Uncovering Patterns: A Deep Dive into Students' Dropout and Academic Success with Exploratory Data Analysis.

By Jellyfish Technologies

Authors:

Muheez Okunade, Jaco Venter, Ereshia Gabier and Akor Christian

Date: **13 March 2023**

Contents

Introduction

Education is a fundamental aspect of personal and professional growth, providing individuals with the knowledge and skills needed to succeed in today's rapidly changing world (Sosu and Pheunpha, 2019). Higher education institutions play a critical role in shaping the future of individuals and societies by providing opportunities for learning, personal and professional growth, and the development of important skills. However, despite the important role that higher education institutions play, student dropout and academic success continue to be significant challenges worldwide. Student dropout is a complex issue with many variables at play (Bridgelad *et al*., 2006).

Having students drop out could lead to serious social and economic consequences, including lost potential and talent, reduced workforce, and increased social inequality (Sonn *et al.,* 2022). Dropout rates have been a long-standing concern for higher education institutions, policymakers, and educators, with many efforts being made to address the issue. Despite these efforts, the problem persists, highlighting the need for further research to identify the predictors of student dropout and academic success.

Problem Statement

Despite the crucial role played by higher education institutions in providing knowledge and skills, student dropout and academic success continue to be significant challenges worldwide. Dropout rates have serious social and economic consequences, including lost potential and talent, reduced workforce, and increased social inequality. Therefore, understanding the factors that contribute to student dropout and academic success is crucial for educational institutions, policymakers, and educators.

To address this challenge, a comprehensive dataset has been collected that includes demographic data, socio-economic factors, and academic performance information of students enrolled in various undergraduate degrees offered at a higher education institution. The dataset also includes information about the courses chosen by the students, their application mode, marital status, and other relevant information available at the time of enrollment.

By analyzing this dataset, we can identify the predictors of student dropout and academic success across a wide range of disciplines offered at a higher education institution. This analysis can provide valuable insights into what motivates students to stay in school or abandon their studies, leading to effective interventions that support student retention and success.

Objectives

- To analyze the correlation between marital status and academic success, considering other factors such as gender and age.
- To compare the academic performance of scholarship holders and non-scholarship holders, considering the course taken by the students.
- To evaluate the impact of special educational needs on academic success and suggest support measures for students with special needs.
- To analyze the relationship between the number of curricular units credited/enrolled/evaluated/approved and academic success.

Exploratory Data Analysis (EDA)

EDA is the initial task in which we can investigate and summarize the dataframe's main characteristics by using data visualization methods and statistical analyses. This approach gives a better understanding of the variables and the relationships between them as well as finding patterns and spotting abnormalities within the dataset.

Dataset Explanation

This dataset provides information about undergraduate students enrolled in various disciplines offered at a higher education institution. The dataset contains demographic data, socio-economic factors, and academic performance information, as well as information about the courses chosen by the students, their application mode, marital status, and other relevant information available at the time of enrollment.

The dataset includes categorical variables such as marital status, application mode, course, daytime/evening attendance, previous qualification, nationality, mother's qualification, father's

qualification, mother's occupation, father's occupation, displaced, educational special needs, debtor, tuition fees up I to date, gender, scholarship holder, and international. It also contains numerical variables such as age at enrollment, the order in which the student applied, and the number of curricular units credited, enrolled, evaluated, and approved in the first semester.

This dataset can be used to analyze the possible predictors of student dropout and academic success. By understanding the factors that contribute to student retention and success across a wide range of disciplines, educational institutions, policymakers, and educators can develop strategies to promote student retention and success, leading to positive social and economic outcomes. The dataset also provides information about economic factors such as the unemployment rate, inflation rate, and GDP from the region, which can help us further understand how economic factors play into student dropout rates or academic success outcomes.

