

No Student Left Behind: Thanks to Data! 😊



🎓 Introduction

"Why do students drop out of school or university?" This is a question that is relevant to all educational systems worldwide. Every year, thousands of students across Portugal begin their educational journeys with the dream of graduating. Yet, a significant percentage never make it to the finish line. One of the most urgent issues facing both schools and universities is dropout, which can be caused by anything from financial difficulties to poor academic performance. In order to identify early indicators, forecast results, and suggest preventative measures, our initiative examines the dropout phenomena from the perspectives of both university and high school students.

💡 Were you aware? In Portugal, academic failure remains the primary cause of school dropouts, while financial difficulties remain the primary cause of university dropouts, according to recent Harvard Kennedy School research.

❓ Problem Statement

Conventional methods of studying dropout rates frequently only look at the school or university level. Furthermore, prior studies have tended to focus on qualitative case studies or simplified statistical summaries. Our goal was to surpass this by developing a data-driven comparison study that makes use of machine learning and visualizations in order to:

- Examine two different age groups: university students (median age 19–20) and schoolchildren (median age 17).
- Make use of Portuguese real-world datasets.
- To identify, forecast, and prevent dropouts, use Tableau dashboards, data science, and machine learning.
- Estimate the risk of student dropout depending on the features that are provided.
- Examine and contrast the elements that influence pupils in school and those at university.
- Make data- and research-supported intervention strategy recommendations.

Our innovative approach allows for a side-by-side analysis of trends at the school and university levels and allows for the establishment of targeted initiatives for all educational levels.

Methodology

a. Data Collection & Tools Used

A big dataset from Kaggle was used to examine the factors that contribute to university student dropout in Portugal. The dataset included more than 4,400 students' academic records, demographic information, socioeconomic indicators, and institutional characteristics. We maintained a final working sample of 3,374 university students with an age range of 18–25 years and the High School Student Dataset with 649 students of median age 17 after filtering the Portuguese nationality.

Tools and Platforms:

- The R programming language was utilized for:
 - Preprocessing and data cleaning
 - Exploratory data analysis (EDA)
 - Outlier detection using the Interquartile Range (IQR) approach
 - Selection and encoding of features
 - Modeling and assessment of machine learning
- Using Tableau:
 - interactive dashboards were created
 - Showing dropout trends and the connections between variables
- **Google Collab (R Kernel)** was the place where the analysis was performed.
- **MS Word/Google Docs** for documentation and reporting.

b. Analysis and Machine Learning Models

The analysis followed a systematic and standardized data analysis pipeline:

1. Preprocessing Steps:

- **Missing Value Check:** There were no missing values in either dataset.
- **Outlier Detection & Removal:** Performed the IQR method to detect and remove outliers in numerical variables such as grades, age, and economic indicators.
- **Correlation Check:** Highly correlated features (correlation > 0.85) were removed to reduce redundancy.
- **Categorical Encoding:** Variables like Gender, Marital Status, and Nationality were encoded into numeric or factor formats.
- **Feature Standardization:** All numeric features were scaled using standardization to prepare for distance-based models like KNN.
- **Filtering:** Focused on Portuguese university students aged between 18 and 25, and school students having median age of 17.

2. Class Distribution Analysis:

In the target variable (goal: Dropout, Enrolled, Graduate), we found a class imbalance (University students). Although dropout was the main emphasis, balanced classification was important. We displayed the distribution and observed the difficulty of identifying "Enrolled" students, who are in transition.

3. Machine Learning Modeling:

Five supervised classification models were trained and tested:

- The interpretable linear model known as **logistic regression**
- **Decision Tree:** A simple, rule-based framework
- For improved generalization and feature significance, we used the **Random Forest** ensemble approach.
- **The Naïve Bayes** classifier is based on probability and assumes feature independence.
- **The K-Nearest Neighbors (KNN)** model is instance-based and depends on distance measures.

The following metrics were used to assess each model after it was trained using an 80/20 train-test split:

Confusion Matrix; Accuracy; Sensitivity (per class: Dropout, Enrolled, Graduate); Interpretability & robustness

-  **Top-Performing Model: Random Forest (~85% accuracy) is the Best Model (University).**

Performance Indicators: 78.03% accuracy and about Class Sensitivity we have:

- Dropout: 75.7%
- Enrolled: 70.6%
- Graduate: 88.9%

These findings show that all three classes performed fairly, with particular emphasis on how well it identified students who were likely to graduate or drop out.

