

Student Performance and Dropout Analysis

Data Visualization Report





Artificial Intelligence department
King Abdullah || School of Information Technology
University of Jordan

Rama Amayrh
Layan Balbisi
Leen Samman
Zaina Abu-Naser

Supervised by
Professor Omar Al-Qadi

Contents

Introduction.....	5
Dataset Description.....	5
GUI Design and Functionality	6
Data Visualization Analysis.....	7
Distribution by Tuition Fees Status	7
Distribution by Scholarship Status.....	7
Distribution by Attendance	8
Distribution by Debtor Status	9
Distribution by International Status	9
Distribution by Educational Special Needs	10
Gender Distribution by Target Categories	11
Target Counts by Gender	11
Scholarship Holders by Target.....	12
Grade Distribution by Target	13
Age at Enrollment by Target.....	13
Daytime vs Evening Attendance.....	14
1 st Semester Student Performance.....	15
2 nd Semester Student Performance	15
1 st Semester Student Performance Distribution	16
2 nd Semester Student Performance Distribution.....	17
Cumulative frequency 1 st Semester Grades	17
Cumulative frequency 2 nd Semester Grades	18
QQ Plot for First Semester Grades	19
QQ Plot for Second Semester Grades	19
Comparison of grade distribution	20
Performance Evaluation by Nationality	21
Global Distribution	22
Top 5 Nationalities by Number of Students Excluding Portuguese.....	23
Target Distribution by Nationality	23
Performance Flow.....	24
Mean Grades by Nationality (1 st Semester)	25
Mean Grades by Nationality (2 nd Semester)	25
1 st Semester Performance Grid	26
2 nd Semester Performance Grid.....	27

Data Columns Overview	27
Challenges and Future Work	28
Conclusion	28
Table of Figures	29

Introduction

The objective of this project is to develop an educational data analytics dashboard that visualizes key aspects of student performance, attendance, and engagement. The aim is to uncover hidden patterns in the data that may influence academic success and to present these insights in a user-friendly graphical user interface (GUI). The dashboard is designed to assist educators and administrators in making data-driven decisions to enhance educational outcomes.

This report provides a comprehensive guide to the developed dashboard and its features. It outlines the design and functionality of the GUI, which was made using the R Shiny framework. Readers will find detailed descriptions of the dashboard's key components, including interactive visualizations that allow users to explore patterns in student performance. The report also highlights significant findings, such as correlations between demographic factors and academic outcomes, and concludes with insights derived from the analysis. Lastly, it includes step-by-step instructions on how to navigate and utilize the application, ensuring accessibility for all users.

Dataset Description

The dataset used for this project was sourced from Kaggle, containing 4,426 records and 35 features in its original form. Key attributes include demographic details (e.g., marital status, nationality, and age at enrollment), academic metrics (e.g., grades, enrollment details, and course performance for both semesters), and external factors like unemployment and inflation rates. These attributes are integral to analyzing and visualizing student performance patterns.

The final dataset, "FinalDataDVv.xlsx," expands on the original data by including two additional derived columns: "First Sem Performance" and "Second Sem Performance." These columns classify students' academic outcomes into four categories: "Failing," "Passing," "Good," and "Excellent," based on their grades in the respective semesters. For "First Sem Performance," grades below 10 are categorized as "Failing," grades between 10 and 12 as "Passing," grades between 12 and 15 as "Good," and grades between 15 and 20 as "Excellent." The same grading thresholds apply to "Second Sem Performance." These derived columns offer an intuitive summary of academic achievements and enhance the dataset's ability to support interactive and insightful visualizations.

This dataset is particularly relevant for the project as it provides a rich and diverse set of attributes critical to understanding factors influencing academic success. The additional performance columns align with the project's objective to develop a data-driven educational dashboard, enabling educators and administrators to make informed decisions through a user-friendly interface and comprehensive visualizations.

GUI Design and Functionality

The graphical user interface (GUI) for the project was built using the **R Shiny** framework. This framework was selected due to its capability to produce interactive web applications with a responsive and user-friendly design. It serves as an excellent platform for visualizing educational data, enabling seamless interaction and analysis for users.

The interface features a structured layout organized with a menu bar containing multiple tabs, each dedicated to a specific type of analysis or visualization. The dashboard includes **30 interactive plots**, categorized under sections such as "Bar Plots," "Performance Plots," "Cumulative," "Nationality Plots," and others. Each tab contains one or more visualizations designed to help users explore and analyze the data intuitively.

The layout consists of well-defined components, including tabs for navigation and buttons for adjusting parameters such as filters and data selection. For example, the "Attendance Histogram" section compares daytime and evening attendance, providing a clear view of patterns and trends. These elements ensure that users can easily access relevant insights.

Users can interact with the application by selecting attributes to filter data or customize visualizations. This interactivity enhances the usability of the dashboard, enabling tailored data exploration. Additional functionality includes filtering options to focus on specific subsets of the data, such as performance metrics or demographic details, and a data export feature that allows users to download filtered datasets or visualizations for further analysis.

Below is an example of the dashboard's interface, featuring a **map visualization**. This map provides a geographic representation of student data, allowing users to analyze patterns across different regions.

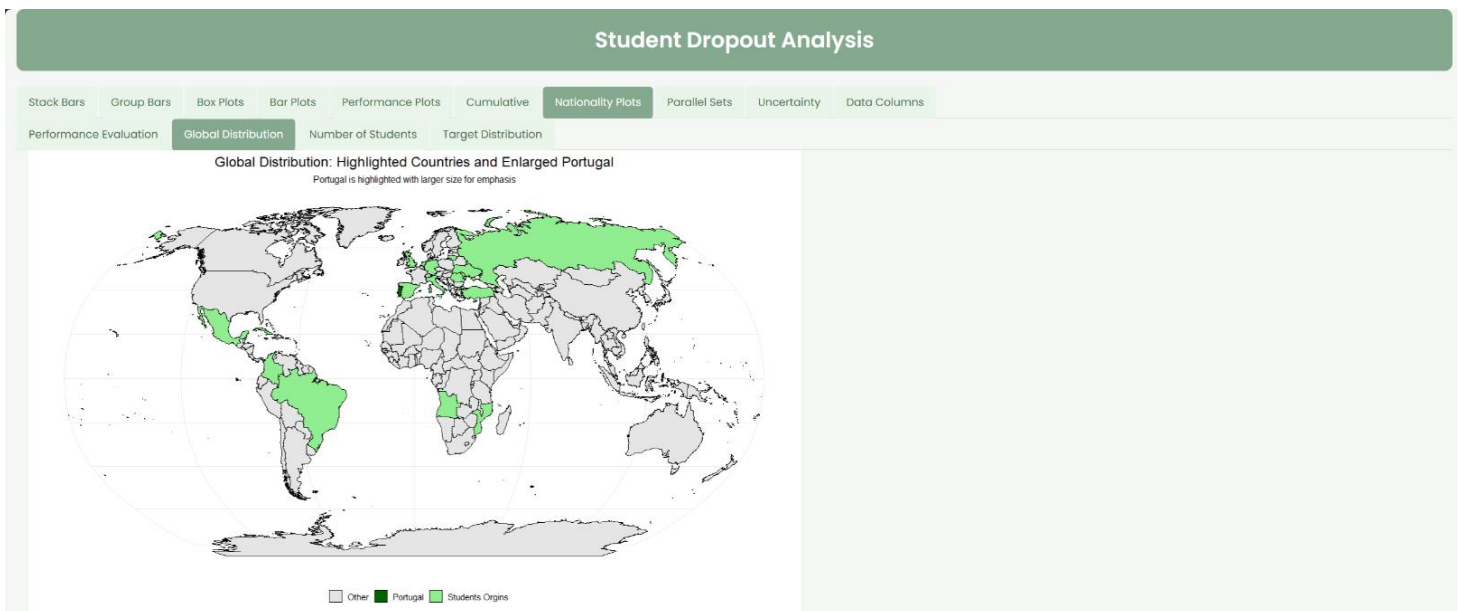


Figure 1: Example of dashboard's interface

The GUI provides a clean, accessible, and dynamic environment for educators and administrators to explore the data. Its design promotes efficiency and clarity, allowing users to uncover meaningful patterns and insights with ease.

Data Visualization Analysis

Distribution by Tuition Fees Status

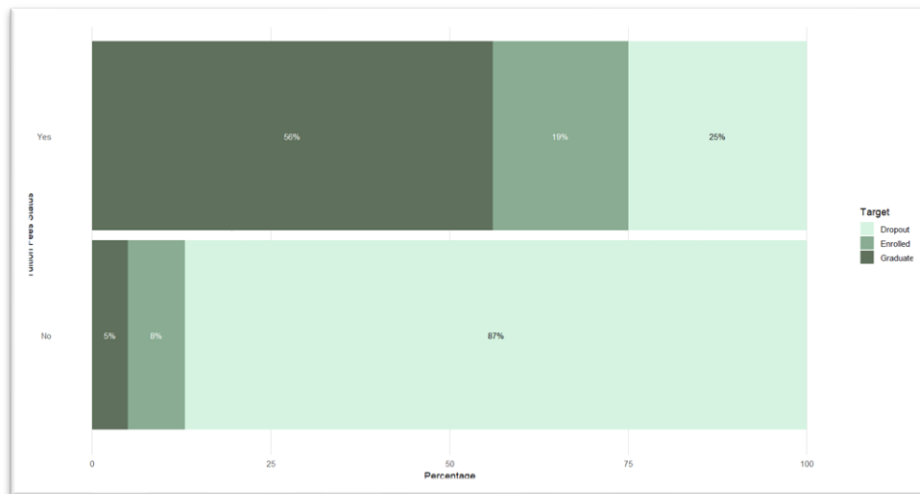


Figure 2: Distribution by Tuition Fees Status

The Distribution by Tuition Fees **Status Stacked Bar Chart** illustrates the percentage distribution of students based on their tuition fee payment status (paid/unpaid) and their final status (enrolled, graduated, or dropped out). The chart was chosen for its clarity in showing the relationship between tuition payment and student outcomes, highlighting significant patterns like the higher dropout rates among students with unpaid fees. The stacked design effectively displays the percentage for each category, making it easy to interpret and visually appealing

Distribution by Scholarship Status

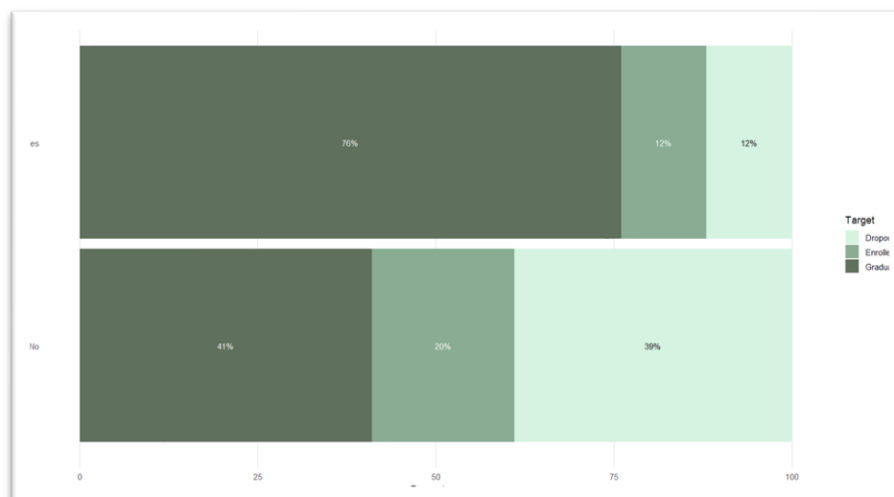


Figure 3: Distribution by Scholarship Status

The Distribution by Scholarship Status **Stacked Bar Chart** represents the percentage distribution of students based on their scholarship status (whether they hold a scholarship or not) and their final academic status (enrolled, graduated, or dropped out). It was selected to highlight the potential relationship between receiving a scholarship and student outcomes. The stacked bar design effectively showcases the proportion of each category, providing clear insights into patterns such as higher retention or graduation rates among scholarship holders. The inclusion of percentages within the bars makes the data easier to interpret at a glance.

