



# Educational System Case Study

Using Python

Presented By  
**Sady García**



# Agenda

1	Project Overview	3
2	Approach	4
3	Dataset Overview	5
4	Univariate Analysis	6
5	Bivariate Analysis	10
6	Multivariate Analysis	16

7	Machine Learning Algorithms	22
8	Insights	26
9	Conclusion	27

# Project Overview

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

- As a Data Scientist intern at Oeson, the goal is to gain valuable insights about the Educational System in Higher Education using a dataset to explore the relevant factors that contribute to higher academic performance.
- The scope is to examine the results from evaluations across different dimensions, including gender, age, scholarship, nationality, among others. Additionally, Supervised Machine Learning Algorithms will be utilized.
- To analyze the data, a combination of Exploratory Data Analysis, Statistical Analysis, and Data Visualization techniques using Python was employed. To predict student performance, Linear Regression Model and Random Forest Model in Python were utilized



# Approach

1

2

3

4

5

6

7

8

9

## Data Understanding

Focusing on relevant columns and understandin the different categories across the dataset.

## Data preprocessing

Identifyng missing values to ensure an accurate analysis

## Exploratory Data Analysis

Using univariate, bivariate and multivariate analysis techniques and proper visualization charts to communicate effectively, the insights.

## Supervised Machine Learning Algorithms

Using Supervise Machine Learning Algorithms to forecast the academic performance



# Dataset Overview

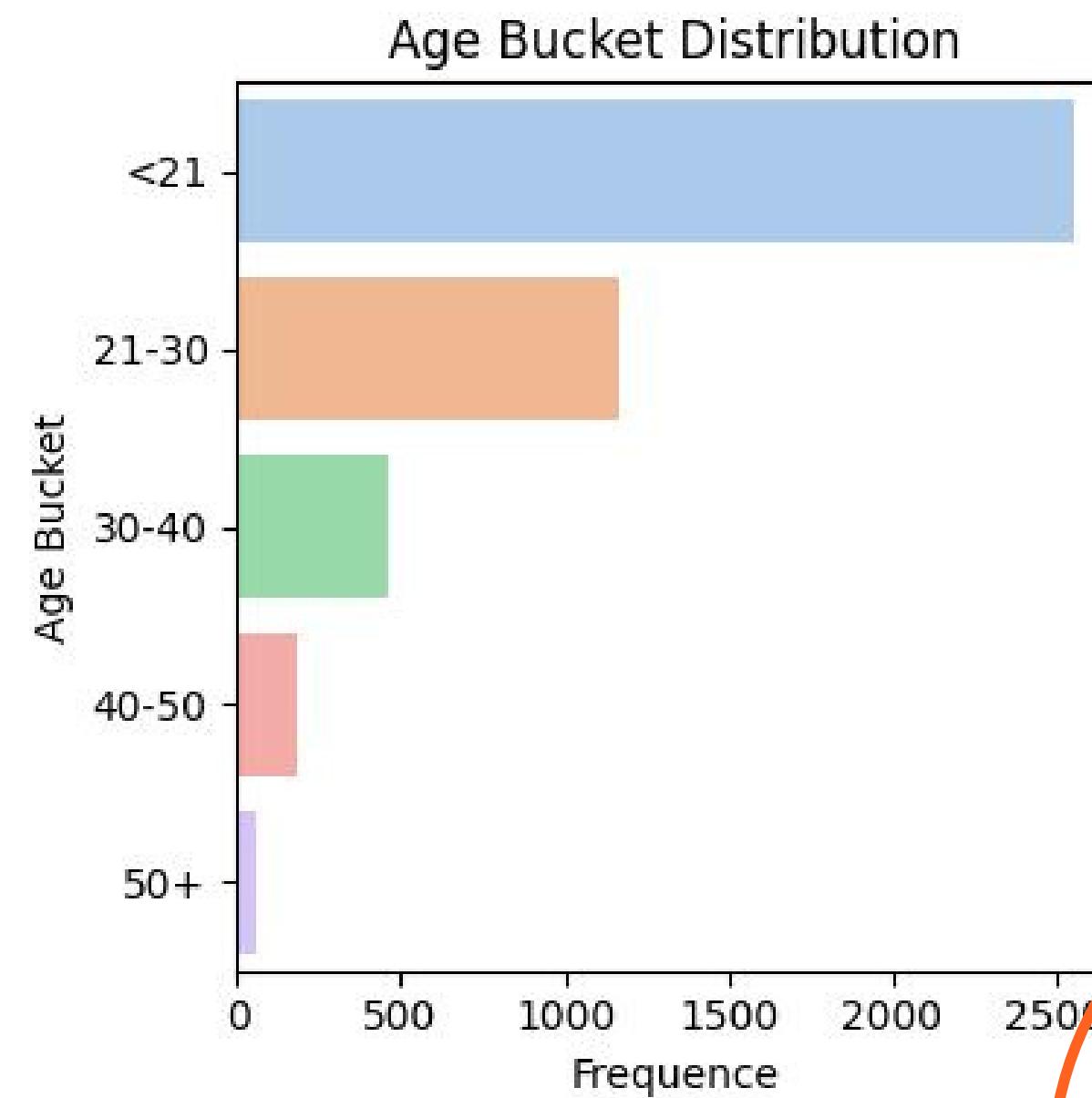
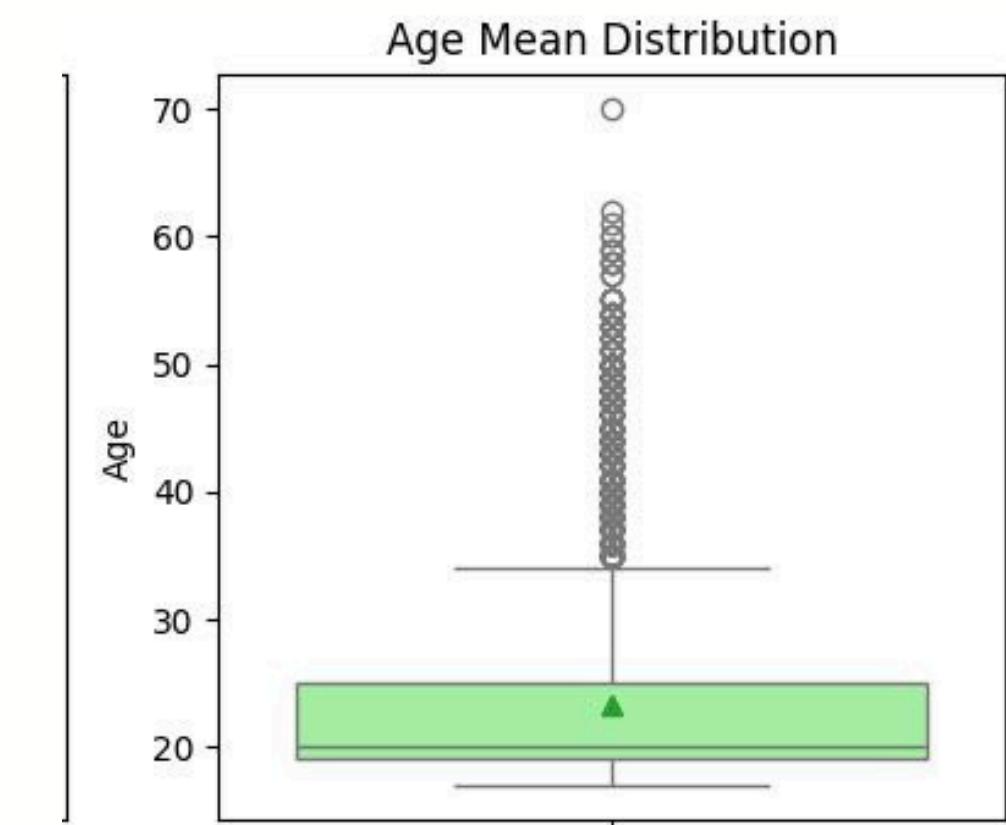
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

The dataset contains 35 columns per more than 4000 rows with information related to :

- Enrollment patterns
- Academic performance
- Economic Indicators

Dimensions were analyzed to infer educational performance.

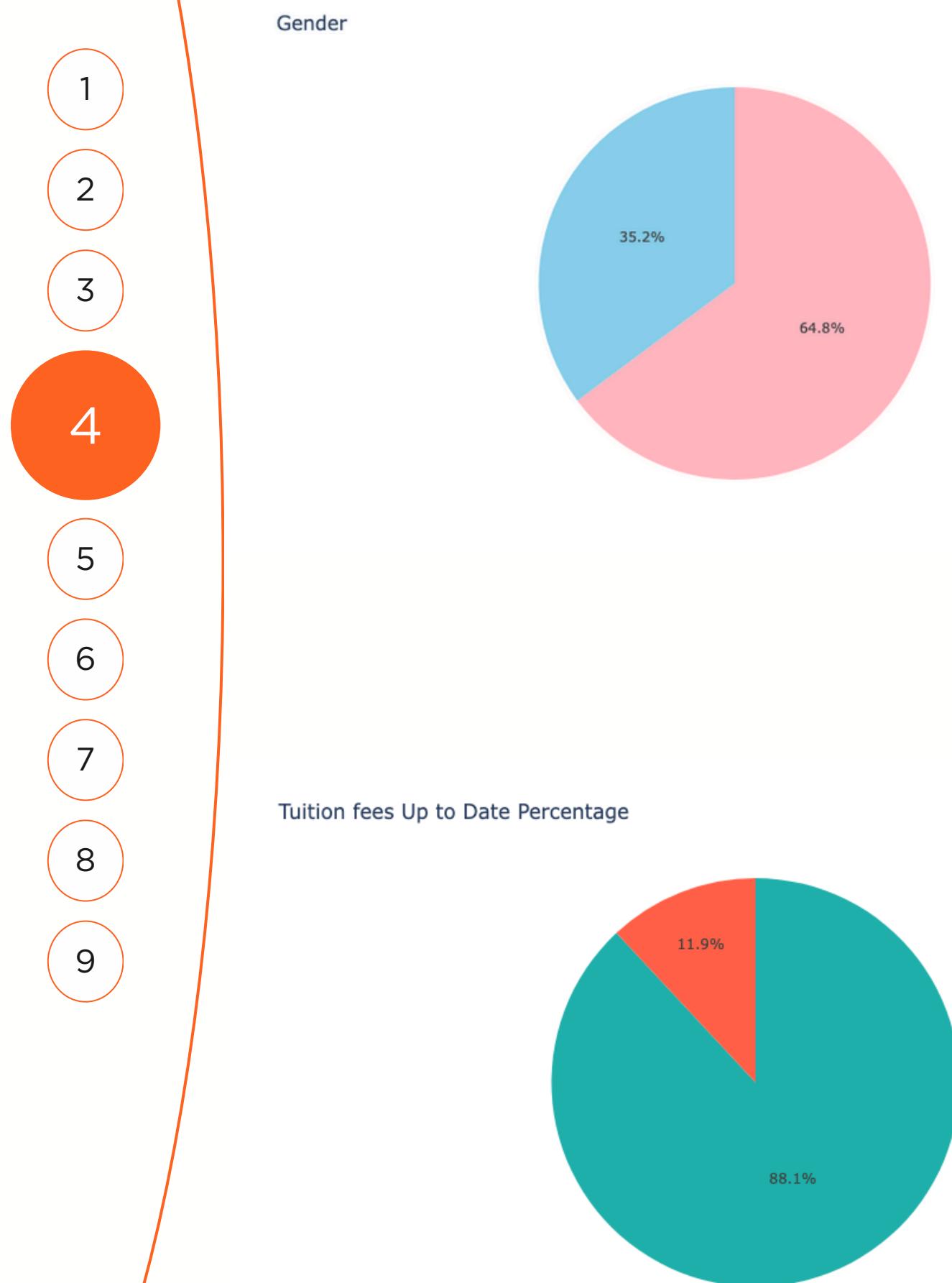
Within this dataset, the Age at Enrollment in Higher Education has a mean age of 20 years.



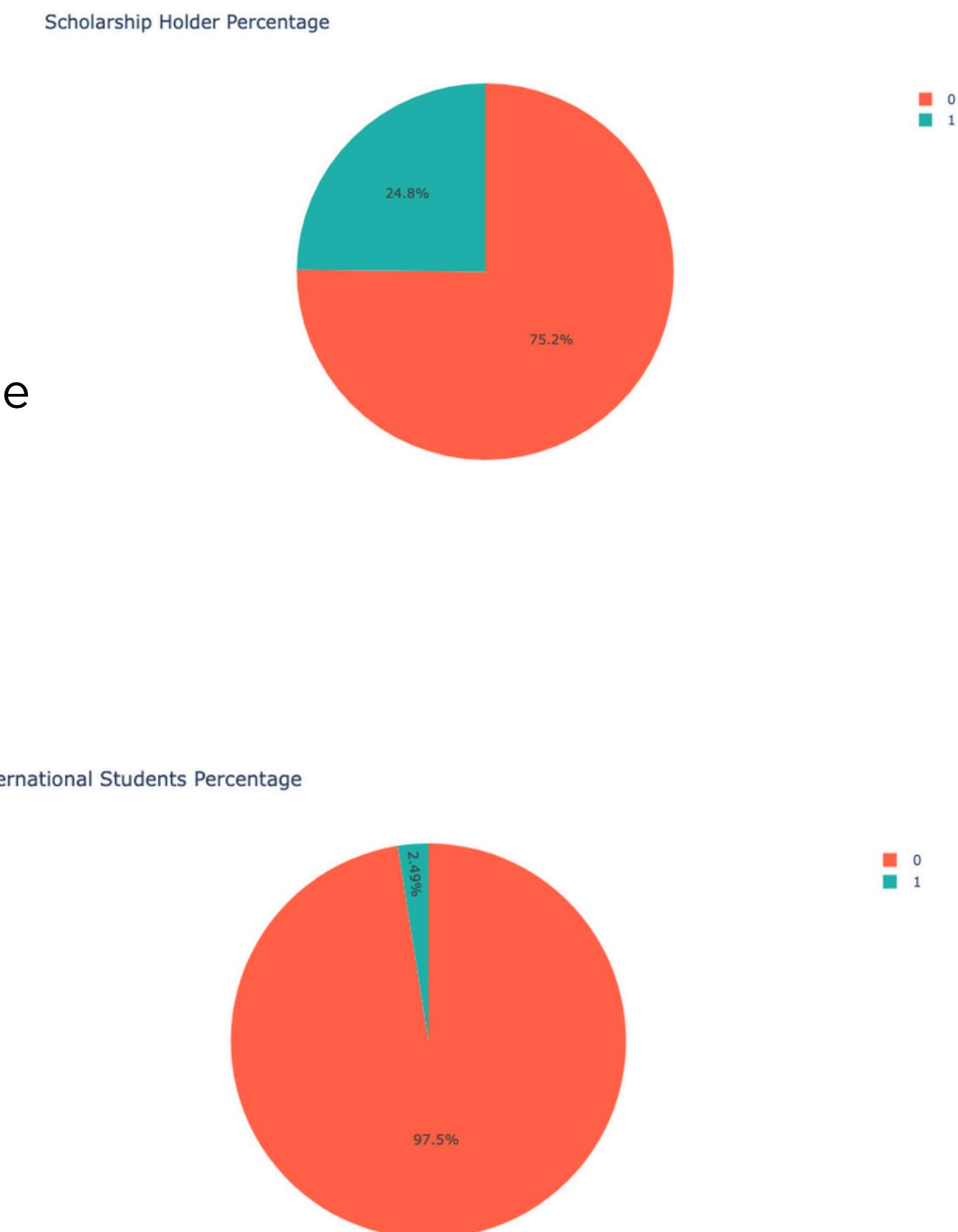
# Univariate Analysis

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

# Enrollment Patterns



- As noticed in the dataset:
- Majority of **women**
  - **88 %** tuition fees Up to Date
  - Only **24%** of students have **scholarhip**
  - Only **2.5%** of students are **Internationals**



