Predict Students' Dropout and Academic Success

IRONHACK Final Bootcamp Project
Mojgan Shabani

About Project

The Data Source for this project has been provided from SATDAP - Capacitação da Administração Pública under grant POCI-05-5762-FSE-000191, Portugal
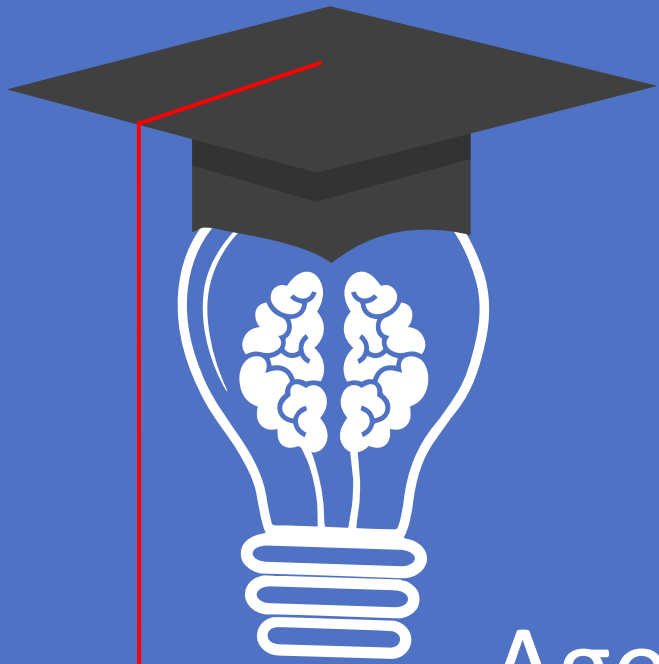
Overview
The dataset was created in a project that aims to contribute to the reduction of academic dropout and failure in higher education, by using machine learning techniques to identify students at risk at an early stage of their academic path, so that strategies to support them can be put into place.

Goal
Predict students' dropout and academic success

Agenda

01 **Data Exploration**

✓ Demographic Information

✓ Course popularity.

✓ Dropout & Graduate

02 **Prediction**

✓ Dropout Prediction.

✓ High score Features in predictions

03 **More insights**

✓ Comparing High scores Features in predictions
✓ Dropout in different Majors

04 **conclusion**

✓ Conclusion & solutions

Prediction Process

01 **Cleaning Data**

We have around 4442 samples with 40 Features. 18 Categorical and 22 Numerical.

02 **Removed Multicollinearity**

With VIF method remove high multicollinearities

03 **Normalized data**

Normalized data with OneHotEncoder and MinMaxScaler
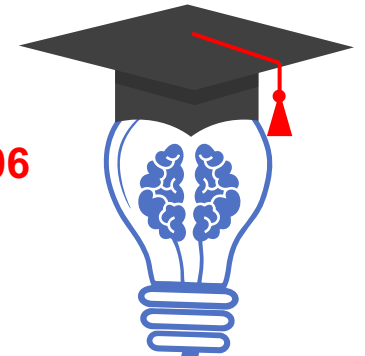
Prediction **05**

Preformed 4 different types of prediction's Models.

Up Sampling **04**

We had class imbalanced, so up sampling for prediction have been performed.

Data Visualization **06**

Using Tableau for Data visualization and gathering insights.