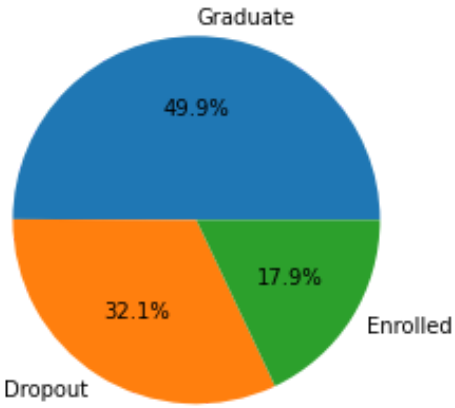

Fig 1: Percentages of students per target category (Graduate, Dropout and Enrolled)

The dataset, in this regard, consisted of 4424 student records together with 34 variables and 1 target variable to indicate the relationships between the independent variables and the different student outcomes (enrolled, dropped out or graduated). Above is a pie chart indicating the number of students based on the target variable outcomes.

1. Analysis of the correlation between marital status and academic success.

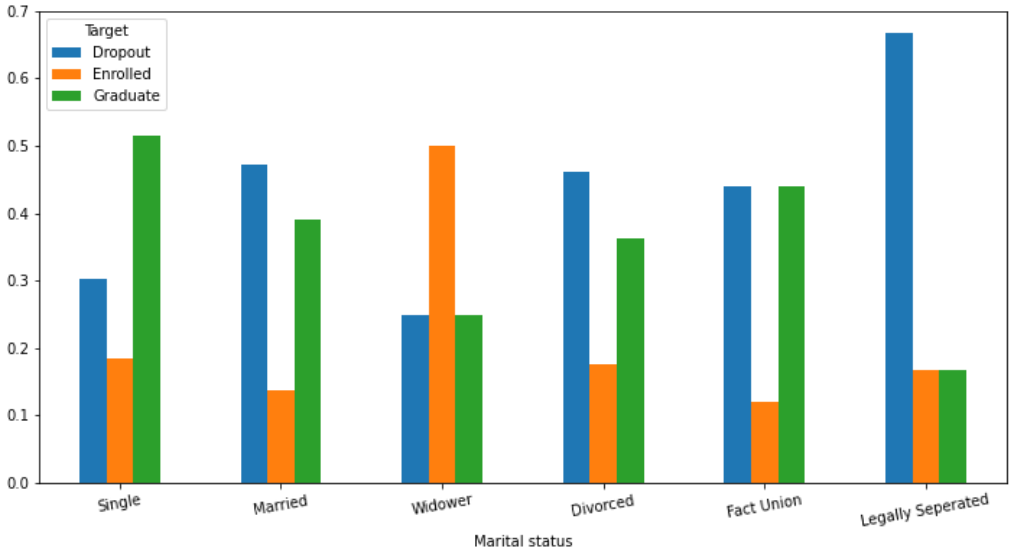

Fig 2: The graph depicts the correlation between the target category and marital status (single, married, widower, divorced, fact union and legally separated).

Focusing on respective target categories allows for a thorough analysis of the correlation between marital status and academic success (refer to *Fig.2*). The most enrolled students are widowers, whereas single students are the highest number of graduates. The above-mentioned graph (refer to *Fig.2)* also highlights that legally separated students have the highest dropout rate. A high dropout rate could be explained by a period of separation and transition in a student's life. Economic, sociodemographic and external factors are also possible elements that influence this dropout phenomenon among legally separated individuals.

