

Data Mining Report

Abdelrahman Ayman 2205033

Arsany Osama 2205122

Malak Mohamed 2205059

Fayda Hesham 2205177

Title: Comprehensive Analysis of High School Student Performance

Table of Contents:

1. Introduction
2. Dataset Overview 2.1 Information about Dataset 2.2 Rationale for Selection
3. Analytical Questions 3.1 Gender Distribution and Success Rates 3.2 Absenteeism Across Subjects and Specialties 3.3 Success Rates Among Different Specializations 3.4 Geographic Variations in Success Rates 3.5 Subject-Specific Success Rates and Top-Performing Students 3.6 Average Grades in Alexandria and Examination Passing Thresholds 3.7 Potential Cheating Rates in Exams Across Different Cities
4. Gender Disparities in Academic Performance 4.1 Gender Distribution and Success Rates 4.2 Examination of Subject-Specific Performance
5. Factors Influencing Student Success 5.1 Application of Data Mining Techniques 5.1.1 Decision Tree Analysis 5.1.2 K-Means Clustering 5.2 Additional Factors Influencing Student Success
6. Recommendations
7. Conclusion
8. Future Directions

1. Introduction

This report aims to provide a comprehensive analysis of high school student performance in Egypt, utilizing a dataset obtained from Kaggle.com. By exploring various aspects of student performance, identifying influential factors, and proposing recommendations for improvement, this report seeks to contribute to the enhancement of educational practices and policies.

2. Dataset Overview

2.1 Information about Dataset

The dataset comprises 153,833 rows and 25 columns, containing detailed information on student demographics, subject scores, and additional variables. It was sourced from Kaggle.com and approved by the Egyptian Ministry of Education for the year 2022.

2.2 Rationale for Selection

The dataset was chosen due to its relevance to our academic journey and familiarity with the data. Its manageable size facilitates efficient analysis, enabling us to explore various dimensions of student performance effectively.

3. Analytical Questions

In this section, we address a series of analytical questions to gain insights into high school student performance:

3.1 Gender Distribution and Success Rates

The analysis reveals the gender distribution among exam takers, with girls constituting the majority. Despite numerical differences, both genders exhibit comparable success rates, highlighting equitable assessment practices.

3.2 Absenteeism Across Subjects and Specialties

An examination of absenteeism rates across subjects and specialties provides insights into patterns of student engagement and attendance.

3.3 Success Rates Among Different Specializations

The analysis of success rates among different specializations sheds light on subject-specific challenges and areas for improvement.

3.4 Geographic Variations in Success Rates

Geographic variations in success rates offer insights into regional disparities in academic performance and potential factors contributing to these differences.

3.5 Subject-Specific Success Rates and Top-Performing Students

A detailed examination of subject-specific success rates and top-performing students provides insights into areas of academic strength and excellence.

3.6 Average Grades in Alexandria and Examination Passing Thresholds

Analysis of average grades in Alexandria and examination passing thresholds offers insights into performance benchmarks and academic standards.

3.7 Potential Cheating Rates in Exams Across Different Cities

Exploring potential cheating rates in exams across different cities provides insights into the integrity of assessment processes and the prevalence of academic misconduct.

4. Gender Disparities in Academic Performance

4.1 Gender Distribution and Success Rates

Further analysis reveals intriguing insights into gender disparities in academic performance, including the examination of success rates across different subjects and specializations.

4.2 Examination of Subject-Specific Performance

A detailed examination of subject-specific performance reveals disparities in success rates and areas for targeted intervention and support.

5. Factors Influencing Student Success

5.1 Application of Data Mining Techniques

5.1.1 Decision Tree Analysis

Decision tree analysis offers valuable insights into the factors influencing student success, including the identification of influential variables and decision-making processes.

5.1.2 K-Means Clustering

K-Means clustering provides insights into patterns and clusters within the data, enabling the identification of groups of students with similar characteristics and performance levels.

5.2 Additional Factors Influencing Student Success

Further analysis explores additional factors influencing student success, including socioeconomic status, parental involvement, and access to educational resources.

5.3 Fuzzy

Here we use fuzzy logic to represent children in categories of good, very good, and excellent, pass, bad, and very bad represent this analysis in the many plots.

