Predicting Students' Dropout and Academic Success

Exploratory Data Analysis (EDA) Report

Golola Charles 2024/HD05/21923U

1. Introduction:

The goal of this analysis is to explore factors that determine student's academic success or failure. The dataset from a higher education institution used contains data related to students enrolled in various undergraduate degrees. such as agronomy, design, education, nursing, journalism, management, social service, and technologies. The dataset includes information known at the time of student enrolment (academic path, demographics, and social-economic factors) and the students' academic performance at the end of the first and second semesters. The data is used to build classification models to predict students' dropout and academic success.

Exploratory data analysis involves different processes which include the following.

1. **Understanding the Dataset.**

   While under this section, these are the questions which required to be attended to:

   - What does the dataset represent?
   - What are the dimensions of the dataset?
   - What types of variables are present in the dataset?
   - Are there any missing values?

   The Dataset illustrates different features which determine the performance of a student. These features can then be based on to Predict Students' Dropout and Academic Success

   It contains 37 features including numerical and categorical features. This is achieved using the code below.

```
df = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/student_data.csv', delimiter=';')
df.shape()
```

- The dataset contains Dropout, Graduate, or Enrolled as the Target Variables.

```
df.columns
```

From the dataset we don’t have any missing values.

df.isnull().sum()

2. Data Wrangling:

Data wrangling is the process of cleaning, Transforming and preparing the data ready for the analysis and ready to be fed into the machine learning model.

The following are the questions which guided me throughout this Activity

- Are there irrelevant or duplicated columns that need to be removed?
- Are there formatting issues like wrong data types?
- Do we need to handle missing values through deletion?

I Checkup for missing values and I found out that no missing values were identified in the dataset.

Data type analysis. The data had a mixture of numerical (int64), string and categorical features. The categorical target variable, which included 3 outcomes: Dropout, Graduate, and enrolled.

df.dtypes

Non-numeric data. For the aim of numerical analysis, the non-numeric columns were excluded.

3. Data Explore:

Sample Exploratory Questions

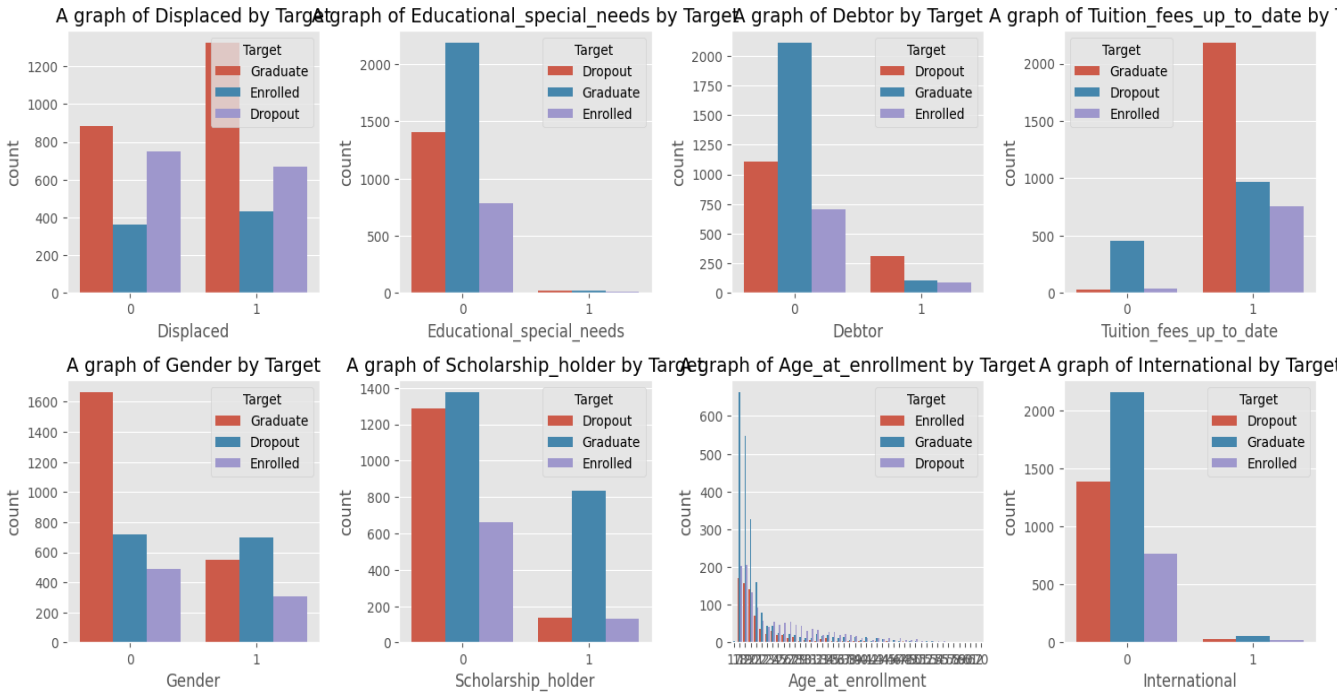
1. What factors contribute most to student dropout?
2. How do academic performance indicators such as Admission Grade or Curricular Units influence academic success or failure?

3. Is there any correlation between socioeconomic indicators and academic outcomes?

According to the analysis on the data frame, The main factors which influence the Success or dropout of a student include the following.
Daytime/evening attendance

- Previous qualification,
- Previous qualification (grade),
- Admission grade,
- Gender,
- Scholarship holder,
- Age at enrollment,
- Unemployment rate,
- Inflation rate,
- GDP

I used different visualisation drawings to show how the different factors collate with the Target ie Success or Dropout



The above graphs shows us how different factors determine /Affect te performance of a student for example

- Special Needs. The graphs reflect that students with special needed preform poorly and many are likely to drop out of school unlike those without special Needs.
- Tuition Fees Up to Date. Its reflected that students whose fees is paid are likely to perform better than those without fees paid.
- Gender: Gender visualisation shows how one gender tend to succeed yet the opposite gender tend to dropout.
- Age at enrollment. Aged students tends to dropout of school unlike those with lower age

Other Visualisations include.

4. Conclusion and Findings / Communication

According to the Exploratory Data Analysis, the main features /Factors to consider while predicting if a student may Dropout or Succeed are:

Daytime/evening attendance, Previous qualification, Previous qualification (grade), Admission grade, Gender, Scholarship holder, Age at enrollment, Unemployment rate, Inflation rate, GDP