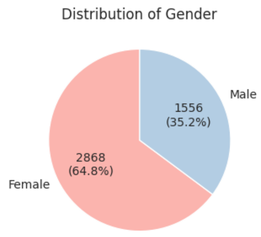
Predict students' dropout and academic success

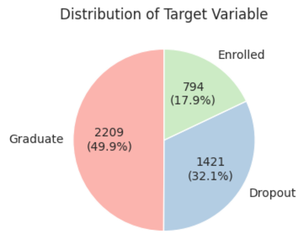
Kelli Gagarina
Liina Hoogand

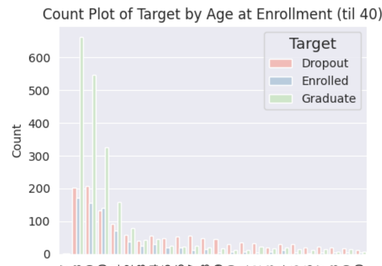
data:https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention

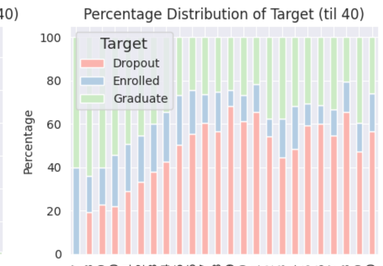
1 Introduction

- Kaggle dataset "Predict students' dropout and academic success," featuring 35 columns and nearly 4,424 rows.
- We aim to uncover the key factors influencing students' dropout rates and find interesting relations.
- Static dashboard on the right displays fundamental insights such as age distribution, gender ratios, and basic correlations.





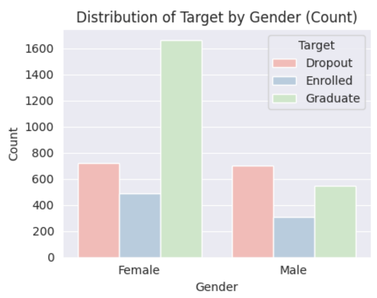


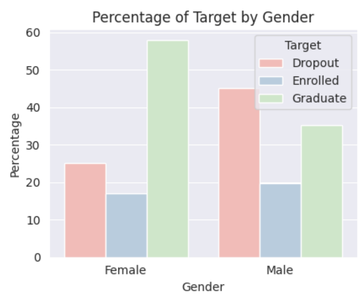


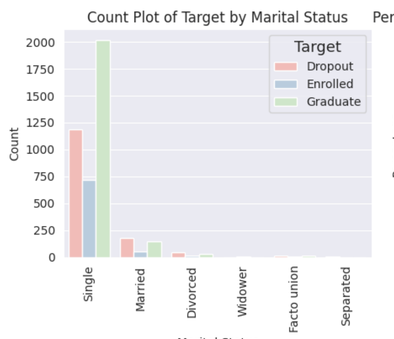
2 Goals

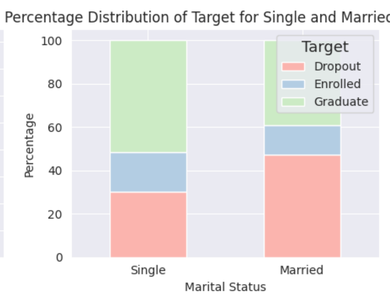
Our goal is to learn from fellow students and optimize academic decisions by

- identifying factors influencing dropout and graduation rates
- extracting insights for understanding of academic success dynamics
- creating predictive model
- testing hypothesis, which is aligned with our interests.



