs | 2209 | 1.041195e-02 | 0.1015293 | 0.00 | 1.00000 |
| Debtor | 2209 | 4.572205e-02 | 0.2089289 | 0.00 | 1.00000 |
| Tuition.fees.up.to.date | 2209 | 9.868719e-01 | 0.1138492 | 0.00 | 1.00000 |
| Gender | 2209 | 2.480761e-01 | 0.4319940 | 0.00 | 1.00000 |
| Scholarship.holder | 2209 | 3.779991e-01 | 0.4849972 | 0.00 | 1.00000 |
| Age.at.enrollment | 2209 | 2.178361e+01 | 6.6932182 | 17.00 | 62.00000 |
| International | 2209 | 2.444545e-02 | 0.1544625 | 0.00 | 1.00000 |
| Curricular.units.1st.sem..credited. | 2209 | 8.474423e-01 | 2.6861484 | 0.00 | 20.00000 |
| Curricular.units.1st.sem..enrolled. | 2209 | 6.669534e+00 | 2.6645791 | 0.00 | 26.00000 |
| Curricular.units.1st.sem..evaluations. | 2209 | 8.276596e+00 | 3.8099079 | 0.00 | 45.00000 |
| Curricular.units.1st.sem..approved. | 2209 | 6.232232e+00 | 2.5833799 | 0.00 | 26.00000 |
| Curricular.units.1st.sem..grade. | 2209 | 1.264365e+01 | 2.6990272 | 0.00 | 18.87500 |
| Curricular.units.1st.sem..without.evaluations. | 2209 | 8.827524e-02 | 0.5896858 | 0.00 | 12.00000 |
| Curricular.units.2nd.sem..credited. | 2209 | 6.668176e-01 | 2.2117090 | 0.00 | 19.00000 |
| Curricular.units.2nd.sem..enrolled. | 2209 | 6.628339e+00 | 2.2971968 | 0.00 | 23.00000 |
| Curricular.units.2nd.sem..evaluations. | 2209 | 8.142146e+00 | 3.2459120 | 0.00 | 33.00000 |
| Curricular.units.2nd.sem..approved. | 2209 | 6.177003e+00 | 2.2686218 | 0.00 | 20.00000 |
| Curricular.units.2nd.sem..grade. | 2209 | 1.269728e+01 | 2.6856076 | 0.00 | 18.57143 |
| Curricular.units.2nd.sem..without.evaluations. | 2209 | 8.057945e-02 | 0.5228722 | 0.00 | 12.00000 |
| Unemployment.rate | 2209 | 1.163934e+01 | 2.6014742 | 7.60 | 16.20000 |
| Inflation.rate | 2209 | 1.197918e+00 | 1.3711161 | -0.80 | 3.70000 |
| GDP | 2209 | 8.183341e-02 | 2.2610812 | -4.06 | 3.51000 |

**Table 4. Descriptive Statistics – Graduate Group**

Students in the Graduate group had the highest average admission grade (mean = 128.79) among all groups, suggesting that stronger academic preparation at entry may be linked to long-term success.

They also enrolled in and passed more courses than other groups. On average, they enrolled in about 6.7 curricular units per semester and successfully completed over 6 units, with high approval rates in both semesters. This reflects consistent academic performance and engagement.

The standard deviations for grades and evaluations are slightly lower than in other groups, suggesting more stable and less varied academic outcomes across the graduate population.

Financial indicators are also positive: tuition fees were up to date for almost all students (mean = 0.99), and the group had the highest percentage of scholarship holders (mean = 0.38), possibly reflecting institutional support and higher motivation.

The average age at enrollment was about 21.8, making this group slightly younger and possibly more likely to be full-time, first-time college students with fewer outside obligations.

*Key finding:*

- Graduates had the highest admission grades, completed more courses, and demonstrated the most consistent performance.
- Their financial readiness (e.g., paying tuition on time) and higher scholarship rates suggest that stability and early preparedness are linked to academic success.

*What is a three-line table format that is commonly used in white papers?*

A three-line table format is a clean and professional style commonly used in white papers and academic reports. It consists of:

- One horizontal line at the top of the table,
- One line below the column headers,
- And one line at the bottom of the table.

There are no vertical lines or internal gridlines between rows. This minimalist layout improves readability and draws attention to key data points.

In R, this format is often produced automatically using functions such as knitr::kable() or xtable() when rendering to Word or PDF.

