Babeș-Bolyai University

Faculty of Mathematics and Computer Science

Statistical Analysis of Baccalaureate Exam Results in Romania

Created by:

Póra Boglárka

Rafain Emőke

Simon Andrea

2024

Contents

Introduction 3

History of the subject 3

Other studies 4

Project objectives 4

Data collection and modeling 5

Descriptive Statistics 6

Inferences 34

Correlation Analysis 34

T-Tests 40

ANOVA 45

Regression Analysis 50

Chi-Square Test 54

Inferences 62

SWOT 64

References 65

Introduction

One of the most important milestones that we reach throughout our lives is the final exam of our high school studies, the baccalaureate exam, as we call it in Romania. It was not a long time ago that we took this exam as well, and it has always excited us to see the results every year, compare them to previous years, check the subjects that got chosen, and see how our schools performed. That is why we decided to start to explore this topic, see what type and amount of data we can find, and work with it. This is how the creation of our Statistical Analysis of Baccalaureate Exam Results in Romania has started.

After cleaning and preparing the collected data, we aimed to get an insight into the basic descriptive statistics of the data, means, variances, and summaries. We wanted to know what type of distributions we are working with, and what we can expect from the respective dataset. However, we considered it crucial to go beyond that and using numerous statistical models, get to know more about the relationships between the dimensions, causes and effects, correlations, the impact of the different variables on the final grade, and what makes a great final exam result.

History of the subject

To provide a short history of the subject, we would like to mention a few words about the origins of the baccalaureate exam. The Romanian Baccalaureate Exam, initially called "Examenul de Maturitate," was introduced in 1925 by Dr. Constantin Anghelescu. It replaced the existing examination system for high school graduates, causing anxiety among students and teachers. The pass rate increased over the years, reaching 70% in 1933.

During the communist era, emphasis shifted to quantity over quality, leading to a decline in educational standards. Candidates were reportedly assisted in passing, and diplomas were sometimes given as gifts.

In contemporary times, measures such as high-tech surveillance cameras and restrictions on mobile phones aim to prevent cheating and maintain the exam's credibility. The pass rate decreased from 80% (2004-2009) to around 55% (2010 onwards) with these measures[1].

Other studies

Before starting to work on our project, we browsed for similar research that deal with the same topic. We found a webpage called BAC+[2], that showcases beautiful visualizations and statistics about the high school final exam results on a national, county, and school level as well. Numerous press articles deal with the topic but none in such a detailed way that we decided to. What distinguishes our idea from the rest is the fact that we aim to examine the data with inferential statistical methods as well, try multiple models, and test hypotheses for significance in the relationships between variables.

Project objectives

We have formulated a complex research question that we aim to answer in the course of our study and draw conclusions from. "To what extent do various factors, including gender, educational form, language, Romanian, native, mandatory, and optional grades, and candidate environment influence the academic performance and promotion outcomes of high school students? Additionally, does the appeal process impact the final scores, and how do these variables collectively contribute to the overall status of candidates, taking into account potential variations in the effects across different candidate groups?"

We plan to study the different variables by analyzing the correlation between the dimensions, running T-tests, ANOVAs, and Chi-Square tests, and also conducting a regression analysis. We build hypotheses around each, and based on the received p-values, we decide whether to keep or reject the null hypotheses in favor of the alternative hypotheses. After that, we examine the results, interpret them, and draw conclusions.

Mainly, we are waiting for conclusions that show a high level of dependence between variables. For instance, we anticipate a positive correlation between the independent grades, if someone performs well in one subject, they probably will get a good grade in the others as well. We also expect that this is not going to be true for each pair because there are students who are only good at stem subjects and not at literature and vice-versa. We also reckon that in many cases appeal grades do not change averages. In most cases, the students judge their performance poorly and there is nothing wrong with the correction of the paper.

Data collection and modeling

The data we used for this study was collected from data.gov.ro[3]. We downloaded the datasets from there, and these datasets are uploaded there by edu.ro[4]. We have four separate datasets that were transformed into pandas data frames, each representing the results in a year: from 2020 to 2023. We were working in Python using Google Collaboratory to be able to keep track of progress simultaneously.

The datasets contain both categorical (for example, candidate code, gender, educational form, final status, etc.) and numerical data (for instance, Romanian grade, native grade, average, etc.).

df2020.shape	df2021.shape	df2022.shape	df2023.shape
(155650, 52)	(133664, 52)	(126453, 74)	(130522, 52)

Figure 1: Dataset dimensions

As we can see in the 1st Figure, originally all four datasets have the same number of columns, except for 2022. The number of rows depends on how many candidates participated in the exam that particular year. However, throughout the data cleaning process, we changed some of the fields, transformed, added new ones, and dropped many of them that were not relevant to our study. In the 2nd Figure, you can see the final variables, all of their names were changed to proper English expressions.

```
df2020_clean.columns
```

```
Index(['unique_candidate_code', 'gender', 'specialization', 'profile',
       'educational_pathway', 'education_form', 'candidate_environment',
       'romanian_sa', 'native_sb', 'hungarian_sb', 'mandatory_sc', 'maths_sc',
       'chosen_sd', 'language', 'promotion', 'ack_grade_a', 'ack_grade_b',
       'ack_grade_c', 'ack_grade_d', 'ack_grade_ea', 'ack_grade_eb',
       'ack_grade_ec', 'ack_grade_ed', 'status_a', 'status_b', 'status_c',
       'status_d', 'status_ea', 'status_eb', 'status_ec', 'status_ed',
       'listening_level', 'reading_level', 'writing_level', 'speaking_i_level',
       'speaking_ii_level', 'romanian_grade', 'native_grade',
       'mandatory_grade', 'chosen_grade', 'appeal_romanian',
       'appeal_romanian_grade', 'appeal_native', 'appeal_native_grade',
       'appeal_mandatory', 'appeal_mandatory_grade', 'appeal_chosen',
       'appeal_chosen_grade', 'status', 'average'],
      dtype='object')
```

Figure 2: Final columns

Descriptive Statistics

After successfully finishing the data cleaning process, we looked at some key measures and began to conduct an Exploratory Data Analysis (EDA) on each of the four datasets to uncover key insights, patterns and trends.

For each of the four datasets we examined measures of central tendency and variability, such as the mean, median, mode, range, and standard deviation of each numerical feature. These measures help with detecting the typical scores candidates would achieve in each subject, to uncover any distribution patterns in the spread of the data.

Measures of Central Tendency and Variability of 2020 Baccalaureate Scores

	Romanian Grade	Native Grade	Mandatory Grade	Chosen Grade	Romanian Grade (a)	Native Grade (a)	Mandatory Grade (a)	Chosen Grade (a)	Average
count	155650.0	7522.0	155650.0	155650.0	23989.0	495.0	20781.0	17635.0	103476.0
mean	6.652607	7.411871	5.935003	6.796577	6.763429	6.700808	5.363560	5.989402	7.905181
std	2.293798	1.797514	3.036192	2.900515	1.908728	1.656187	2.361584	2.187129	1.262246
min	-2.0	-2.0	-2.0	-2.0	1.0	2.5	1.0	1.0	5.0
25%	5.3	6.35	4.2	5.25	5.15	5.45	3.6	4.2	6.85
50%	6.8	7.6	6.35	7.6	6.9	6.65	5.0	5.70	8.08
75%	8.5	8.7	8.45	9.1	8.5	8.1	7.45	7.9	9.0
man	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0

Figure 3: Basic Statistics Table of 2020 Baccalaureate Scores

The above figure shows the key measures for all numerical features of the 2020 Exam Score Dataset. Looking at the Romanian Written Exam Grades, the average grade achieved by the 155650 candidates who took the exam was 6.65.

```
[5]  df2020["romanian_grade"].median()

     6.8

[4]  df2020["romanian_grade"].mode()

     0    5.0
     Name: romanian_grade, dtype: float64
```

Figure 4: Median and Mode of Romanian Grades in 2020

As for the median and mode, in the case of the Romanian Written Exam Grades from the 2020 session, the median was 6.8, meaning that there is a rough symmetrical distribution. The mode was 5.0, meaning that the passing grade occurs most frequently amongst these grades. The 2.29 standard deviation indicates that the data points on average deviate by 2.29 from the mean, meaning a wider range of data. When it comes to the quartiles, the first 25% of the data is below 5.3. The median is 6.8, indicating that half of the candidate scores fall under 6.8, half are higher scores, and the last quarter shows that 75% of the data is below 8.5. The data appears to be moderately spread, ranging between -2 and 10. Looking at the strange minimum value of the Romanian Grades, the value -2 appears to be an outlier that will not be considered further on into the Exploratory Data Analysis. One possible explanation for the appearance of this negative value (the Romanian grading system is between 1 and 10) can be that some schools or even some counties mark the absence of students, or any kind of attempt at cheating by using this negative number as a sign of immediate disqualification. Further on into our analysis, we decided to only look at data concerning students who passed, the passing grade subject-by-subject being a 5 and by average score being a 6.

```
[10] df2020["average"].median()

     8.08

[11] df2020["average"].mode()

     0    9.13
     Name: average, dtype: float64
```

Figure 5: Median and Mode of Average Grades in 2020

Taking a closer look at the average, the mean of 7.9 and median of 8.08 yet again suggests a quite symmetrical distribution, the mode being a 5 meaning the passing grade is still the most common grade for the average as well. Knowing this, there is a strong relationship between the Romanian grade and the Final Score.

Frequency Distributions of 2020 Baccalaureate Scores & Candidate Information

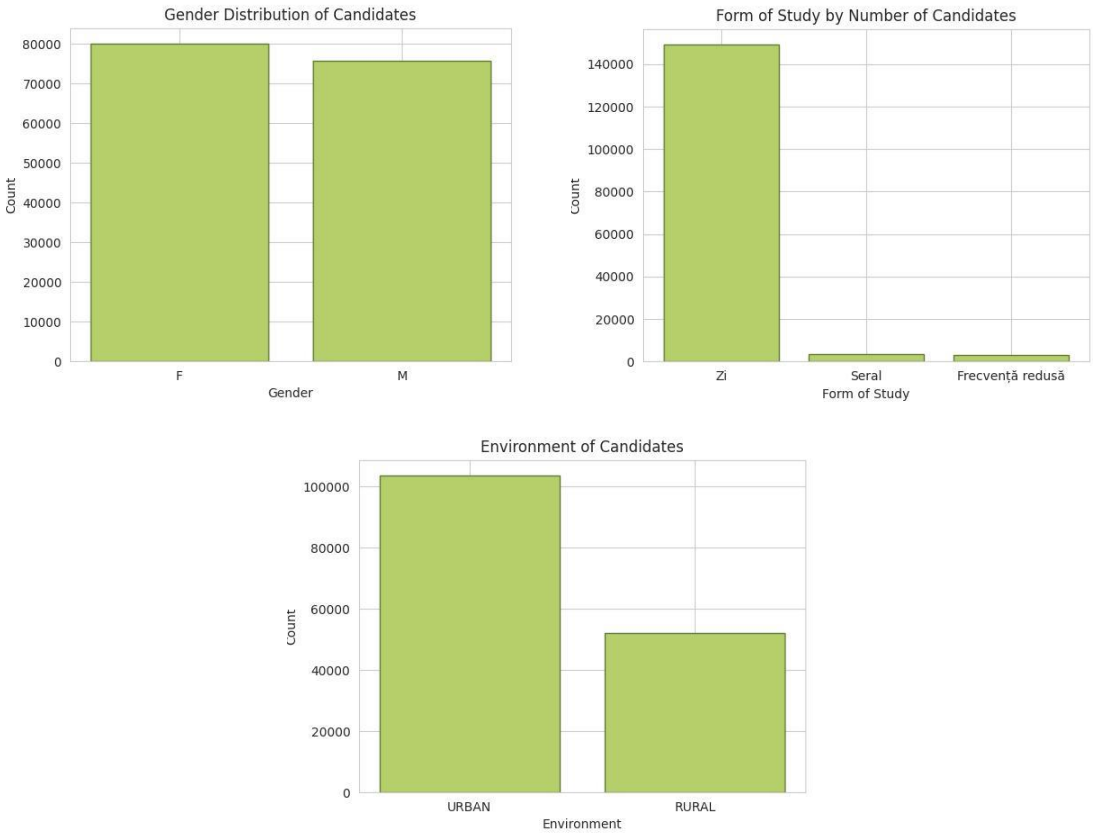


Figure 6: Distribution of Gender, Form of Study and Environment of Candidates

The above figure shows that the 2020 candidates are almost equally distributed between the two genders, most being from an urban environment and attending school by day.