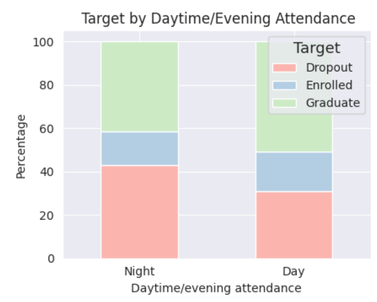


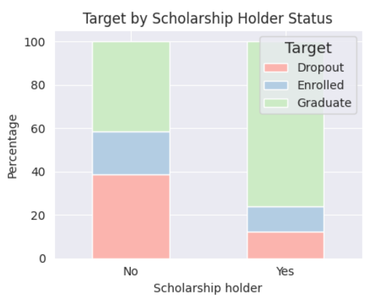


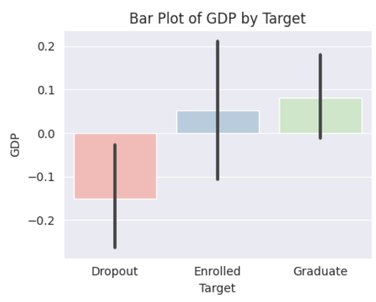


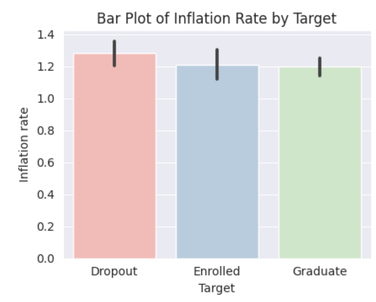
3 Methods used

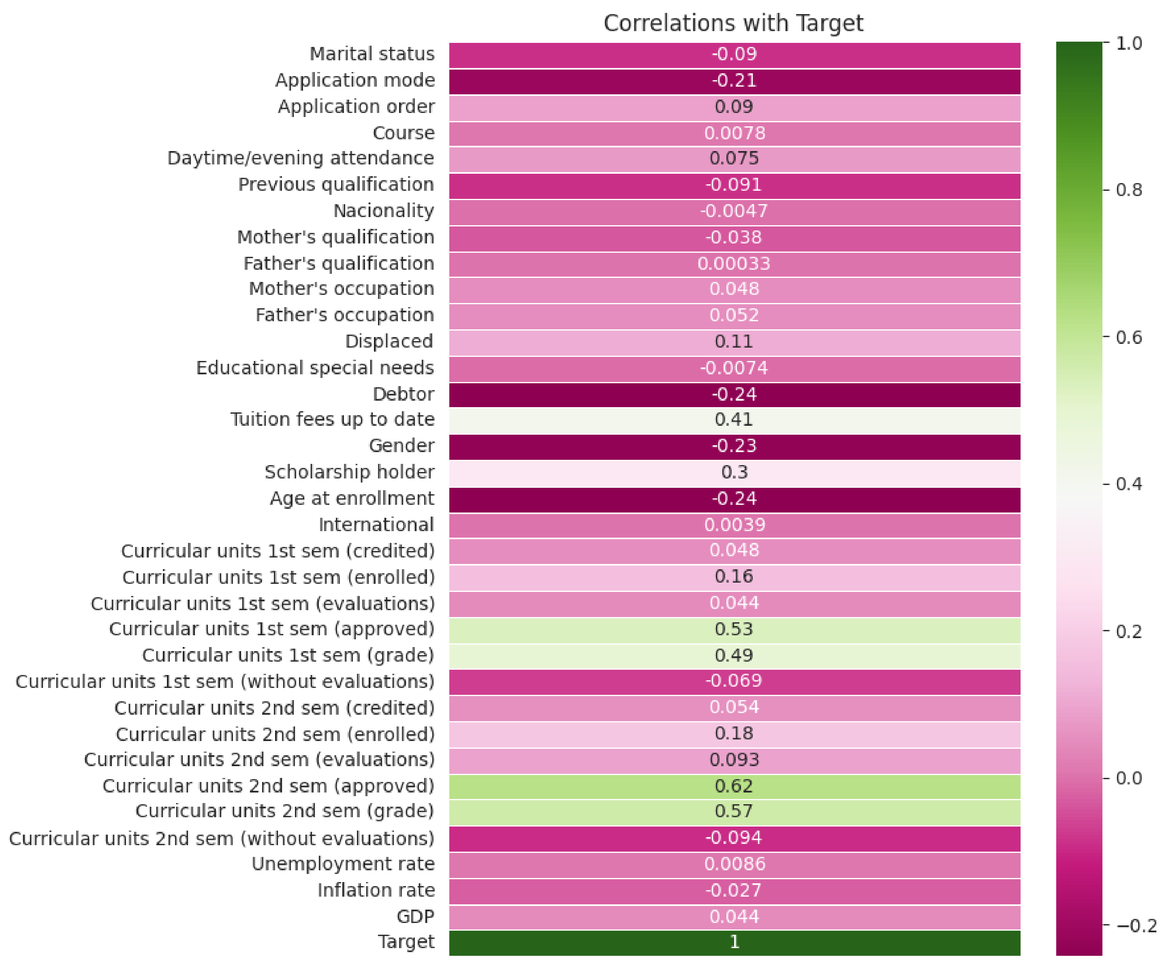
- Visualize demographics in static **dashboards**.
- Use a **heatmap** to identify influential factors for student outcomes.
- Evaluate **Random Forest**, **Logistic Regression**, and **SVM**
- **Chi-squared test** for testing the hyphotesis.
- Create an initial dynamic dashboard for interactive learning.











4 Results

Prediction model

- Random Forest Classifier emerges as the top-performing model with accuracy of 80%.
- Evaluated through cross-validation - 76.4%.
- Randomized search caused decline in accuracy of the model(overfitting).
- Including all columns increases accuracy by 3-4%.

Hyphotesis

- **Null Hyphotesis:** There is no significant association between the qualification of parents and the academic success of their children.

- Chi-squared Test - Rejecting the null hyphotesis:
  Mother's Qualification Chi-squared value: 225.0 P-value: 3.175e-19
  Father's Qualififcation Chi-squared value: 217.95 P-value: 5.824e-21

- The father's qualification displayed a more substantial connection with the target value than the mother's qualification.

Categorization