Distribution by Attendance

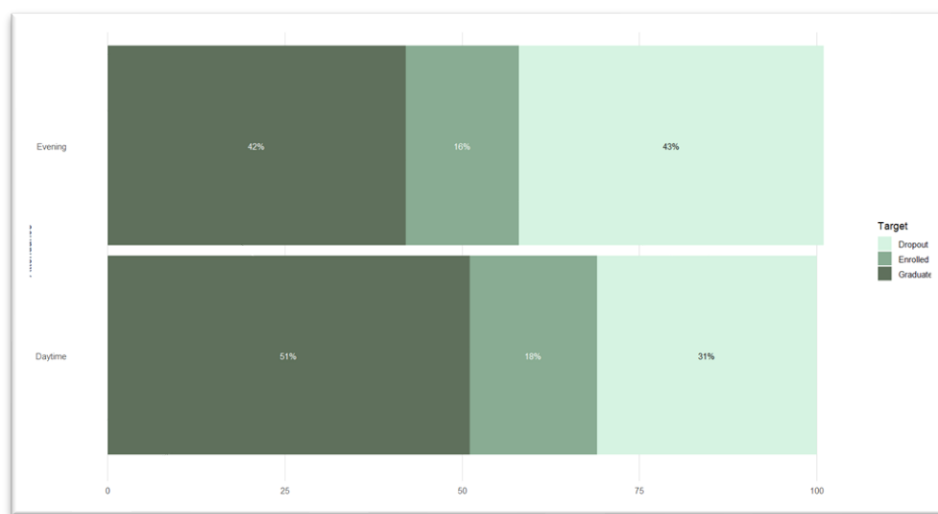


Figure 4: Distribution by Attendance

The Distribution by Attendance **Stacked Bar Chart** visualizes the percentage distribution of students based on their attendance type (daytime or evening) and their final academic status (enrolled, graduated, or dropped out). It was chosen to explore how attendance patterns may influence student outcomes. The stacked bar chart effectively highlights the differences between attendance types, showcasing potential trends such as higher dropout rates among specific attendance categories. The design ensures clarity by including percentage labels within the bars, enabling quick and intuitive interpretation of the data.

Distribution by Debtor Status

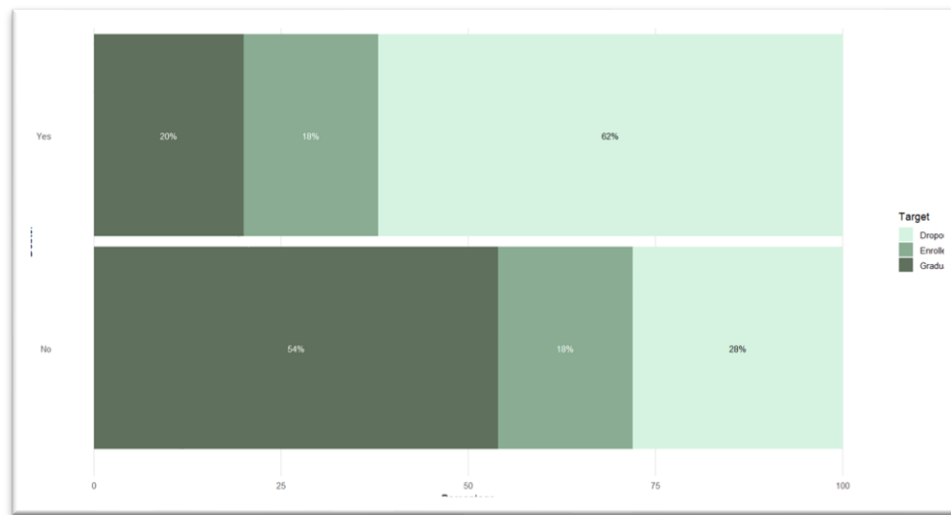


Figure 5: Distribution by Debtor Status

The Distribution by Debtor Status **Stacked Bar Chart** displays the percentage distribution of students based on their debtor status (whether they owe fees or not) and their final academic status (enrolled, graduated, or dropped out). It was selected to examine how being a debtor impacts student outcomes, such as the likelihood of dropping out or graduating. The stacked bar design provides a clear visualization of the proportions within each debtor category, helping to uncover significant patterns. The inclusion of percentage labels within the bars ensures the data is both accessible and easy to understand.

Distribution by International Status

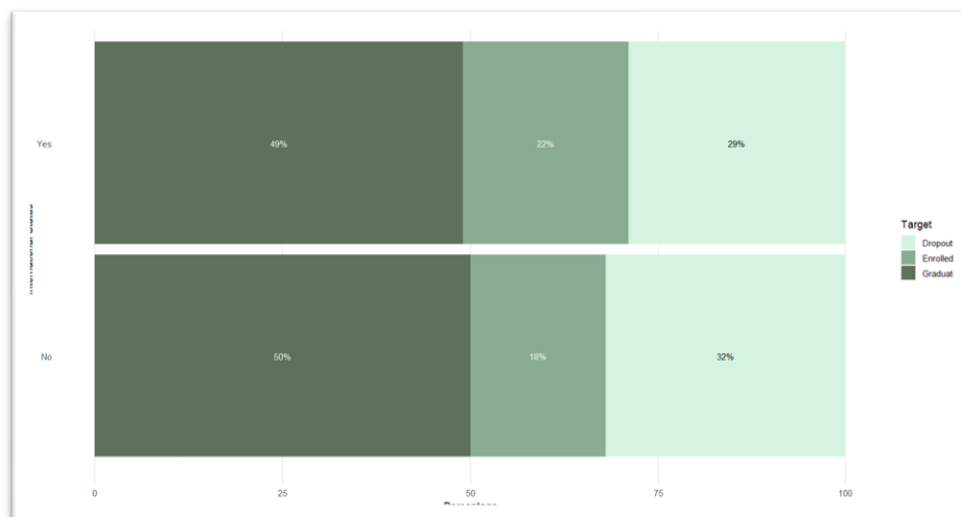


Figure 6: Distribution by International Status

The Distribution by International Status **Stacked Bar Chart** illustrates the percentage distribution of students based on their international status (whether they are domestic or international students) and their final academic status (enrolled, graduated, or dropped out). The visualization was chosen to explore how being an international student may influence academic outcomes, shedding light on potential differences in retention or success rates. The stacked bar design allows for a clear comparison between categories, while the inclusion of percentage labels makes the data easier to interpret and visually engaging.

Distribution by Educational Special Needs

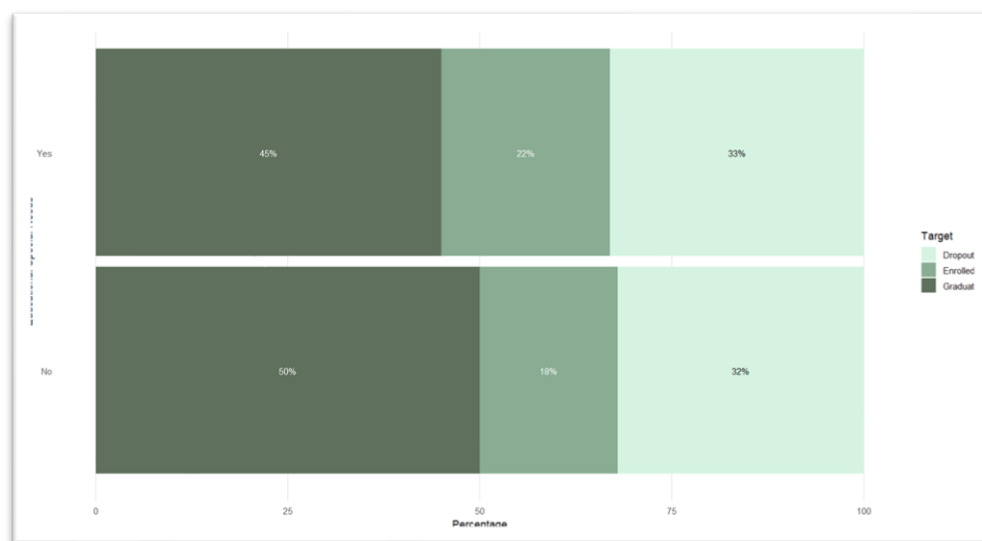


Figure 7: Distribution by Educational Special Needs

The Distribution by Educational Special Needs **Stacked Bar Chart** shows the percentage distribution of students with and without educational special needs, categorized by their final academic status (enrolled, graduated, or dropped out). The plot was chosen to examine how educational special needs may impact student outcomes, highlighting potential differences in retention and success. The stacked bar chart design effectively displays the proportions within each category, making it easy to compare the outcomes of students with special needs to those without. The inclusion of percentage labels within the bars ensures the data is easily interpretable and visually appealing.