Random Forest is the best model (school) with 100% accuracy.

Random Forest: Why??

- It is resistant to overfitting, especially when the dataset contains many variables, and it captures intricate, non-linear interactions between features.
- It provides transparency and interpretability through built-in feature importance.

Top Dropout Predictors (by feature importance):

- **Debtor Status:** Students who owed money for tuition had a far higher chance of dropping out.
- **Approved Curricular Units:** Dropping out was significantly associated with fewer courses completed.
- **Status of Tuition Payment:** Students who had out-of-date records of their tuition fees were more vulnerable.
- **Gender:** Male students were more prone to drop out compared to females, which is consistent with both the statistics and literature.

This model is a crucial decision-support tool for early intervention efforts because it not only provides actionable insights but also significant predictive power.

Findings and Outcomes

a- University Students (Filtered):

- 3,374 in total.
⇒ There were 371 (11%) dropouts; 1,676 graduates (49.7%) and 517 (15.3%) enrolled.

b- School Students (Filtered):

- 649 in total.
⇒ There were Dropouts: 100 (15.4%) ; 549Enrolled (84.6%) and no Graduates.

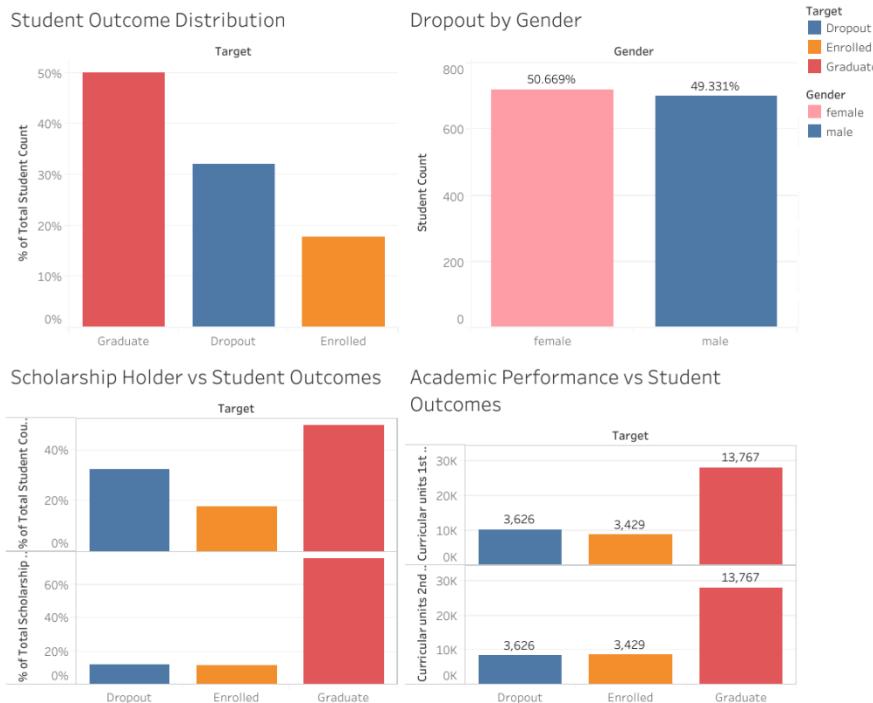
Common Dropout Patterns:

- **Gender:** At the university level, males are more likely to drop out, however trends vary by school.
- **Financial Aspects:** Dropout rates were noticeably higher among students who owed money.
- **Parents Education:** Students were more vulnerable if their parents had less education.
- **Academic Achievement:** Dropout was strongly associated with subpar performance in core areas.

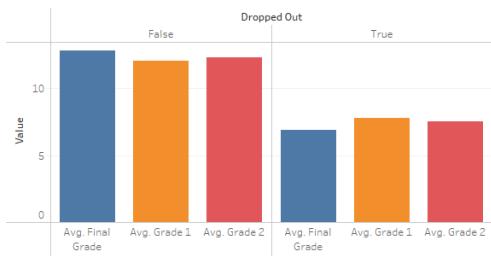
Visual Dashboards

- To help us identify trends more quickly and explain the findings to non-technical audiences, we developed dashboards in Tableau that represented:
 - Dropout rates by gender, qualification, and curriculum performance.
 - Financial and parental factors.
 - Comparisons between school and university dropout profiles.