## PART 2

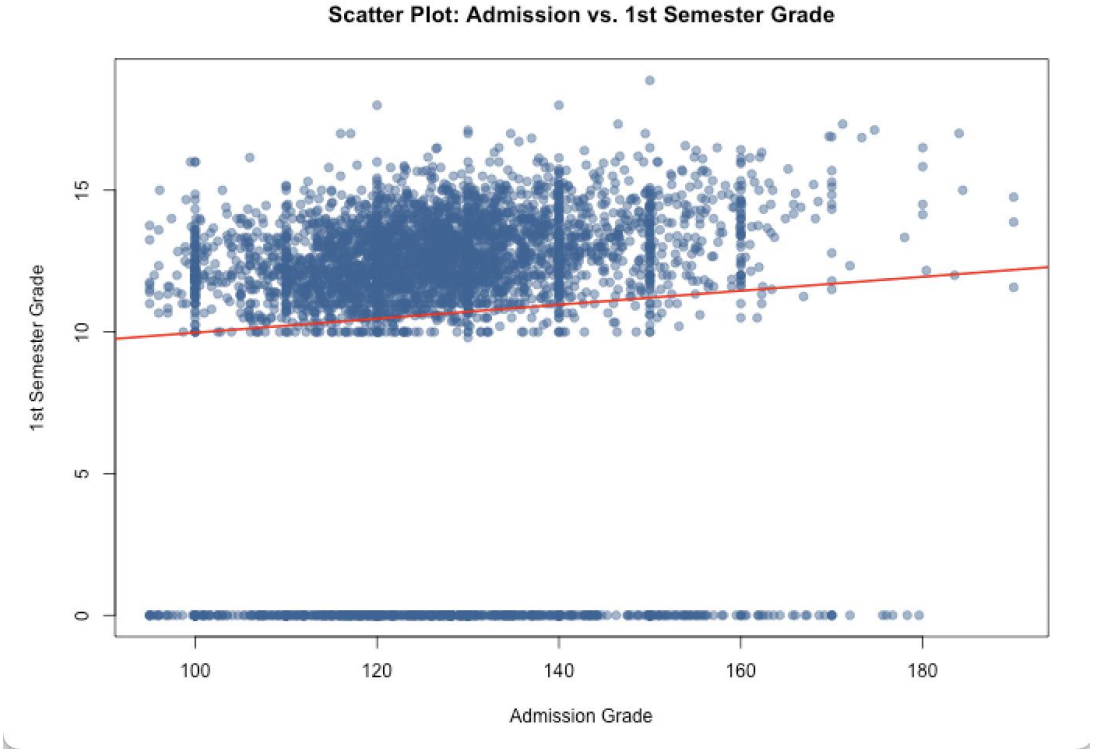
The use of png() in R enables the redirection of plot output to a PNG image file. This approach facilitates reproducibility and professional presentation of statistical visualizations in reports.

*1. Scatter Plot: Admission Grade vs. 1st Semester Grade*

This scatter plot shows the relationship between students' admission grades and their academic performance in the first semester. Each dot represents one student. A linear regression line (in red) is added using abline().

Purpose:

To examine the relationship between students' admission grades and their first-semester academic performance, and explore whether early academic readiness is associated with later success.




This scatter plot visualizes the relationship between Admission Grade (x-axis) and 1st Semester Grade (y-axis). Each dot represents a student. The red line is a linear regression line, added using abline(), which helps identify trends in the data.

*Key observations:*

- There is a slight positive linear trend — as admission grades increase, first-semester grades also tend to increase.
- The data points are densely packed between admission grades of 100 to 160, and first-semester grades of 10 to 15.

- A cluster of zeros in the bottom of the plot suggests a notable number of students with very low performance or possible dropouts.
- The overall weak slope of the red line indicates that although there is some correlation, it is not very strong, suggesting other factors may influence first-semester performance beyond just admission grades.

*Key finding:*

- There is a weak but positive correlation between admission scores and first-semester grades.
- However, the presence of many low-performing students with decent admission grades suggests that academic readiness alone is not a strong predictor of early success — other support factors likely play a critical role.

*2. Jitter Plot*

This jitter plot visualizes the distribution of scholarship holders (0 = No, 1 = Yes) across the three target groups: Dropout, Enrolled, and Graduate. We use jitter() to spread overlapping points, and par() to adjust margins.