Gender Distribution by Target Categories

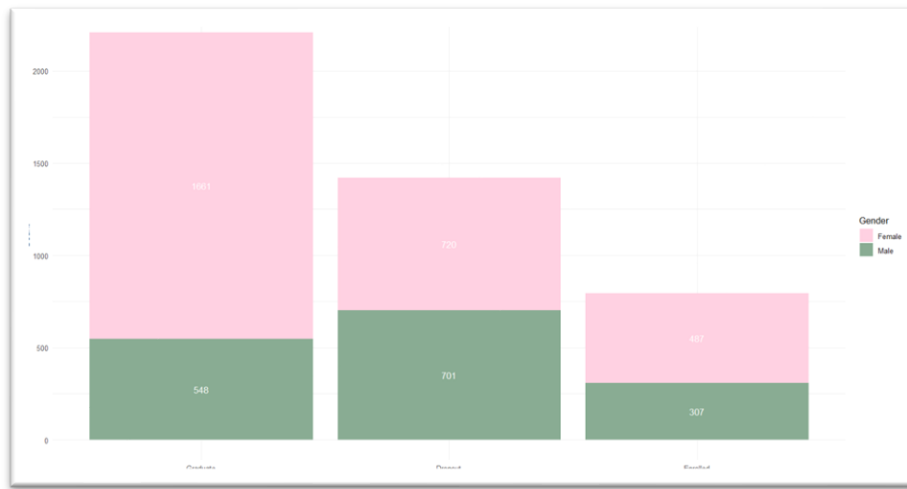


Figure 8: Gender Distribution by Target Categories

The Gender Distribution by Target Categories **Stacked Bar Chart** shows the distribution of students by gender across different target categories, highlighting the number of individuals within each category based on gender. The target categories are ordered according to the total number within each category, providing insights into the gender distribution. The stacked bar chart design was chosen to effectively display the proportions within each category, making it easy to compare outcomes between male and female students. The inclusion of percentage labels within the bars ensures that the data is easily interpretable and visually appealing.

Target Counts by Gender

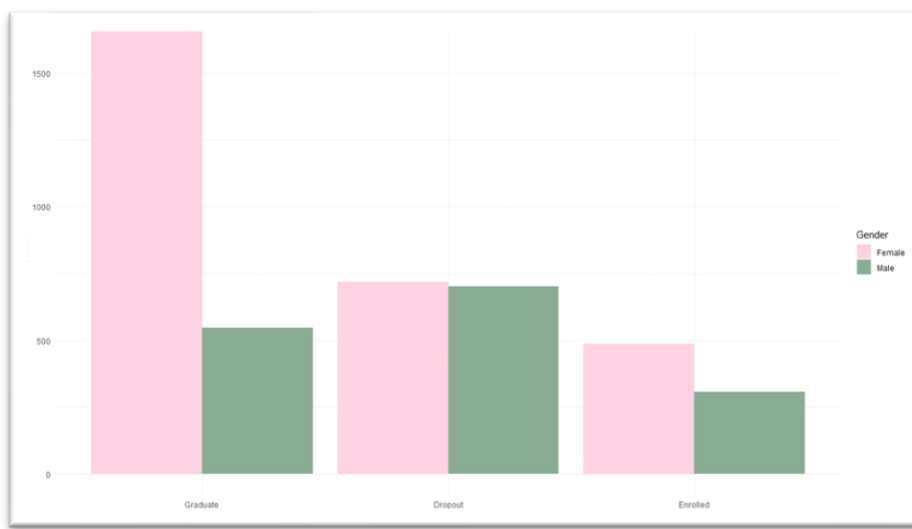


Figure 9: Target Counts by Gender

The Target Counts by Gender **Dodged Bar Chart** displays the counts of students in different target categories, segmented by gender. The bars represent the number of students within each target category, with male and female students differentiated by color. The chart provides a clear comparison of the gender distribution across the various target categories, allowing us to examine any potential disparities in the number of male and female students within each category. The use of distinct colors ensures easy identification, while the inclusion of percentages can further help in understanding the proportional representation of each gender in the categories.

Scholarship Holders by Target

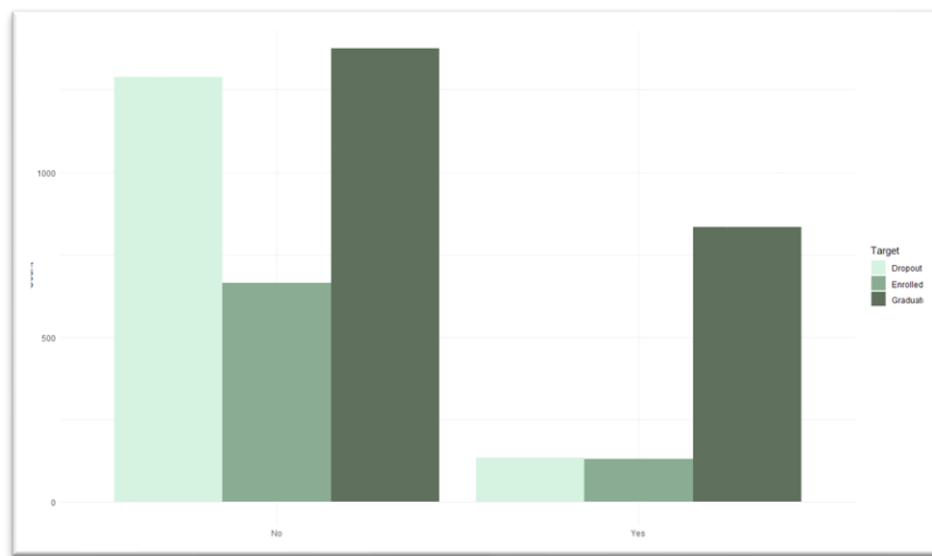


Figure 10: Scholarship Holders by Target

The Scholarship Holders by Target **Dodged Bar Chart** illustrates the distribution of scholarship holders across different target categories (Graduate, Dropout, Enrolled). The bars represent the count of scholarship holders, with separate bars for each target category. This design allows for a clear comparison of the scholarship status within each category, showcasing whether scholarship holders are more prevalent among graduates, dropouts, or enrolled students. The color-coding ensures the distinction between the target categories, providing a visually appealing and easy-to-understand comparison.

Grade Distribution by Target

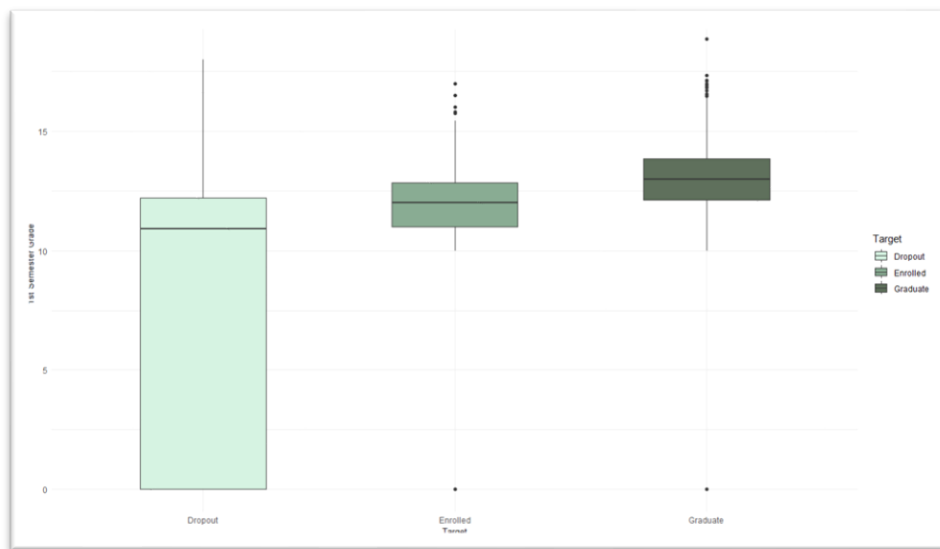


Figure 11: Grade Distribution by Target

The Grades Distribution by Target **Box Plot** displays the distribution of first-semester grades across different target categories (Graduate, Dropout, Enrolled). The box plot allows for a clear visual representation of the grade distributions, including the median, interquartile range, and any outliers for each target category. By comparing the grade distributions, we can analyze how academic performance varies among students with different outcomes. The color-coding of the target categories enhances the clarity and visual appeal, making it easy to distinguish between the groups.

Age at Enrollment by Target

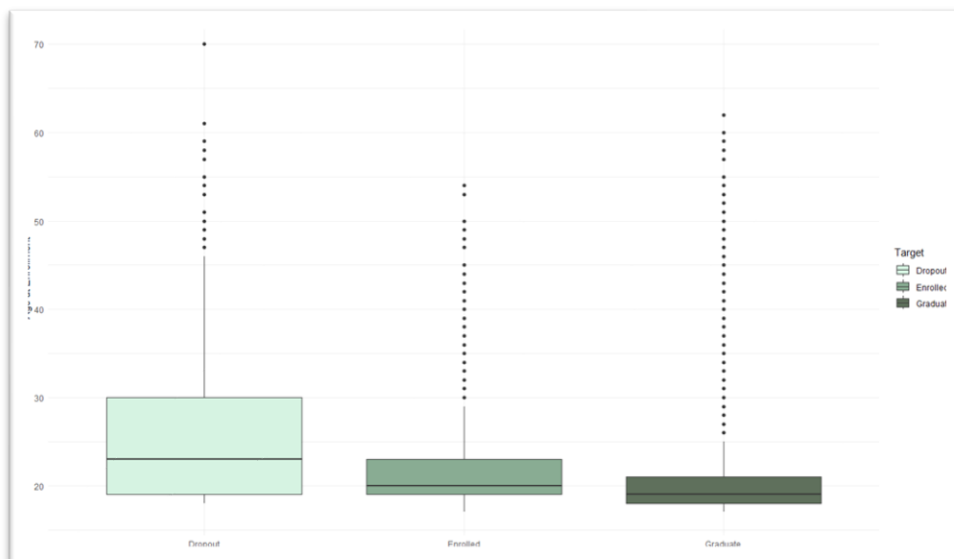


Figure 12: Age at Enrollment by Target

The Age at Enrollment by **Target Box Plot** shows the distribution of students' ages at the time of enrollment, categorized by their target status (Graduate, Dropout, Enrolled). The box plot highlights the variation in age for each group, showing the median, interquartile range, and potential outliers. By comparing the distributions, we can examine how age at enrollment differs across students with various outcomes. The color coding enhances the clarity, making it easier to distinguish between the target categories and analyze the relationship between age and academic status.