Student Outcome Overview

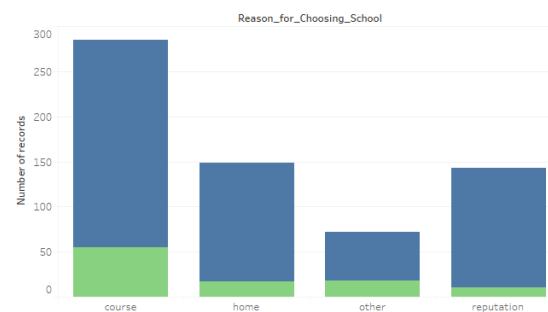


Dropout rates vs grades



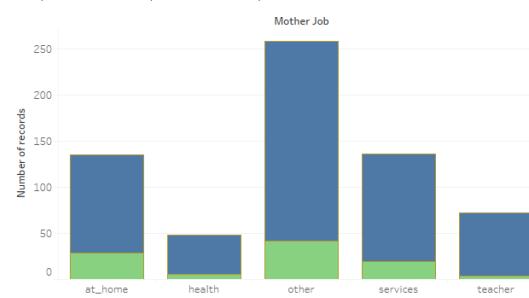
Measure Names
 Avg. Final Grade
 Avg. Grade 1
 Avg. Grade 2

Dropout rates vs reason of choosing school



Dropped Out
 False
 True

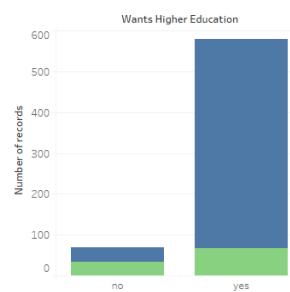
Dropout rates vs parent's occupation



We can deduce a few insights from this dashboard:

- 1- Dropout rates are always associated with lower grades.
- 2- Dropout rates are the lowest when parents are in the teaching/ health sector
- 3- dropout rates tend to increase dramatically when students pick schools based on a course.
- 4- dropout rates show a significant decrease when students have goals of pursuing higher education

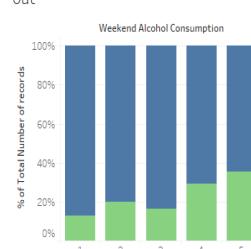
Dropout rates vs wanting higher education



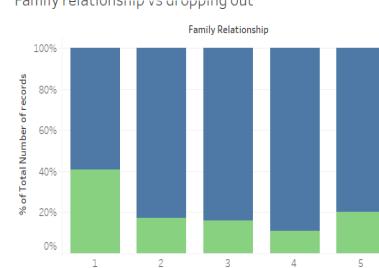
Mother's education vs dropping out



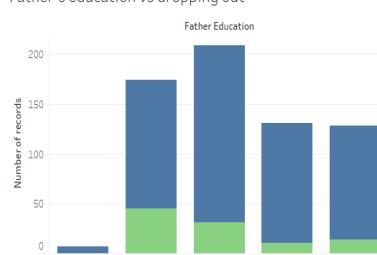
Alcohol consumption vs dropping out



Family relationship vs dropping out



Father's education vs dropping out



We can determine from these graphs that:

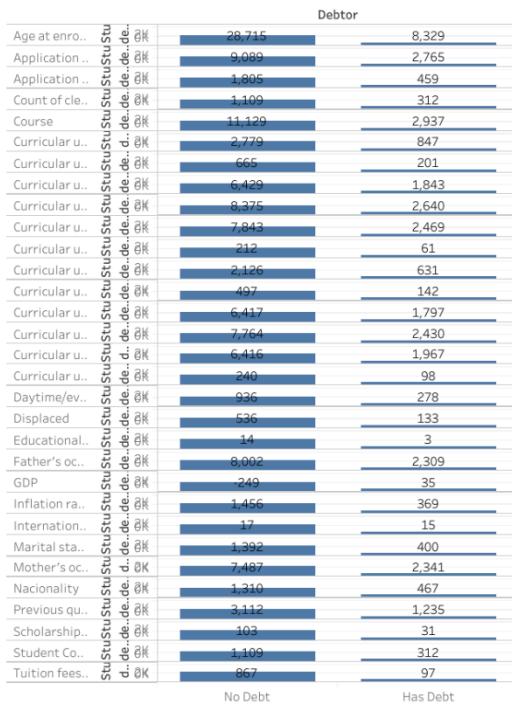
The higher the mother's and father's education, the less likely a student will drop out.
 The higher the alcohol consumption, the more likely a student will drop out.
 Good family relationships decrease chances of dropping out, and high previous class failures increase them.