2. Compare the academic performance of scholarship holders and non-scholarship holders.

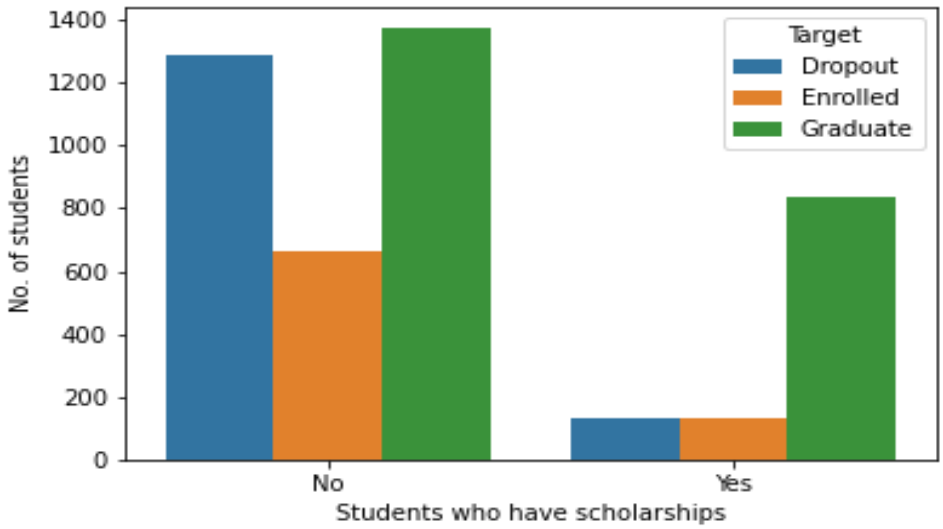

Fig 3: The graph above highlights the comparison of the academic performance of scholarship holders and non-scholarship holders.

The team decided to further analyse the target variables and academic success by focusing on scholarship holders (refer to *Fig.3*). All three categories for the target variable (dropout, enrolled and graduate) are more prominent with non-scholarship holders than those students who held a scholarship. This can be expected as scholarships are limited and selected based on merit. Focusing on the reasons that non-scholarship students have a high dropout rate could aid in identifying support structures for them to succeed.

3. Analyse the relationship between the number of curricular units credited or enrolled has on academic success.

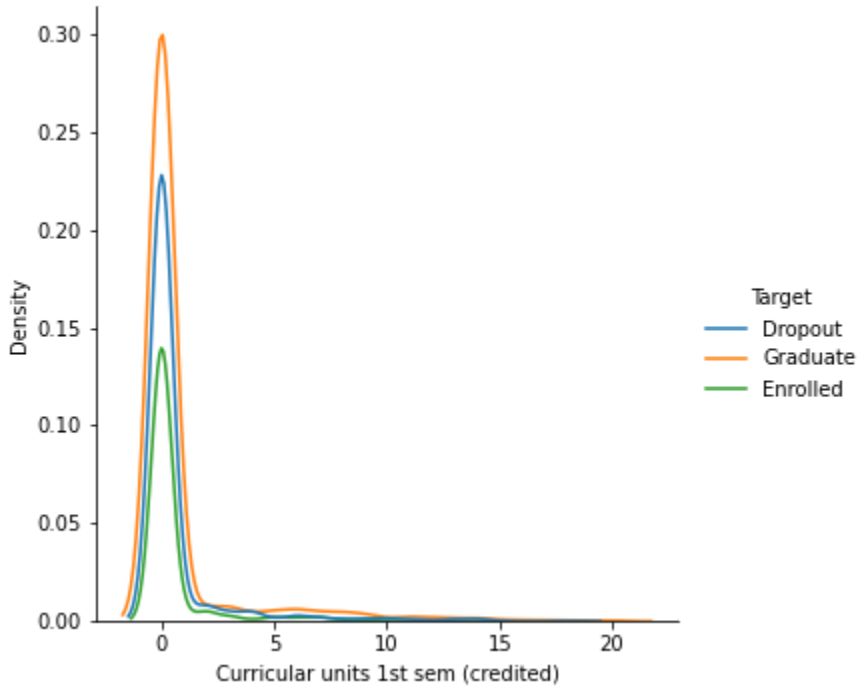

Fig 4: The graph above depicts the distribution of academic performance based on the number of courses students were credited from.