Daytime vs Evening Attendance

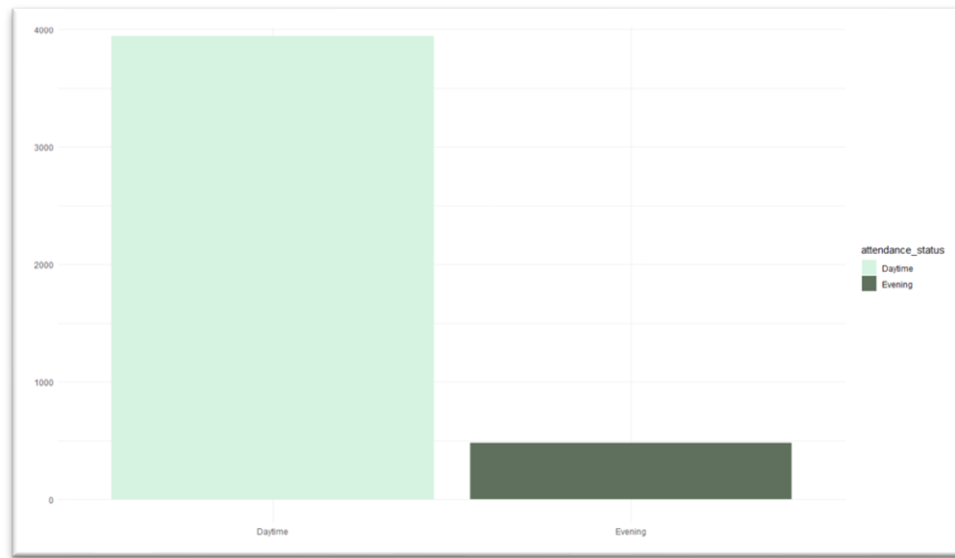


Figure 13: Daytime vs Evening Attendance

The Daytime vs Evening Attendance **Bar Plot** illustrates the distribution of students based on their attendance status, categorized into Daytime and Evening attendance. The bar plot represents the number of students in each attendance category, allowing for a clear comparison between the two groups. The color coding helps distinguish between Daytime and Evening attendance, providing an easy-to-interpret visual of the count of students in each category. This plot offers valuable insights into the distribution of attendance preferences within the student population.

1st Semester Student Performance

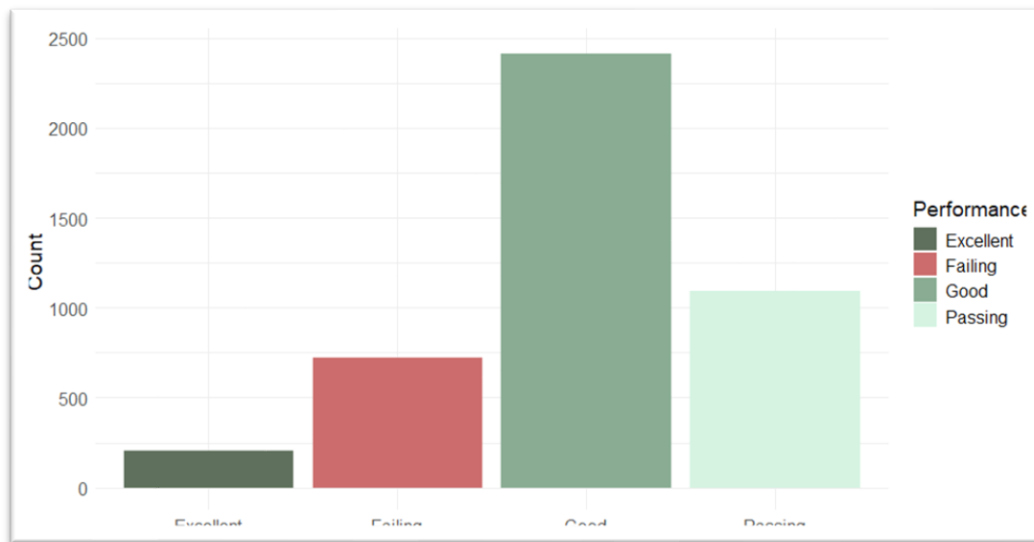


Figure 14: 1st Semester Student Performance

The Performance Distribution (1st Semester) **Bar Plot**, this bar plot shows how students performed in the first semester. The performance is divided into four categories: Failing, Passing, Good, and Excellent. The color-coded bars make it easy to see how many students are in each category, giving a quick overview of their academic success in the first half of the year.

2nd Semester Student Performance

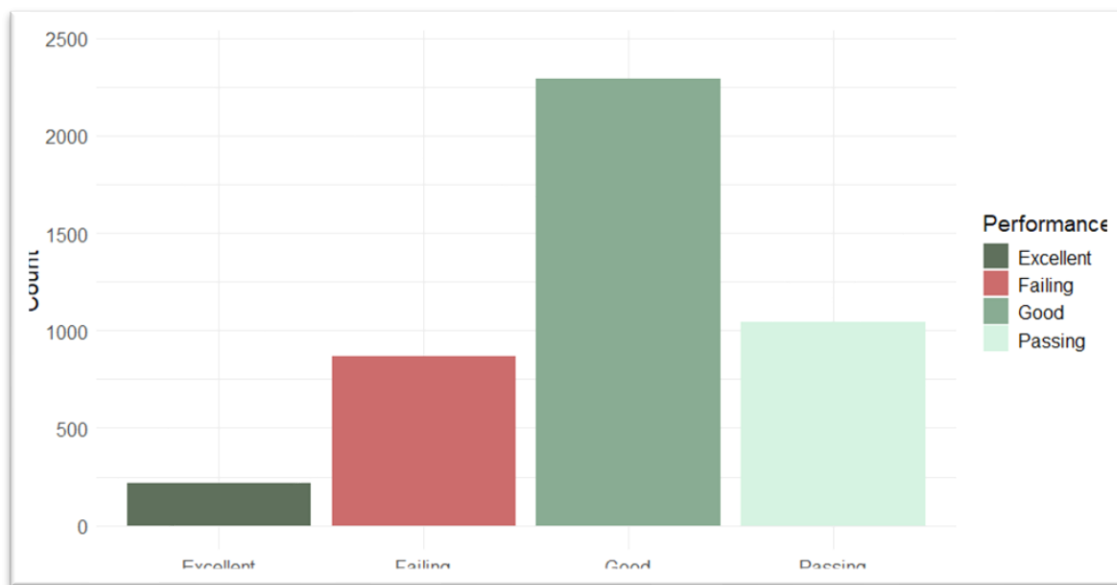


Figure 15: 2nd Semester Student Performance

The Performance Distribution (2nd Semester) **Bar Plot**, this plot shows student performance in the second semester using the same categories. Compared to the first semester, there are noticeable changes. For instance, the number of students in the Failing category may have decreased, while more students moved into the Good and Excellent categories. This comparison helps highlight academic progress or areas that need improvement between the two semesters.

1st Semester Student Performance Distribution

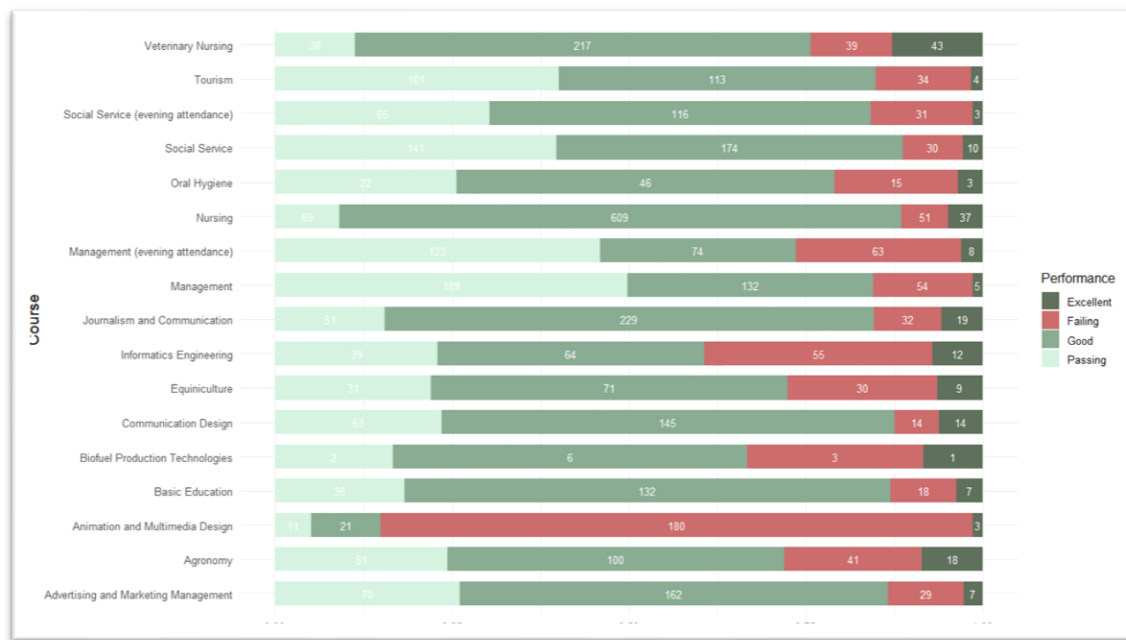


Figure 16: 1st Semester Student Performance Distribution

The Performance Distribution by Course (1st Semester **stacked proportional bar chart** visualizes the distribution of student performance in the first semester across different courses. The stacked bar chart shows the proportion of students in each performance category (Failing, Passing, Good, Excellent) for each course. The color-coding makes it easy to identify the performance trends within each course, while the percentage labels ensure the data is clear and interpretable. The chart provides insights into how students performed in each course during the first semester, helping to identify courses with higher or lower success rates.

2nd Semester Student Performance Distribution

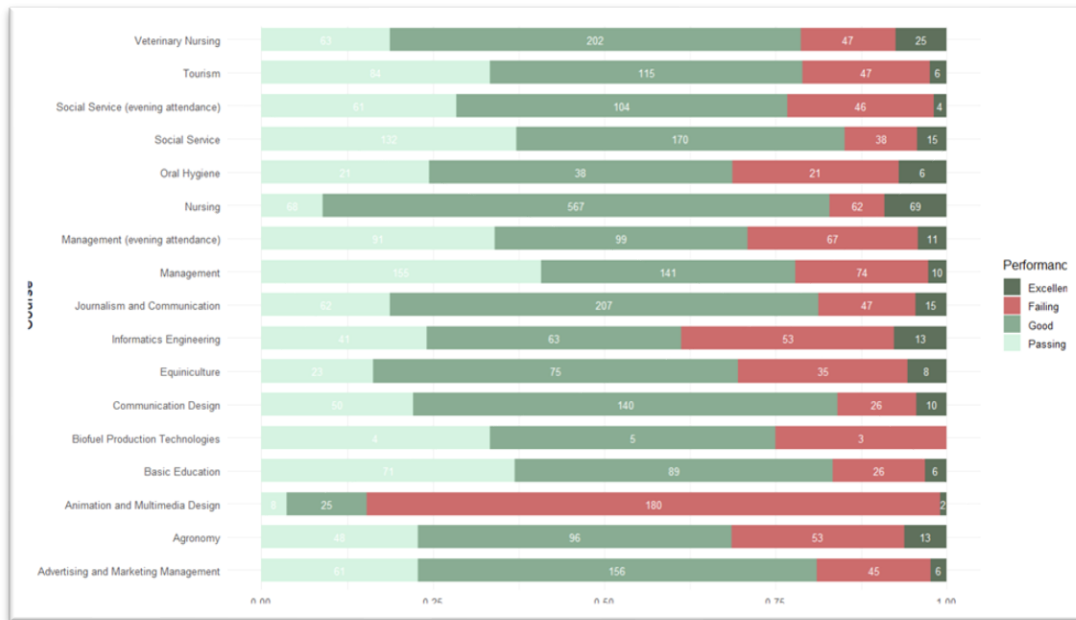


Figure 17: 2nd Semester Student Performance Distribution

The Second Semester Performance Distribution plot is a **stacked proportional bar chart** that displays how students performed in different courses during the second semester. Each bar represents a course and is divided into sections for performance levels: Excellent, Good, Passing, and Failing. The proportions make it easy to compare performance across courses, and the numbers inside the bars show the exact count of students in each category. This chart helps identify courses where students improved or continued to face challenges.