Purpose:

To visualize the distribution of scholarship recipients across different academic outcomes (Target groups). The jittering helps reveal patterns in dense, overlapping data when using binary and categorical variables.


This jitter plot shows the relationship between students' scholarship status (0 = No, 1 = Yes) and their academic outcome (1 = Dropout, 2 = Enrolled, 3 = Graduate).

*Key observations:*

- Jittering was applied to both axes to reduce overplotting and make individual points easier to see.
- The data show a higher density of scholarship holders (1) in the Graduate group (3) compared to the Dropout group.
- The Dropout group appears to have more students without scholarships.
- This visual pattern suggests a potential positive relationship between receiving a scholarship and completing the program.
- The Enrolled group (2) shows a mixed distribution, likely including students still on their academic path.

This chart gives an intuitive visual cue that scholarship support may be associated with greater academic persistence and success.

*Key finding:*

- A visual pattern suggests that scholarship holders are more likely to graduate, while non-recipients are more common among dropouts.
- This supports the idea that financial support is a protective factor against dropout and may contribute to student success.

*When do you use a jitter chart?*

A jitter chart is used when there is a high chance that data points will overlap—for example, when variables are binary or categorical and many observations share the same values. In such cases, a regular scatter plot would make it hard to see the actual data distribution.

In this case:

- The x-axis (Scholarship Holder) is binary (0 = No, 1 = Yes), and the y-axis (Target) is categorical with three levels (1 = Dropout, 2 = Enrolled, 3 = Graduate).
- Without jittering, most points would be stacked on top of each other, making patterns invisible.
- By applying jitter() to both axes, this chart spreads the points slightly, allowing us to clearly observe:
    - More scholarship holders in the Graduate group
    - Fewer scholarship holders in the Dropout group
    - A mixed distribution in the Enrolled group

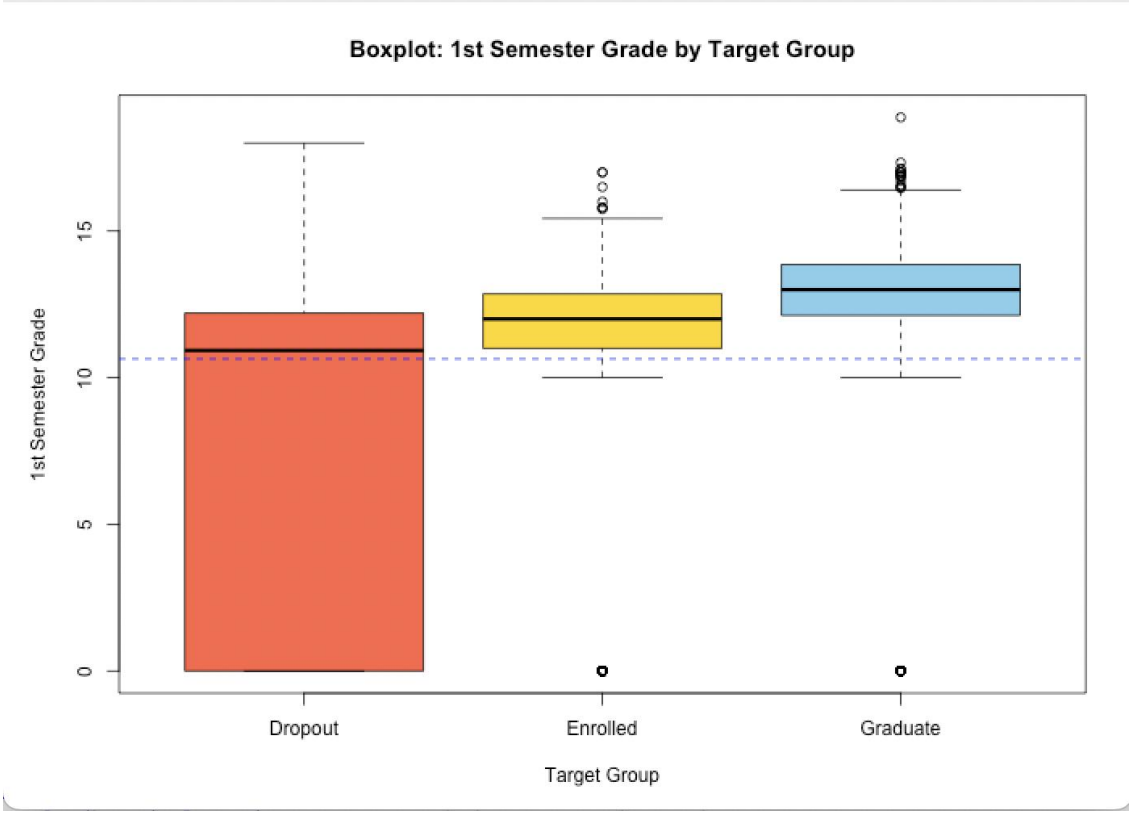
This visual reveals meaningful trends that would otherwise be hidden due to overplotting, making jitter charts ideal for this type of data structure.