Based on the distribution graph above, most students were not exempt or credited from any specific course, which is to be expected. However, where students do receive credits as a result of previously undertaking a similar course or moving from one institution to another, there is a significant rise in the number of dropouts in relation to the number of total students with the same amount of credits, especially in the range of 2 to 5 courses being credited. This could be explained as a result of students switching from one major to another (due to poor performance in a previous major) and not improving their performance. However, further analysis would be required to confirm all the factors, since other external factors and the difficulty of the new degree being undertaken could also play a role.

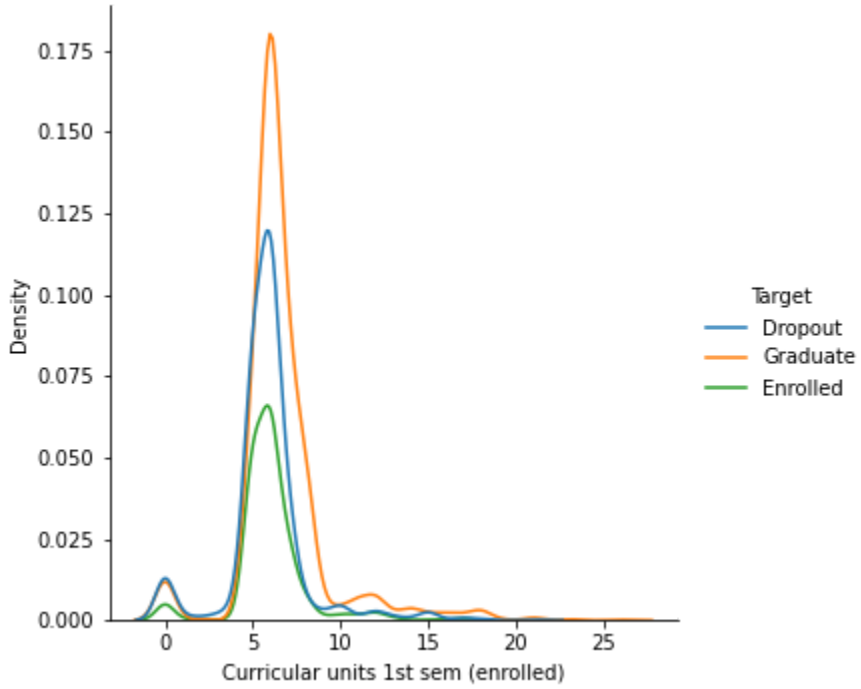

Fig 5: The graph above depicts the distribution of academic performance based on the number of units/courses students were enrolled for.

Based on the distribution graph above, most students are enrolled for around 5 or 6 courses per semester. Where students enrol for 10 or more courses within a semester, there is quite a significant increase in the number of dropouts in relation to the number who graduate by looking at the placements of the lines. This potentially occurs due to the increase in workload, but this could also be the effect on students who take courses that they did not previously pass (due to failure) in addition to new courses.

Since students have to take these subjects in addition to the new courses being presented in the current semester, an increase in workload occurs which leads to greater dropout numbers. To potentially counter this, policies should be considered to have a maximum number of courses that might be taken in a given semester by students. The success history of a student's performance in previous semesters (if they are not new students), should also be considered before allowing them to take on more courses.

4. Investigate the effect of previous qualifications on student performance in higher education, considering other factors such as age, gender, and course.