Cumulative frequency 1st Semester Grades

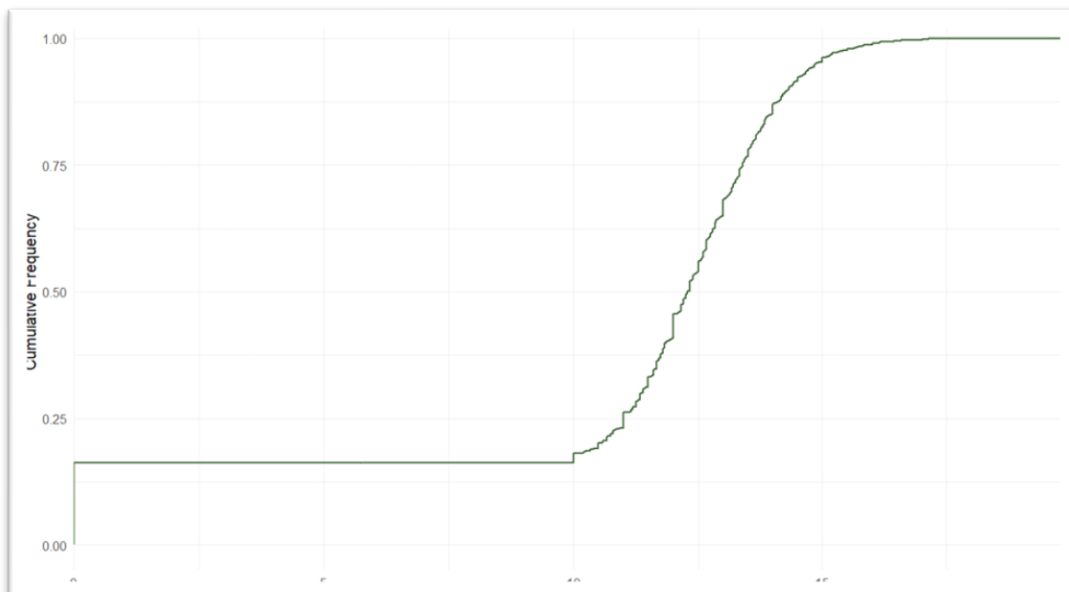


Figure 18: cumulative frequency First Semester Grades

This plot is a cumulative frequency graph (**Empirical Cumulative Distribution Function - ECDF**). It shows the distribution of first-semester grades. The x-axis represents grades (out of 20), and the y-axis indicates the percentage of students who scored at or below each grade. A steep rise between grades 10 and 15 highlights that most students scored within this range.

The sharp increase in this range shows the concentration of grades, with fewer students scoring below 10 or above 15. The plot effectively visualizes where the majority of student grades lie.

Cumulative frequency 2nd Semester Grades

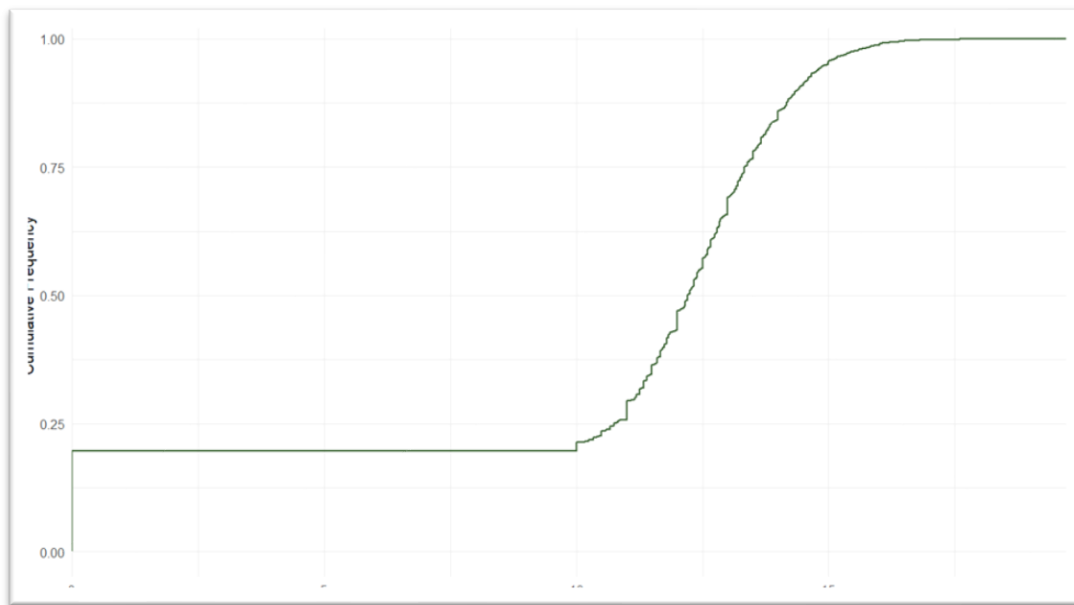


Figure 19: Cumulative Frequency Second Semester Grades

Similar to the first semester, the x-axis shows grades, and the y-axis shows the percentage of students scoring at or below each grade.

The shape of the curve reveals how grades are spread and allows for a comparison with the first semester. Steeper sections indicate more students scoring similar grades, while flatter areas show more varied performance.

QQ Plot for First Semester Grades

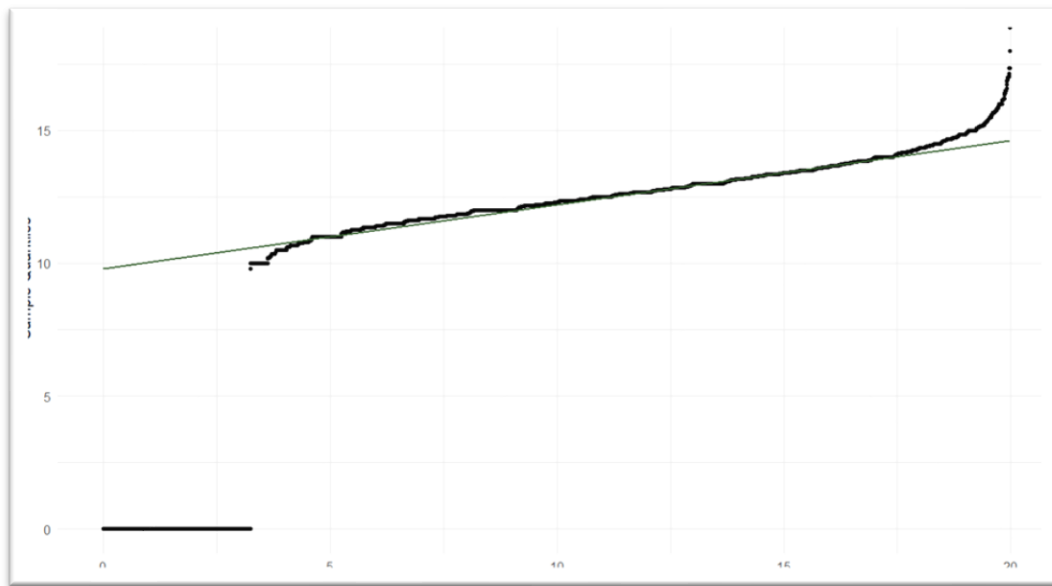


Figure 20: QQ Plot for First Semester Grades

The **Q-Q plot** for the first semester shows how the grades compare to a uniform distribution ranging from 0 to 20. The black points represent the actual grades, while the green line shows the expected pattern if the grades were evenly distributed. If the points follow the line closely, it means the grades are spread out evenly. Deviations from the line, especially at the ends, show differences from the expected distribution. A special thing about this plot is that it uses a uniform distribution instead of the more common normal distribution, making it unique for evaluating how grades are distributed across their entire range.

QQ Plot for Second Semester Grades

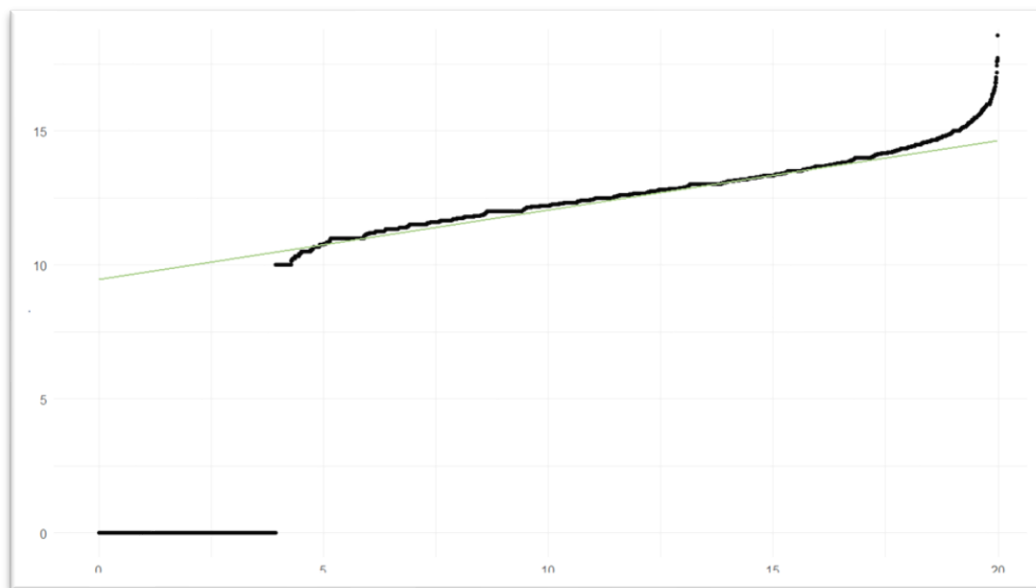


Figure 21: QQ Plot for Second Semester Grades

The **Q-Q plot** for the second semester also compares the grades to a uniform distribution between 0 and 20. Like the first semester, the black points are the grades, and the green line is the expected pattern. However, compared to the first semester plot, this one shows slightly more deviation at the ends, suggesting that the second semester grades are less evenly distributed. A unique feature of this plot is the use of a uniform distribution as the reference, which is less common but helpful for understanding if grades are spread out evenly or if there are patterns like clustering or outliers.

