

January 2024

Academic Drop Out Analysis - Introduction and Exploratory Analysis

Diagnostic and Predictive Analysis of Student Dropouts and Academic Success

Python, Microsoft Excel, Tableau

Introduction:

The dataset contains information on students and their marital status, application details, previous academic qualifications, socioeconomic status, admission grades, and academic performance. The target variable indicates whether a student dropped out or achieved academic success. Using this dataset, a predictive model will be developed tailored to forecast dropout rates among incoming students, in order to facilitate support to those who may be at risk. The data also incorporates economic metrics such as unemployment rate, inflation, and GDP. This model will contribute to the understanding of these factors influencing educational outcomes.

Objectives:

1. Comprehensive Exploratory Data Analysis to understand significant factors influencing academic performance
2. Create classification model for predicting student dropout and academic success
3. Create a predictive model that can be used on incoming students

Data Collection:

Dataset from Kaggle:

<https://www.kaggle.com/datasets/missionjee/students-dropout-and-academic-success-data-set>

Most of the data is categorical converted to numerical, VLOOKUP tables used for data cleaning for Tableau (ex. '1' displayed in column is equivalent of "Secondary Education")

About the data:

4424 entries, 37 features

Dataset includes information about:

Students' demographic

- Age
- Gender
- Marital status
- Nationality

- Displaced (Y/N)
- International (Y/N)

Students' Socioeconomic information includes

- Mother and Father occupation and qualification
- Educational special needs
- Debtor (Y/N)
- Tuition fees up to date
- Scholarship holder

Students' academic information includes

- Application mode
- Application order
- Course
- Daytime/evening attendance
- Previous qualification and grade (last level of education completed)
- Admission grade
- 1st semester units credited, enrolled, evaluations, approved, grade
- 2nd semester units credited, enrolled, evaluations, approved, grade
- Dropout or Graduate

Macroeconomic information includes

- Unemployment rate
- Inflation rate
- GDP

Data Overview

> *print(acaddropout.head())*

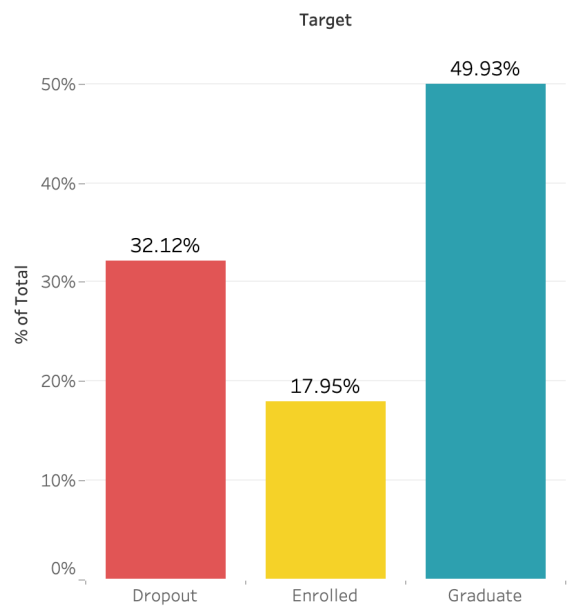
	Marital status	Application Mode	...	GDP	Target
0	Single	Short cycle diploma holders	...	1.74	Dropout
1	Single	Technological specialization diploma holders	...	0.79	Graduate
2	Single	1st phase - general contingent	...	1.74	Dropout
3	Single	Short cycle diploma holders	...	-3.12	Graduate
4	Married	Change in institution/course (International)	...	0.79	Graduate

Data Cleaning and Preprocessing:

For Tableau and Python

1. VLOOKUP tables for categorical columns, convert numerical categorical to string categorical

Exploratory Data Analysis:
50% Graduates, 32% Dropouts, 18% Enrolled



Males have a higher dropout rate than females

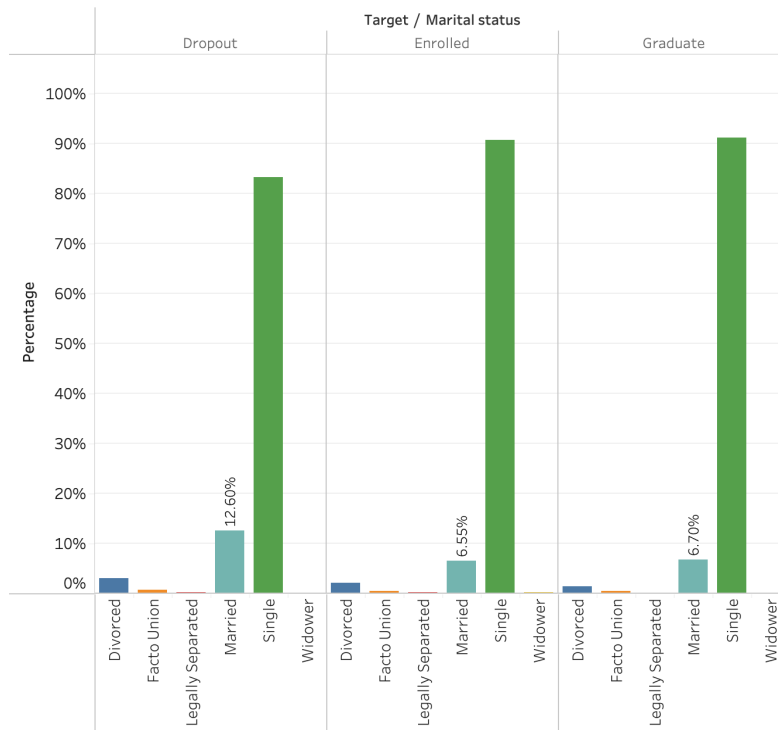
Gender	Target		
	Dropout	Enrolled	Graduate
Female	25.10%	16.98%	57.91%
Male	45.05%	19.73%	35.22%