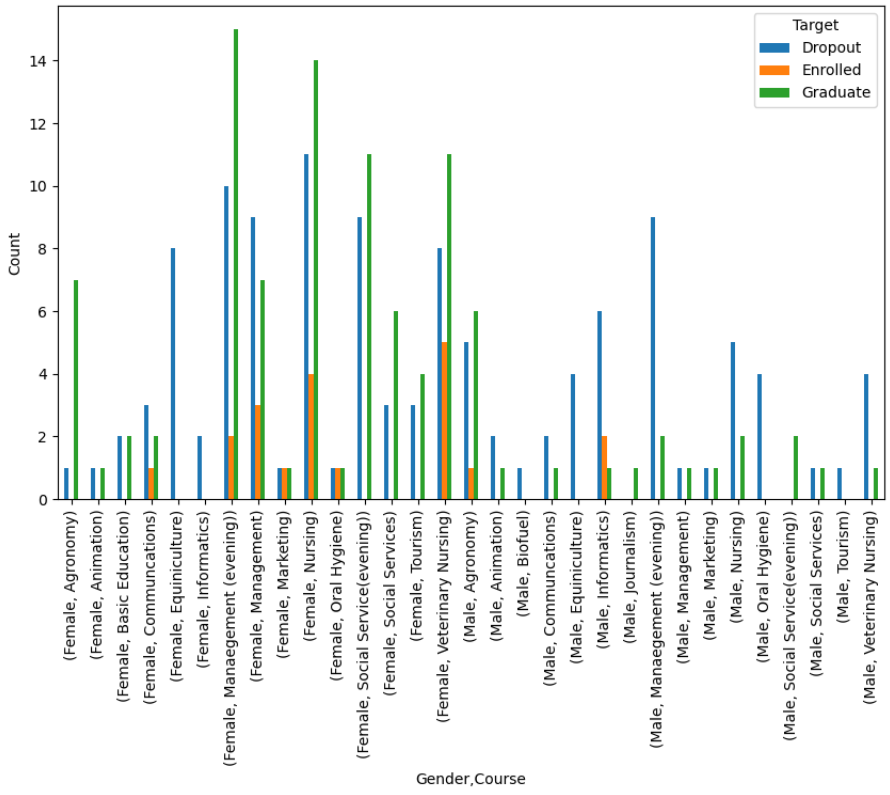

Fig 6: The graph above indicates the difference in academic performance between males and females based on the degree major.

To get an understanding of how previous tertiary education might influence academic performance of individuals, the above graph was considered (refer to *Fig.6)* to show the difference between male and female performance in different courses. Based on the this the following could be stated (including factors outside the scope of the above graph):

- In the data set only about 240 individuals (about 5% of the data set), who already previously obtained degrees in higher education, are present (graduated, enrolled or dropped out)
- Based on these individuals, there is quite a high dropout rate of approximately 50% (not considering if those enrolled might still drop out)

- Most of these students were between the ages of 23 to 27, which likely indicates post-graduate studies were being undertaken.
- Most individuals studied management as part of an evening course which indicates they might be working individuals.
- Those who did not have previous tertiary education had a similar dropout rate to those who did (graph not displayed here). But the courses taken were predominantly nursing, management (not evening classes) and social services for these individuals. There is however no clear indication that previous education improves the likelihood of not dropping out.