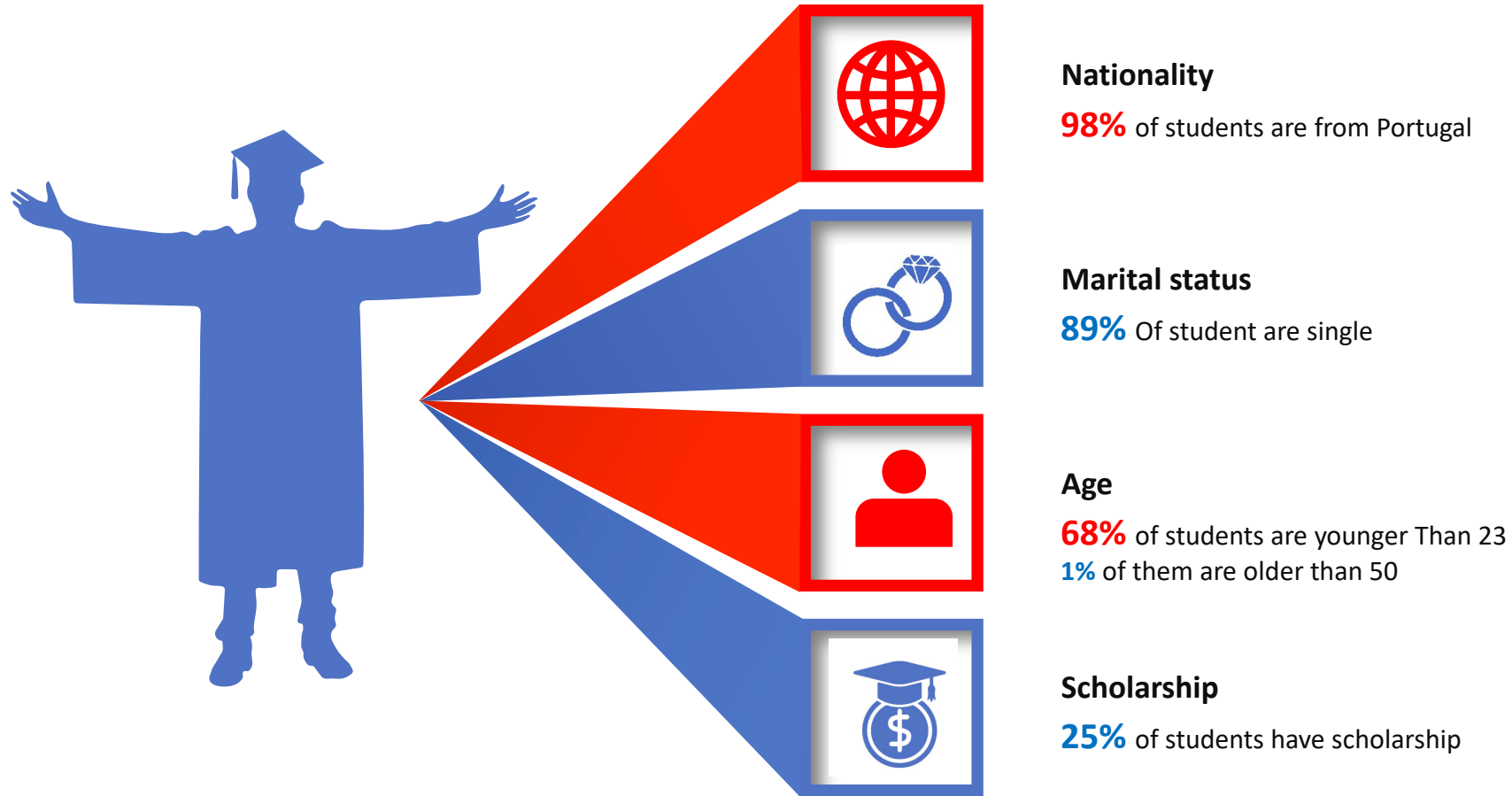# Demographic Information

65%

Female

Youngest age is 17

Oldest age is 70

Male

35%

Demographic Information

Nationality

98% of students are from Portugal

Marital status

89% Of student are single

Age

68% of students are younger Than 23
1% of them are older than 50

Scholarship

25% of students have scholarship

Major's Popularity

Nursing, 17.3%

Management, 8.6%

Social Service, 8.0%

Veterinary Nursing, 7.6%

Journalism and Communication, 7.5%