# Academic Performance

1

2

3

4

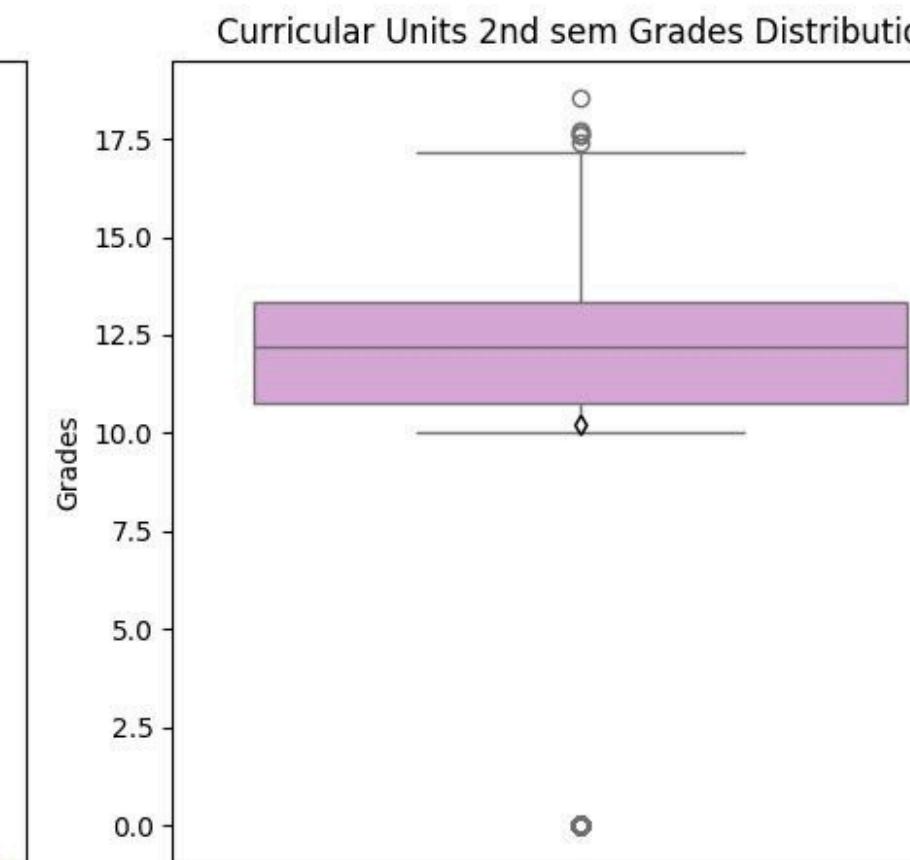
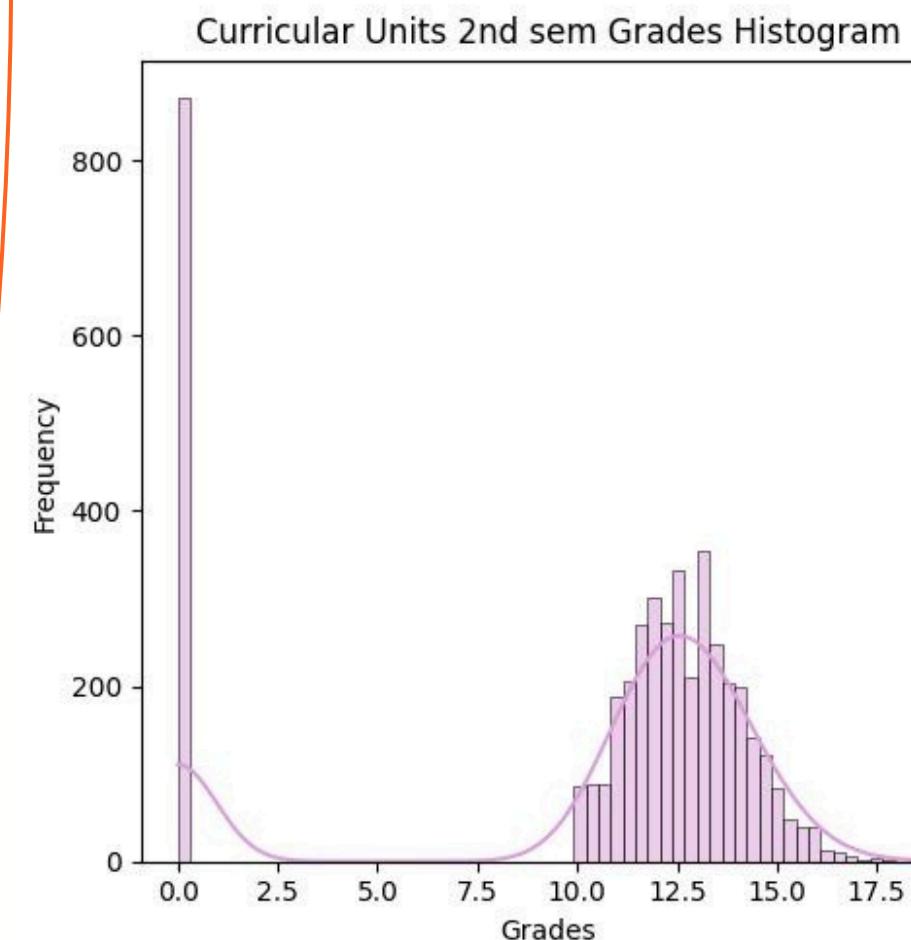
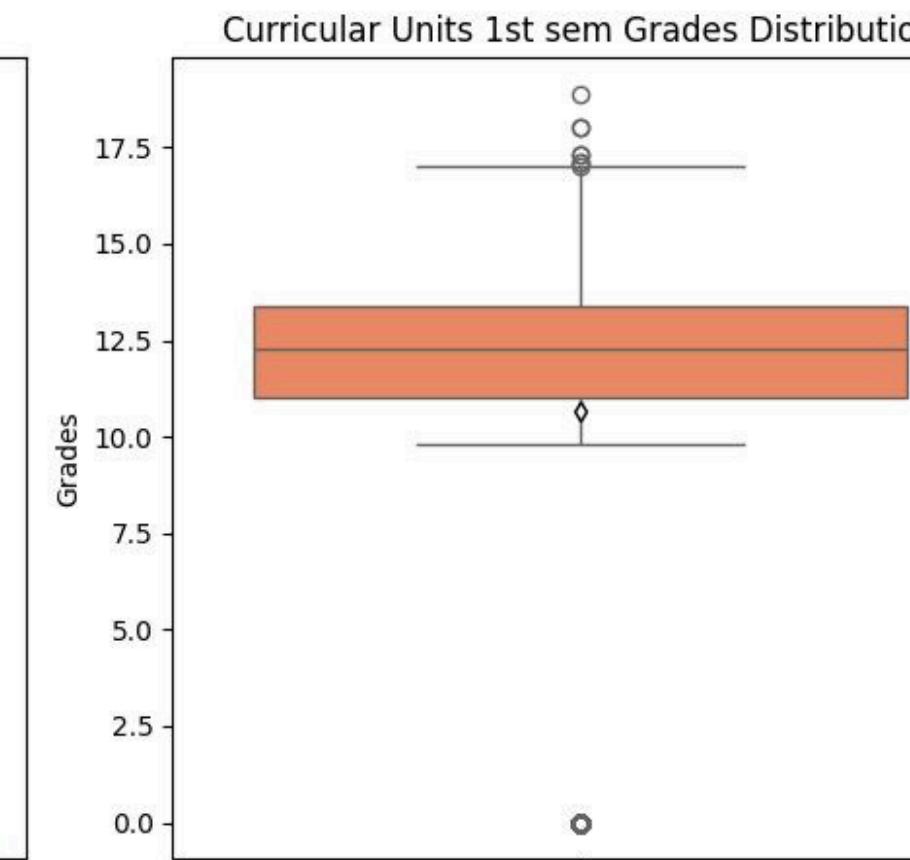
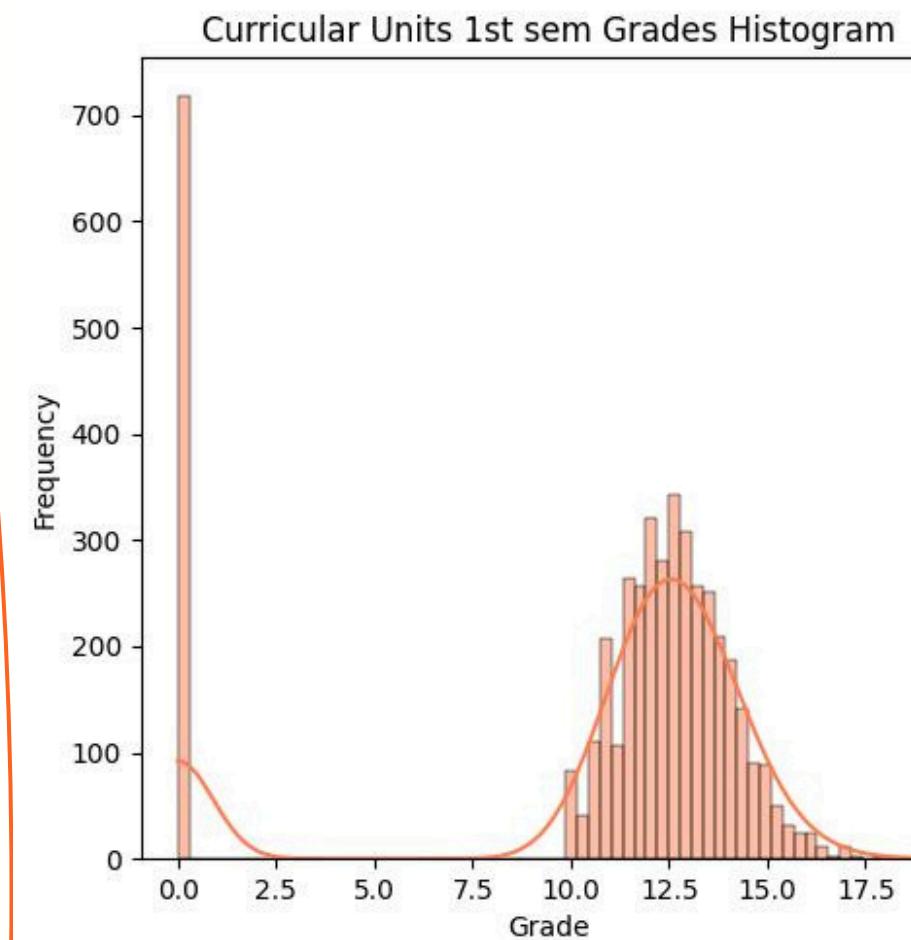
5