*3. Boxplot*

The boxplot compares first-semester grades among Dropout, Enrolled, and Graduate students. It clearly shows differences in median, range, and the presence of outliers. A horizontal dashed line (abline()) represents the overall mean grade.

*Purpose:*

To compare the distribution of first-semester grades across the three outcome categories: Dropout, Enrolled, and Graduate. This allows us to detect differences in academic performance and identify outliers that may signal students at risk.




This boxplot compares the 1st semester grades of students across three groups based on academic outcomes:

- Dropout (left, red)
- Enrolled (middle, yellow)
- Graduate (right, blue)

*Key observations:*

- The Graduate group has the highest median grade, with most data concentrated in the upper grade range, indicating strong academic performance.
- The Dropout group shows the lowest median and a very wide range, with many students having grades near zero — suggesting academic struggles.
- The Enrolled group lies in between, with a more moderate and narrower distribution.
- Outliers are visible in the Enrolled and Graduate groups, shown as small dots above and below the whiskers.

- A horizontal dashed line (blue) represents the overall mean grade across all students, providing a reference point.

This visualization highlights how academic performance in the first semester is closely associated with students' eventual outcomes, with clear separation among the three groups.

*Key finding:*

- Academic performance in the first semester is strongly linked to student outcomes.
- Graduates had the highest and most consistent grades, while dropouts had the lowest and widest-spread performance. First-semester grades could be a valuable early indicator for identifying at-risk students.

*How can you use boxplots to detect outliers?*

Boxplots are useful for detecting outliers because they visualize:

- A boxplot shows the interquartile range (IQR) — the middle 50% of the data (from Q1 to Q3).
- The whiskers typically extend to 1.5 times the IQR beyond the first (Q1) and third quartiles (Q3).
- Any data point that lies outside the whiskers is considered an outlier and is plotted as a separate dot or circle.

In this case, in this boxplot of 1st Semester Grades by Target Group, outliers appear as individual dots plotted above or below the whiskers of each box. These points fall outside the typical range of values (i.e., beyond 1.5× the interquartile range from the box edges).

In this case:

- The Enrolled group shows several low outliers, including students with 0 grades, which may indicate risk of dropout or academic probation.
- The Graduate group has a few high outliers, which may reflect exceptional academic performance.
- Dropout students don't have visible outliers above, but show a cluster of very low or zero grades, which aligns with their status.

This use of the boxplot provides a quick visual flag for extreme values, helping institutions:

- Identify students who may be struggling early (e.g., 0-grade students in the Enrolled group).
- Detect patterns of exceptional or at-risk performance for intervention or support.

## Summary

The analysis highlights clear patterns across the three student outcome groups:

- Dropout students entered with lower admission grades, passed fewer courses, and showed inconsistent academic performance. They were also older on average and more likely to face financial delays.
- Enrolled students demonstrated moderate progress with mixed results. Many paid tuition on time, but grade performance varied, indicating they are at a turning point where support could influence final outcomes.
- Graduate students performed well academically, passed most courses, and maintained financial stability. They also received more scholarships, suggesting strong early preparation and institutional support.

Visual evidence from scatter plots, jitter charts, and boxplots confirms that early academic and financial indicators strongly correlate with final outcomes. First-semester performance, in particular, offers a reliable signal for identifying students at risk.

Overall, the findings support the use of predictive models based on enrollment-time data to guide early interventions. This approach can help institutions improve retention and student success rates through timely, targeted support.

## References

**UCI Machine Learning Repository**. (2021). *Predict students dropout and academic success*. University of California, Irvine, School of Information and Computer Sciences. https://archive.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success