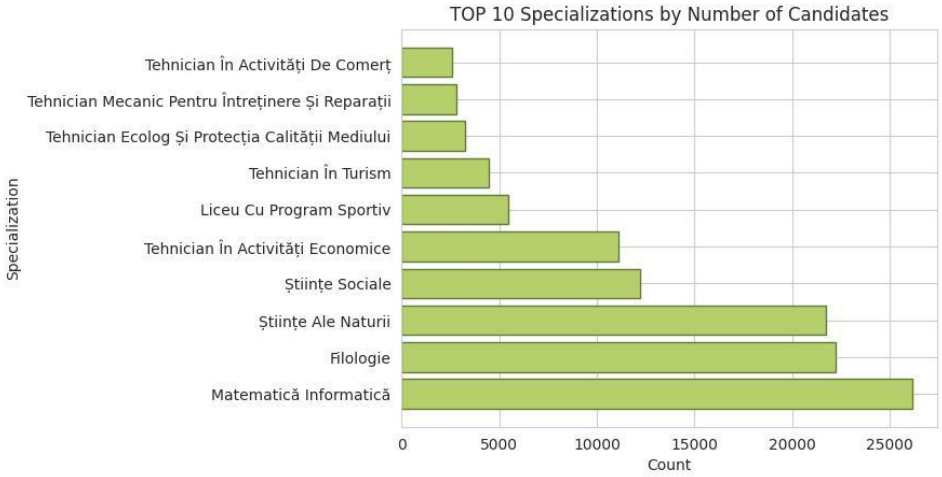


Figure 7: Distribution of Most Frequented Specializations by Number of Candidates

The above diagram shows that the most frequented specializations are Mathematics and Informatics, Philology and Natural Sciences.

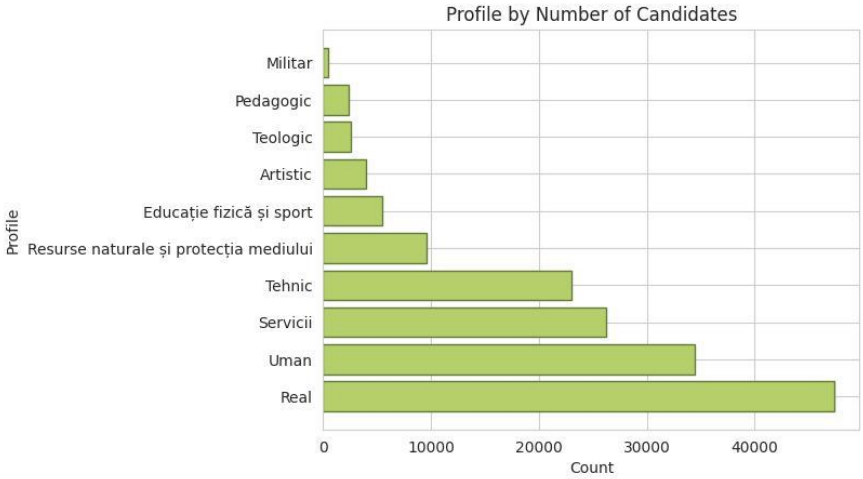


Figure 8: Distribution of Profiles

As far as the profiles go, most specializations fall under the STEM category, and most majors are theory focused.

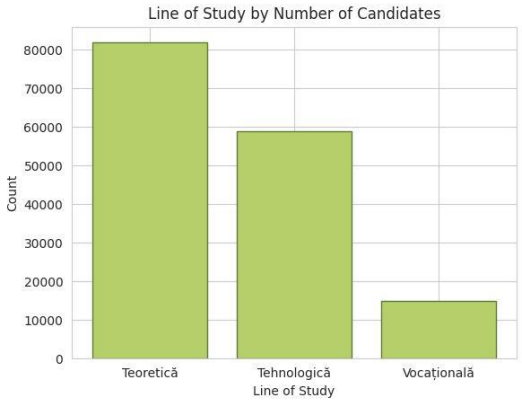


Figure 9: Distribution of Educational Pathways

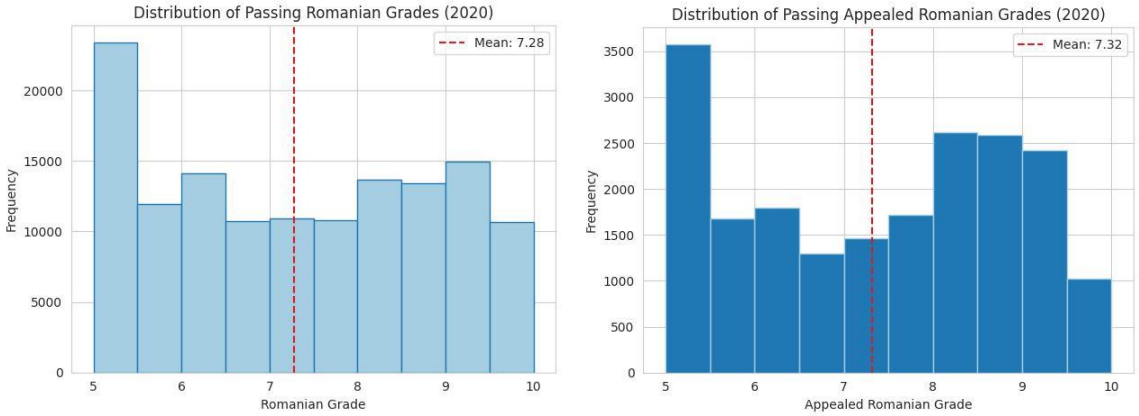


Figure 10: Distribution of Romanian Grades and Their Appeals

Taking a first look at the Distribution of Passing grades for the Romanian subject, a majority of grades are the passing grade 5, the rest of the grades having a quite uniform distribution, with an average grade of 7.28. The scores received after the appeals were quite similar, with a mean of 7.32, overall meaning that some students did manage to get their grades raised. Looking at the figure above it looks true for students with already higher grades.

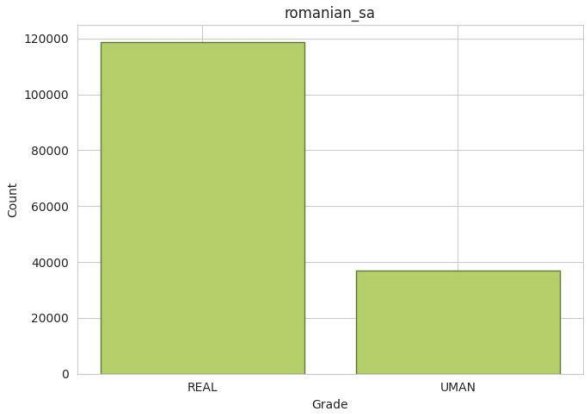


Figure 11: Distribution of Romanian Papers

Most of these students got to take the 'Real' paper, meaning that the profile of their class was most likely in the STEM category as well, although in some cases this could differ.

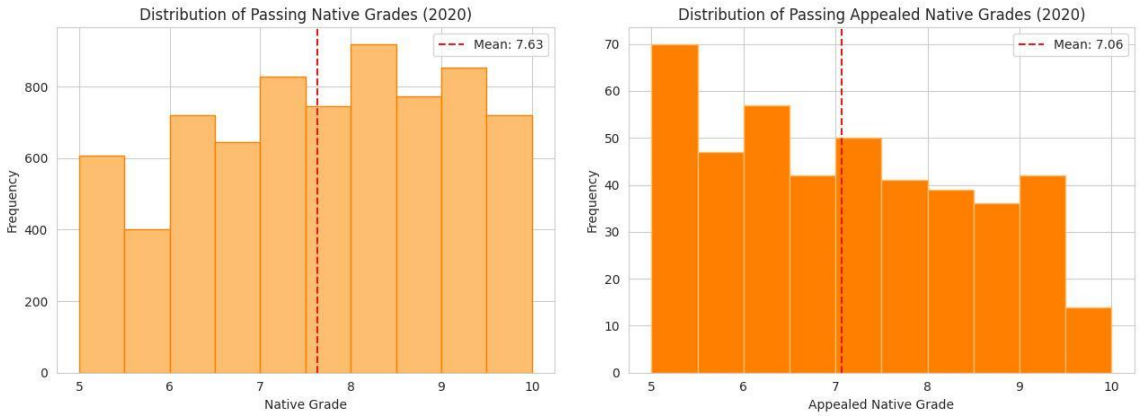


Figure 12: Distribution of Native Grades and Their Appeals

Whilst all candidates are required to take the Romanian written exam, only those with a mother tongue other than Romanian are eligible to take the native written exam. The scores have a far different distribution, with an overall higher average of 7.63 compared to their Romanian native peers, with a spread slightly left-skewed. When it comes to the appeals, the grades were significantly different, having a lower average of 7.06 that could suggest that students actually lost points appealing.

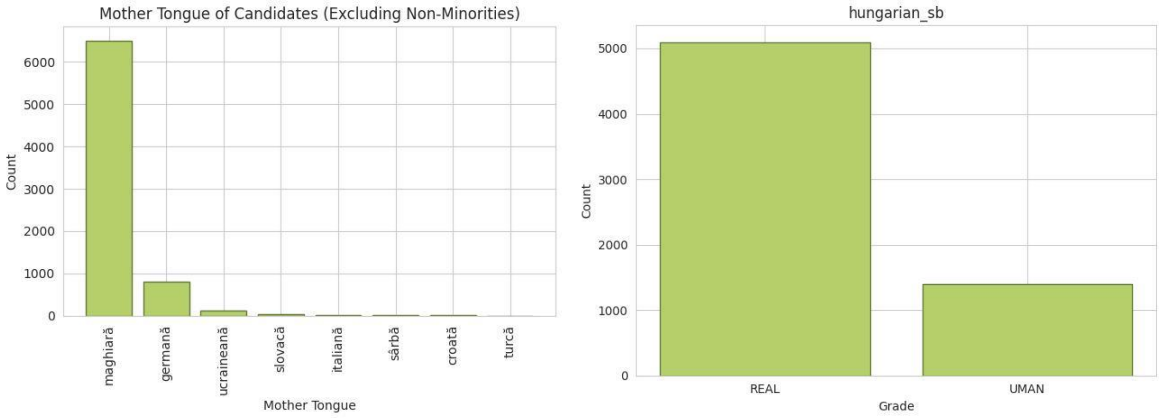


Figure 13: Distribution of Non-Romanian Natives and Hungarian Papers

Looking at the native tongues of students there is an overwhelming majority of Hungarian students who, just as well as the Romanian-native students took the 'Real' papers.

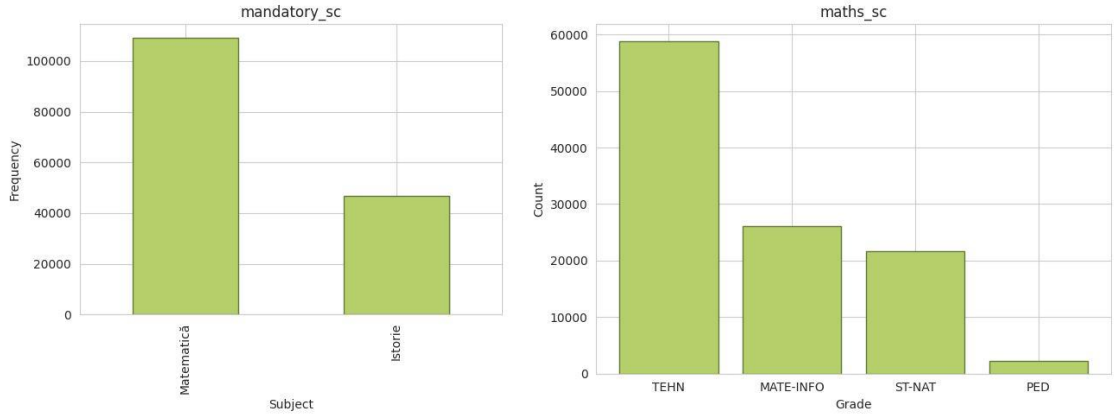


Figure 14: Distribution of Mandatory Subjects and Papers

The third paper of the Baccalaureate exam is a mandatory subject, and usually, although exceptions exist, the STEM profile takes a mathematics, whilst the humanities takes a history exam. The mathematics papers this year were mostly technical.