Dropout rate increases as age increases

Age at enrollment (2 yr bins)	Target		
	Dropout	Enrolled	Graduate
16		40.00%	60.00%
18	21.01%	16.90%	62.10%
20	24.54%	22.80%	52.66%
22	35.11%	21.63%	43.26%
24	45.98%	22.77%	31.25%
26	57.84%	16.76%	25.41%
28	61.74%	13.42%	24.83%
30	63.46%	12.50%	24.04%
32	50.00%	12.26%	37.74%
34	53.21%	15.60%	31.19%
36	57.14%	10.39%	32.47%
38	55.22%	13.43%	31.34%
40	44.44%	16.67%	38.89%
42	47.37%	13.16%	39.47%
44	53.49%	13.95%	32.56%
46	56.67%	6.67%	36.67%
48	58.33%	16.67%	25.00%
50	60.87%	8.70%	30.43%
52	27.27%	9.09%	63.64%
54	58.33%	8.33%	33.33%
56	50.00%		50.00%
58	66.67%		33.33%
60	33.33%		66.67%
62			100.00%
70	100.00%		

Marital status - double the normal dropout rate among married students

Marital Status



Scholarship Holder - students holding scholarships have lower dropout rate

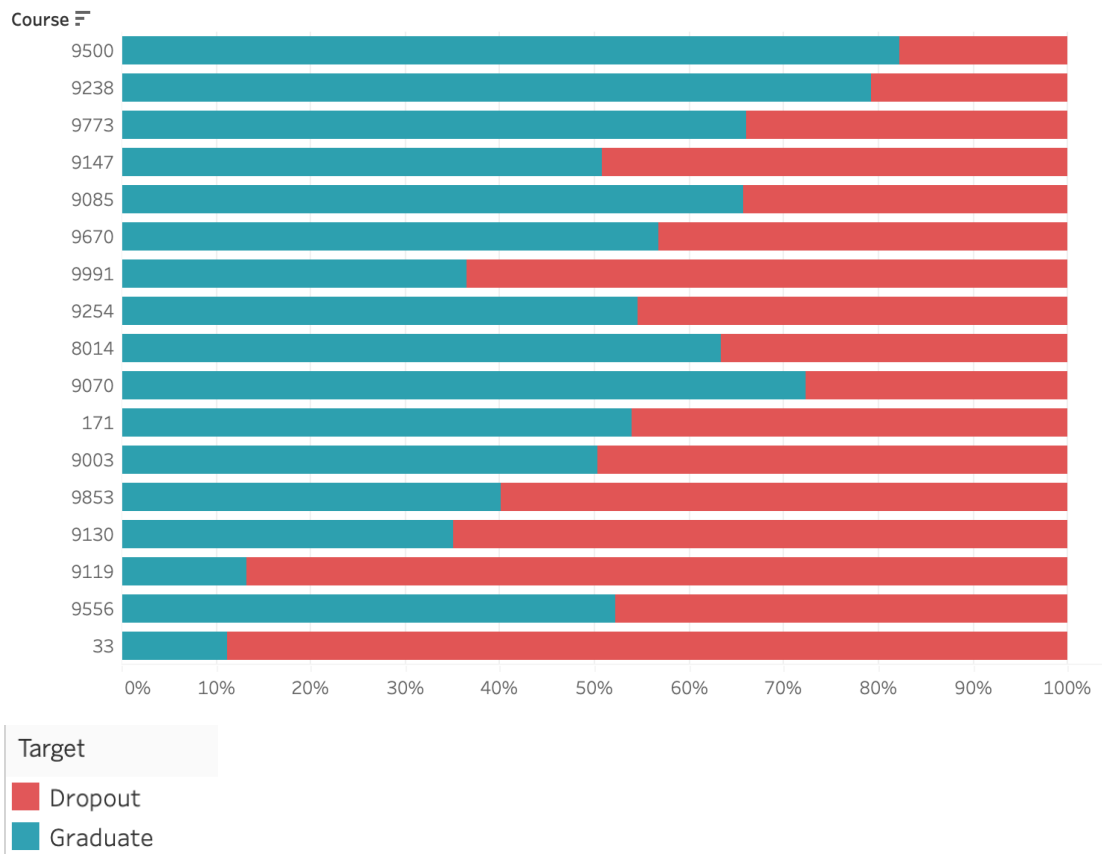
Scholarship holder	Target		
	Dropout	Enrolled	Graduate
No	38.71%	19.97%	41.32%
Yes	12.19%	11.83%	75.98%