Management (evening attendance), 6.1%

Advertising and Marketing Management, 6.1%

Tourism, 5.7%

Communication Design, 5.1%

Social Service (evening attendance), 4.9%

Agronomy, 4.7%

Animation and Multimedia Design, 4.9%

Basic Education, 4.3%

Informatics Engineering, 3.8%
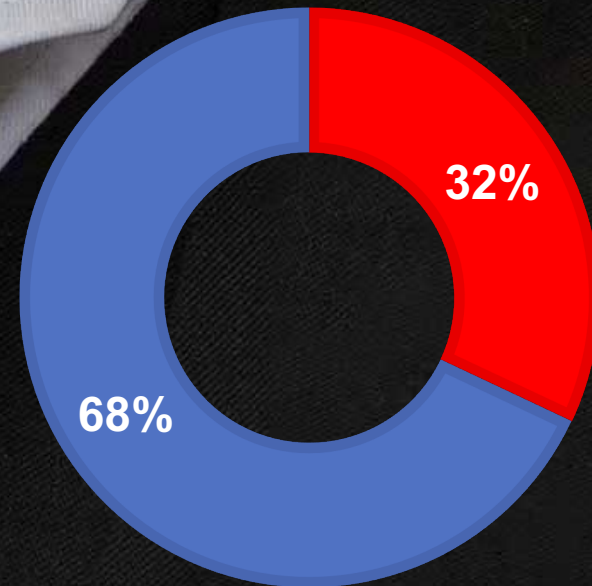
Equinculture, 3.2%

Oral Hygiene, 1.9%

Biofuel

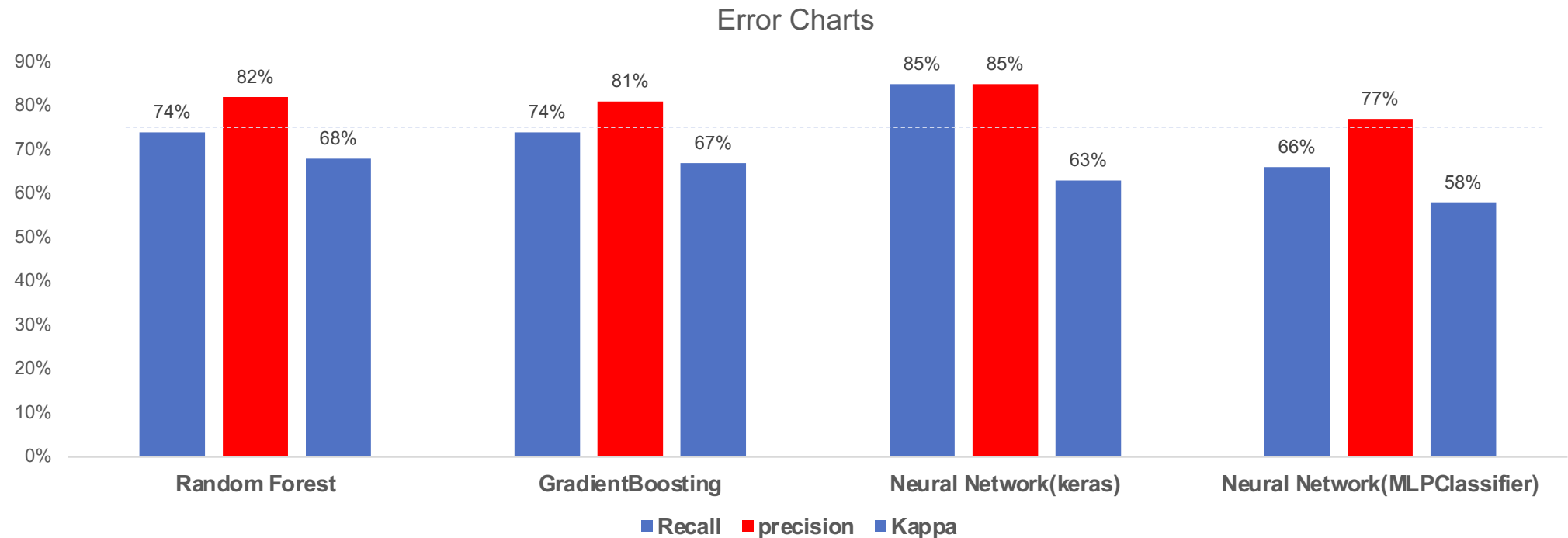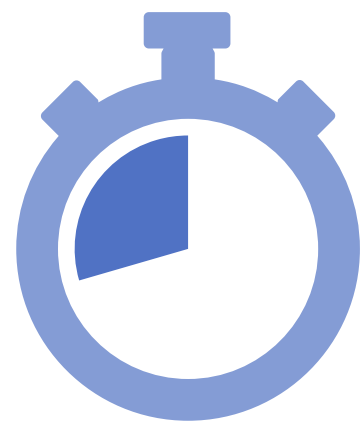# Graduate /Dropout