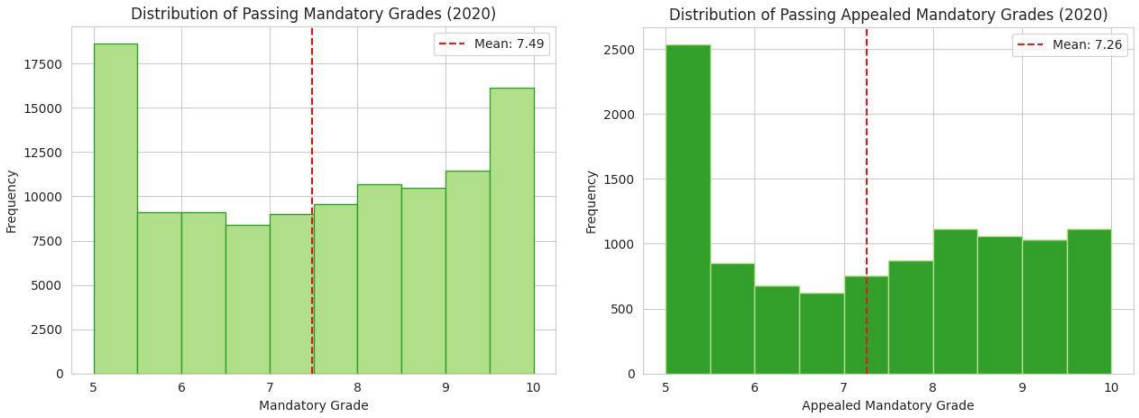


Figure 15: Distribution of Mandatory Grades and Their Appeals

When it comes to the grades from these subjects, there appeared to be a lot of passing scores, as well as quite a few maximum scores, the rest distributed uniformly, with a higher number of grades between 9.5 and 10, having an average of 7.49. The average of the appeal was 7.26, a majority of students getting the passing grade.

The last written exam of the Baccalaureate in Romania involves a chosen subject that depends on the student's specialization.

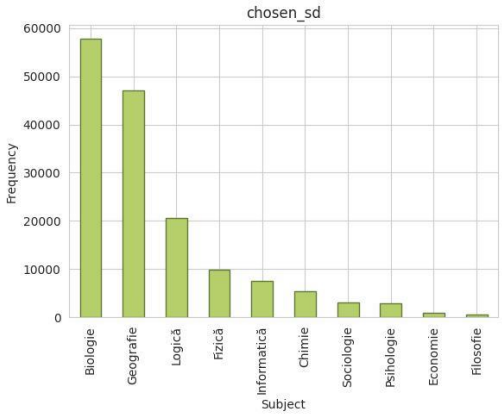


Figure 16: Distribution of Chosen Subjects

The most popular subjects to choose were Biology and Chemistry, strongly linked to the majority of students who studied mathematics and sciences or attended a technical major.

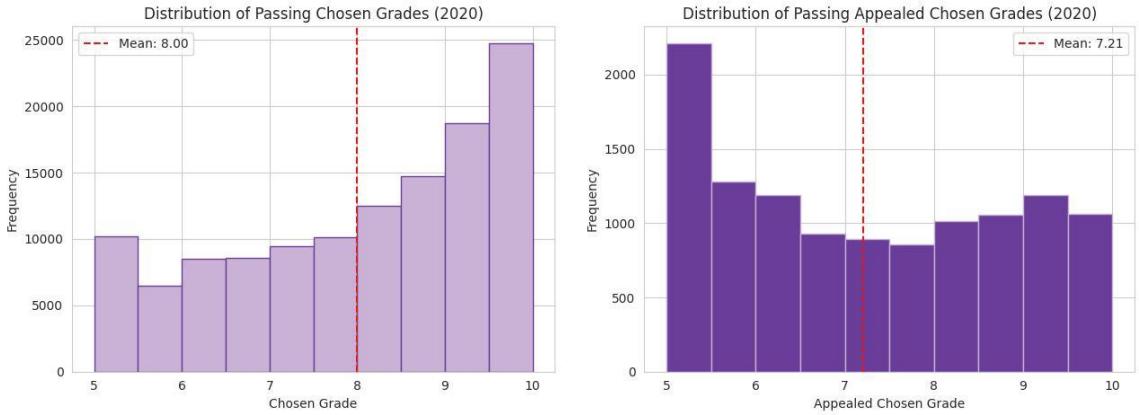


Figure 17: Distribution of Chosen Grades and Their Appeals

Looking at the chosen grades, there is a strong left-skewed spread, most candidates achieving outstanding marks. The average was an 8, whilst the appeal was a 7.21. When it comes to the appeals, a lot of candidates achieved the passing grade on this paper as well.

Students must obtain a foreign language qualification. They can achieve this either at the exams or provided by an institution such as Cambridge for English qualifications.

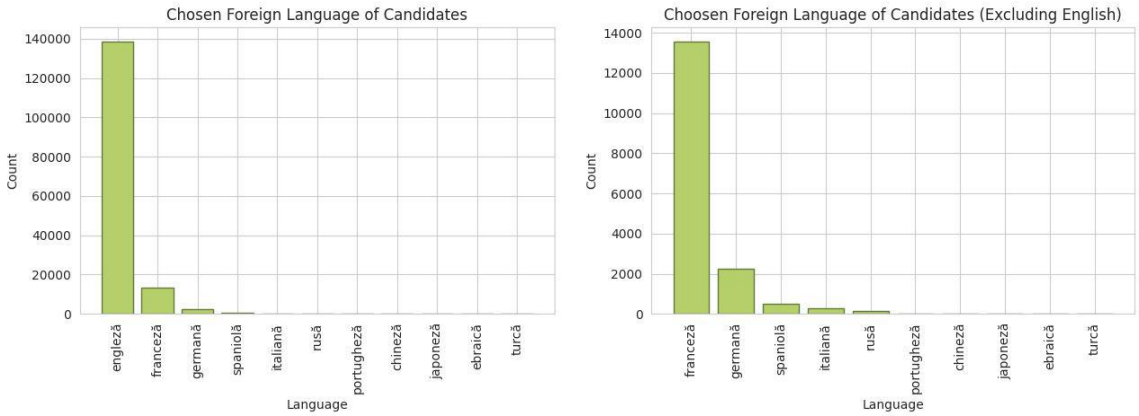


Figure 18: Distribution of Foreign Languages of Candidates

It comes as no surprise that the most popular choice of foreign language to study by candidates was English, the not-so-close second being French.

It is also required from the students to take a set of oral exams to prove their proficiency, but since 2020 was the first year affected by the COVID-19 virus, this year these speaking exams were not held, in respect to the health regulations, and instead candidates got marked by their grades obtained throughout high school.

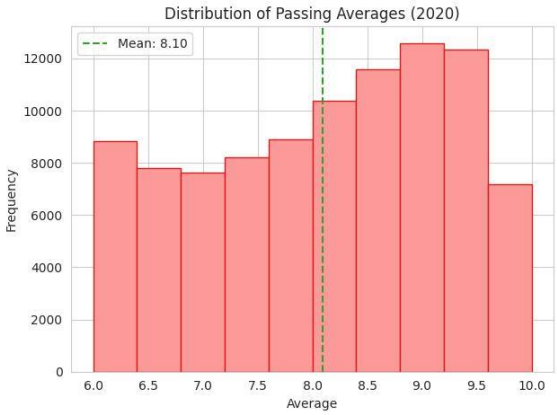


Figure 19: Distribution of Passing Averages

When it comes to the distribution of final grades (only looking at passing scores), the average was 8.1, the distribution is slightly left-skewed.

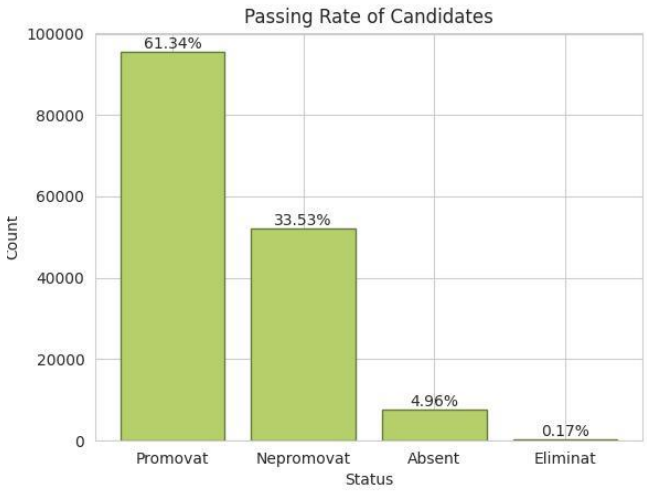


Figure 20: Rate of Passing in 2020

Looking at the above diagram, the overall rate of passing candidates was 61.34%, meaning that circa 40% of students failed to have a chance at higher education.

Comparative Analysis of 2020 Grades, Average and Influencing Features

Next, we covered the influence of important categorical features on the four written exam grades and the final grade (average). We looked at six important factors to consider when testing for inferences, such as gender, profile, specialization, form of study, environment of candidate and native language.

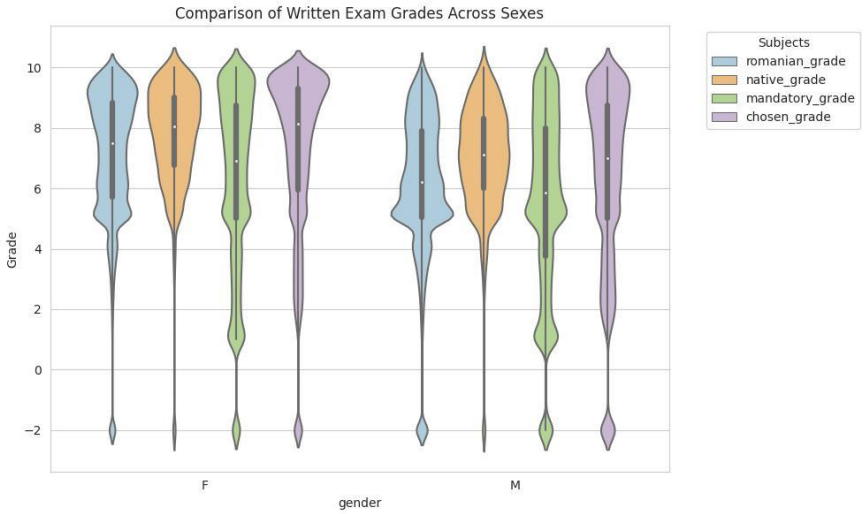


Figure 21: Comparison of Written Exam Grades and Gender

When it comes to gender there were more female students in the year 2020 that achieved a higher score from all four subjects, and male students represented more of the passing grades.

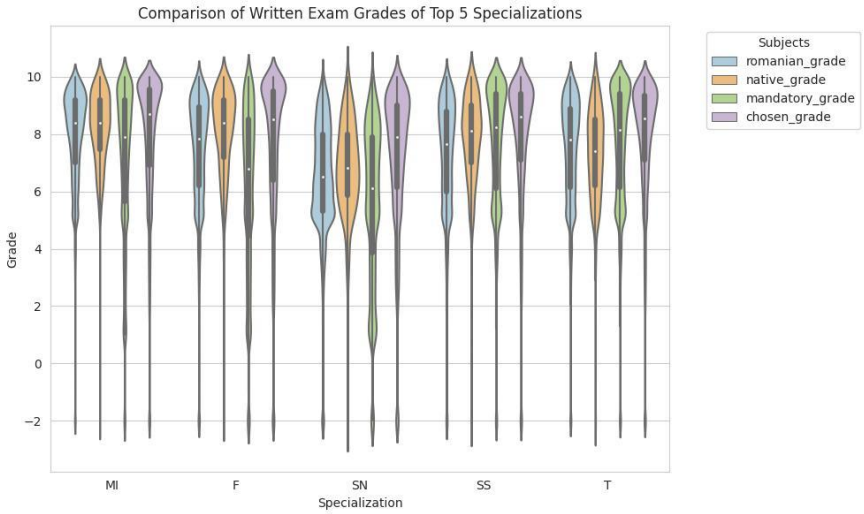


Figure 22: Comparison of Written Exam Grades and Top 5 Specializations

Across the specializations the most passing grades from Romanian were scored by Natural Science students, and most students with mother tongues different than Romanian fall into the Natural Science specialization.

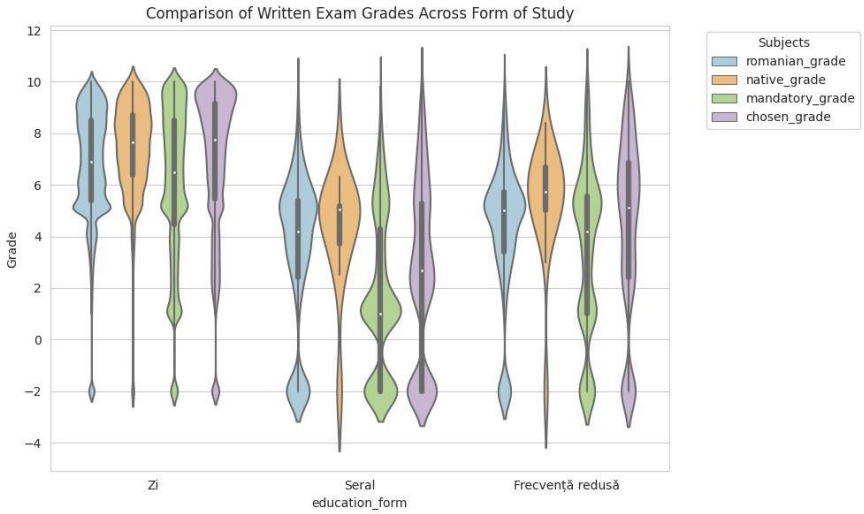


Figure 23: Comparison of Written Exam Grades and Form of Study

The above figure shows that the students who go to school by day achieved most of the higher scores compared to their peers who go to study at night or are in distance learning.