Number of failures vs dropping out

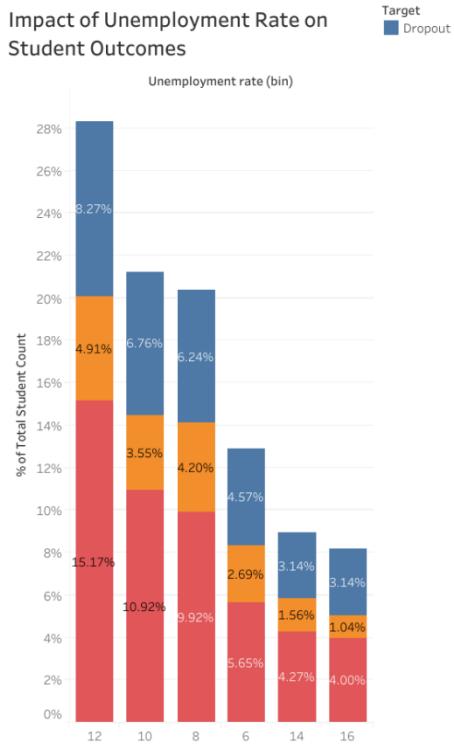


Financial & Economic Factors

Financial Burden Analysis

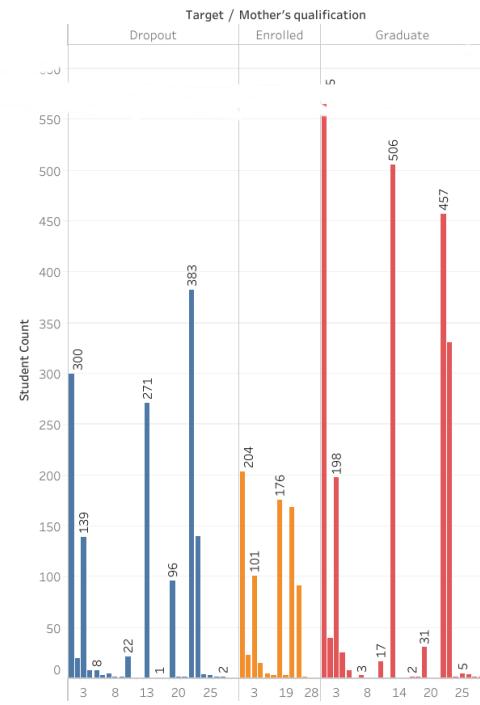


Impact of Unemployment Rate on Student Outcomes

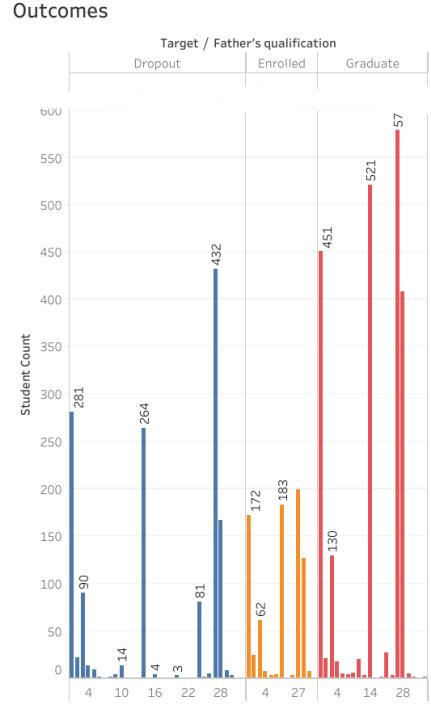


Parental Education & Student Success

Parental Qualification vs. Student Outcomes



Parental Qualification vs. Student Outcomes





The “Don’t Panic, Just Do This Survival Kit” List

For University Students:

- 📌 Create early warning systems according to debt and grades.
- 📌 Expand financial aid and scholarship opportunities.
- 📌 Establish academic and psychosocial support facilities.