Comparison of grade distribution

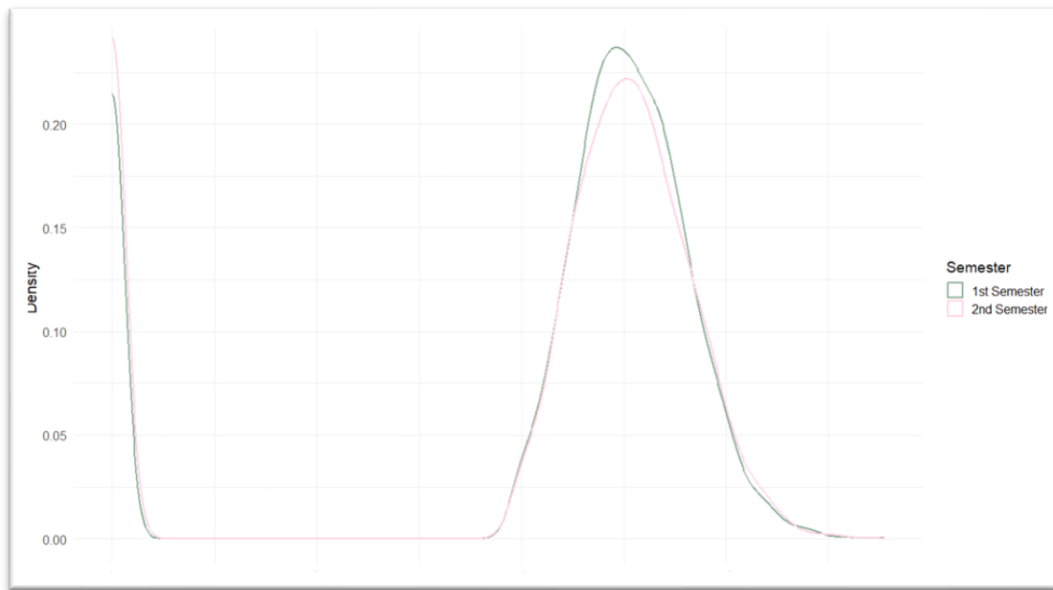


Figure 22: Comparison of grade distribution

The plot is a **density plot** comparing student grade distributions between the 1st and 2nd semesters. It shows how grades are distributed, with the x-axis representing grades and the y-axis showing density. The green and pink curves represent the 1st and 2nd semesters, making it easy to compare grade patterns.

A special aspect of using this plot is that it provides a smooth, continuous view of the grade distributions, unlike a histogram. This makes it easier to visually spot subtle differences between the semesters, such as shifts in performance trends. The use of distinct custom colors further emphasizes these comparisons.

Performance Evaluation by Nationality

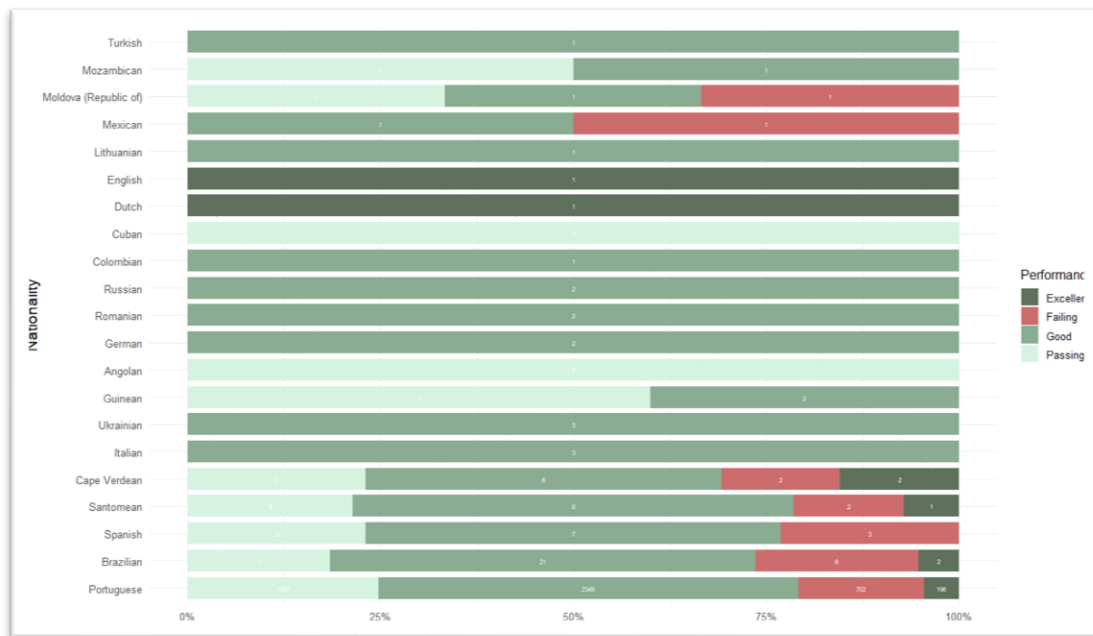


Figure 23: Performance Evaluation by Nationality

This visualization is a **stacked bar chart** that represents the performance evaluation of students by nationality, specifically for the 1st semester. Each bar represents a nationality, and the different segments within the bars correspond to the performance levels ("Excellent," "Good," "Passing," and "Failing"). The chart uses color coding to differentiate these levels, providing a clear visual comparison of student outcomes across nationalities.

A special feature of this plot is that it displays the performance distribution as percentages, allowing a proportional comparison across nationalities, regardless of the total number of students from each group. This makes it easy to spot which nationalities have higher proportions of students in each performance category. The labels within the bars further help in identifying exact counts.

Global Distribution

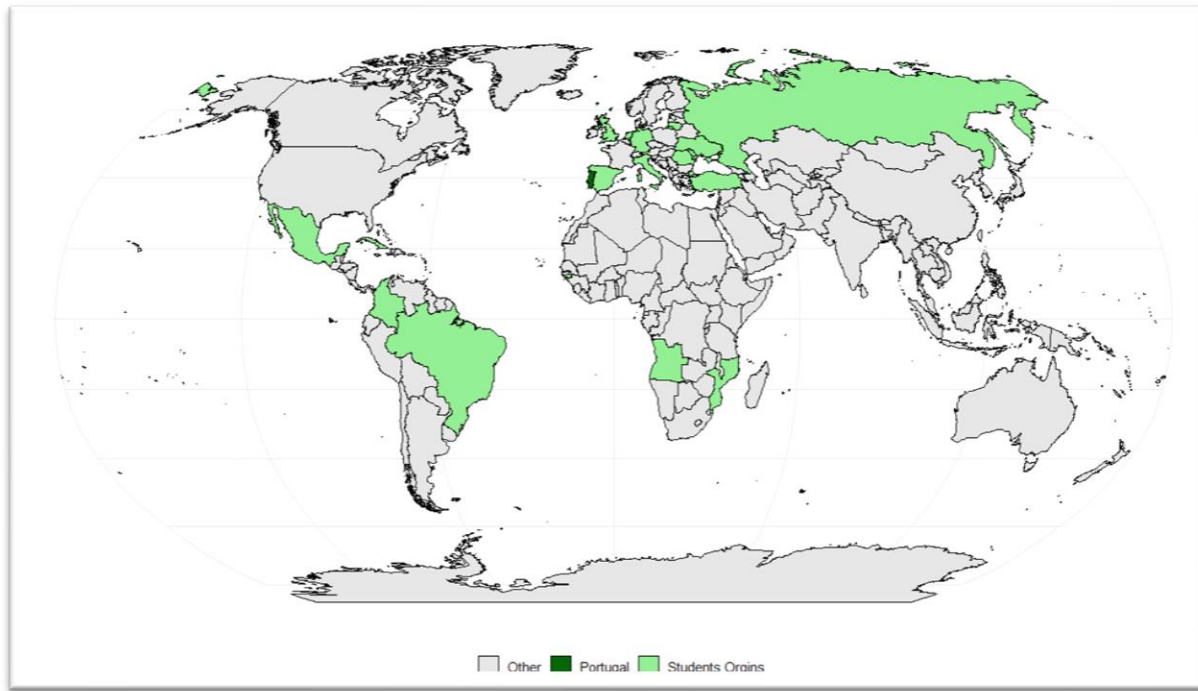


Figure 24: Global Distribution

The plot presents a **global distribution map** of student origins, with a specific emphasis on Portugal. Titled "Global Distribution: Highlighted Countries and Enlarged Portugal," it highlights in light green the countries where students originate, while Portugal is distinctly emphasized by being colored in dark green and enlarged for visual prominence. Countries not part of the dataset are shaded in gray. The map employs the Robinson projection, which balances distortions in size and shape, providing a more visually pleasing representation of global geography. A legend at the bottom explains the color coding used for Portugal, student origin countries, and the rest of the world. This visualization effectively showcases the geographical diversity of students while drawing particular attention to Portugal.

Top 5 Nationalities by Number of Students Excluding Portuguese

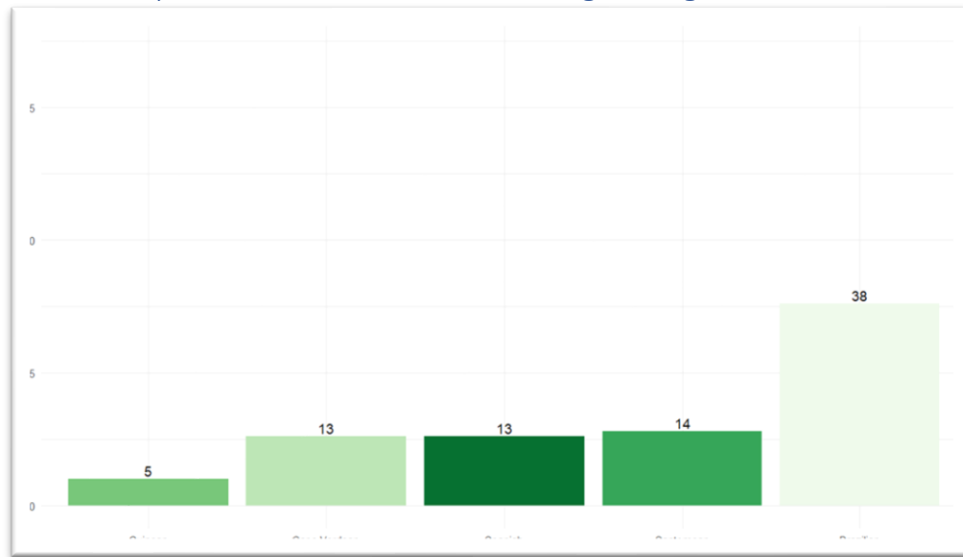


Figure 25: Top 5 Nationalities by Number of Students Excluding Portuguese

The plot shown is a **bar chart** representing the top five nationalities of students by count, excluding Portuguese nationals. Each bar corresponds to a nationality, with its height reflecting the number of students, while the exact counts are displayed above the bars for clarity. The green color palette, with varying shades, enhances visual distinction between nationalities. The exclusion of Portuguese nationals emphasizes the diversity among other groups, and the y-axis is dynamically extended to prevent label overlap. This plot effectively highlights the contribution of non-Portuguese nationalities in the dataset.