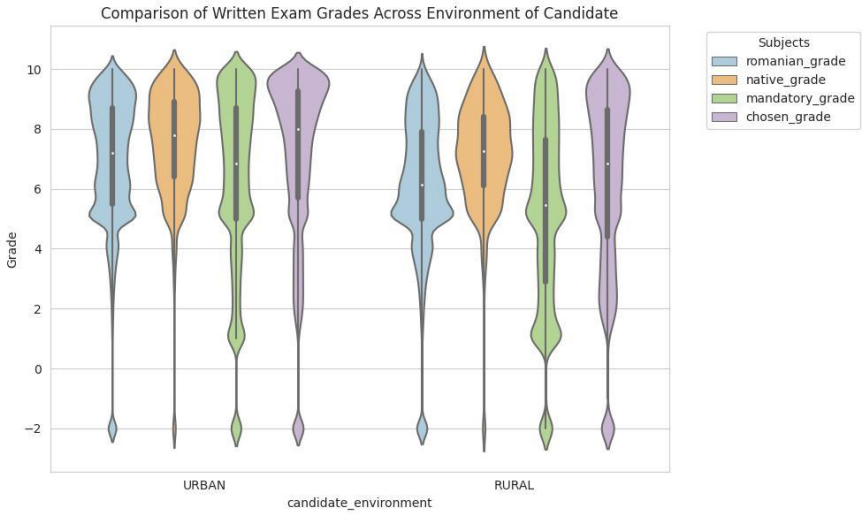


Figure 24: Comparison of Written Exam Grades and Environment of Candidates

The majority of passing grades from Romanian fell into the rural category, the urban one achieving higher scores from their chosen subject.

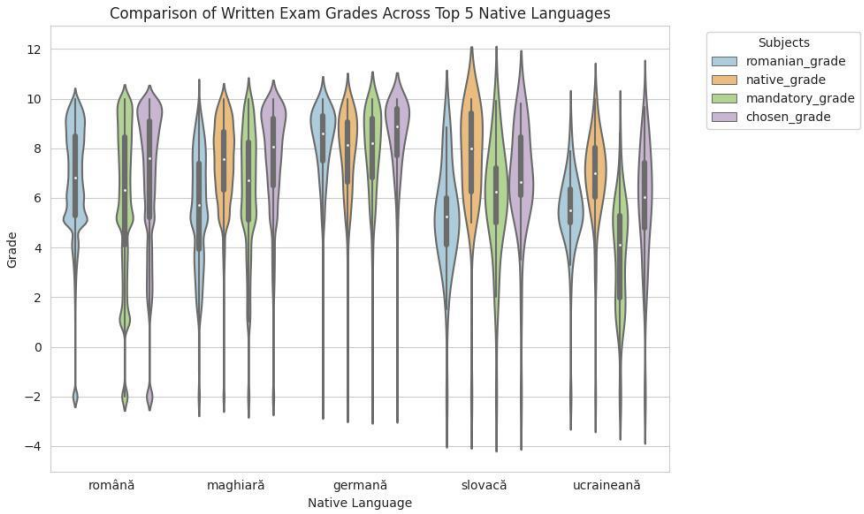


Figure 25: Comparison of Written Exam Grades and Top 5 Native Languages

Showcasing the difference in mother tongue, there is an obvious and significant difference in the Romanian grades of native Romanian speakers, compared to their minority peers, Slovakian students having the majority of lower grades. Ukrainian and Romanian students achieved more of the lowest scores in their mandatory exams.

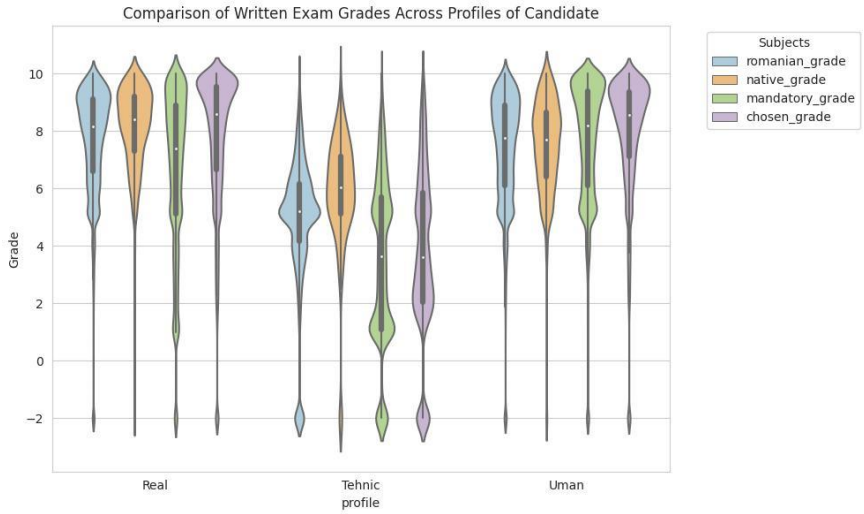


Figure 26: Comparison of Written Exam Grades and Profiles

When it comes to profiles, the technical students achieved an apparent lower score from all four subjects, the other two profiles performing rather similarly, the humanities students achieving more of the higher scores from History than their peers from Mathematics.

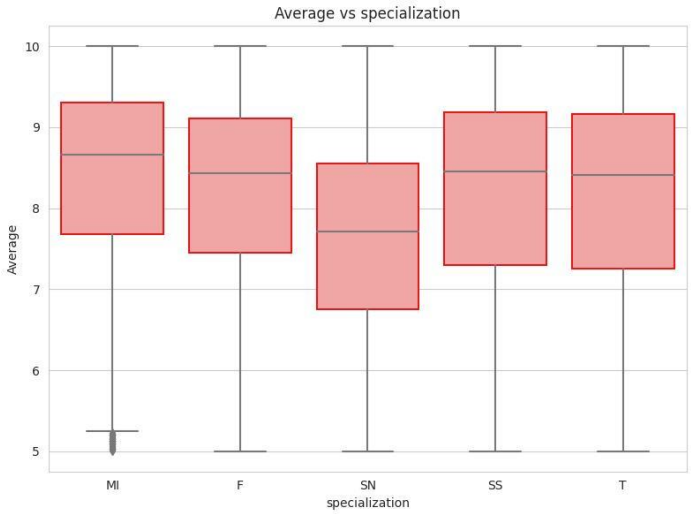


Figure 27: Box Plot of 2020 Averages by Specialization

Looking at the above figure, in the top five specializations a lower range of final grades were achieved by the Natural Sciences specialization, the rest of students achieving quite similar results, the highest range of grades were written by the Mathematics and Informatics students. The final grades have an overall normal distribution.

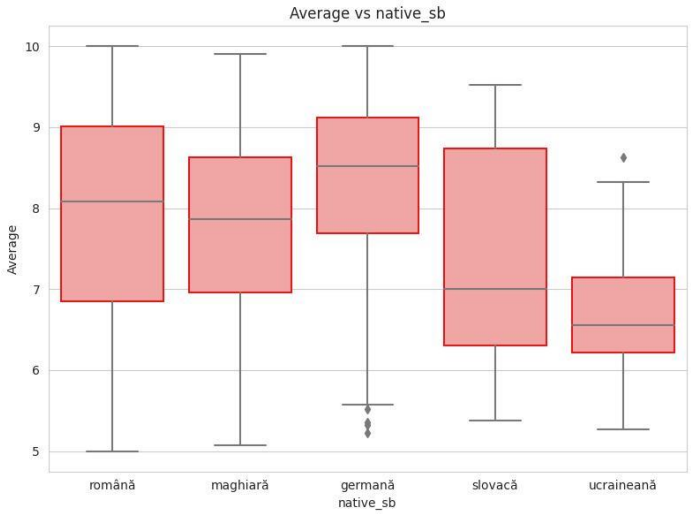


Figure 28: Box Plot of 2020 Averages by Native Language

The native language of students is also quite influential, as discussed before, but when it comes to final grades, the highest-ranking group was the native German candidates, also with the highest median, the lowest range being the Ukrainian students.

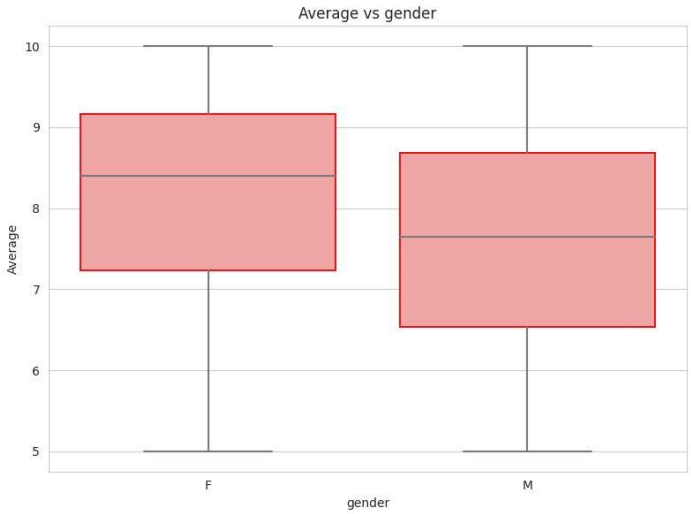


Figure 29: Box Plot of 2020 Averages by Gender

The male candidates performed better with a higher median and achieved a range of better scores overall.

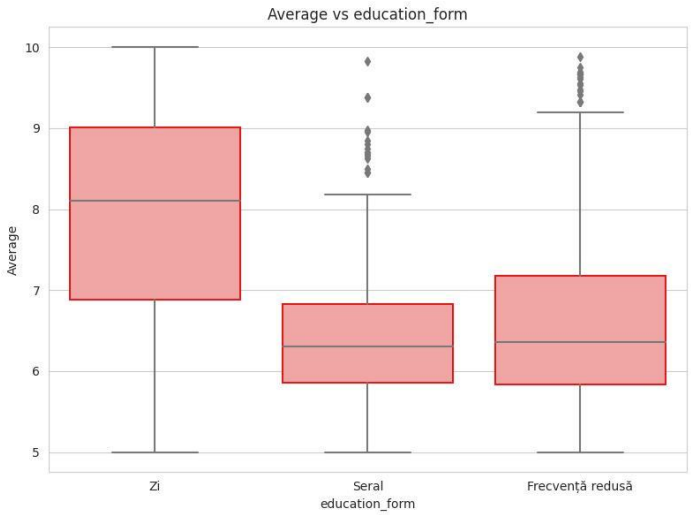


Figure 30: Box Plot of 2020 Averages by Education Form

When it comes to form of study, in the final grade as well, most students achieved a significantly higher score when attending school by day instead of night or by participating in distance studies. Not only is the median of day students higher but most night students couldn't even reach the majority of day student's grades.

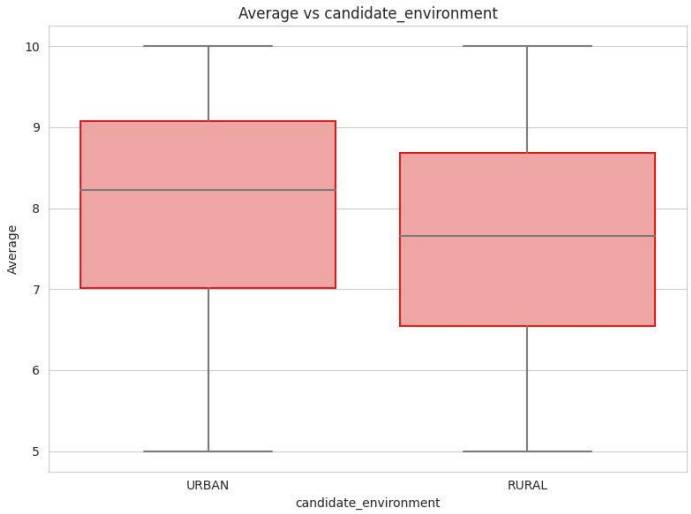


Figure 31: Box Plot of 2020 Averages by Candidate Environment

Last, but not least, looking at candidate's environments, the majority of urban students scored better on their exams.

Overall, it is safe to say that gender, profile, specialization, form of study, environment of candidate and native language greatly influence the Romanian, native language, mandatory subject and chosen subject grades, thus having an influence on the final grade (average) as well.

Measures of Central Tendency and Variability of 2021 Baccalaureate Scores

After gaining some valuable insights related to exam performance of 2020 candidates, we conducted a similar analysis on each of the following 3 years to see if there were any improvements or quite the contrary.