6

7

8

9



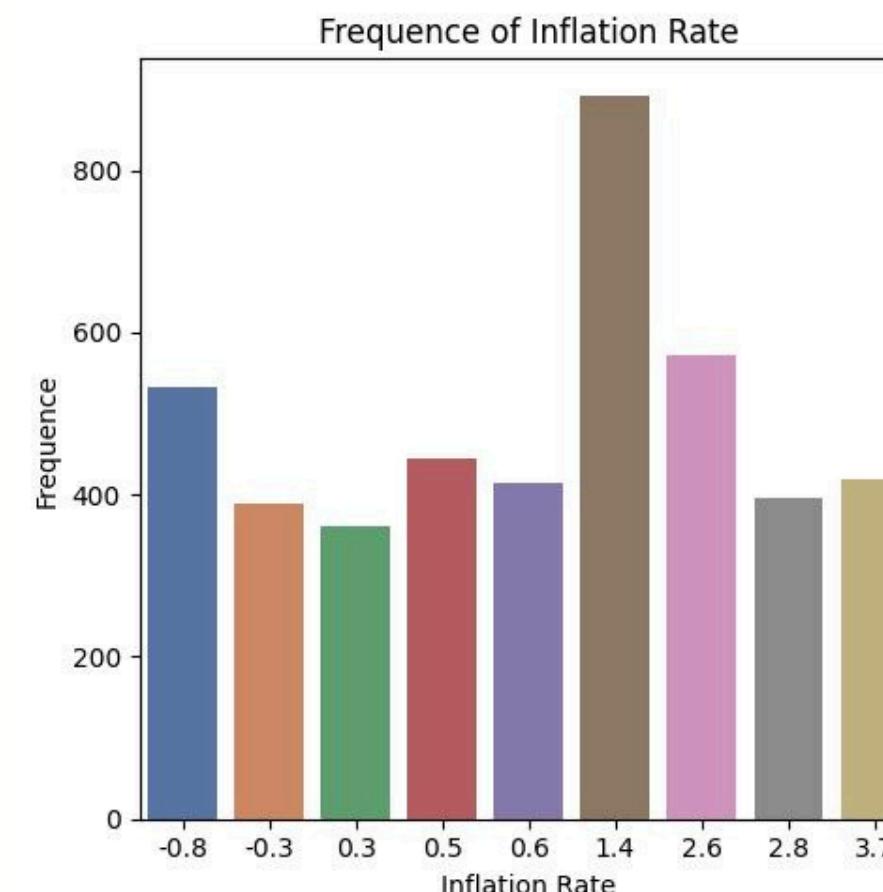
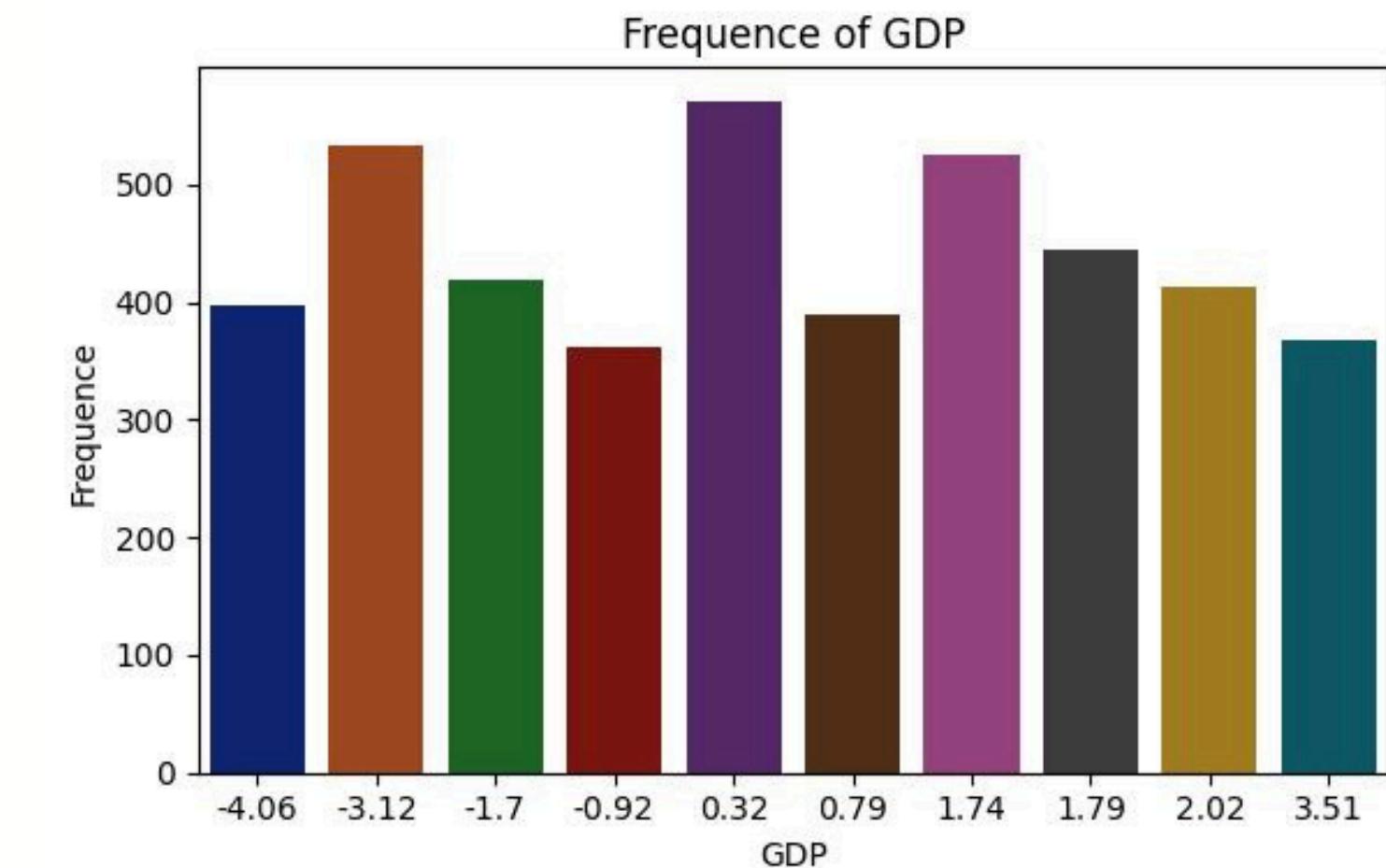
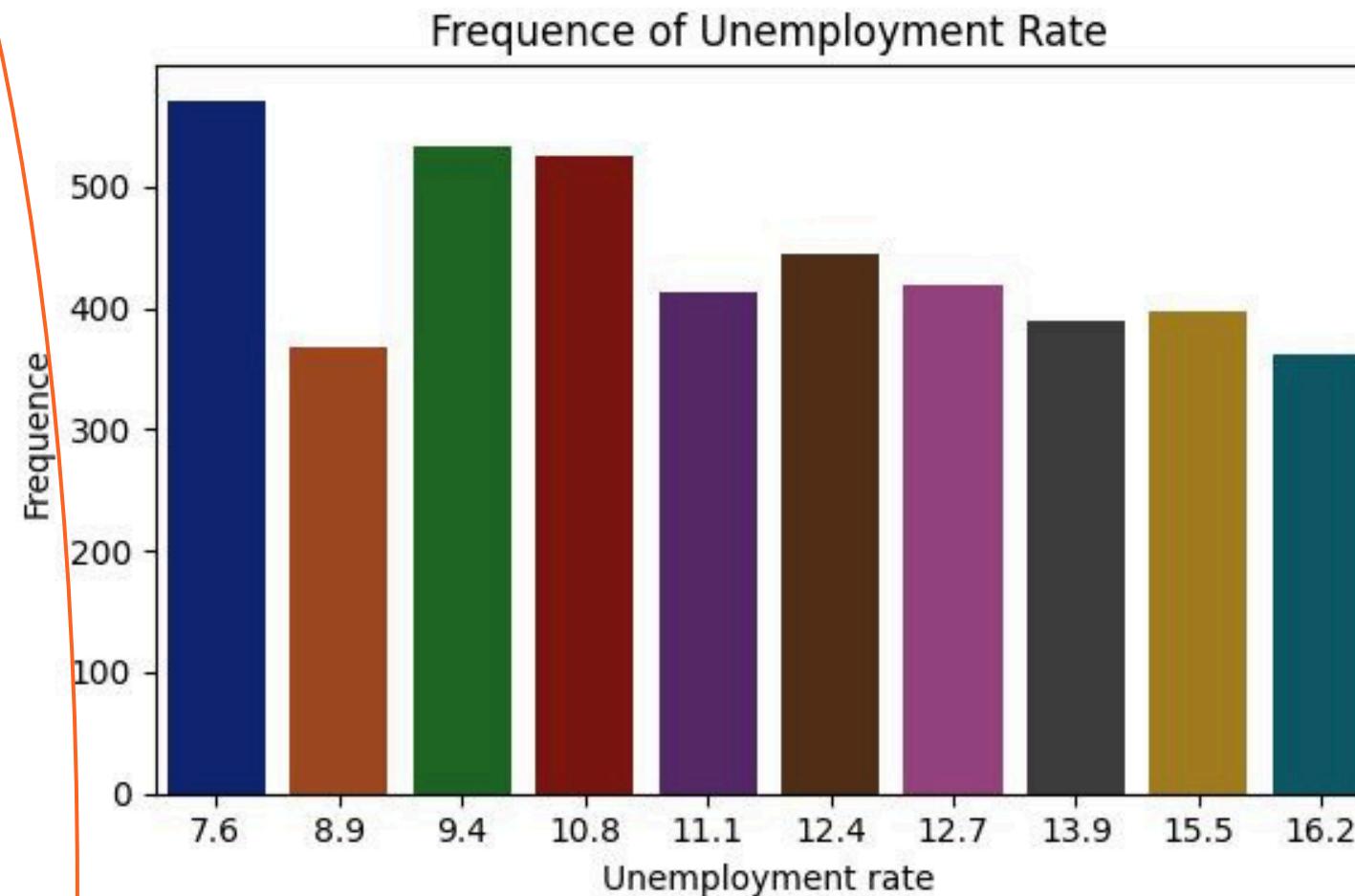
- During **1st** semester :
  - Mean Grade : **8.29**
  - Median Grade : 8.0
- During **2nd** semester
  - Mean Grade : **8.06**
  - Median grade : 8.0

In both semesters the grades have a similar behaviour.

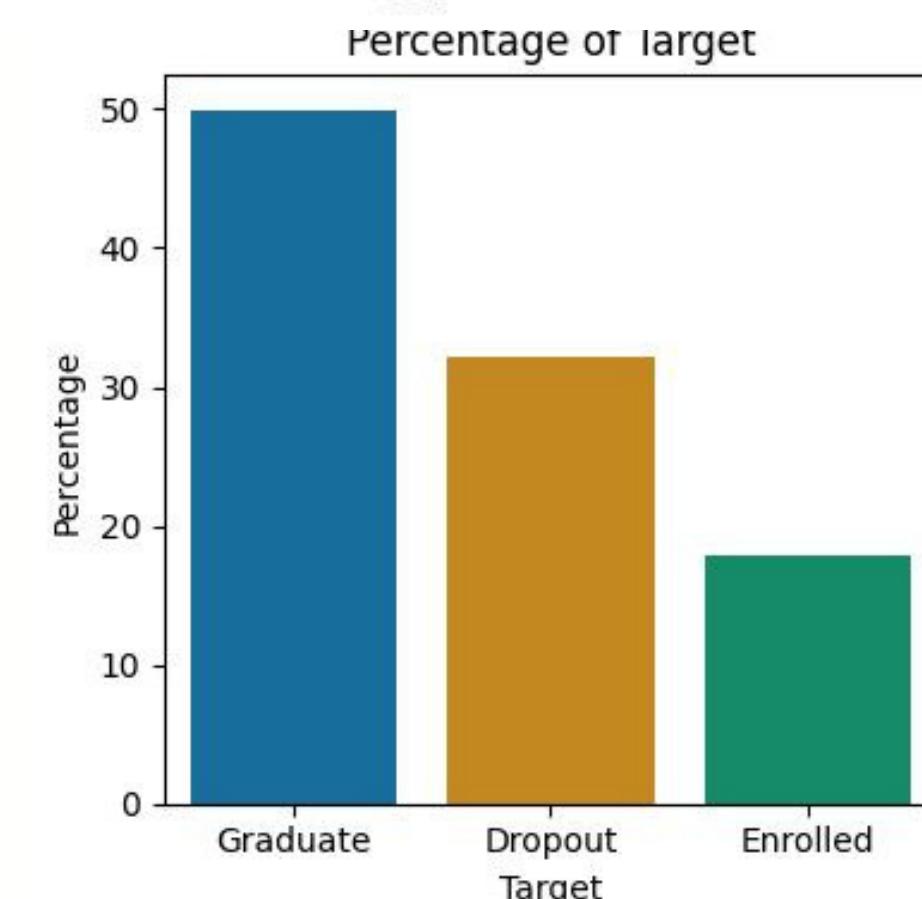
- Around **89 % attended** courses

# Economic indicators

1  
2  
3  
4  
**5**  
6  
7  
8  
9



Economic indicators such as Unemployment rate, GDP and Inflation rate works to understand the influence of the economic pattern in the academic performance. Almost **50%** of the population in the dataset were **graduated**

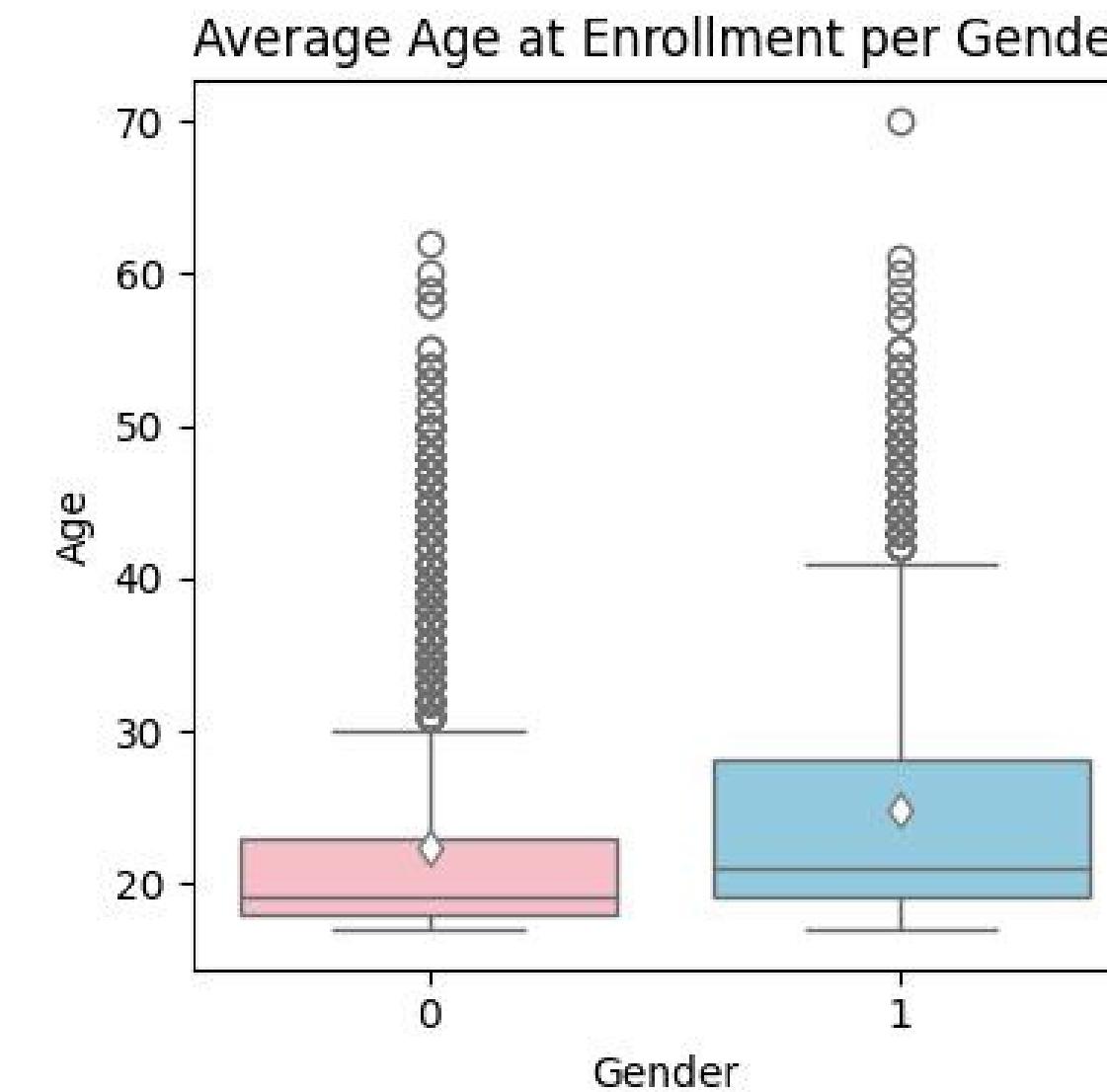
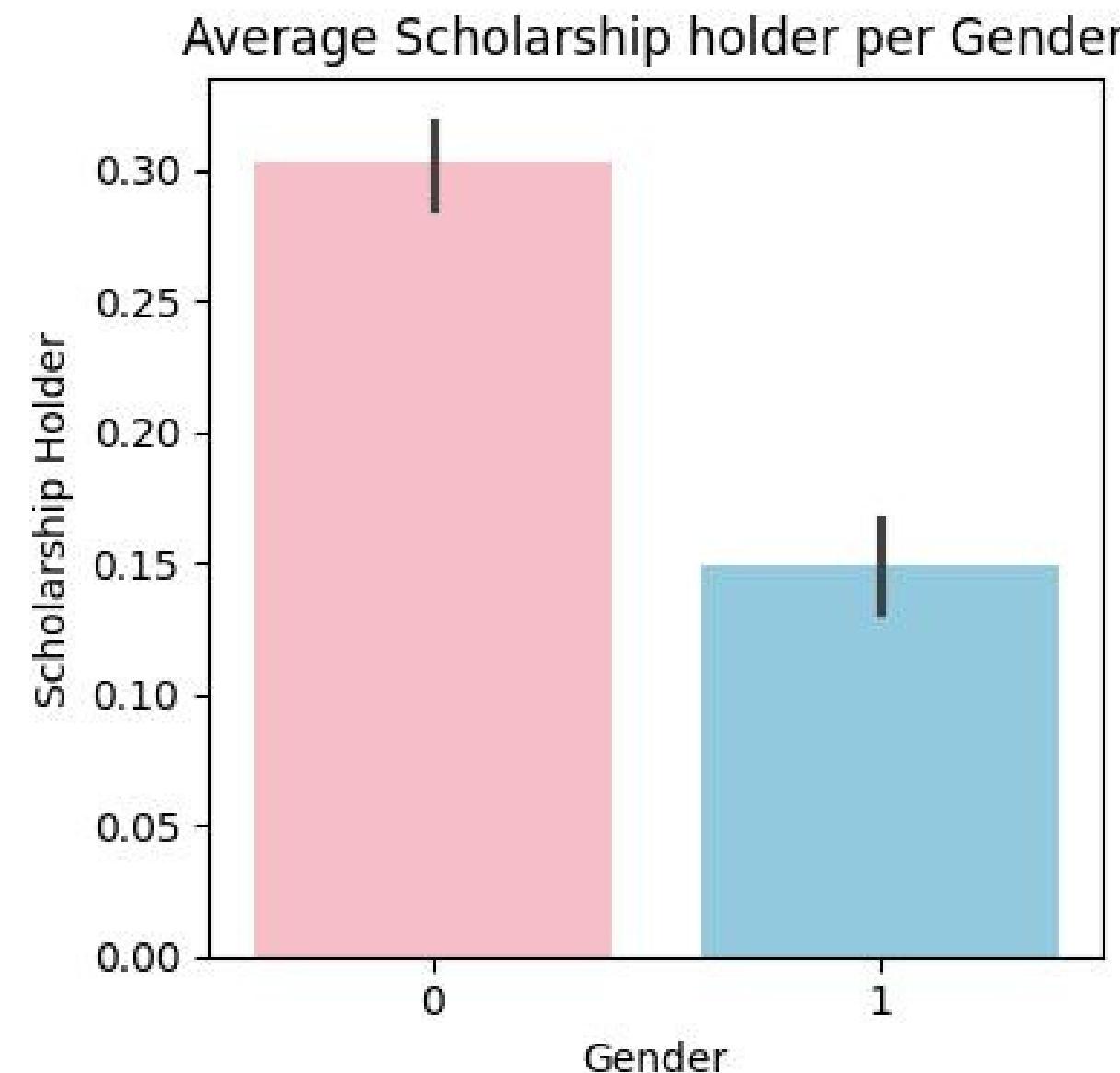