68%

32%

Prediction

For this project, four different type of prediction models have been performed and Based On the Error Metrics the best models is **Random Forest**.
Based on this Model, Dropout Can be predicted by 74% of success

Error Charts

Legend: Recall | precision | Kappa

Model	Recall	precision	Kappa
Random Forest	74%	82%	68%
GradientBoosting	74%	81%	67%
Neural Network(keras)	85%	85%	63%
Neural Network(MLPClassifier)	66%	77%	58%

High score Features in predictions

Grade average in the 2nd semester

Grade average in the 1st semester

Paid Tuition fees

17%

10%

8%

Average Grade in 2nd semester for dropout student is half of Graduated ones
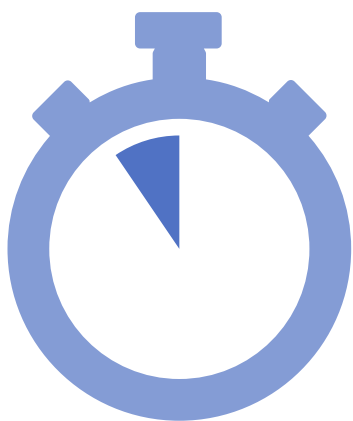Dropout   Grade(Avg): **6**
Graduated Grade(Avg): **12**

Average Grade fin 1st semester is dropout student is half of Graduated ones
Dropout   Grade(Avg):**7**
Graduated Grade(Avg): **12**
.