	Romanian Grade	Native Grade	Mandatory Grade	Chosen Grade	Romanian Grade (a)	Native Grade (a)	Mandatory Grade (a)	Chosen Grade (a)	Average
count	133664.0	6573.0	133664.0	133664.0	23158.0	421.0	17200.0	15202.0	95289.0
mean	6.690776	7.638270	6.148013	7.071971	6.866236	6.837340	5.703823	6.543786	7.862957
std	2.293798	1.797514	3.036192	2.900515	1.908728	1.656187	2.361584	2.187129	1.262246
min	-2.0	-2.0	-2.0	-2.0	1.1	2.9	1.0	1.0	5.0
25%	5.4	6.7	5.0	5.8	5.3	5.6	3.95	5.0	6.86
50%	7.0	8.0	6.5	7.9	7.2	6.7	5.05	6.45	8.0

75%	8.45	9.0	8.4	9.2	8.5	8.3	7.9	8.6	8.91
man	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0

Figure 32: Basic Statistics Table of 2021 Baccalaureate Scores

The number of candidates partaking in the exams decreased from previous years, and there have been changes in the grades as well. Taking into account the previous year's 7.9 average, the year stayed quite consistent with an average of 7.86, with a higher average in mandatory and chosen subject grades. As for the appeals, each grade increased compared to the previous year.

Frequency Distributions of 2021 Baccalaureate Scores & Candidate Information

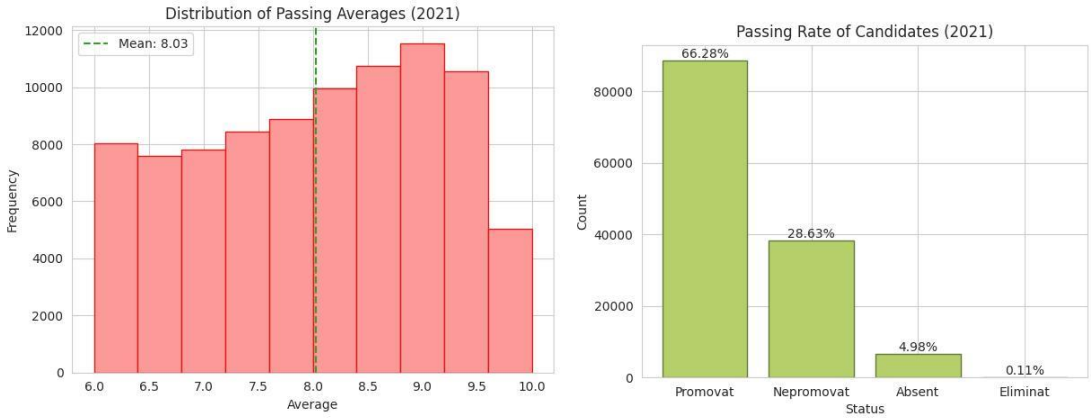


Figure 33: Distribution of Final Grades and Passing Rate of 2021

Compared to last year's results the distribution of final grades stayed relatively the same, with a new average of 8.03, the passing rate of candidates however increased to 66.28%.

Comparative Analysis of 2021 Grades, Average and Influencing Features

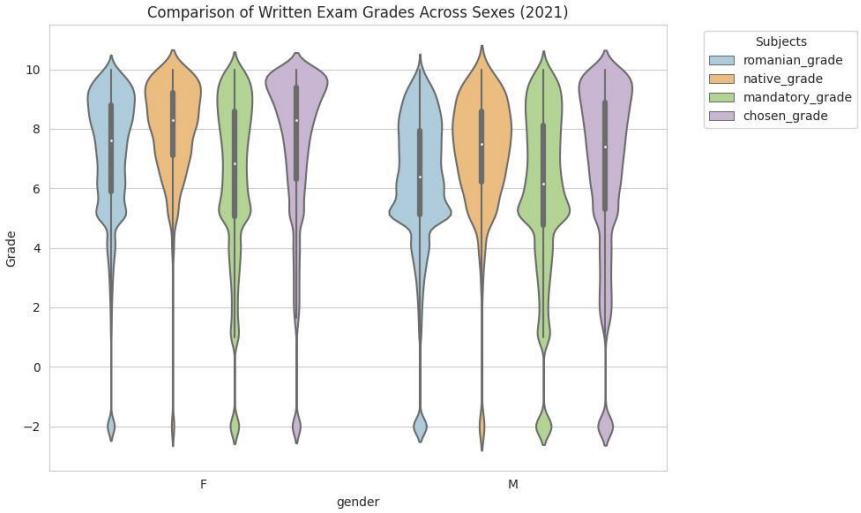


Figure 34: Written Exam Grades by Gender in 2021

Looking at the 2021 results, the female population of candidates managed to score higher more times than the male population, who managed to score most of the passing grades.

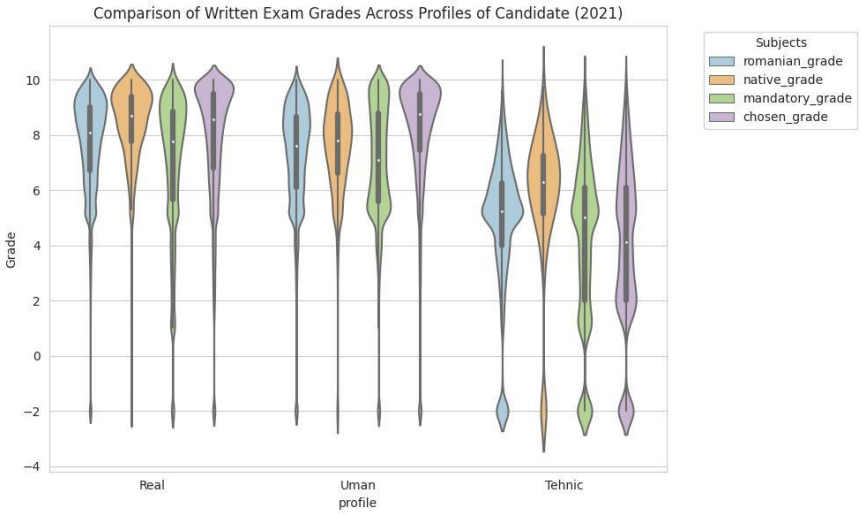


Figure 35: Written Exam Grades by Profile in 2021

Similarly to a year before, the technical majors managed to score below the STEM and humanities profiles from all subjects.

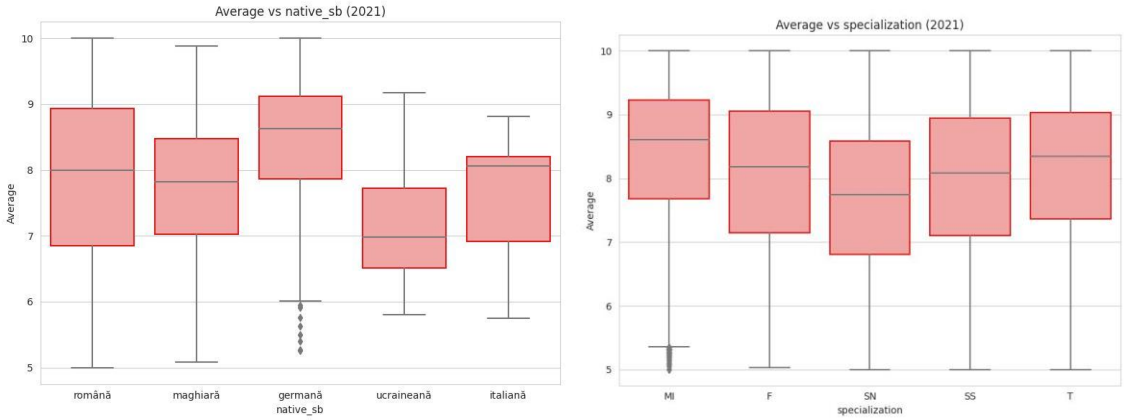


Figure 36: Average Scores by Native Language and Specialization in 2021

The final grades ranging the highest belong to native German candidates, also with the highest median, the lowest median being the Ukrainian students.

In the top five specializations a lower range of final grades were achieved by the Natural Sciences specialization, the highest range of grades were written by the Mathematics and Informatics students.

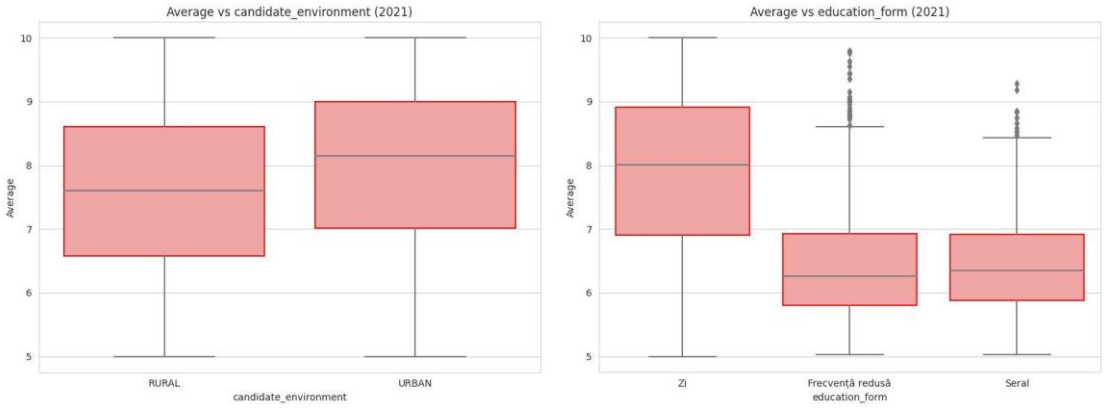


Figure 37: Average Scores by Candidate Environment and Form of Study in 2021

Looking at the candidate's environments, the majority of urban students did better overall.

When it comes to form of study, in the final grade as well, most students achieved a significantly higher score when attending school by day.

Measures of Central Tendency and Variability of 2022 Baccalaureate Scores

	Romanian Grade	Native Grade	Mandatory Grade	Chosen Grade	Romanian Grade (a)	Native Grade (a)	Mandatory Grade (a)	Chosen Grade (a)	Average
count	126453.0	6321.0	126453.0	126453.0	22185.0	423.0	13226.0	13292.0	126453.0
mean	6.676278	7.635437	6.856939	7.386546	6.648826	6.742908	6.114895	6.698834	5.486605
std	2.159785	1.650552	2.703356	2.504983	1.869696	1.635314	2.336604	2.117041	4.634584
min	-2.0	-2.0	-2.0	-2.0	1.15	1.2	1.0	1.3	-3.0
25%	5.4	6.7	5.35	6.3	5.1	5.5	4.15	5.0	5.63
50%	6.75	7.9	7.45	8.1	6.75	6.8	5.55	6.65	7.41
75%	8.35	8.85	9.05	9.2	8.3	8.1	8.4	8.65	8.75
man	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0

Figure 38: Basic Statistics Table of 2022 Baccalaureate Scores

Looking at the 2022 results, the final grade average had a significant drop from last year's 7.86 to a 5.48, and the number of candidates decreased as well. When it comes to the average grades, they also suffered a slight drop from the previous year's results. The appeals have improved from the mandatory and chosen subject, but compared to last year the Romanian and native language grades are lower.

Frequency Distributions of 2022 Baccalaureate Scores & Candidate Information

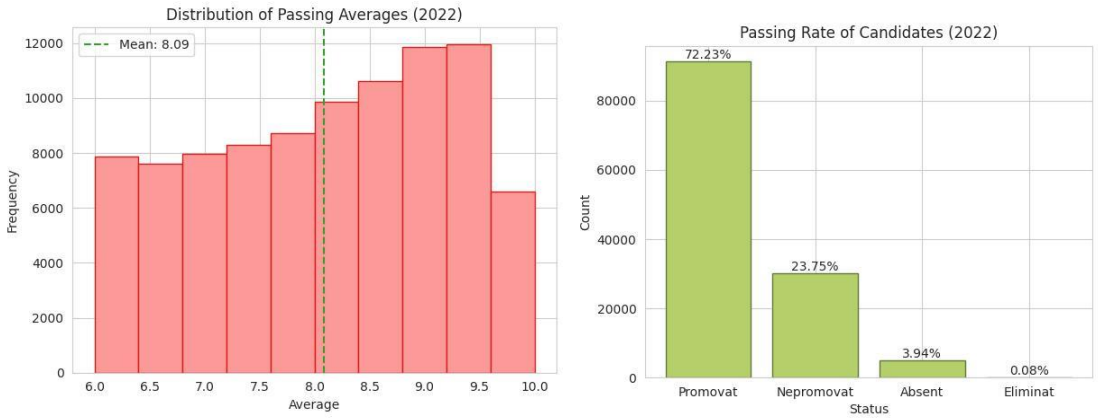


Figure 39: Distribution of Final Grades and Passing Rate of 2022

The distribution of final grades stayed relatively the same, with a new average of 8.09, the passing rate of candidates however had a significant increase from the previous 66.28% to 72.23%, which can be due to the fact that student no longer had to face COVID-19 restrictions and attended face-to-face classes. Considering all factors, the achieved passing rate is a great improvement that can further be studied.