# Bivariate Analysis

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

# Enrollment Patterns

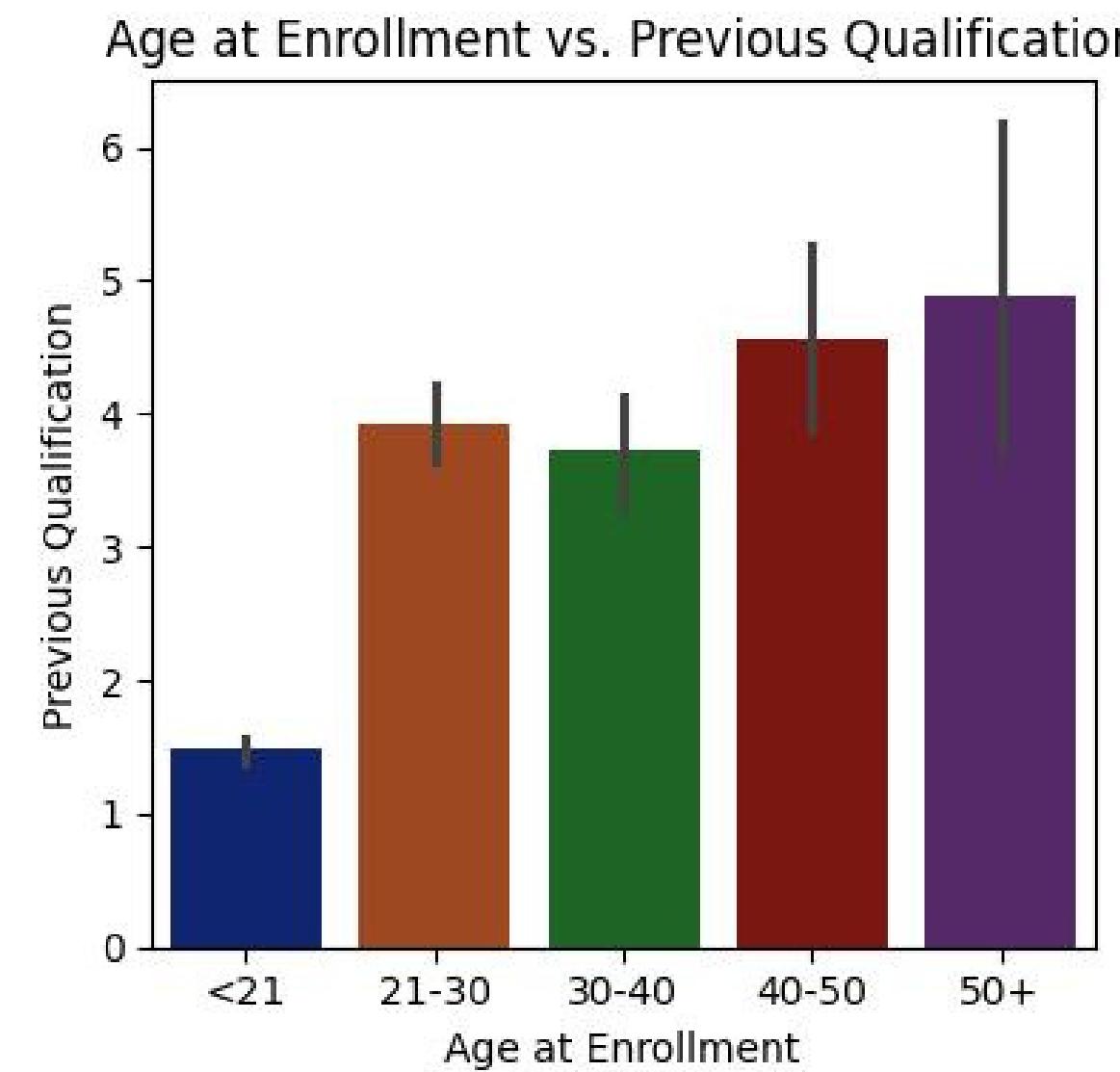
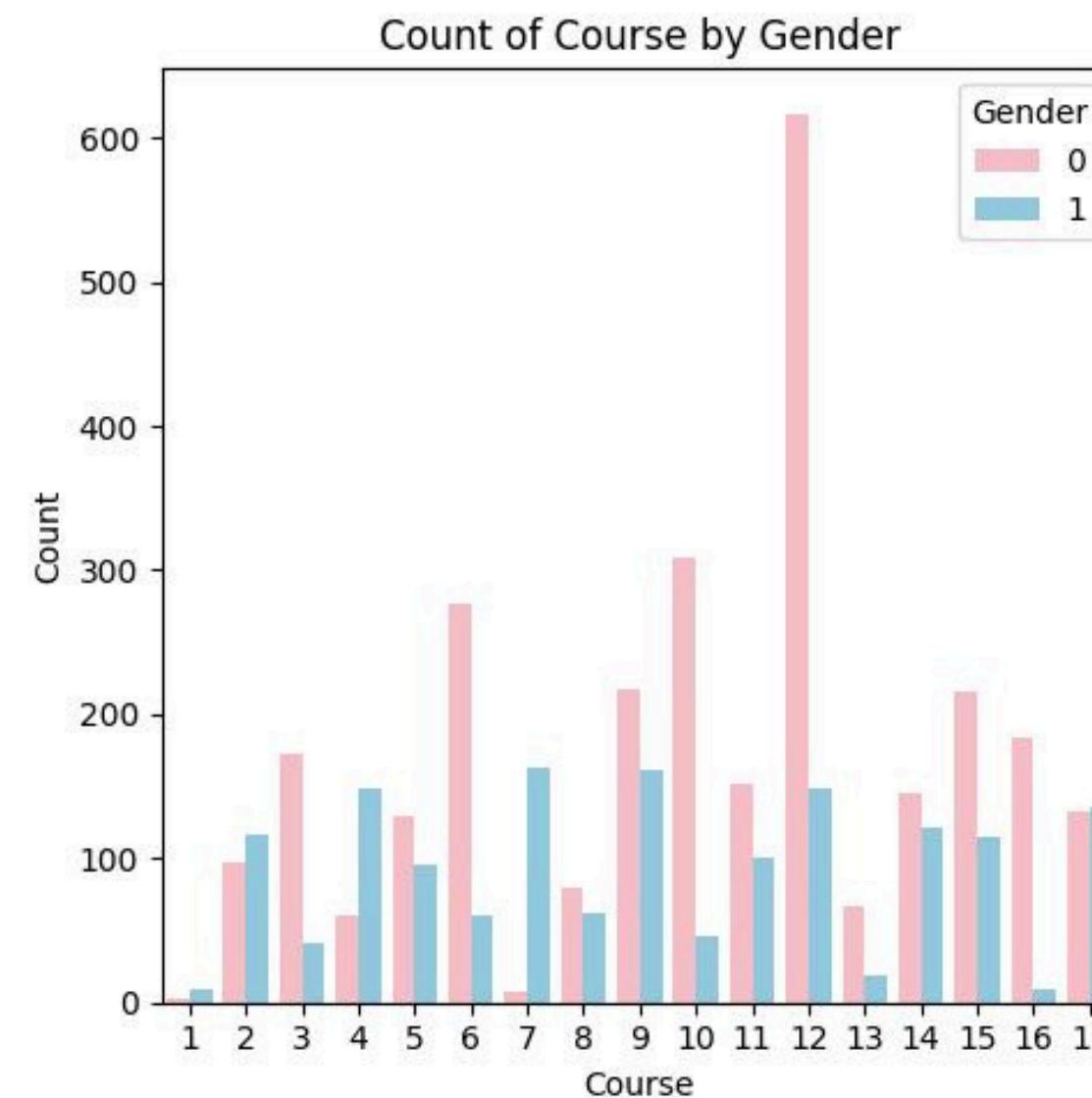
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9



- **Women** obtain more **scholarships** to sustain their studies.
- Additionally, they **start** their studies at a younger age, **22** years old, compared to **24** for men.
- The **widest range** of age at enrollment is for **men**, with a maximum of **70**, compared to **62** for women.

