STATISTICAL ANALYSIS REPORT ON IDENTIFYING STUDENTS AT RISK OF DROPPING OUT

Introduction

The study aimed to investigate the relationship between the educational process of students and dropout rates. The statistical analysis was based on variables describing education status, financial aid status, Admission grades, and socio-economic status

Methods

Descriptive statistics such as mean, median, standard deviation (SD), and the first and third quartiles (IQR) were used to characterize the studied population. For continuous data, the range was also employed to describe the characteristics of the study group.

	Marital status	Application mode	Application order	Course	Daytime/evening attendance\t	Previous qualification	Nacionality	Mother's qualification	Father's qualification	Mother's occupation
count	4424	4424	4424	4424	4424	4424	4424	4424	4424	4424
unique	6	18	8	17	2	17	21	29	34	32
top	0	0	1	11	1	0	0	0	26	9
freq	3919	1708	3026	766	3941	3717	4314	1069	1209	1577

	Father's occupation	Displaced	Educational special needs	Debtor	Tuition fees up to date	Gender	Scholarship holder	International	Target
count	4424	4424	4424	4424	4424	4424	4424	4424	4424
unique	46	2	2	2	2	2	2	2	3
top	9	1	0	0	1	0	0	0	2
freq	1010	2426	4373	3921	3896	2868	3325	4314	2209

	Previous qualification (grade)	Admission grade	Unemployment rate	Inflation rate	GDP	Total Curricular units (credited)	Total Curricular units (enrolled)	Total Curricular units (evaluations)	Total Curricular units (approved)	Total Curricular units (grade)	Total Curricular units (without evaluations)
count	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000
mean	132.613314	126.978119	11.566139	1.228029	0.001969	1.251808	12.502712	16.362342	9.142405	20.871049	0.287975
std	13.188332	14.482001	2.663850	1.382711	2.269935	4.220239	4.608821	7.664986	5.960612	9.637080	1.285682
min	95.000000	95.000000	7.600000	-0.800000	-4.060000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	125.000000	117.900000	9.400000	0.300000	-1.700000	0.000000	10.000000	13.000000	5.000000	22.000000	0.000000
50%	133.100000	126.100000	11.100000	1.400000	0.320000	0.000000	12.000000	16.000000	10.000000	24.500000	0.000000
75%	140.000000	134.800000	13.900000	2.600000	1.790000	0.000000	14.000000	20.000000	12.000000	26.532500	0.000000
max	190.000000	190.000000	16.200000	3.700000	3.510000	39.000000	46.000000	72.000000	43.000000	36.570000	24.000000

Observations

There are more Graduate students than Dropouts
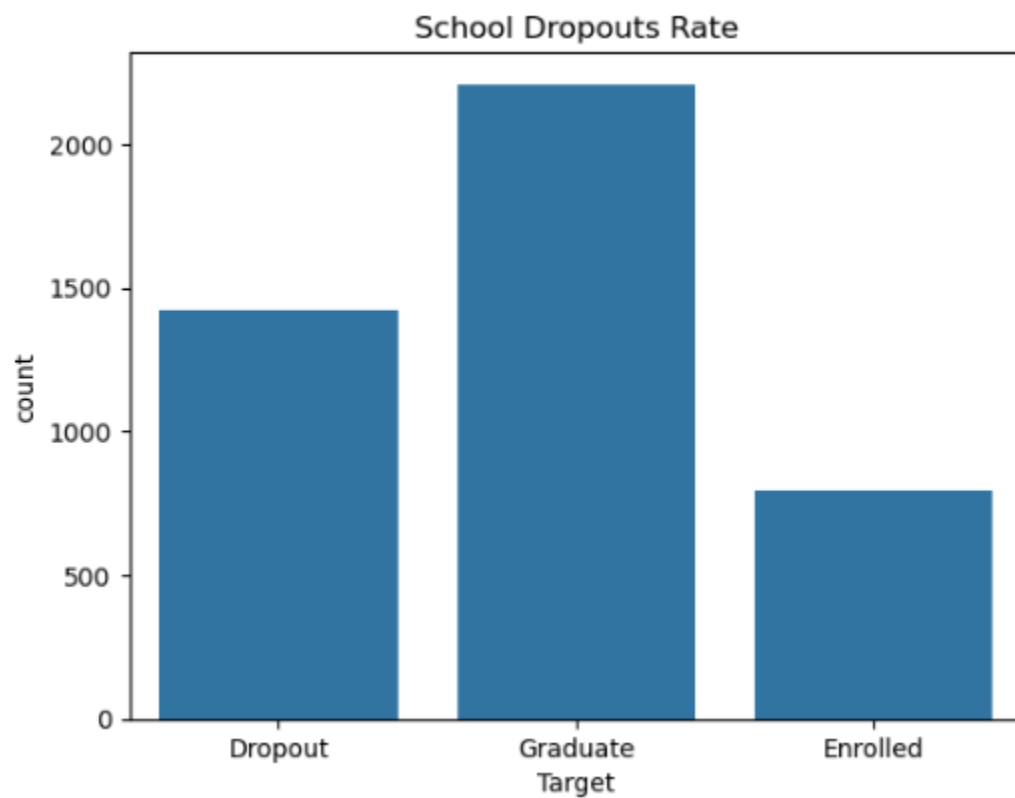
Hypothesis Testing and Result