Comparative Analysis of 2022 Grades, Average and Influencing Features

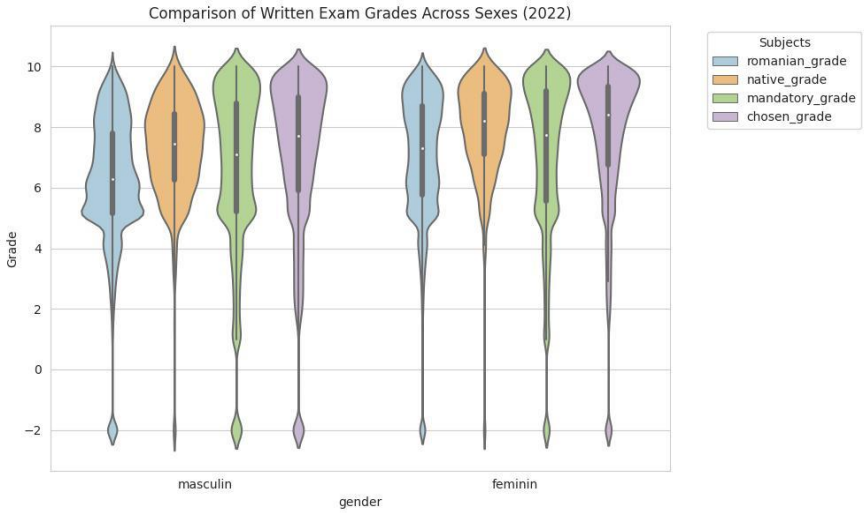


Figure 40: Written Exam Grades by Gender in 2022

Looking at this year's results, the distribution of grades across genders is the same as last year's. The female population of candidates managed to score higher more times, with male candidates scoring passing grades more often than their female peers for most subjects, excluding the chosen subject, where grade distribution is quite similar.

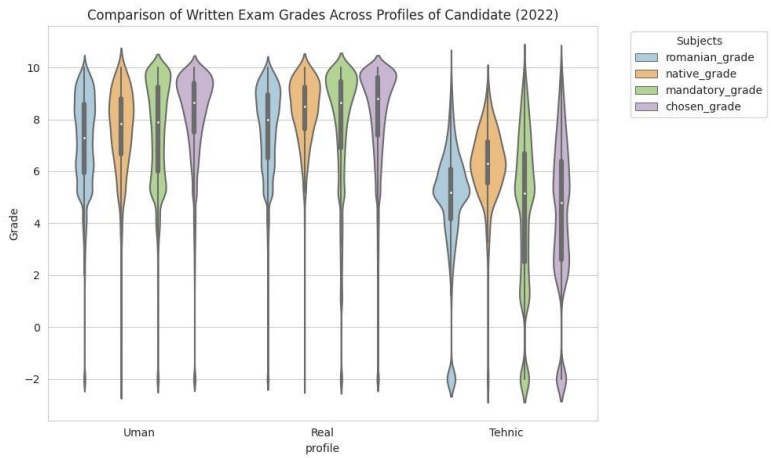


Figure 41: Written Exam Grades by Profile in 2022

When it comes to profiles, just like the year before the Technical majors managed to score below the ‘Real’ and ‘Uman’ profiles from all subjects.

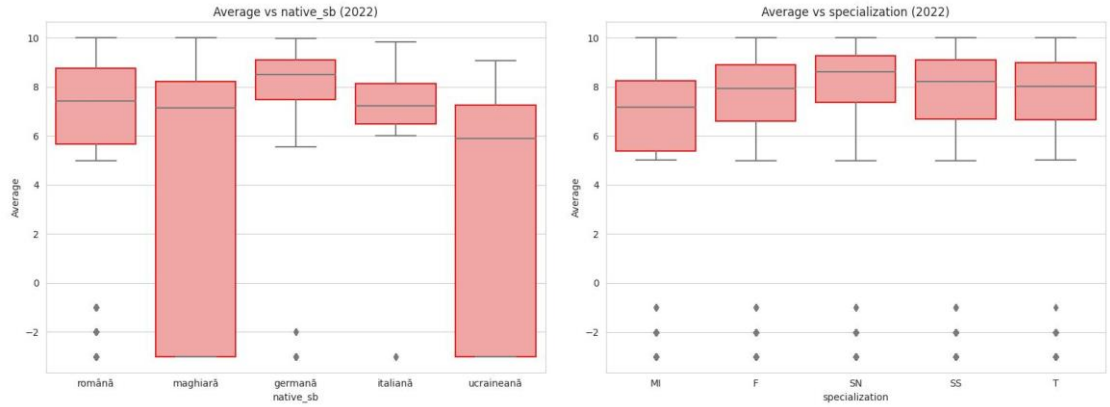


Figure 42: Average Scores by Native Language and Specialization in 2022

The final scores are quite different this year, as noted looking at the averages in the previous section of the analysis, where more native Hungarian and Ukrainian students scored significantly lower as most of their grades fell below the median values. As mentioned before, the values of -2 are outliers, possibly due to the fact that these candidates were exempt from taking the exam. Surprisingly, the specialization grouping is also distributed differently from the previous year. This time the highest average was scored by students from the Natural Sciences specialization, with overall more students achieving higher final grades.

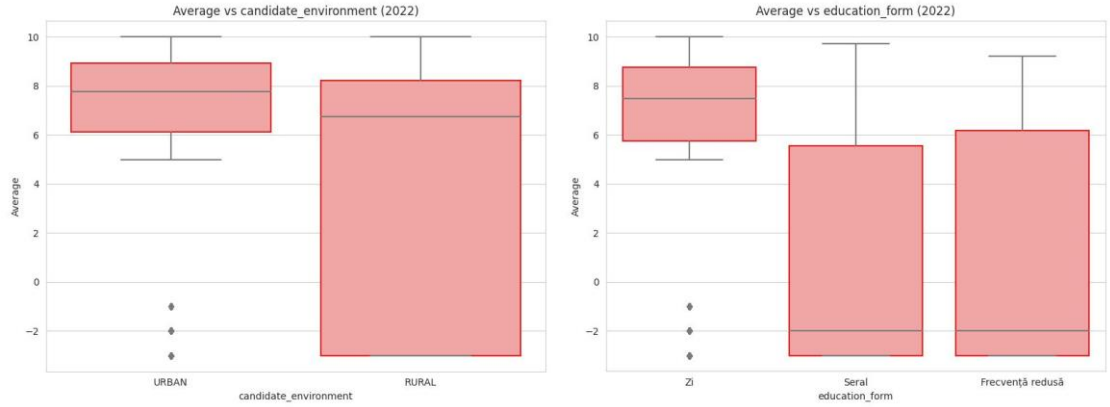


Figure 43: Average Scores by Candidate Environment and Form of Study in 2022

When it comes to candidate’s environments, majority of urban students did better overall, having a higher median as well. Most of the rural candidates scored lower than the median. Same as for last year’s results, for the form of study, most students achieved a significantly

higher score when attending school by day, the median far higher than for other forms of education, although most night and distance students achieved higher grades than the median.

Measures of Central Tendency and Variability of 2023 Baccalaureate Scores

	Romanian Grade	Native Grade	Mandatory Grade	Chosen Grade	Romanian Grade (a)	Native Grade (a)	Mandatory Grade (a)	Chosen Grade (a)	Average
count	130522.0	6671.0	130522.0	130522.0	26189.0	488.0	13386.0	14577.0	101454.0
mean	6.661397	7.571249	6.744083	7.173927	6.774468	6.848566	6.227742	6.555625	7.805532
std	2.145408	1.711849	2.657173	2.534455	1.742377	1.578529	2.254574	2.052865	1.196263
min	-2.0	-2.0	-2.0	-2.0	1.0	1.6	1.0	1.4	5.0
25%	5.4	6.6	5.4	6.1	5.25	5.8	4.3	5.0	6.83
50%	6.95	7.8	7.25	7.85	7.1	6.9	5.75	6.5	7.93
75%	8.25	8.8	8.8	9.0	8.25	8.1	8.35	8.45	8.83
man	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0

Figure 44: Basic Statistics Table of 2021 Baccalaureate Scores

This year's average final grade had a slight decrease from last year's at 7.8. When it comes to the written exam grades, all average scores slightly decreased from the previous averages. Taking a look at the appeals however, one appeal increased, that being the average score of the chosen subject.

Frequency Distributions of 2023 Baccalaureate Scores & Candidate Information

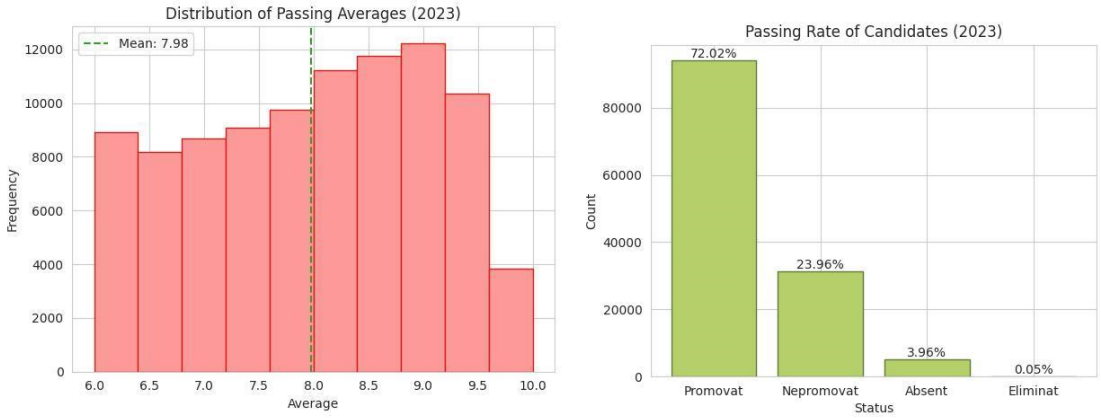


Figure 45: Distribution of Final Grades and Passing Rate of 2023

The distribution of final grades is relatively the same compared to a year before, with a new average of 7.89. The passing rate of candidates had a slight decrease to 72.02%.

Comparative Analysis of 2023 Grades, Average and Influencing Features

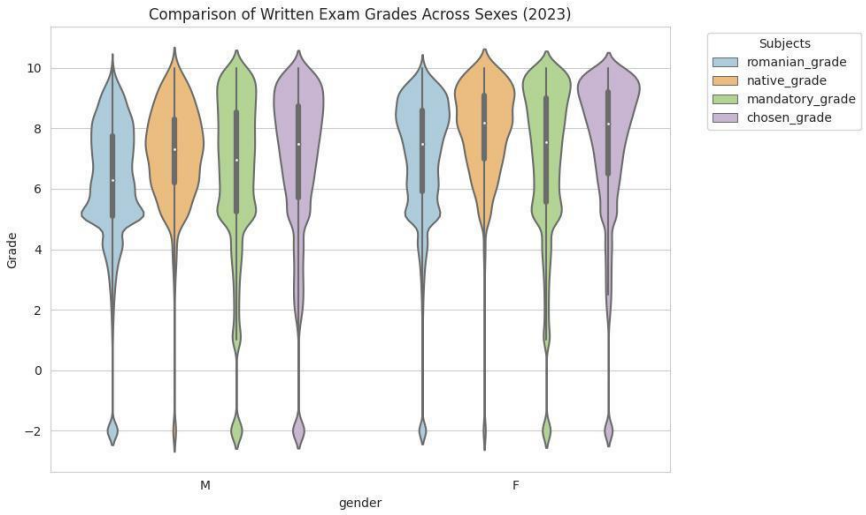


Figure 46: Written Exam Grades by Gender in 2023

Just as the year before, we concluded that the female population of candidates managed to score higher more times, with male candidates scoring passing grades more often, for most subjects. The female students did significantly better on the Native Language exam, out of all exams.

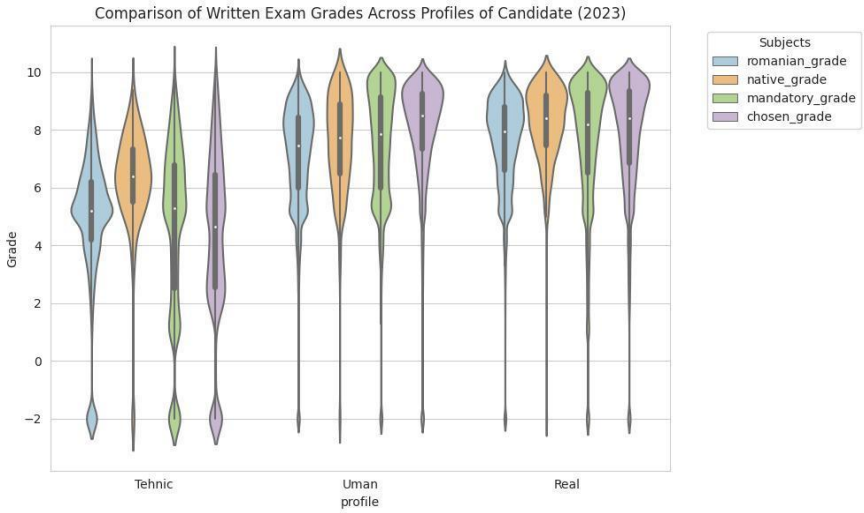


Figure 47: Written Exam Grades by Profile in 2023

In terms of profiles, just like the year before, the technical majors managed to score below the ‘Real’ and ‘Uman’ profiles from all subjects, the students having more higher grades in the ’Real’ profile from Romanian and Native Language.