For School Students:

- 📌 Family counseling services are offered at neighborhood community centers or schools.
- 📌 To create a steady support network, pair students without parent guardians with committed mentors or "school buddies."
- 📌 Lead peer or local professional-led school-based substance abuse awareness campaigns.



Three-Year Retention Framework



Based on the findings from our predictive analysis and literature study, we offer a complete 3-year-data-driven retention strategy for higher education institutions in Portugal

- ◆ **First Year: Proactive Recognition:**
- **Put in place Early Alert Systems** that identify at-risk pupils based on demographic, economical, and academic characteristics utilizing predictive models (such as Random Forest).

- **Institutional Screening:** Consistent audits to identify deteriorating academic performance and financial risk (such as unpaid tuition).

- ◆ **Second Year: Specific Actions**

- **Tailored Academic Support:** Peer-assisted learning and tutoring for kids who perform poorly in classes.
- **Debt Relief & Financial Aid Navigation:** Helping students who are in debt find assistance or restructure their payments.
- **Mentorship Programs:** To increase motivation and a sense of community, pair at-risk students with faculty advisors or senior peers.

- ◆ **Third Year: Systemic Change and Policy**

- **Increase Grant Accessibility:** Update the requirements for scholarships to include more at-risk and middle-class students.
- **Require Data Tracking:** For retention analytics, mandate that colleges keep complete, anonymized student datasets.
- **Regional Assistance Centers:** To offer financial and educational support, set up community-based outreach centers in areas with high dropout rates.



Final Thoughts Before We Drop Out :p

One key aspect in dropouts, as demonstrated by numerous school and university dropout insights, is a **lack of guidance**. Because they chose the incorrect major or, worse, the improper course of study, students frequently end up dropping out of college. Important implementations that could stop university dropouts from beginning their education will be covered in this section.



Find Your Path, No Map Required



As a high school student, contemplating your future and profession might be, to put it mildly, daunting. Without much direction, students will frequently choose the most popular majors with a strong market presence as the safest option. But they might not be passionate about that.



WHAT WE CAN DO? Here's the Plan (Spoiler: It's Good)



A student-centered career advisory program is vital, especially for senior students. Students are more likely to make well-informed judgments that complement their long-term objectives and passions if they are exposed to a variety of industries, career

routes, and real-world applications of diverse degrees at an early age. Furthermore, our research indicates that students who aspire to pursue higher education are much less likely to drop out. This could be accomplished by letting students develop life goals based on their preferences and investigate career options early in life.   

Additionally, some students do not find themselves qualified for university at all. Even after dropout rates have been reduced to a minimum, some students will still leave because they feel like they belong somewhere else. For students who would rather pursue careers in community colleges, blue-collar jobs, or even entrepreneurship, a career-guidance program should also offer an alternative route.

 In conclusion, creating a guidance program for high school students has several advantages, such as:

- ⇒ Lessen students' anxiety and uncertainty about potential career paths.
- ⇒ Promotes alignment between students' academic objectives and their interests.
- ⇒ Introduces students to a variety of professional pathways and real-world applications.
- ⇒ Reduces college dropout rates by assisting students in making educational decisions early.
- ⇒ Assist students in establishing long-term, purpose-driven objectives for their lives.



References

- Harvard Kennedy School. (n.d.). *Addressing the completion challenge in Portuguese higher education*. Retrieved from <https://www.hks.harvard.edu/>
- Macau Business. (2022, March 7). *Portugal: Only very poor get scholarships so many students give up study*. Retrieved from <https://www.macaubusiness.com/portugal-only-very-poor-get-scholarships-so-many-students-give-up-study/>
- Moreira da Silva, D. E., Solteiro Pires, E. J., Reis, A., de Moura Oliveira, P. B., & Barroso, J. (2022). Forecasting students dropout: A UTAD University study. *Future Internet*, 14(3), 76. <https://doi.org/10.3390/fi14030076>
- ResearchGate. (n.d.). *Higher education dropout in northern Portugal*. Retrieved from <https://www.researchgate.net/>
- European Commission. (2024). *Education and Training Monitor 2024*. EUR-Lex. Retrieved from <https://eur-lex.europa.eu/>
- Organisation for Economic Co-operation and Development. (n.d.). *Education at a glance*. Retrieved from <https://www.oecd.org/education/education-at-a-glance/>