# Enrollment Patterns

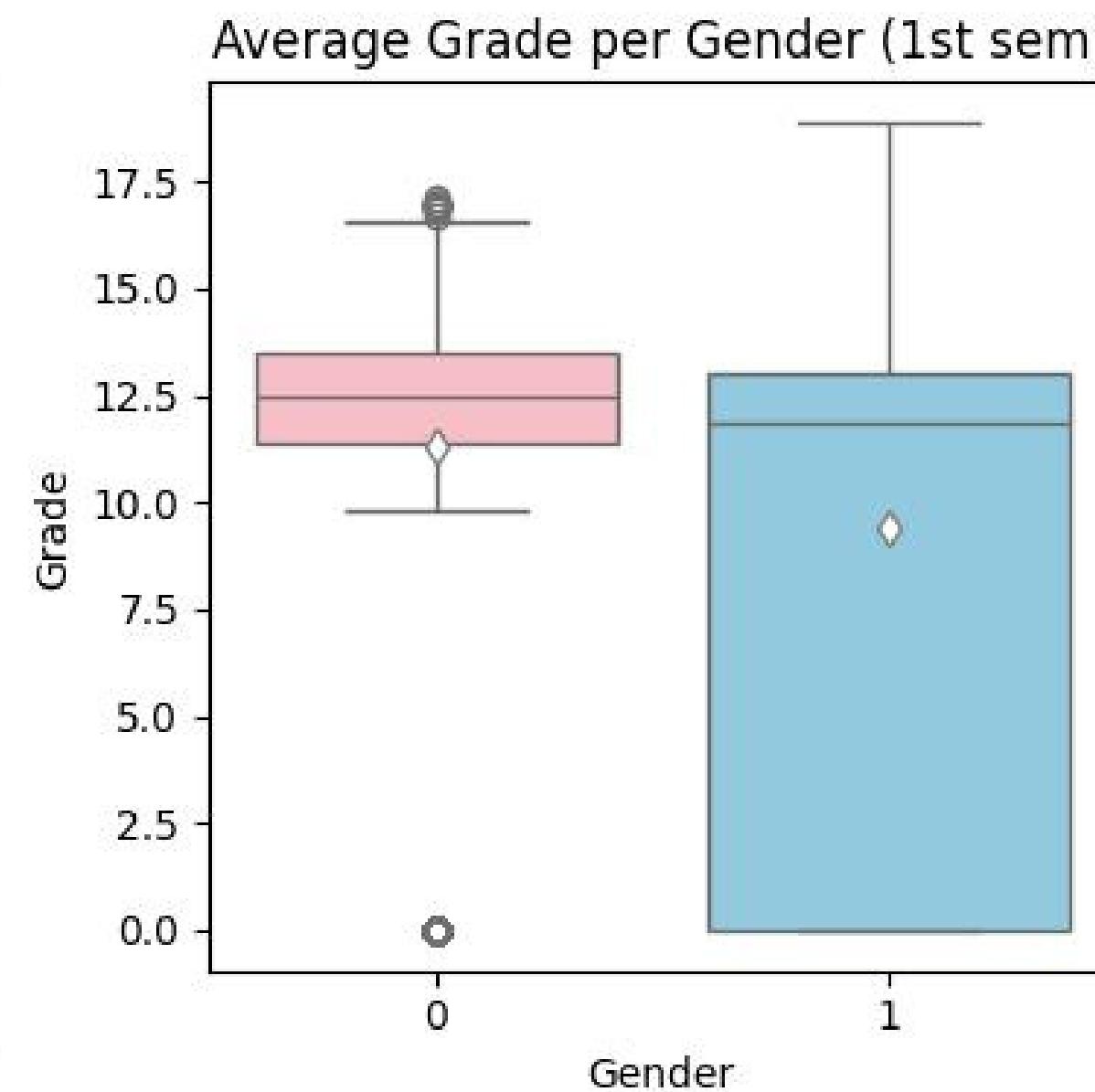
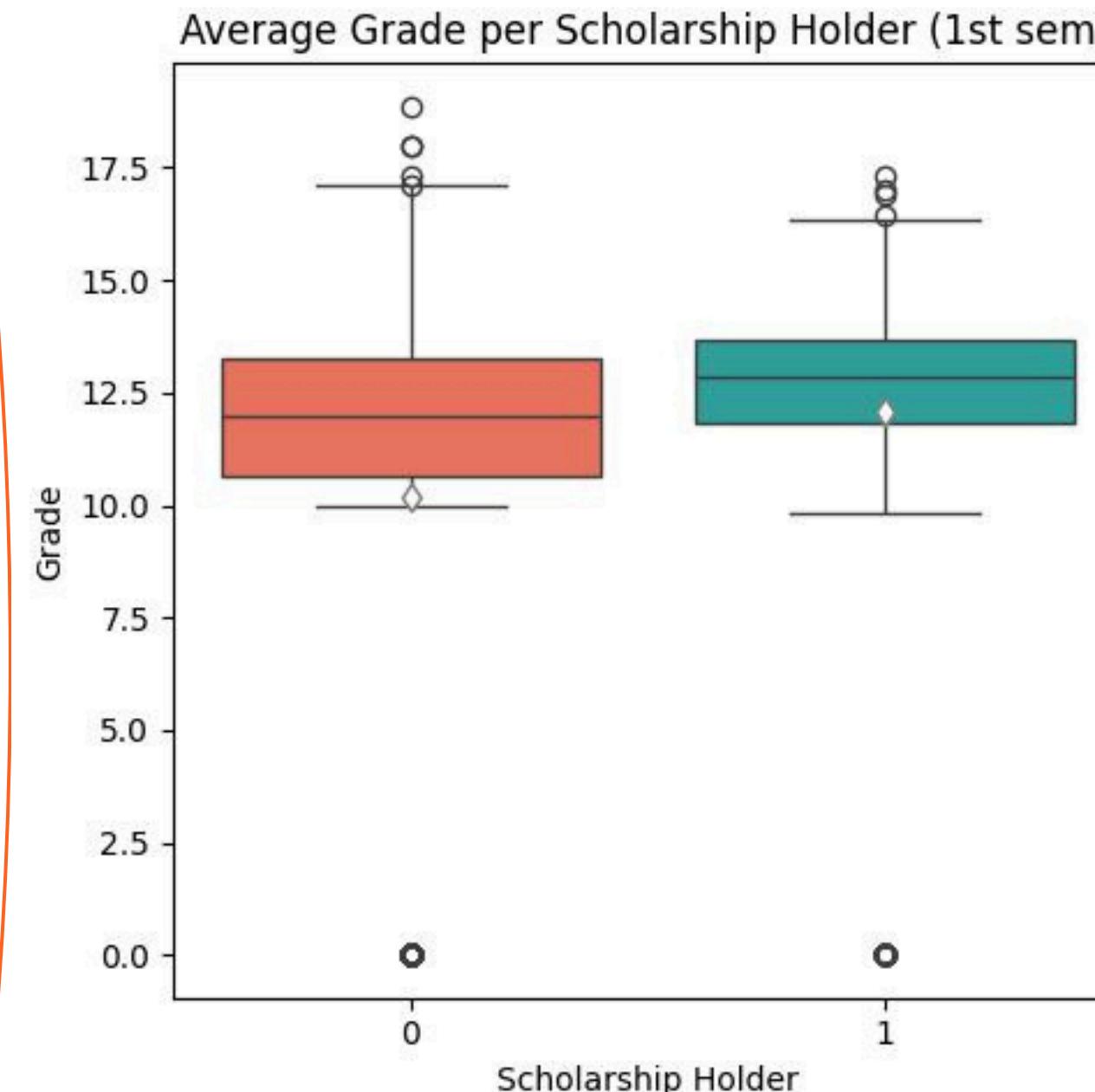
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9



- Women exhibit a higher presence in almost all courses.
- At older ages, the range of previous qualifications required to enroll in a certain course is wider compared to younger individuals.

# Academic Performance

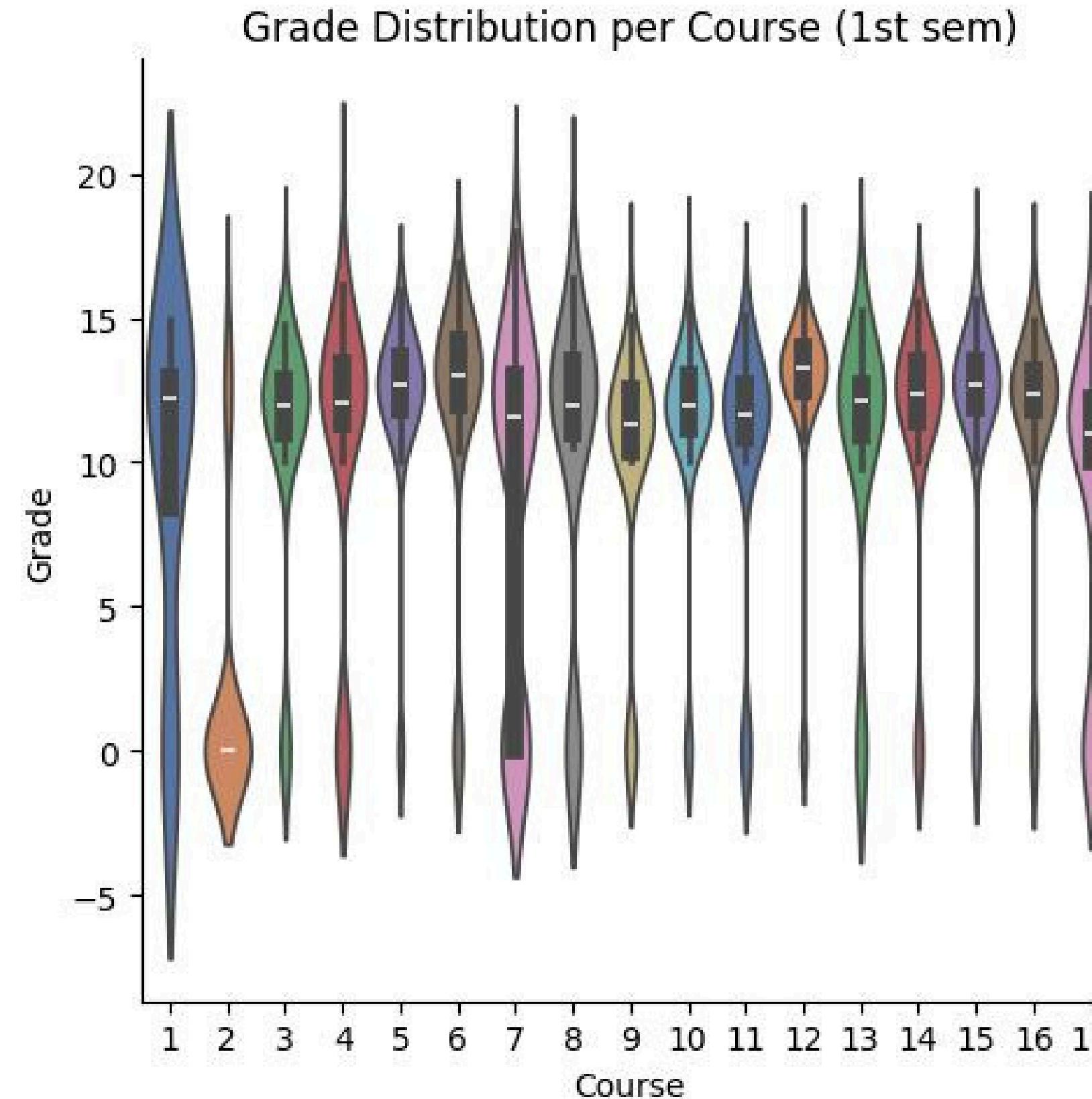
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9



- **Women** exhibit a higher average grade **11.31** compared with **9.39**, but the range is wider for men as well the max grade
- **Scholarship holders** exhibit a better academic performance with an average grade of **12.05**

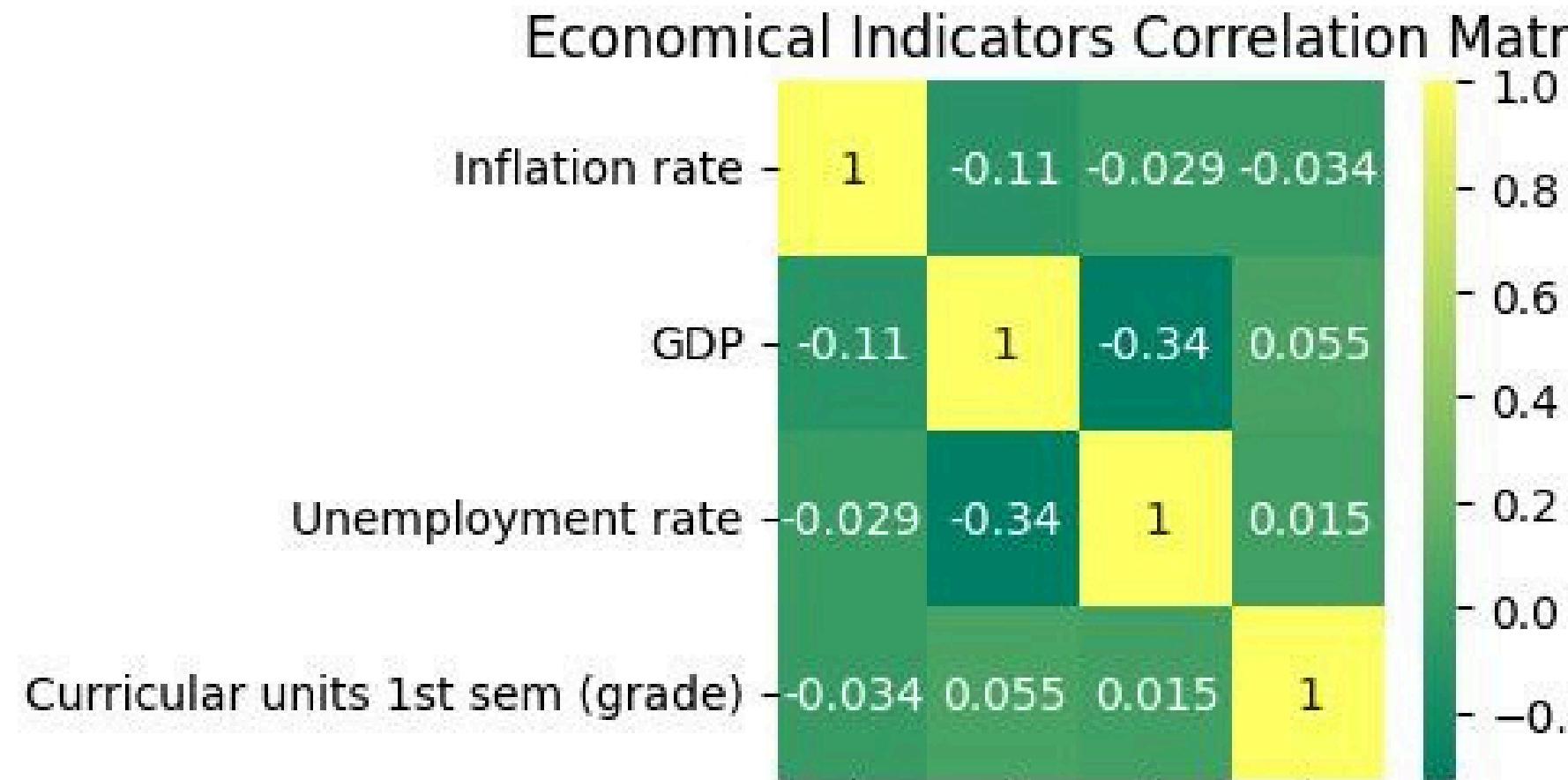
# Academic Performance

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9



- **Grades** distribution across **Course** have a similar average grade between **9 to 12.4**, except for the **course 2** where the average grade is **2.06**

# Economic Indicator



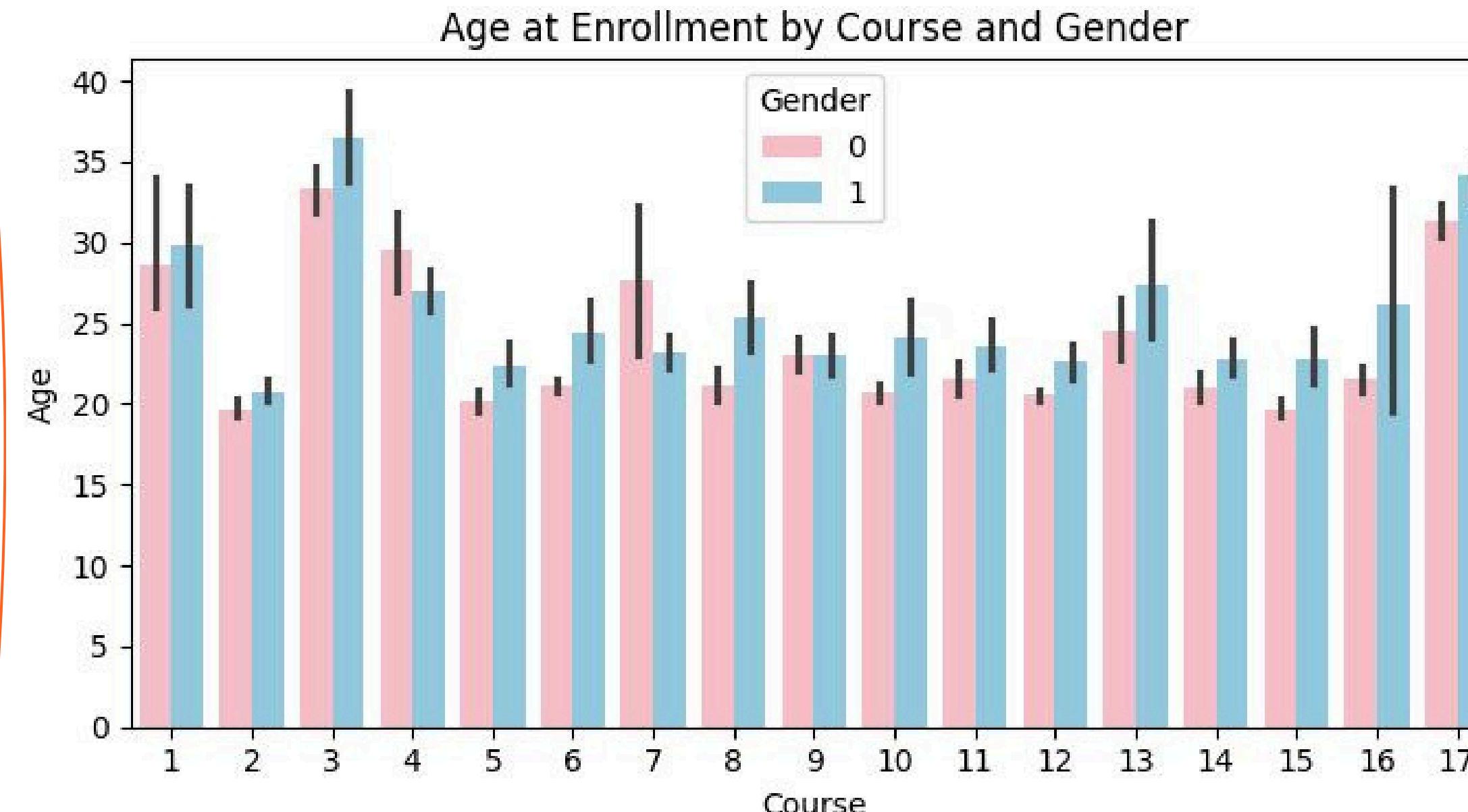
The correlation matrix between the economic indicators and academic performance during the 1st semester concludes that the variables are not related.