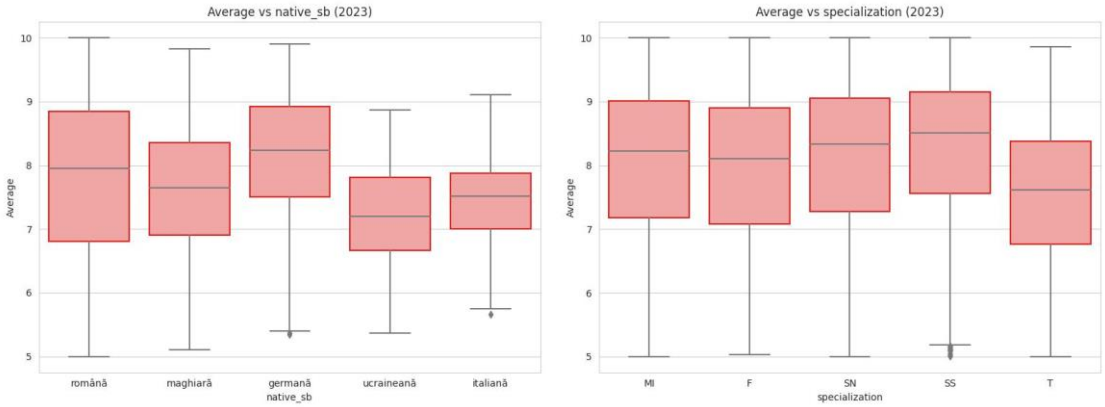


Figure 48: Average Scores by Native Language and Specialization in 2023

Looking at average scores the German Native students, yet again, managed to have the highest median final scores. By specialization, this year the highest median was achieved by Social Sciences, closely followed by the other Specializations.

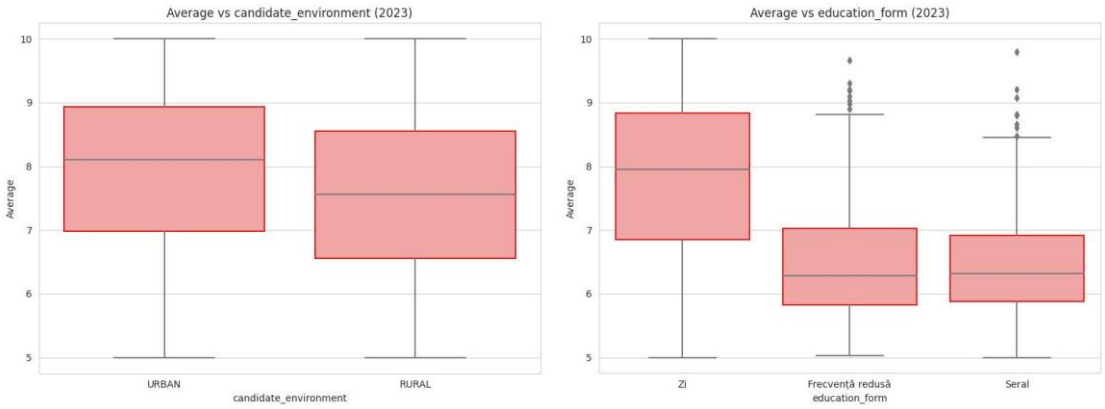


Figure 49: Average Scores by Candidate Environment and Form of Study in 2023

Yet again, the urban candidates did better, only slightly, having a higher median final grade. Same as for last year's results, most students achieved a significantly higher score when attending school by day, the median far higher than for other forms of education.

Overview of Exam Results between 2020 and 2023

Last but not least, we compared some of the features across all datasets to see possible patterns or trends related to the Romanian, Native Language, Mandatory and Chosen Subject Grades, as well as Average Scores.

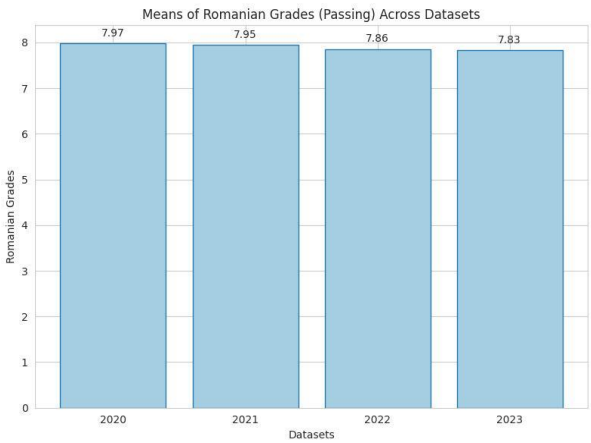


Figure 50: Average Romanian Scores between 2020 and 2023

When it comes to the Romanian Grades, over the years there is a slight linear decrease in average scores.

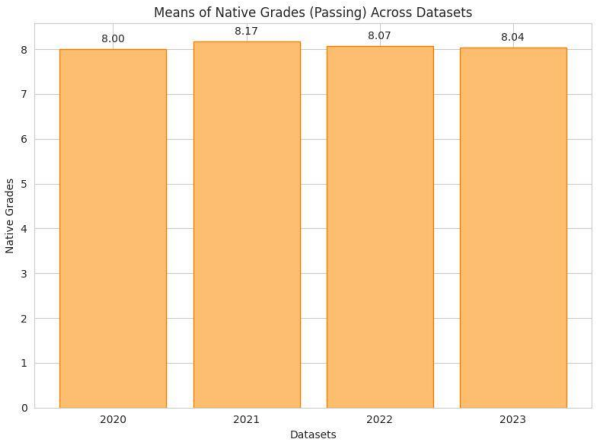


Figure 51: Average Native Scores between 2020 and 2023

After the year 2020, there was a slight increase in average Native grades, however from 2021 on, there is a slight decrease.

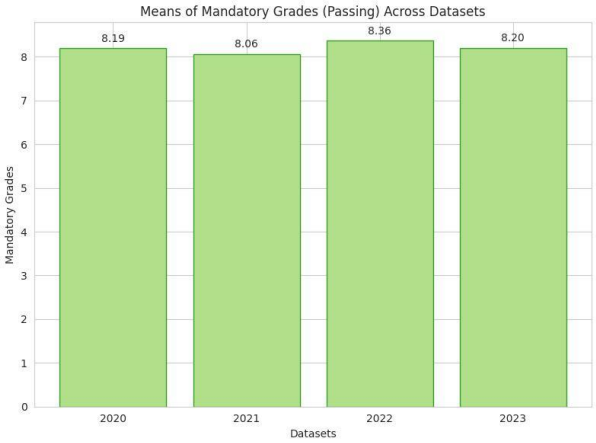


Figure 52: Average Mandatory Scores between 2020 and 2023

There is no linear trend when it comes to Mandatory average grades.

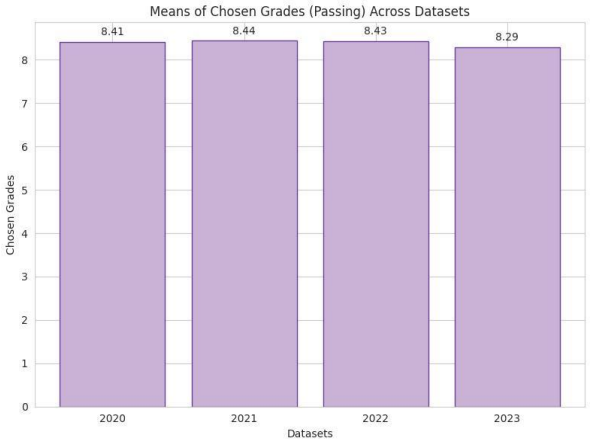


Figure 53: Average Chosen Scores between 2020 and 2023

Just like the native grades, the chosen grades averages also experienced a slight increase in 2021 and continued to decrease after.

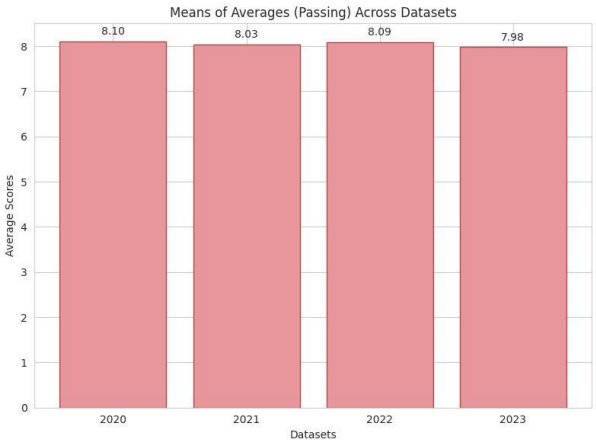


Figure 54: Average Final Scores between 2020 and 2023

There also seems to not be any linear trend in the increase and decrease of Final Grade averages.

For deducting inferences, we decided to closely monitor the number of female and male candidates throughout the years.

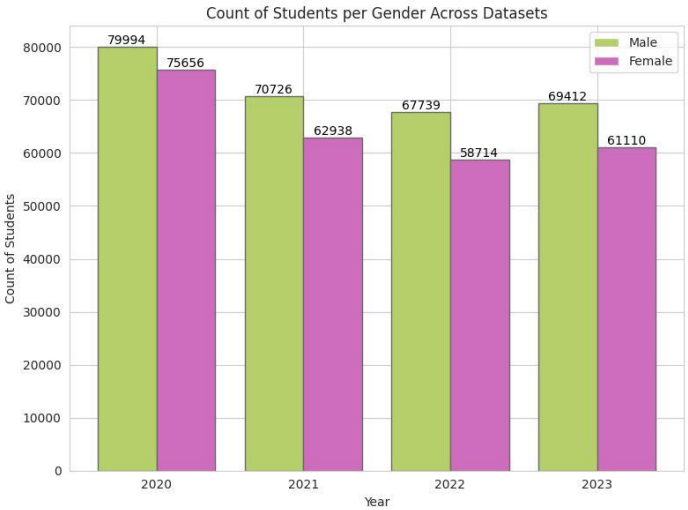


Figure 55: Count of Candidates by Gender between 2020 and 2023

Up until the recent year (2023) there has been a linear decline of attending candidates from both genders, having a slightly more male population, but in the most recent round of examinations, the number of candidates increased.

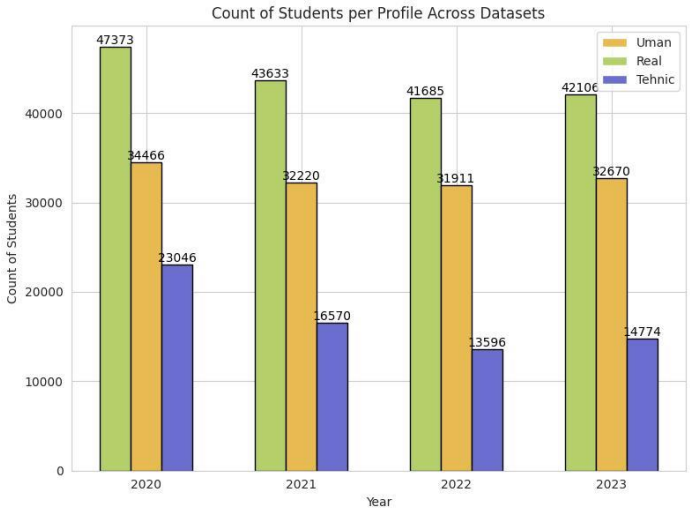


Figure 56: Count of Candidates by Profile between 2020 and 2023

When it comes to the profiles of candidates, the same can be said about the linear decrease of candidates, having most students from ‘Real’ profiles, followed by ‘Uman’ and lastly ‘Tehnic’ students, but with the recent session the numbers of attendees slightly increased.

Inferences

Correlation Analysis

Correlation Analysis of 2020 Baccalaureate Grades

Hypothesis Testing

H0: There is no linear correlation between the written exam grades and their appeals.

H1: There is a strong linear correlation between the written exam grades and their appeals.