Just **25%** of Dropout students has paid tuition but for graduated ones is **75%** ,we will see the related graph in next page
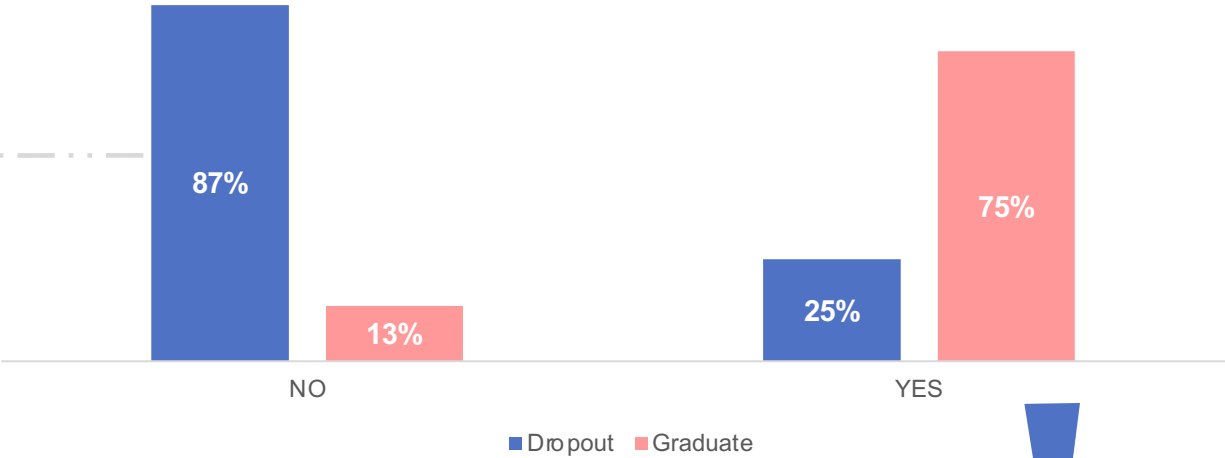
⚠️ The average of Previous qualification grade((Dropout:131,Graduate:133) and admission grade(Dropout:125,Graduate:128) is almost the same for Both Groups so something else must has been the main driving factor.
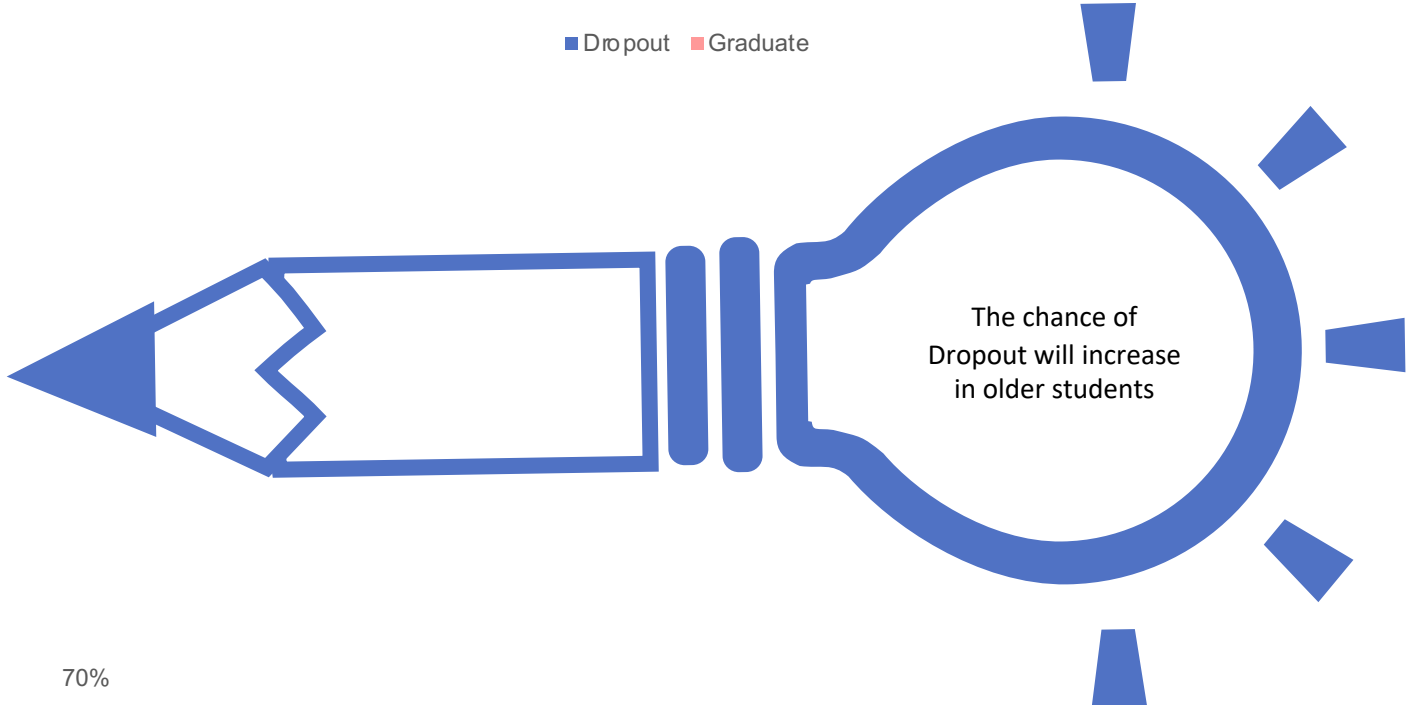
Comparing Other Features

Paid Tuition fees

It shows that 87% of students hadn't paid their tuition fees, so it seems they had doubt about finishing university

87%

13%

25%

75%

NO

YES

■ Dropout ■ Graduate

% of Dropout in each age range

Age range	%
50+	53%
45-50	58%
40_45	46%
35-40	57%
30-35	55%
25-30	58%
20-25	29%
17-20	21%