In the study, the chi2_independence, and ttest were used to check if the dropout rates are dependent on the admission grades and to test the means of 2 different categories.

The following results were found:

- H_o: Students with all grades is the same for all education status
  H_a: Students with higher admission grades are less likely to drop out.

Alpha=0.1

Results: Since the p-value gotten was less than the significant level, we reject the hypothesis and conclude that Students with higher admission grades are less likely to drop out.

This shows the proportions of each grades

```
Target    Admission grade
Dropout   120.0              0.055595
          130.0              0.052076
          140.0              0.045039
          100.0              0.041520
          110.0              0.024631
                               ...
Graduate  176.0              0.000453
          178.0              0.000453
          184.0              0.000453
          184.4              0.000453
          190.0              0.000453
Name: proportion, Length: 1292, dtype: float64
```
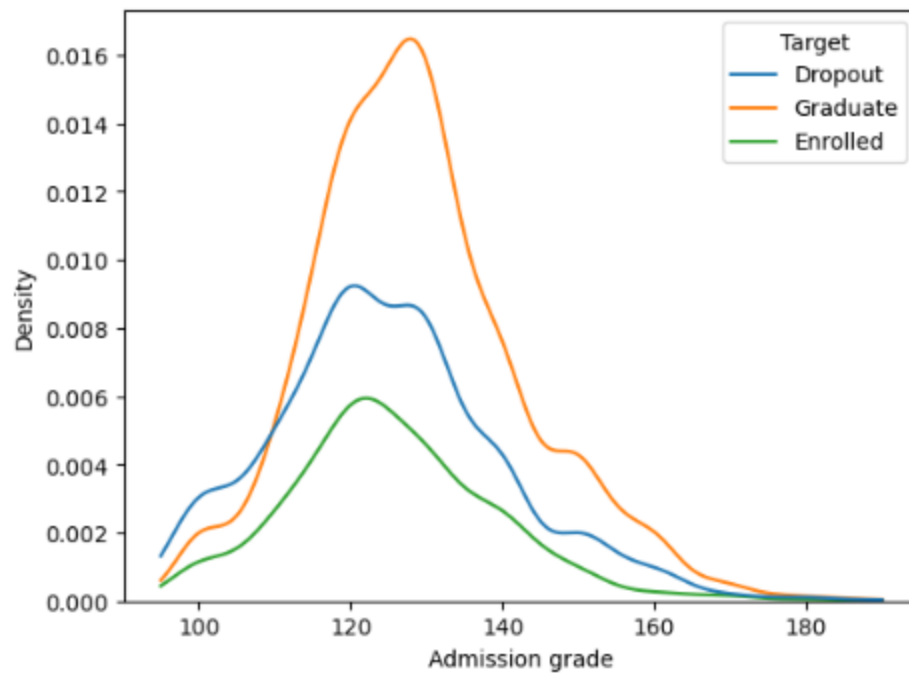
- H$_0$: Dropout rates are the same among students receiving financial aid or scholarships and those not receiving financial aid or scholarships. H a: Dropout rates are lower among students receiving financial aid or scholarships
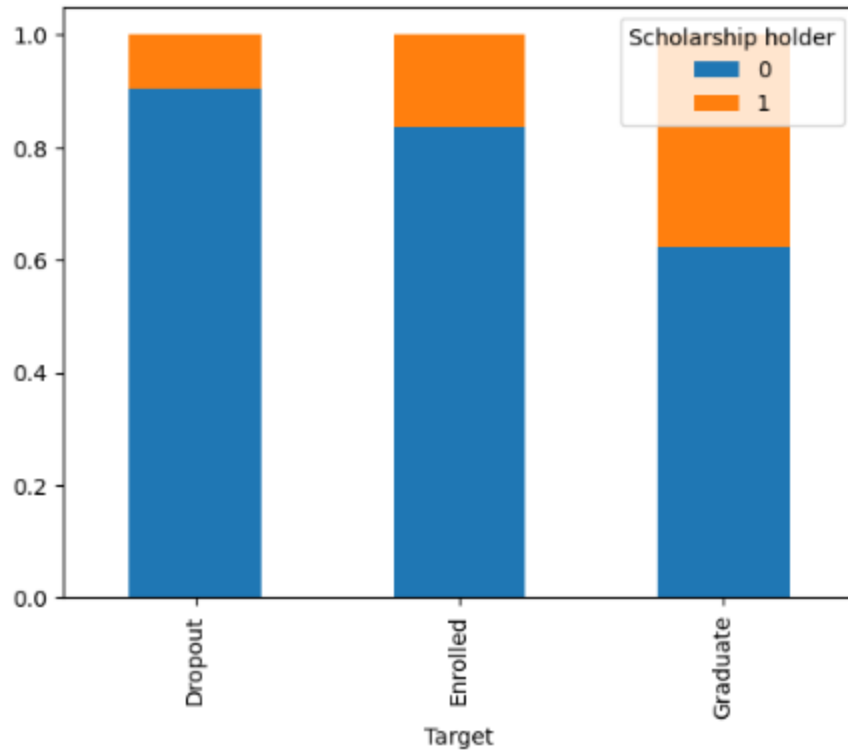
Alpha=0.1