Target Distribution by Nationality

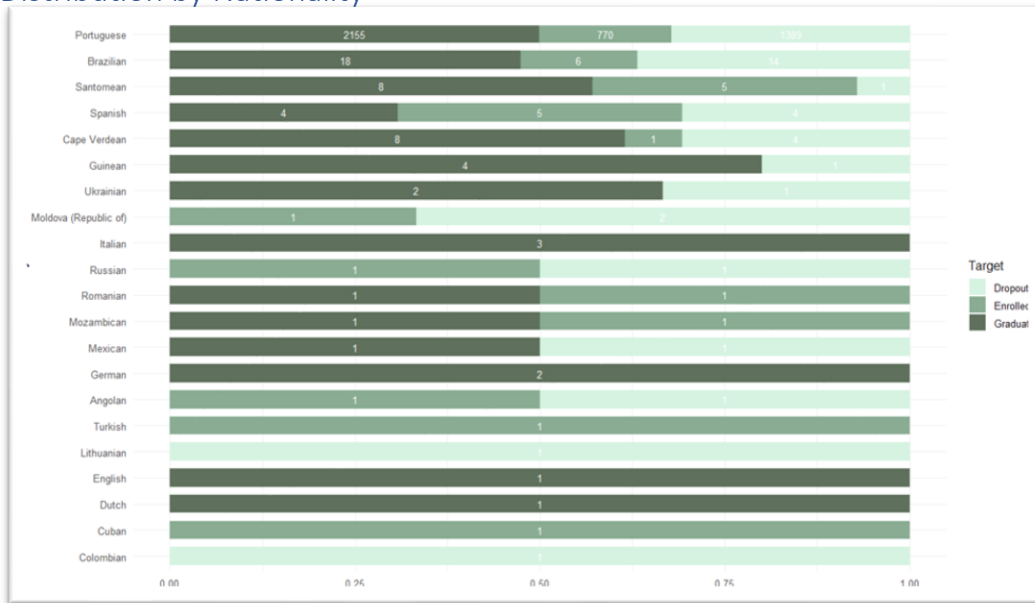


Figure 26: Target Distribution by Nationality

The plot shown is a **stacked bar chart** illustrating the target distribution (Dropout, Enrolled, Graduate) of students by nationality. Each bar represents a nationality, divided into proportions corresponding to the three target categories, with the total bar length normalized to 1 (100%). The varying shades of green differentiate the target categories, allowing easy visual comparison of their relative proportions across nationalities. This plot highlights not just the number of students but their distribution across outcomes, offering valuable insights into educational trends for different nationalities.

Performance Flow

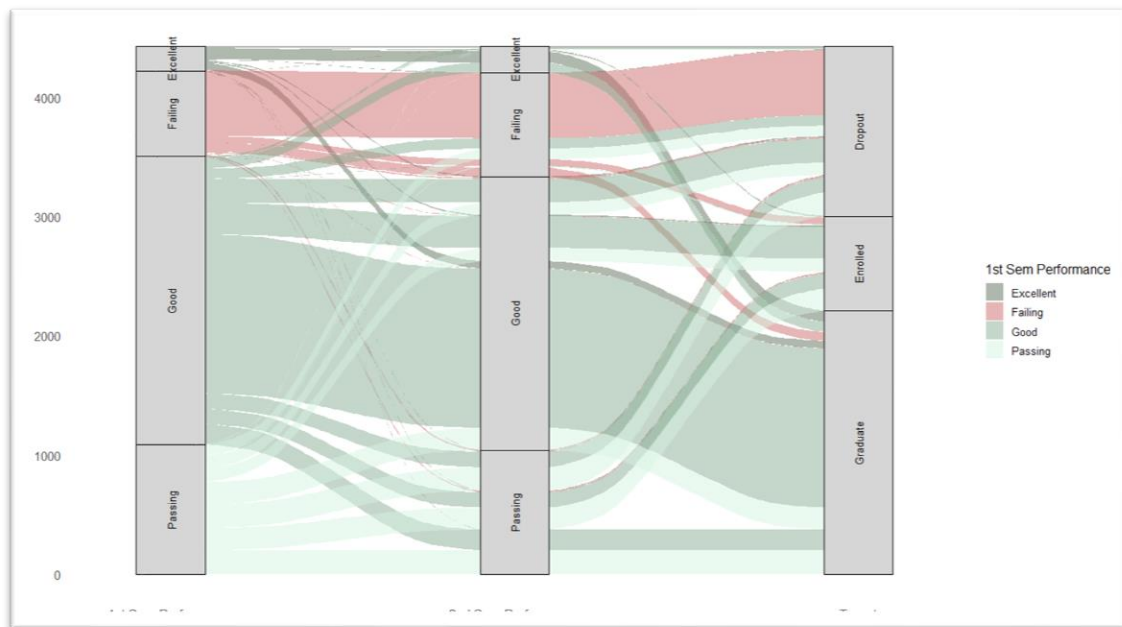


Figure 27: Performance Flow

The **parallel set chart** shows how students move through three stages: 1st-semester performance, 2nd-semester performance, and their final outcome (Dropout, Enrolled, Graduate). The width of each segment represents the number of students in that group, and the flows between stages show changes, like students improving from "Failing" to "Passing" or "Good." Custom colors highlight 1st-semester performance levels, making it easier to spot trends and changes. This chart helps identify where students might need support to improve their outcomes.

Mean Grades by Nationality (1st Semester)

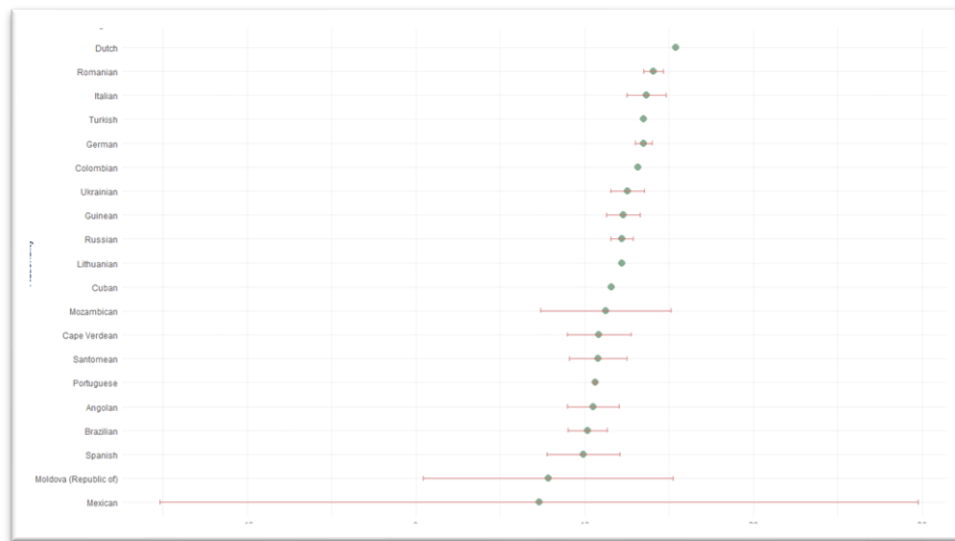


Figure 28: Mean Grades by Nationality (1st Semester)

For the 1st semester, the **dot plot with error bars** illustrates the mean grades of students categorized by nationality, with horizontal error bars showing the variability in the data, such as standard deviations or confidence intervals. The grades for most nationalities are clustered within a similar range, but some groups exhibit higher variability, as indicated by wider error bars. Notably, certain nationalities, such as Mexican and Moldovan students, show extreme values with larger variability, suggesting either outliers or broader disparities in performance within these groups.

Mean Grades by Nationality (2nd Semester)

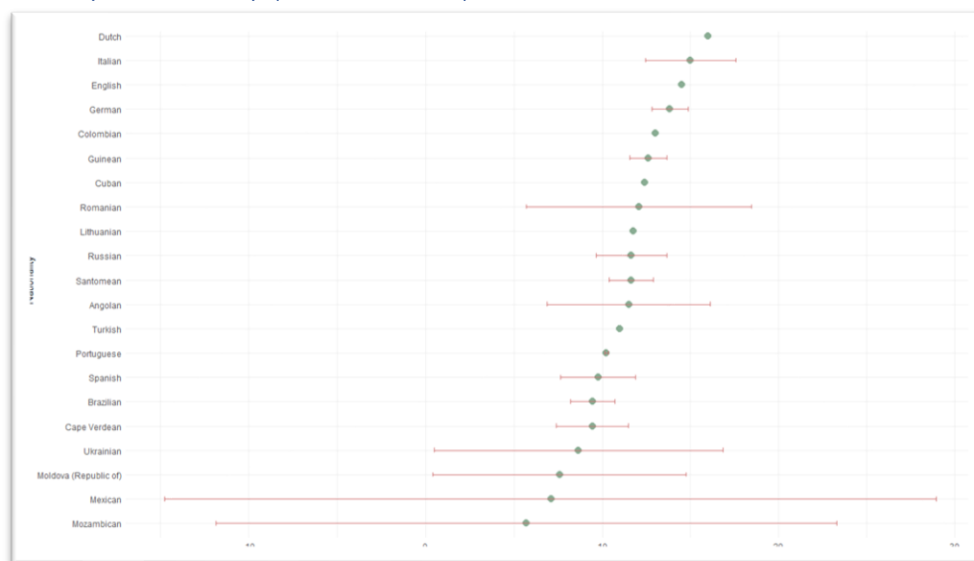


Figure 29: Mean Grades by Nationality (2nd Semester)

For the 2nd semester, the plot similarly displays mean grades by nationality but reveals a shift in performance trends compared to the 1st semester. While many nationalities maintain consistent patterns, others show changes in mean grades or variability. For instance, some nationalities that previously had broader error bars exhibit more compact ranges in the 2nd semester, suggesting reduced variability. A key difference is the ordering of nationalities, with some groups experiencing noticeable improvements or declines in their average grades.