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

# Multivariate Analysis

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

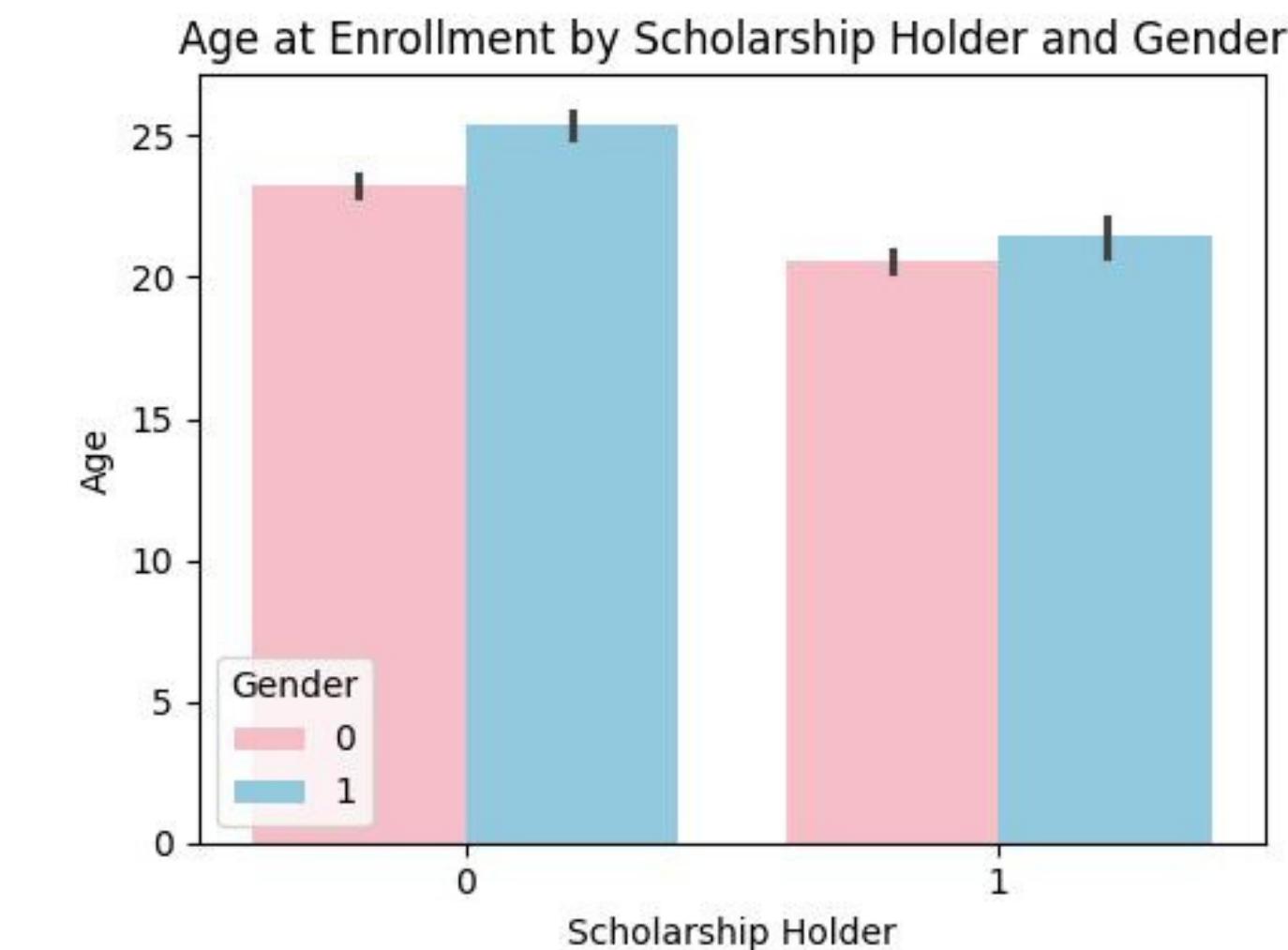
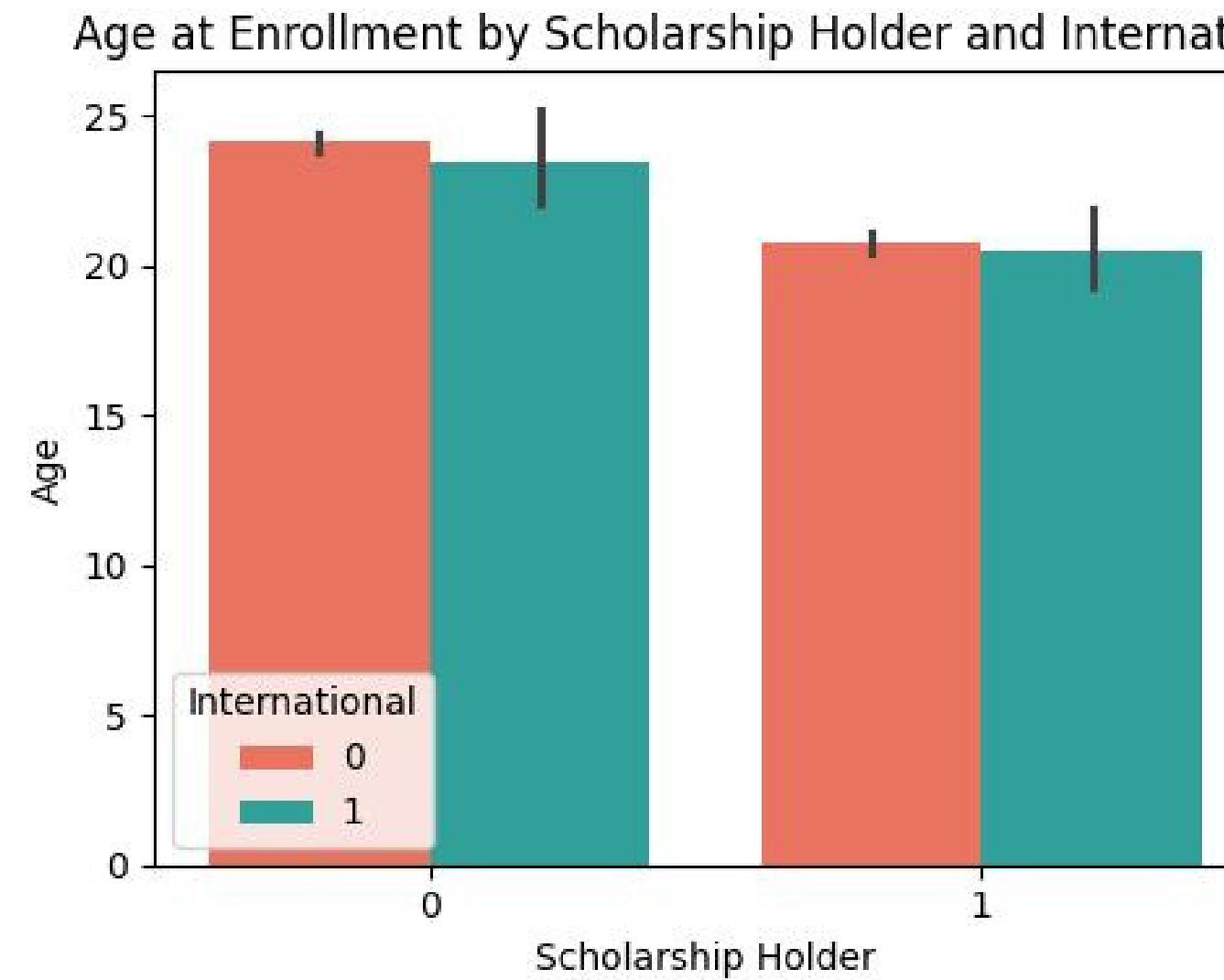
# Enrollment Patterns



The graph shows the relation between age, course and gender. The men present a higher range of Age at enrollment compared with women across the different courses.

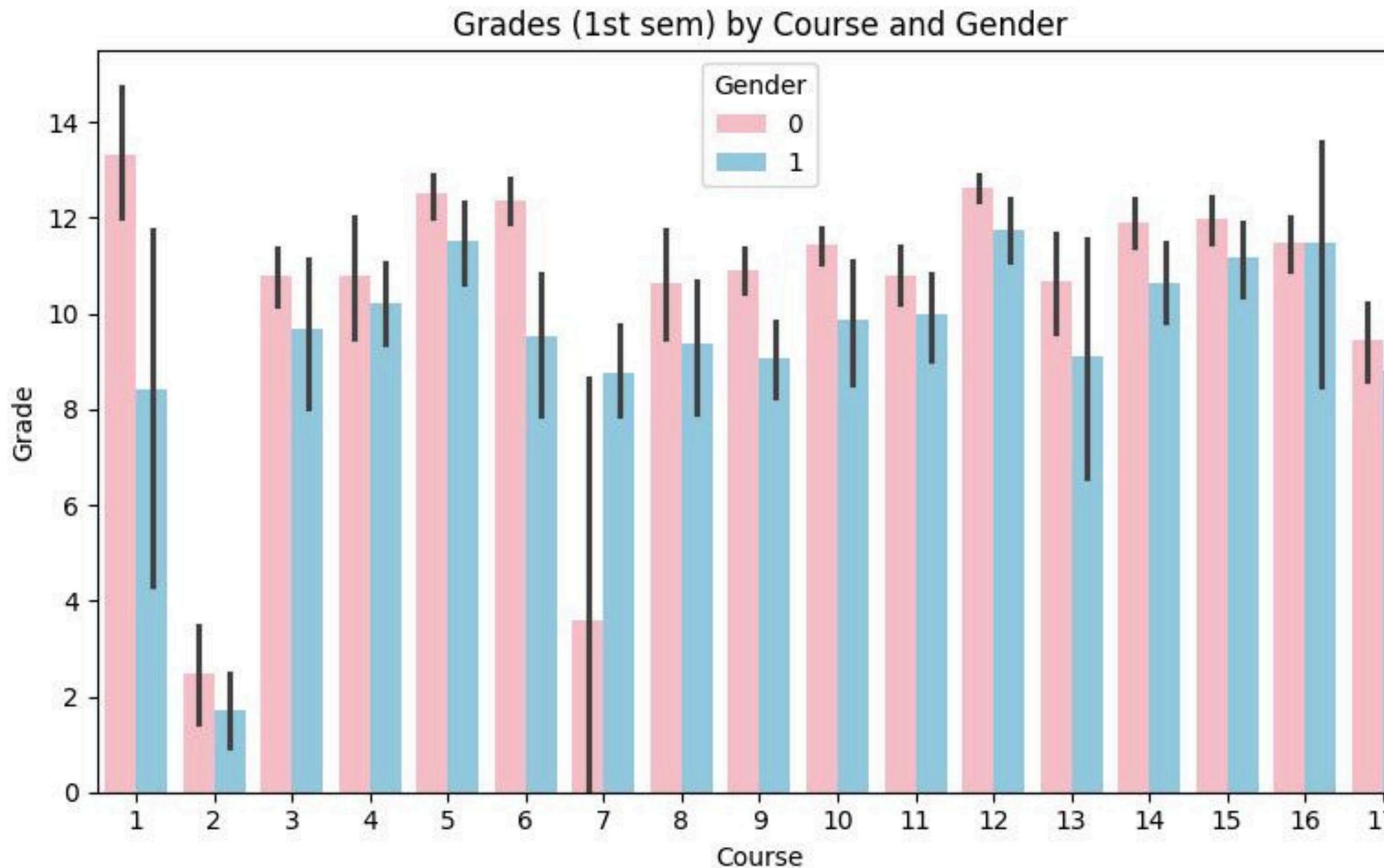
# Enrollment Pattern

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9



- In the graph depicting age, scholarship, and international student status, international students tend to study at younger ages, as do scholarship holders.
- In the graph illustrating age, scholarship, and gender, women with scholarships enter university at younger ages.