Results: Since the p-value gotten was less than the significant level, we reject the hypothesis and conclude that Dropout rates are lower among students receiving financial aid or scholarships.

```
Target     Scholarship holder
Dropout   0                     0.905700
          1                     0.094300
Enrolled  0                     0.836272
          1                     0.163728
Graduate  0                     0.622001
          1                     0.377999
Name: proportion, dtype: float64
```

<Axes: xlabel='Target'>
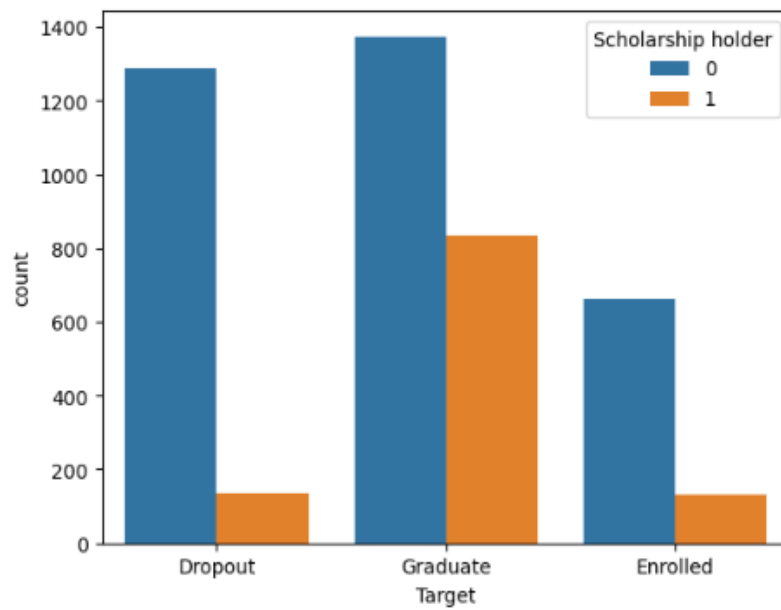
- Higher socio-economic status correlates with lower dropout rates.
- H_o: There's no correlation between socio-economic status and dropout rates.
  H a: Higher socio-economic status correlates with lower dropout rates.
  The diagram shows no correlation with the socio-economic status

Correlation Matrix Heatmap