5.3 Hierarchical

here we make a correlation matrix, which shows the previous graph, but with numbers that we normalized using StandardAero, and information can be extracted from it more easily. The strongest relationships:

Pure and applied mathematics, Chemistry with physics, Biology with geology, The four literary subjects, The three subjects outside the final total

6. Recommendations

Based on the analysis, several recommendations are proposed to enhance educational practices and improve student outcomes:

1. Implement targeted interventions to support students in challenging subjects, particularly those with lower success rates.
2. Strengthen examination oversight mechanisms to prevent cheating and maintain the integrity of assessment processes.
3. Promote gender equity through the implementation of inclusive educational policies and programs.
4. Foster collaboration between educational stakeholders to address regional disparities and promote equitable access to educational resources and opportunities.

7. Conclusion

In conclusion, this report provides valuable insights into high school student performance in Egypt, addressing gender disparities, subject-specific challenges, and factors influencing academic outcomes. By leveraging these insights and recommendations, educational stakeholders can work collaboratively to enhance educational practices and improve student outcomes.

8. Future Directions

Future research endeavors should focus on longitudinal studies to track student progress over time and assess the long-term impact of interventions on academic outcomes. Additionally, exploring the efficacy of targeted support programs and

interventions in addressing gender disparities and subject-specific challenges warrants further investigation.

This extended report offers a comprehensive analysis of high school student performance, encompassing detailed insights into gender disparities, subject-specific performance, and factors influencing academic outcomes. Through rigorous analysis and evidence-based recommendations, it aims to inform educational stakeholders and drive positive change in educational practices and policies.

Report on Grade Adjustment Analysis

Introduction: This report presents the results and analysis of the grade adjustment process conducted using the K-medoids clustering algorithm. The aim was to identify failing grades in various subjects and raise them to passing grades based on certain thresholds, thereby improving the overall performance of the students.

Methodology:

- The **K-medoids algorithm** was applied to the sample data to reduce computational complexity.
- The Silhouette method was utilized to evaluate students' performance in clusters for each subject, testing the range of k from 2 to 10 to determine the **optimal** number of clusters.
- Students who did not have the subject prescribed were removed from the analysis.
- Failing grades were identified within specific score ranges for each subject.
- Students scoring above the failing threshold were raised to passing grades.

Results: A summary of the adjustments made for each subject is as follows:

- **Arabic:** Grades in the range of 20 to 31 were failing. Students scoring 32 or above were raised to a passing grade of 40 (+8).

- **First Foreign Language:** Grades in the range of 8 to 15 were failing. Students scoring 16 or above were raised to a passing grade of 25 (+9).
- **Second Foreign Language:** Grades in the range of 6 to 13 were failing. Students scoring 14 or above were raised to a passing grade of 20 (+6).
- **History:** Grades in the range of 9 to 14 were failing. Students scoring 15 or above were raised to a passing grade of 30 (+15)
- **Geography:** Grades in the range of 9 to 23 were failing. Students scoring 24 or above were raised to a passing grade of 30 (+6)
- **Philosophy:** Grades in the range of 4 to 16 were failing. Students scoring 17 or above were raised to a passing grade of 30 (+13).
- **Psychology:** Grades in the range of 10 to 19 were failing. Students scoring 20 or above were raised to a passing grade of 30 (+10).
- **Chemistry:** Grades in the range of 9 to 22 were failing. Students scoring 23 or above were raised to a passing grade of 30 (+8).
- **Physics:** Grades in the range of 4 to 18 were failing. Students scoring 19 or above were raised to a passing grade of 30 (+11).
- **Biology:** Grades in the range of 13 to 19 were failing. Students scoring 20 or above were raised to a passing grade of 30 (+10).
- **Geology:** Grades in the range of 12 to 18 were failing. Students scoring 19 or above were raised to a passing grade of 30 (+11).
- **Pure Mathematics:** Grades in the range of 11 to 20 were failing. Students scoring 21 or above were raised to a passing grade of 30 (+9).
- **Applied Mathematics:** Grades in the range of 9 to 15 were failing. Students scoring 16 or above were raised to a passing grade of 30 (+14).

Conclusion: The grade adjustment process has resulted in significant improvements in the academic performance of students across various subjects. By identifying failing grades and raising them to passing grades based on predetermined thresholds, the analysis has highlighted the effectiveness of the K-medoids clustering algorithm in dealing with outliers and improving overall student outcomes. The adjustments made demonstrate a substantial increase in the number of passing grades, indicating the successful application of data-driven strategies to enhance educational outcomes