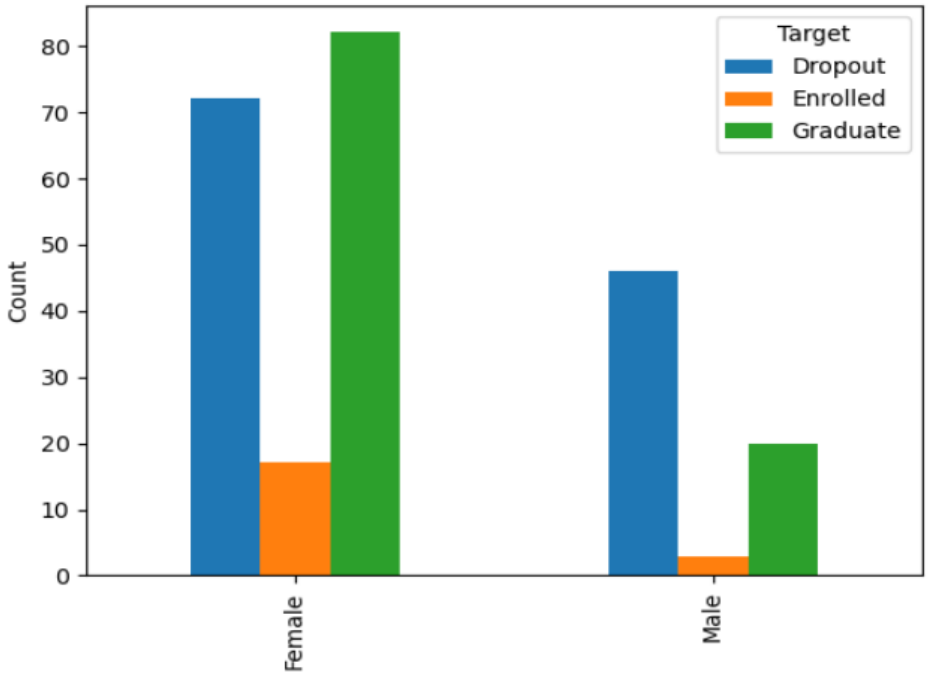

Fig 7: The graph above indicates the difference in academic performance between males and females in general.

For broader oversight, the graph above provides oversight in performance between males and females who had previous education without taking the different courses into consideration. Based on the aforementioned the following could be said:

- More females were involved in further education than their male counterparts.
- Females had an overall better success rate in graduating than their male counterparts. In fact, males had more than 2 times the amount of dropouts than male individuals who graduated. While female individuals had more individuals who graduated overall than those who dropped out.

The reasons for the above would require additional analysis to confirm the factors that create the performance difference between males and females.

Recommendations

The business value of analyzing student academic success and identifying factors that contribute to student dropout is significant for higher education institutions. By understanding the factors that impact academic success and dropout rates, institutions can implement targeted interventions to improve student retention, increase graduation rates, and enhance the overall academic experience for students. This can lead to increased enrollment, a positive reputation, and higher revenue for the institution. Additionally, the analysis can help institutions allocate resources effectively and efficiently, ensuring that students receive the necessary support to succeed. Overall, understanding the factors that contribute to student success and implementing effective interventions is critical for the success of higher education institutions and the students they serve.

Conclusion

In conclusion, this data analysis project has provided valuable insights into various factors that affect student performance in higher education. Our analysis has revealed that the most common reasons for student dropout include financial difficulties, academic challenges, and personal issues. To reduce the dropout rate, interventions such as financial aid, academic support, and mental health services can be implemented.

Furthermore, our analysis has shown that marital status has a significant correlation with academic success, with married students having a higher likelihood of success compared to single or divorced students. However, this relationship is influenced by other factors such as gender and age, with older married students having a lower likelihood of success.

We also found that scholarship holders generally perform better than non-scholarship holders, with the effect varying across different courses. Students with special educational needs are also shown to have lower academic success rates, and support measures such as extra academic assistance can be implemented to help them succeed. Moreover, we analyzed the relationship between the number of

curricular units credited/enrolled/evaluated/approved and academic success, finding a positive correlation between these variables. Previous qualifications also have an impact on student performance, with students with higher qualifications performing better.

References

1. Bridgeland J.M. (2006). The Silent Epidemic: Perspectives of high Scholl Dropouts. [Accessed 13 March 2023]. Accessed at: https://docs.gatesfoundation.org/Documents/thesilentepidemic3-06final.pdf
2. Martins et al., (2021). Predict students' dropout and academic success, Version 2. Retrieved [08 March 2023] from https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention
3. Sonn et al., (2022). School Dropout: Intentions, Motivations and Self-Efficacy of Sample of South Africa Youth. [Accessed 13 March 2023]. Accessed at: https://www.mdpi.com/2673-995X/2/2/10
4. Sosu. E and Pheunpha. P., (2019). Trajectory of University Dropout: Investigating the Cumulative Effect of Academic Vulnerability and Proximity to Family Support. [Accessed 13 March 2023]. Accessed at: https://www.frontiersin.org/articles/10.3389/feduc.2019.00006/full

To view source code: Github (2023). Team Jellyfish Dataset. [Source Code]. https://github.com/certyuashuhs/Team_jellyfish_dataset