A special feature of this plot is its ability to integrate mean grades with variability in a compact and visually interpretable manner. The horizontal error bars not only indicate the spread of grades but also highlight differences in data reliability and consistency across nationalities. Furthermore, the side-by-side analysis of two semesters allows for a clear temporal comparison, enabling the identification of performance trends, emerging disparities, or the impact of interventions over time.

1st Semester Performance Grid

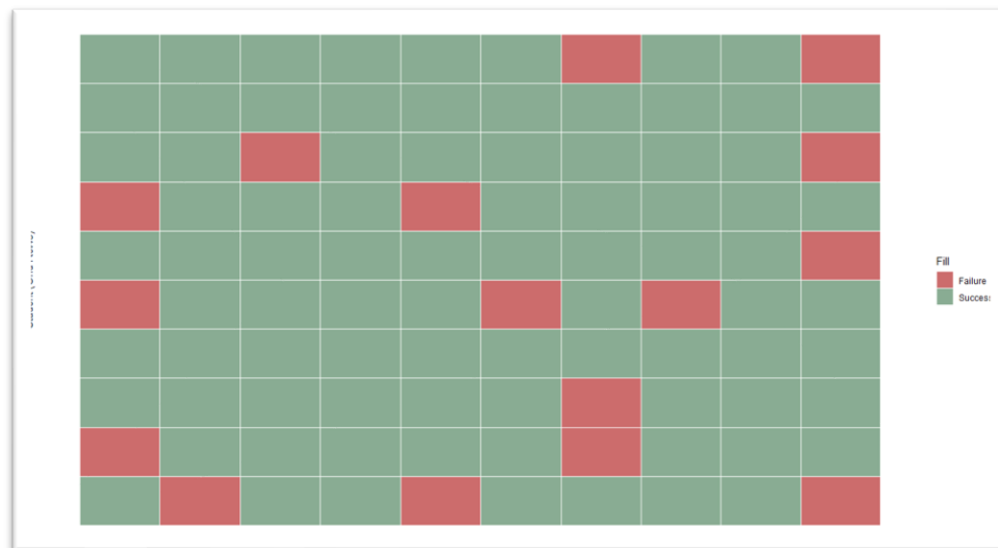


Figure 30: 1st Semester Performance Grid

The first semester **grid plot** visualizes student performance using color-coded cells, where green represents success and red indicates failure. This shows that **16% of students struggled** during the first semester. The scattered distribution of red cells suggests no particular subgroup is disproportionately affected, indicating uniform challenges across the cohort.

This concise visualization highlights the proportion of struggling students, enabling educators to quickly identify the need for support. The intuitive design simplifies complex data, making it easy to assess overall performance at a glance.

2nd Semester Performance Grid

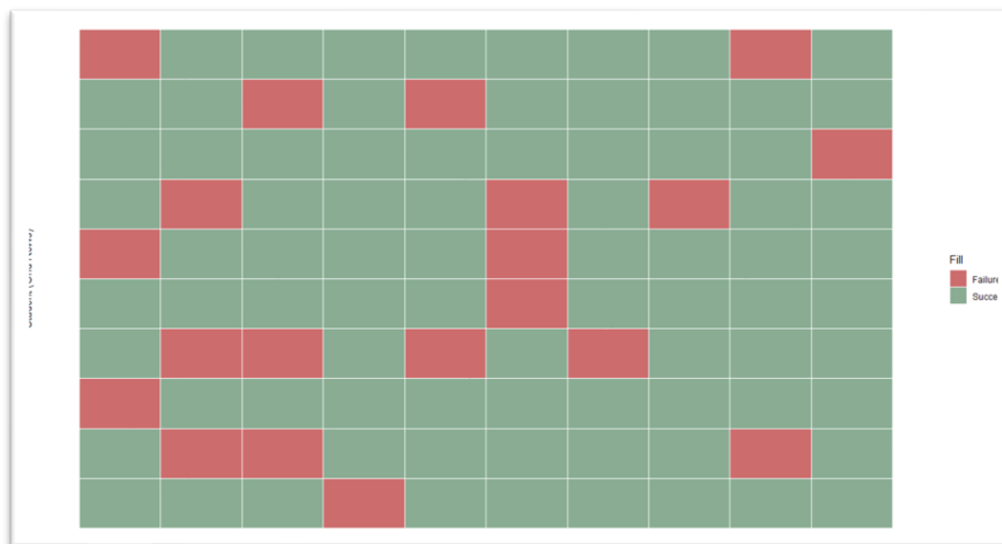


Figure 31: 2nd Semester Performance Grid

The second semester **grid plot** uses the same format but reveals a rise in the failure rate to 20%, suggesting increased challenges or insufficient interventions. The continued scattering of red cells indicates that struggling students remain spread across the cohort, without clustering in specific groups.

The comparison of both semesters highlights a 4% increase in the failure rate, signaling a decline in performance. This trend underscores the need for more proactive and effective measures to address challenges faced by students early on.

Data Columns Overview

[1] "Marital_Status"	"Application_Make"	"Application_Order"
[4] "Course"	"Attendance_Status"	"Previous_qualification"
[7] "Nationality"	"Mother's_qualification"	"Father's_qualification"
[10] "Mother's_occupation"	"Father's_occupation"	"Displaced"
[13] "Educational_special_needs"	"Debtor"	"Tuition_fees_up_to_date"
[16] "Gender"	"Scholarship_holder"	"Age_at_enrollment"
[19] "International"	"Curricular_units_1st_sem_(credited)"	"Curricular_units_1st_sem_(enrolled)"
[22] "Curricular_units_1st_sem_(evaluations)"	"Curricular_units_1st_sem_(approved)"	"Curricular_units_1st_sem_(grade)"
[25] "Curricular_units_1st_sem_(without_evaluations)"	"Curricular_units_2nd_sem_(credited)"	"Curricular_units_2nd_sem_(enrolled)"
[28] "Curricular_units_2nd_sem_(evaluations)"	"Curricular_units_2nd_sem_(approved)"	"Curricular_units_2nd_sem_(grade)"
[31] "Curricular_units_2nd_sem_(without_evaluations)"	"Unemployment_rate"	"Inflation_rate"
[34] "GDP"	"Target"	"First_Sem_Performance"
[37] "Second_Sem_Performance"	"First_Sem_Category"	"Second_Sem_Category"

Figure 32: Data Columns Overview

This panel lists the various data columns used in the Student Dropout Analysis, categorizing a wide range of variables relevant to student performance and dropout risks. These columns include demographic attributes like "Marital_status," "Nationality," and "Gender", academic performance indicators such as "Curricular_units" and "First_Sem_Performance", and socio-economic factors like "GDP" and "Employment_rate." The inclusion of variables such as "Attendance_status" and "Scholarship_holder" highlights their potential influence on student outcomes.

Challenges and Future Work

Throughout this project, our team encountered several challenges that required collaboration and coordination to overcome. One of the main challenges was the design of the user interface using Shiny. We faced difficulties in organizing and distributing the plots across different sections of the app, which required extra effort to ensure smooth navigation and that the app could handle large datasets without performance issues. Additionally, selecting the best visualization methods for the data was not always straightforward. We had to experiment with several types of charts to ensure the data was clear, easy to understand, and visually appealing.

Looking Future Enhancements, we have several ideas for improving and expanding the project. We plan to incorporate more advanced, interactive visualizations, such as scatter plots and correlation matrices, to uncover deeper insights and patterns within the data. We also intend to integrate machine learning models that can predict student performance based on their current grades, which would help improve student outcomes by identifying those who need additional support. Furthermore, we want to enhance the app to support real-time data updates, allowing users to instantly see new trends and changes as data is added. Lastly, we aim to add customizable filters to allow users to focus on specific courses, semesters, or student demographics, making the app more interactive and better suited to various user needs.

Conclusion

This project has provided valuable insights into student performance analysis using data visualization and Shiny. By categorizing students' performance across different semesters and visualizing it through stacked bar charts, we were able to identify clear trends and patterns. The most notable insight was the noticeable differences in student success rates based on their grades, with clear distinctions between "Failing," "Passing," "Good," and "Excellent" performance categories.

The use of Shiny allowed us to create an interactive and user-friendly interface, making the visualizations easily accessible to anyone interested in exploring the data. The GUI was designed to be intuitive, offering smooth navigation between different semesters and performance categories. This feature enhanced the overall usability, allowing users to engage with the data in a dynamic way.

Overall, the project successfully demonstrated the power of data science and visualization in understanding educational trends. The results not only highlight the importance of data-driven decision-making but also emphasize the need for tools like Shiny that facilitate easy access to complex data. This project serves as a foundation for future work, including the integration of predictive models and further improvements to the user interface.

Table of Figures

Figure 1: Example of dashboard's interface	6
Figure 2: Distribution by Tuition Fees Status	7
Figure 3: Distribution by Scholarship Status	7
Figure 4: Distribution by Attendance	8
Figure 5: Distribution by Debtor Status	9
Figure 6: Distribution by International Status	9
Figure 7: Distribution by Educational Special Needs	10
Figure 8: Gender Distribution by Target Categories	11
Figure 9: Target Counts by Gender	11
Figure 10: Scholarship Holders by Target	12
Figure 11: Grade Distribution by Target	13
Figure 12: Age at Enrollment by Target	13
Figure 13: Daytime vs Evening Attendance	14
Figure 14: 1st Semester Student Performance	15
Figure 15: 2nd Semester Student Performance	15
Figure 16: 1st Semester Student Performance Distribution	16
Figure 17: 2nd Semester Student Performance Distribution	17
Figure 18: cumulative frequency First Semester Grades	18
Figure 19: Cumulative Frequency Second Semester Grades	18
Figure 20: QQ Plot for First Semester Grades	19
Figure 21: QQ Plot for Second Semester Grades	19
Figure 22: Comparison of grade distribution	20
Figure 23: Performance Evaluation by Nationality	21
Figure 24: Global Distribution	22
Figure 25: Top 5 Nationalities by Number of Students Excluding Portuguese	23
Figure 26: Target Distribution by Nationality	23
Figure 27: Performance Flow	24
Figure 28: Mean Grades by Nationality (1st Semester)	25
Figure 29: Mean Grades by Nationality (2nd Semester)	25
Figure 30: 1st Semester Performance Grid	26
Figure 31: 2nd Semester Performance Grid	27
Figure 32: Data Columns Overview	27