# Academic Performance

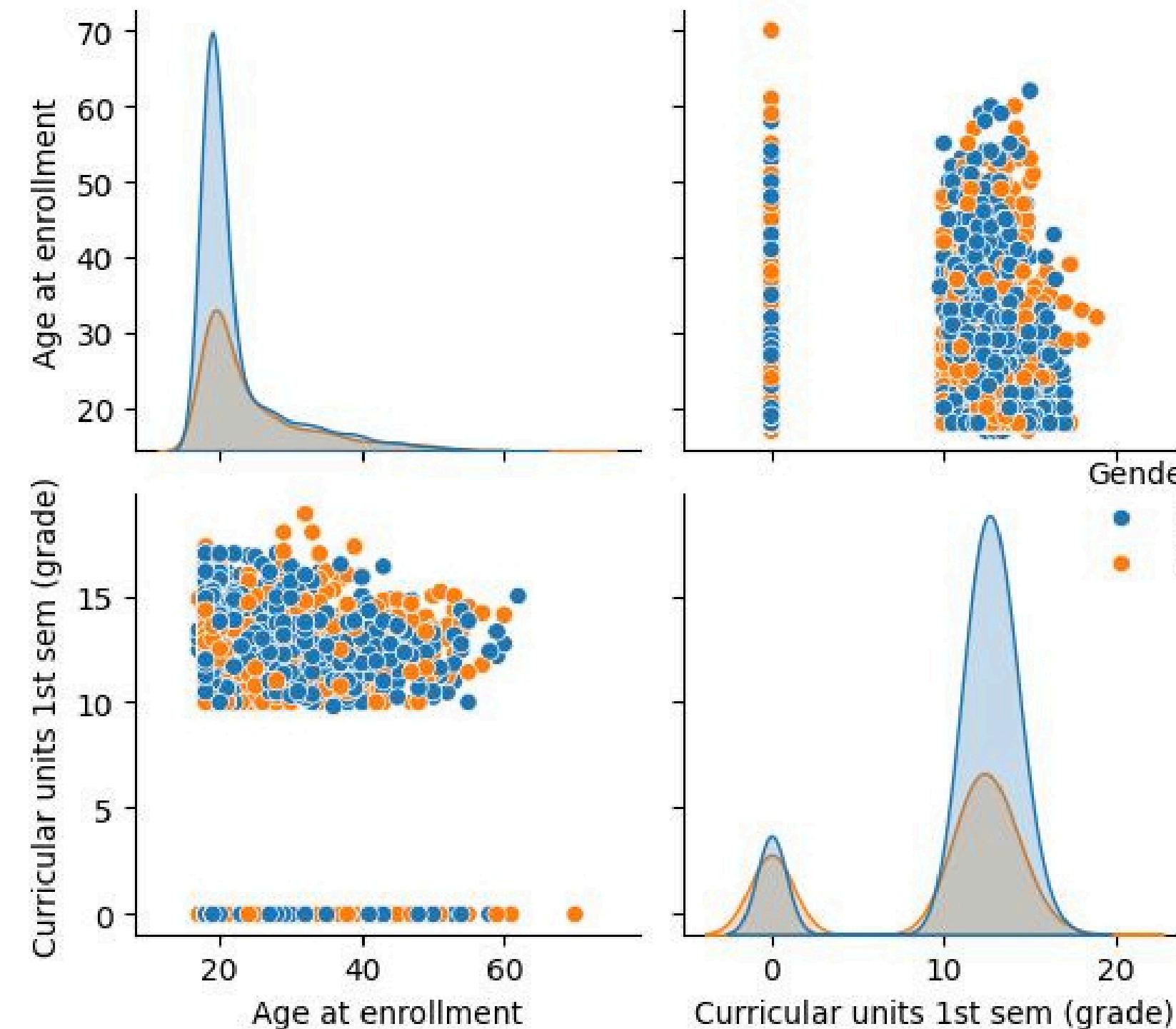


In the graph, women tend to have better grades compared to men across all course distributions, except for one course where they exhibit one of the lower performances.

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

# Academic Performance

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

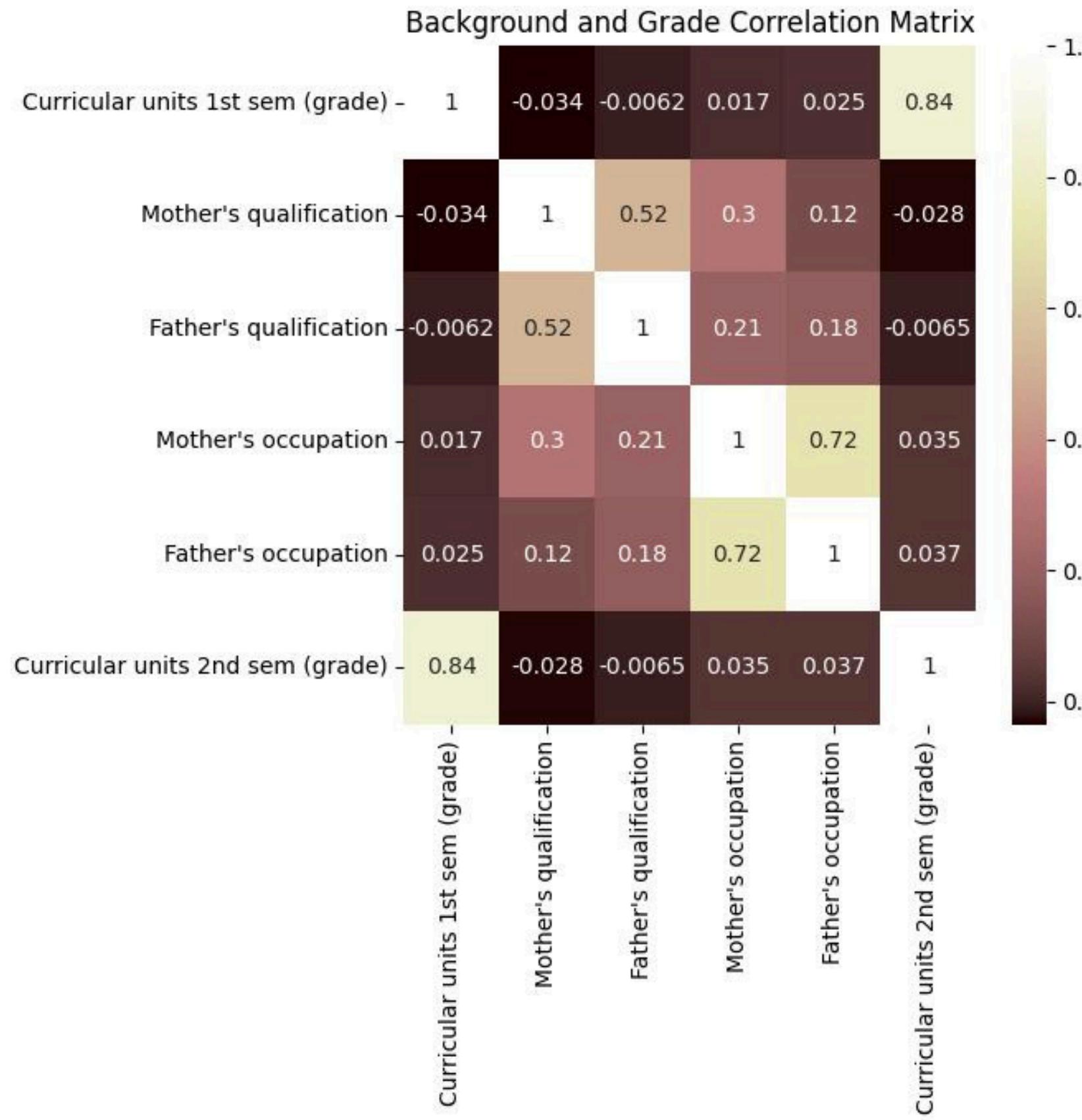


In this graph, an apparent relation between age at enrollment and grades emerges, where performance tends to be lower at older ages.

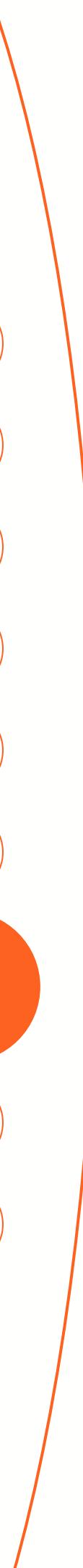
Therefore, this relation will be explored in the next section using Supervised Machine Learning Algorithms.

# Academic Performance

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9



The correlation matrix reveals a relationship between the grades during the 1st and 2nd semesters, which will be studied in the next section. Additionally, the correlation matrix indicates that the family background of the students does not significantly affect their performance.



# **Supervised Machine Learning Algorithms**

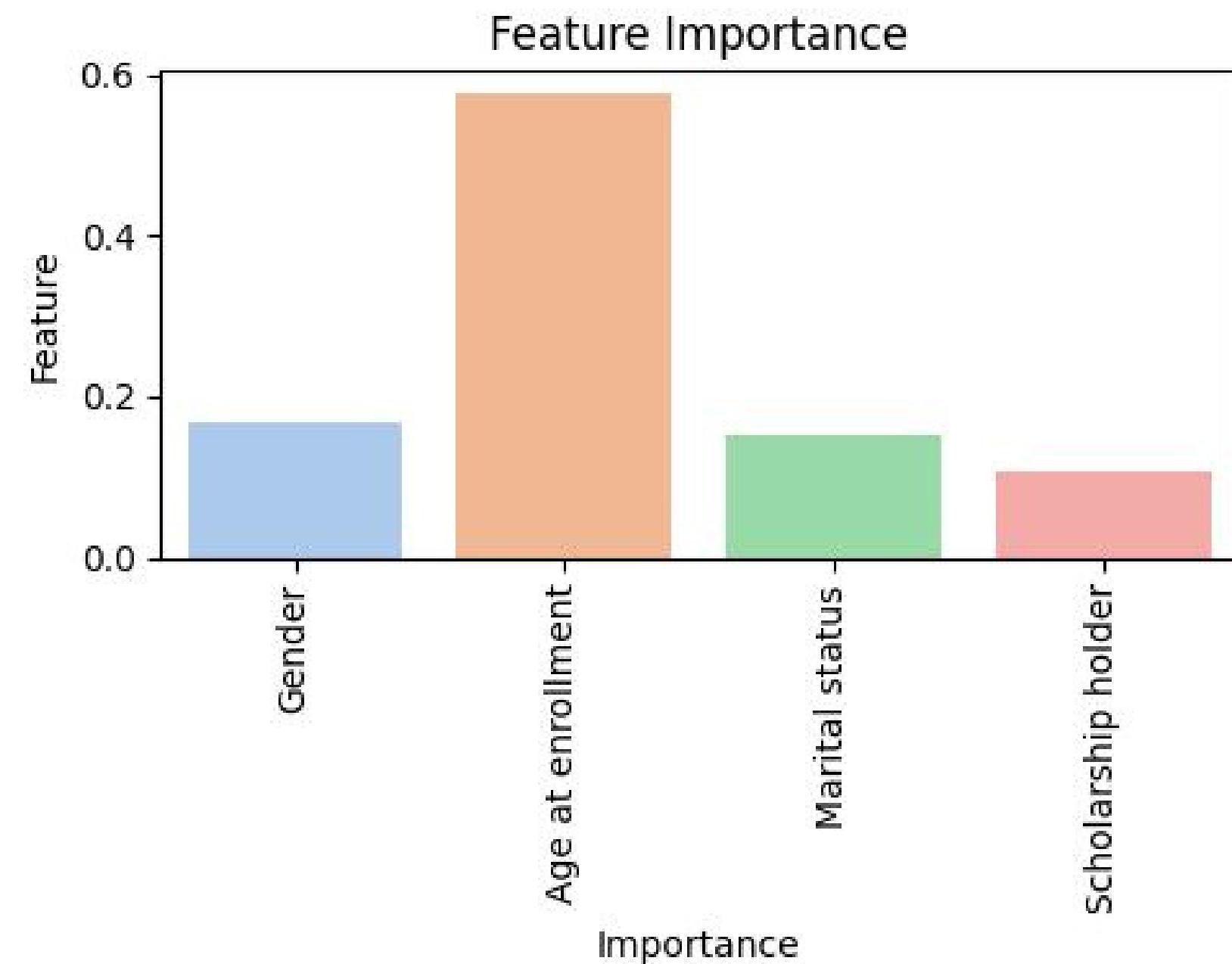
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