0% 10% 20% 30% 40% 50% 60% 70%

The chance of Dropout will increase in older students
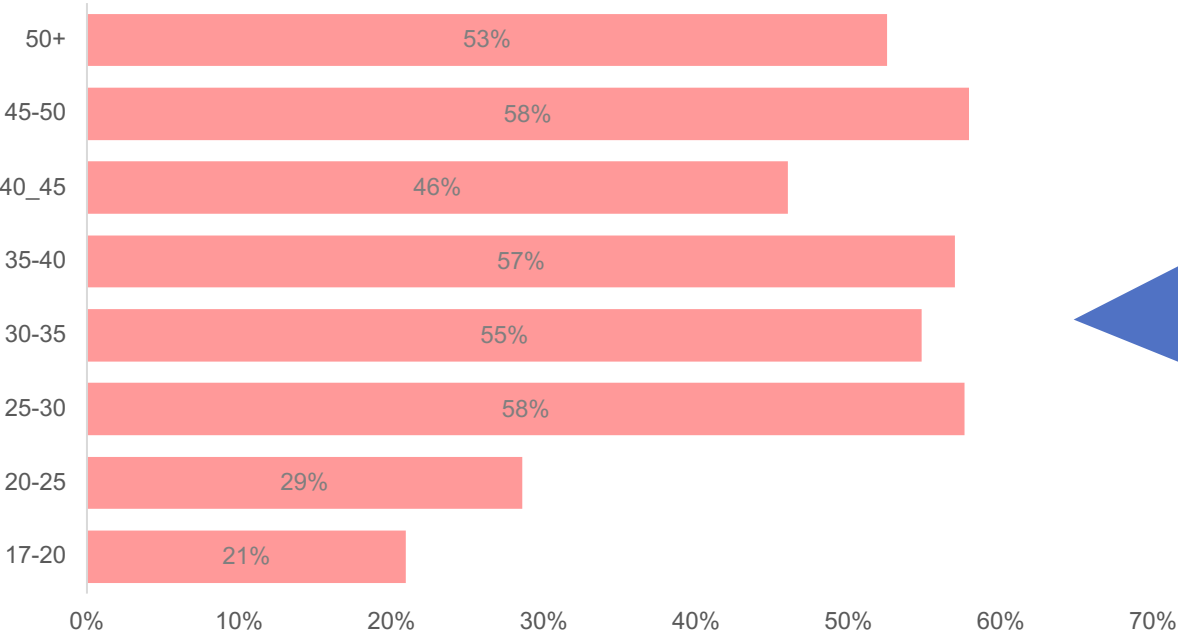
Dropout in different Majors

% of Dropout in each Major

Major	%
Nursing	15%
Social Service	18%
Communication Design	23%
Veterinary Nursing	27%
Journalism and Communication	31%
Social Service (evening attendance)	33%
Management	35%
Advertising and Marketing Management	35%
Tourism	38%
Animation and Multimedia Design	38%
Oral Hygiene	38%
Agronomy	41%
Basic Education	44%
Management (evening attendance)	51%
Informatics Engineering	54%
Equinculture	55%
Biofuel Production Technologies	67%

%Dropout distribution based on major in total

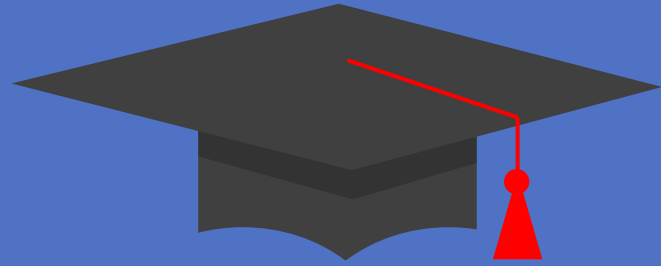| Major	%
Biofuel Production Technologies	1%
Oral Hygiene	2%
Communication Design	4%
Social Service	5%
Social Service (evening attendance)	5%
Equinculture	5%
Animation and Multimedia Design	6%
Basic Education	6%
Agronomy	6%
Veterinary Nursing	6%
Informatics Engineering	6%
Advertising and Marketing Management	7%
Tourism	7%
Journalism and Communication	7%
Nursing	8%
Management	9%
Management (evening attendance)	10%

- Managements(evening attendance) has highest proportion between majors in total dropout of university.
- However significant number of Biofuel Production Technologies tends to drop out with 67% dropout rate!!!