Debtor - students in debt are at higher risk of dropout

Debtor

Debtor	Target		
	Dropout	Enrolled	Graduate
No	28.28%	17.95%	53.76%
Yes	62.03%	17.89%	20.08%

Courses vs Graduate and Dropout Percentage - Course 33 and 9119 have a high dropout ratio



Displaced - displacement does not significantly affect dropout, however students who are not displaced have a higher dropout rate

Displaced

Displaced	Target		
	Dropout	Enrolled	Graduate
No	37.64%	18.07%	44.29%
Yes	27.58%	17.85%	54.58%

Daytime/Evening Attendance - attendance time does not significantly affect dropout, however evening students have a higher rate

D/E attendance

Daytime/ev..	Target		
	Dropout	Enrolled	Graduate
Daytime	30.80%	18.24%	50.95%
Evening	42.86%	15.53%	41.61%

Admission Grades (in 5 point bins) - The higher the admission grade, the lower the dropout rate

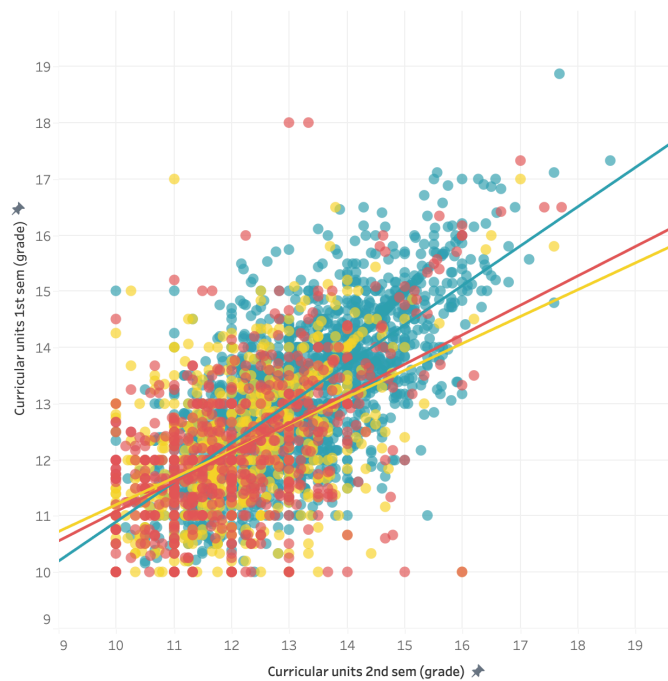
Admission grade (bin)	Target		
	Dropout	Enrolled	Graduate
95	54.00%	18.00%	28.00%
100	51.78%	18.27%	29.95%
105	42.60%	22.49%	34.91%
110	36.83%	19.69%	43.48%
115	31.43%	18.48%	50.10%
120	32.65%	21.99%	45.36%
125	26.59%	14.77%	58.63%
130	30.32%	16.77%	52.90%
135	25.18%	17.02%	57.80%
140	32.62%	20.31%	47.08%
145	16.53%	16.53%	66.94%
150	34.46%	12.99%	52.54%
155	24.59%	4.92%	70.49%
160	32.91%	10.13%	56.96%
165	25.00%	6.25%	68.75%
170	21.05%	26.32%	52.63%
175	50.00%		50.00%
180	25.00%	25.00%	50.00%
185			
190	33.33%	33.33%	33.33%

Age at Enrollment (2 year bins) - dropout rate correlates positively with age at enrollment

Age at enrollment (2 yr bins)	Target		
	Dropout	Enrolled	Graduate
16		40.00%	60.00%
18	21.01%	16.90%	62.10%
20	24.54%	22.80%	52.66%
22	35.11%	21.63%	43.26%
24	45.98%	22.77%	31.25%
26	57.84%	16.76%	25.41%
28	61.74%	13.42%	24.83%
30	63.46%	12.50%	24.04%
32	50.00%	12.26%	37.74%
34	53.21%	15.60%	31.19%
36	57.14%	10.39%	32.47%
38	55.22%	13.43%	31.34%
40	44.44%	16.67%	38.89%
42	47.37%	13.16%	39.47%
44	53.49%	13.95%	32.56%
46	56.67%	6.67%	36.67%
48	58.33%	16.67%	25.00%
50	60.87%	8.70%	30.43%
52	27.27%	9.09%	63.64%
54	58.33%	8.33%	33.33%
56	50.00%		50.00%
58	66.67%		33.33%
60	33.33%		66.67%
62			100.00%
64			
66			
68			
70	100.00%		

1st semester Grade vs 2nd semester Grade - graduated students tended to have higher grades in both semesters

1st sem vs 2nd sem Grade



Target

- Dropout
- Enrolled
- Graduate

Discussion:

Limitations:

Classification model's accuracy was not very strong, however there are many outside factors that may affect a student's dropout or academic success which may have not been recorded.

Future Research:

Other data to consider:

- Student mental health
- Family and community support
- School environment and engagement
- Peer Influences
- Extracurricular activities