# Random Forest

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

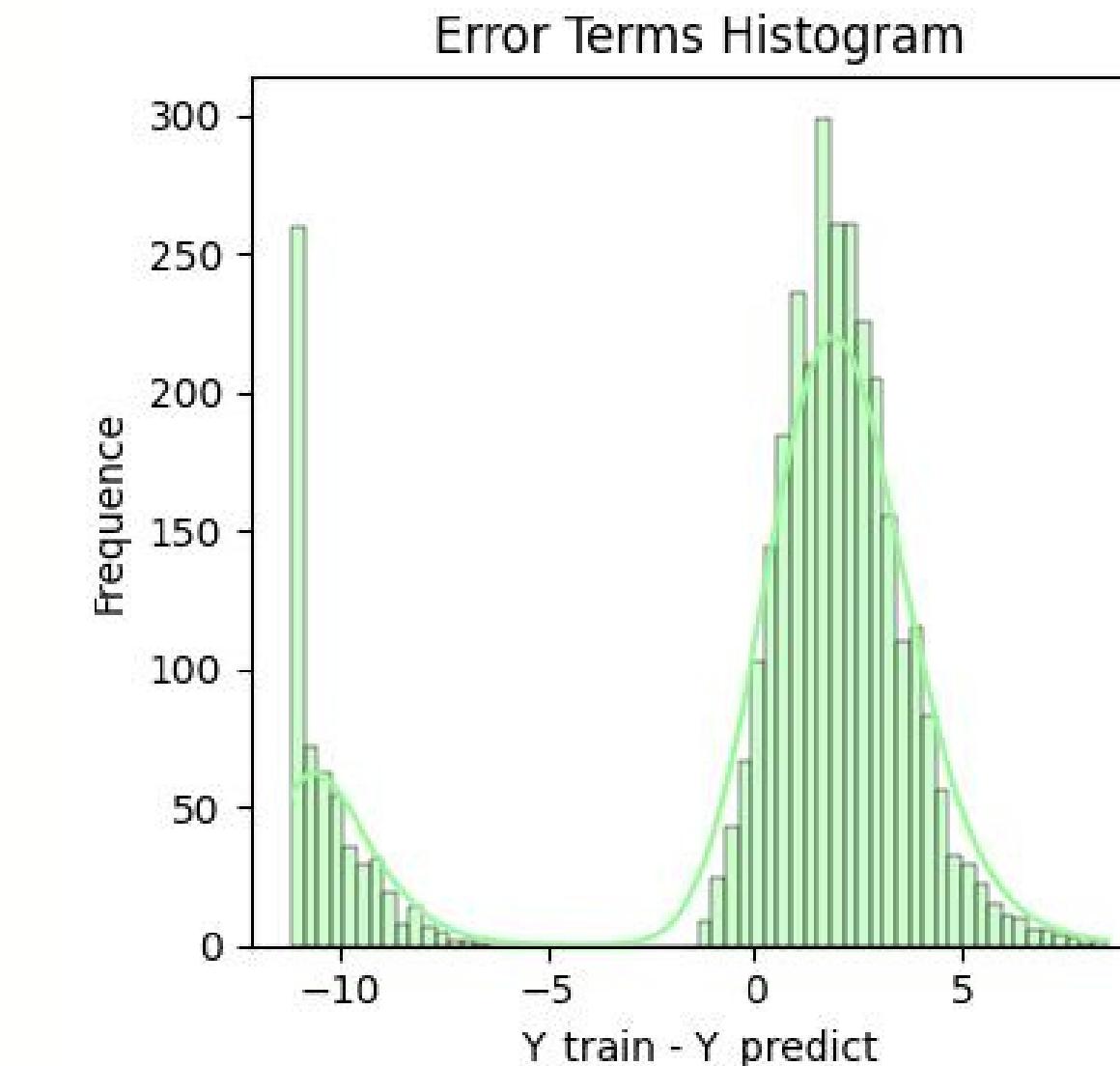
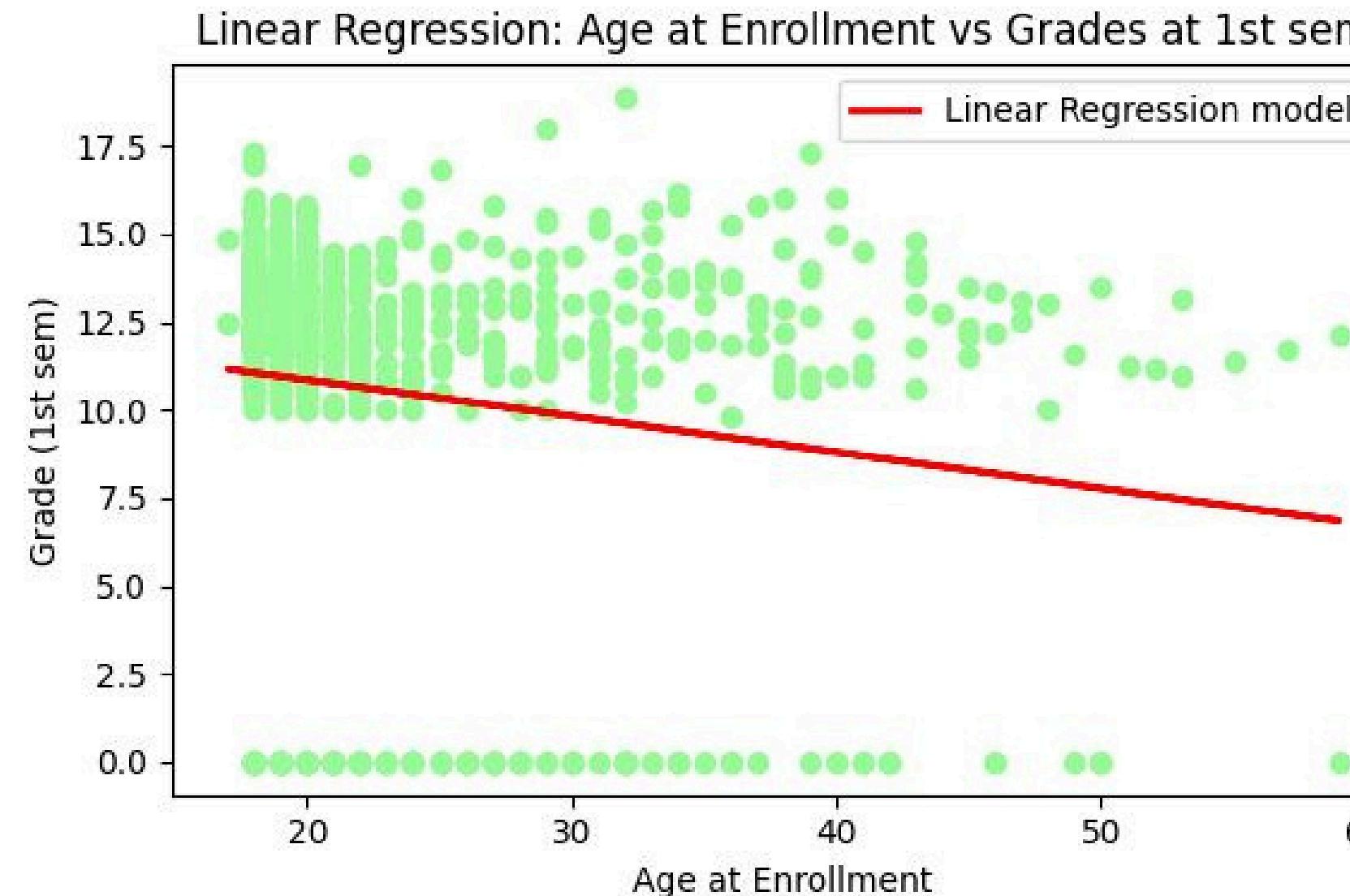
In applying the Random Forest algorithm to the dataset with variables 'Gender', 'Marital status', and 'Scholarship holder' alongside 'Age at enrollment', it becomes evident that age emerges as the most influential feature in predicting 'grade'.

This underscores the significant role age plays in determining academic performance during the first semester.



# Linear Regression

- 1
- 2
- 3
- 4
- 5
- 6
- 7



**Grades  $12.912 - 0.102 * \text{Age}$**   
**R-squared: 0.0010**

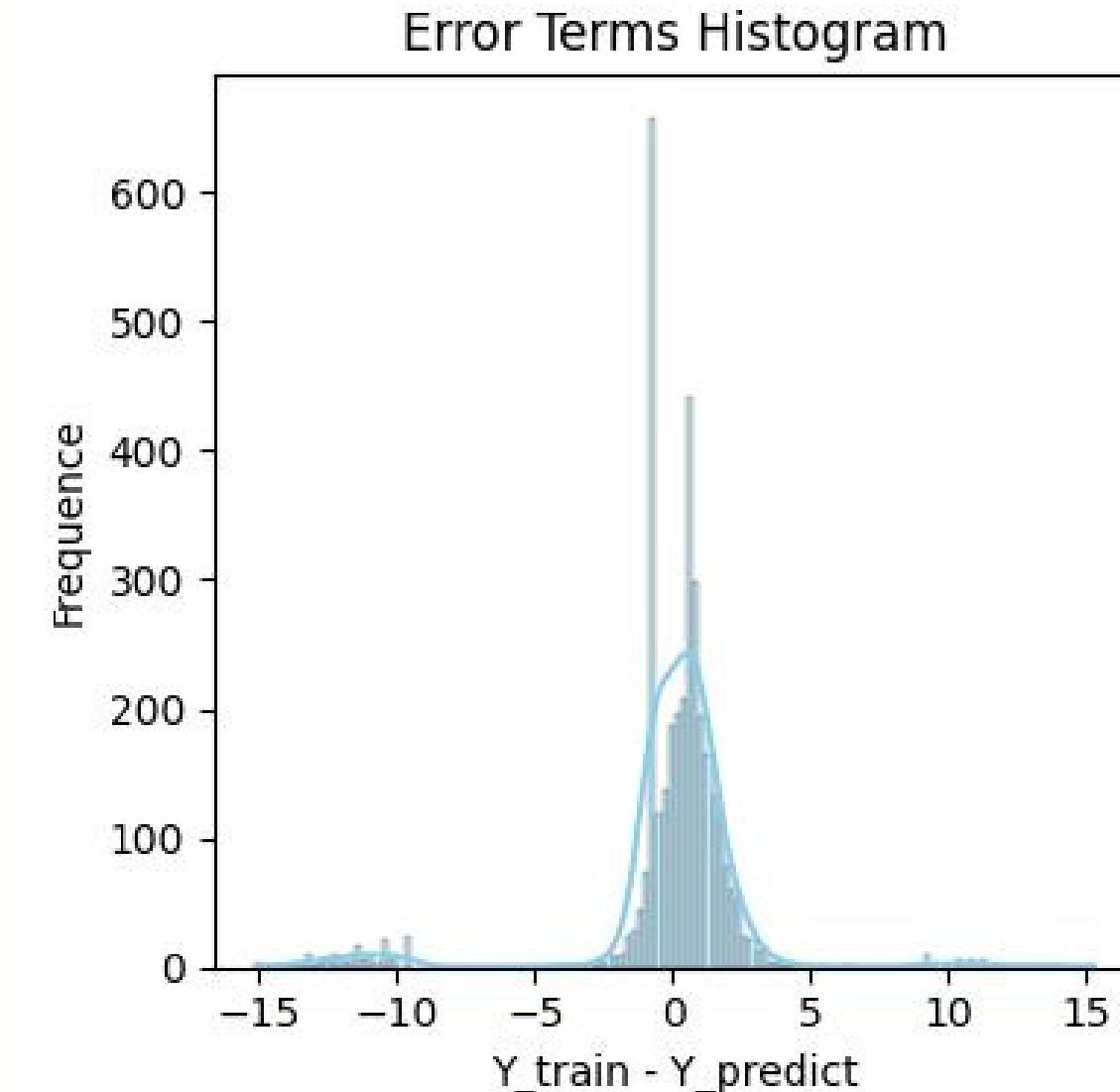
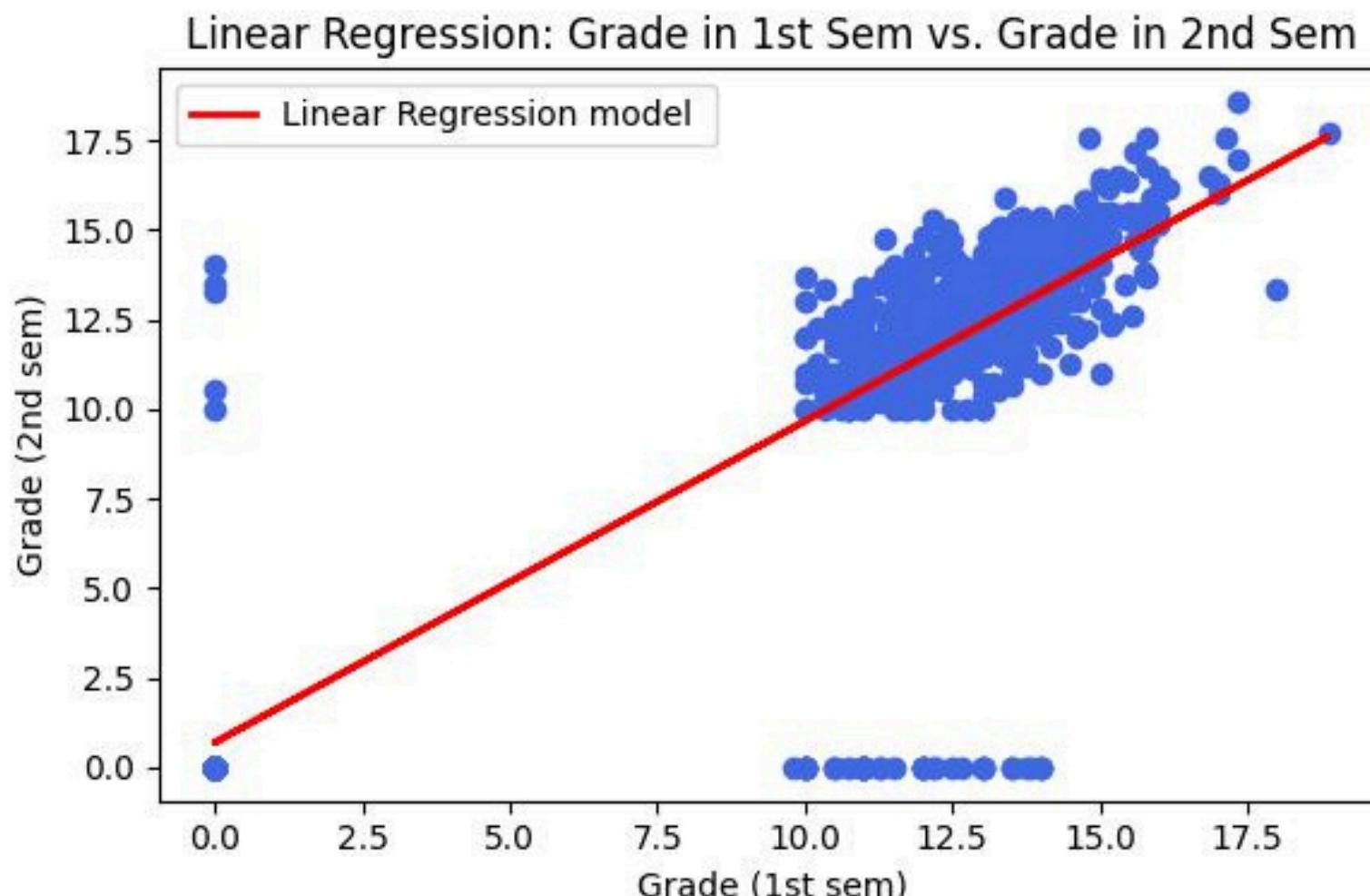
A linear regression model was evaluated between Age and Grade (1st semester), but the Linear Regression was not the most appropriate choice, as the R-squared value was low at 0.001. This could be attributed to outliers, such as grades equal to 0.

Additionally, a negative correlation was observed, indicating that academic performance tends to decrease with older ages

# Linear Regression

**Grades (2nd sem) = 0.69 + 0.896 \* Grades 1st sem**  
**R-squared : 0.692**

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9



A linear regression model was evaluated between Grades in both semesters, but the Linear Regression was not the most appropriate choice, as the R-squared value was low at 0.69. This could be attributed to outliers. Additionally, a positive correlation was observed, indicating that academic performance tends to be similar in the next semester

# Insights



- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

1. **Gender Disparity:** Women appear to be more prevalent across all courses, suggesting a potential gender disparity in enrollment rates within the educational system.
2. **Age and Enrollment:** There is a wide range of ages at enrollment, indicating that individuals enter higher education institutions at various stages of life. Additionally, older individuals tend to enroll in courses requiring a wider range of previous qualifications compared to younger individuals.
3. **Scholarship Impact:** Scholarship holders and international students tend to start their studies at younger ages compared to non-scholarship holders and non-international students. This suggests that scholarships may facilitate access to education for younger individuals.
4. **Family Background and Performance:** The dataset suggests that the familial background of students does not significantly impact academic performance. This indicates a level playing field in terms of performance regardless of family circumstances.
5. **Economic Indicator:** The dataset suggests that the economic indicators such as GDP, Unemployment rate and Inflation rate does not affect the academic performance.

# Conclusions

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

- Gender disparities exist in enrollment rates, while age at enrollment varies widely, with scholarships potentially facilitating access for younger individuals.
- Family background does not seem to significantly affect academic performance, indicating a level of equality in educational opportunities.
- Based on the analysis, it is evident that while a linear regression model can provide some insights into the relationship between grades in both semesters, it may not be the most suitable approach due to the presence of outliers and the limited explanatory power of the model.
- Further investigation using more advanced modeling techniques or addressing issues such as outliers could potentially improve the accuracy of predicting academic performance across semesters.



# Further details

## GitHub

<https://github.com/SadyGarcia/Educational-Case-Study.git>

## LinkedIn

[www.linkedin.com/in/sadymonserrat-garcía-delgadillo](https://www.linkedin.com/in/sadymonserrat-garcía-delgadillo)