Conclusion
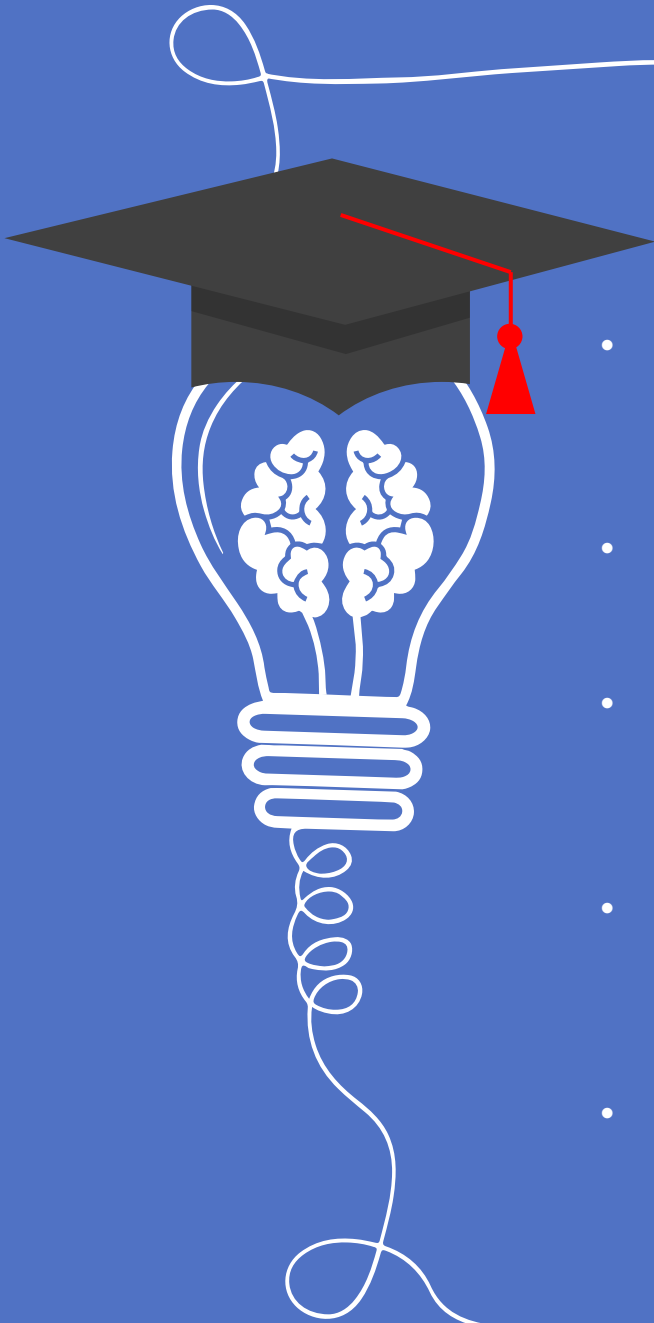
- Based on This Sample data , Students who preform well in 1st and 2nd semester will be unlikely to Dropout from the university ,so further study and data collection is required to identify drivers for lack of performance in the initial semesters.

- In the mean time universities can help reinforce studies for those students who doesn't performed well in the 1st semester.

- As tuition payment correlates with the increase in commitment and a reduction in the dropout, the universities can further study the reason for non-payment or drivers that could facilitate tuition payment.

- As highlighted in the findings, older students would be more likely to drop out; hence, the universities can create coaching/guiding programs to navigate through academic semesters for this demographics.

- For those majors which has high rate of dropout another research is recommended to identify missing link the reasoning.

THANK YOU