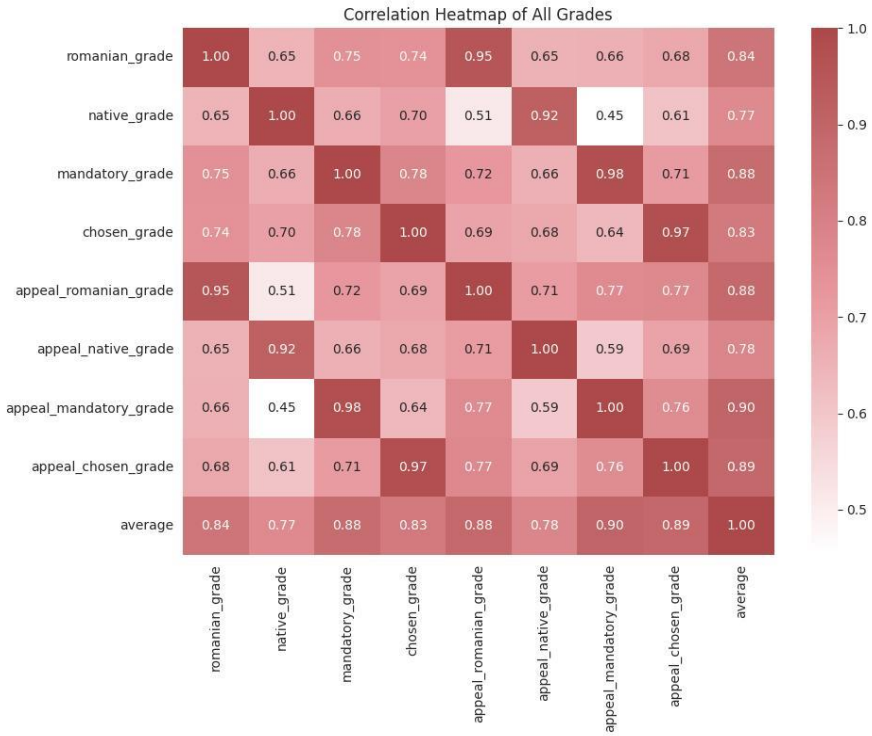


Figure 57: Correlation Heatmap of 2020 Exam Grades

There is a high correlation between the four written exams and their appeal grades. Most likely, the teachers grading the appeals agreed with the previous scores and there were no significant increases or decreases in the scores after appealing. The appeal grade seems to follow the grade closely, showing a strong linear relationship.

When it comes to the final score, the most influential written exam is the mandatory subject, followed by Romanian, chosen and native language. The most influential appeal is also the mandatory subject, followed this time by the chosen subject. All correlation coefficients have a positive value, meaning that with the increase of a respective grade, the other increases as well and vice versa.

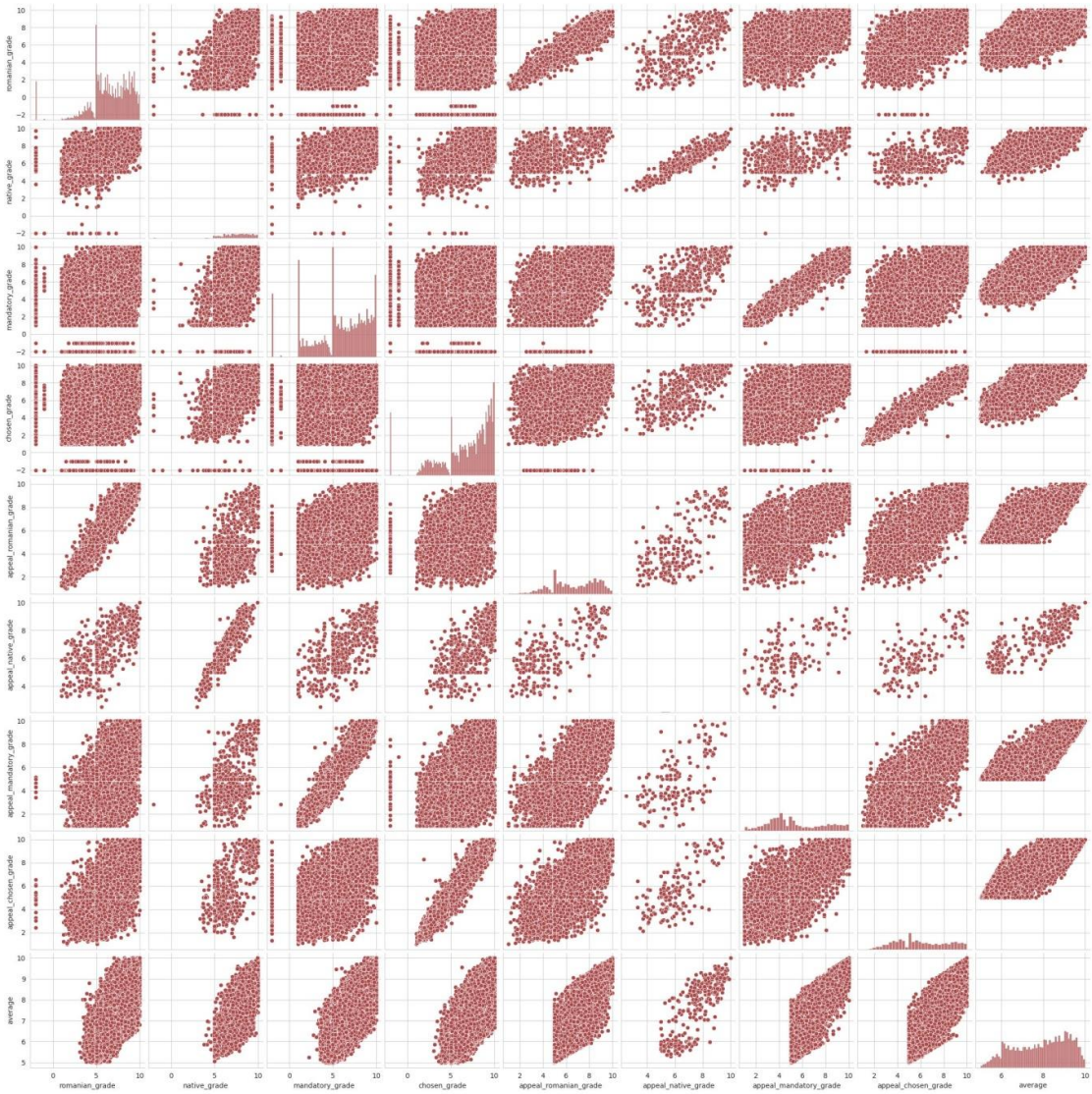


Figure 58: Paired Scatter Plot of 2020 Exam Grades

The same inferences can be made from the paired plot above, showing the strong linear correlation between written grades and their appeals.

Keeping in mind that a correlation coefficient with a value higher than 0.6 results in a strong linear relationship between two variables, we reject the null hypothesis and accept the alternative hypothesis, since all correlation coefficient values fall in the range of 0.90-0.99.

Correlation Analysis of 2021 Baccalaureate Grades

Hypothesis Testing

H0: There is no linear correlation between the written exam grades and their appeals.

H1: There is a strong linear correlation between the written exam grades and their appeals.

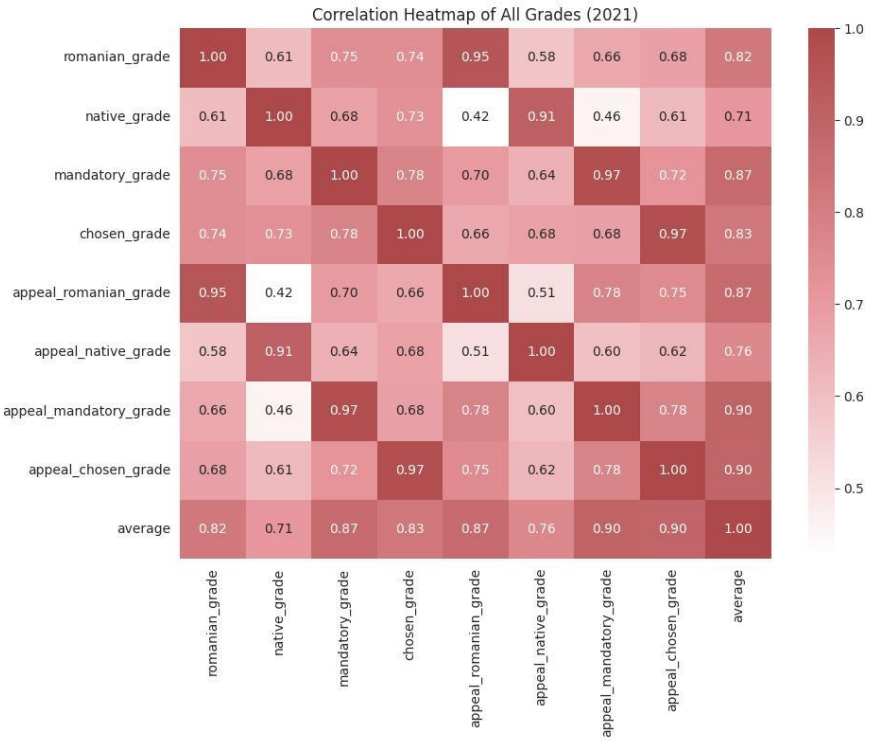


Figure 59: Correlation Heatmap of 2021 Exam Grades

The results for the 2021 grades correlation analysis match the previous year's results, with only slight changes, as in there is a high correlation between the four written exams and their appeal grades. The appeal grades seem to follow the written grades closely suggesting a strong linear relationship between the features.

When it comes to the final score, the most influential written exam is yet again the mandatory subject, followed by chosen, Romanian and native language. The most influential appeal is the mandatory and chosen subject with the same correlation coefficient. The correlation coefficients in this case are also all positive.

With all correlation coefficient values being way above 0.6 (in the range of 0.90-0.99) in the case of the 2021 results as well, we came to the same conclusion as for the year before. There is a strong linear relationship between written grades and appeal grades, so we reject the null hypothesis and accept the alternative hypothesis.

Correlation Analysis of 2022 Baccalaureate Grades

Hypothesis Testing

H0: There is no linear correlation between the written exam grades and their appeals.

H1: There is a strong linear correlation between the written exam grades and their appeals.

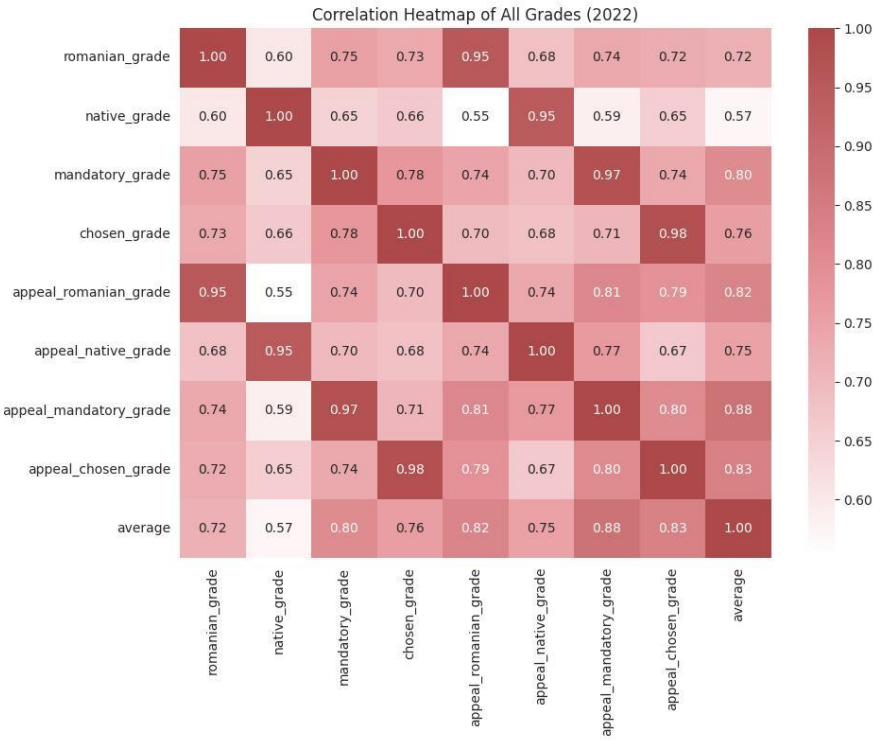


Figure 60: Correlation Heatmap of 2022 Exam Grades

The same conclusions can be made for the 2022 grades as well, meaning that the written grades and their appeals are strongly correlated. In this case, the final grade is influenced the strongest by the mandatory grade, and in terms of appeals, the mandatory appeal grade. This year's Native grade seems to have a lot less of an influence on the Final Grade than in the previous two years. All correlation coefficients are positive this time again.

Yet again, there is a strong linear relationship between written grades and appeal grades, so we reject the null hypothesis and accept the alternative hypothesis.

Correlation Analysis of 2023 Baccalaureate Grades

Hypothesis Testing

H0: There is no linear correlation between the written exam grades and their appeals.

H1: There is a strong linear correlation between the written exam grades and their appeals.

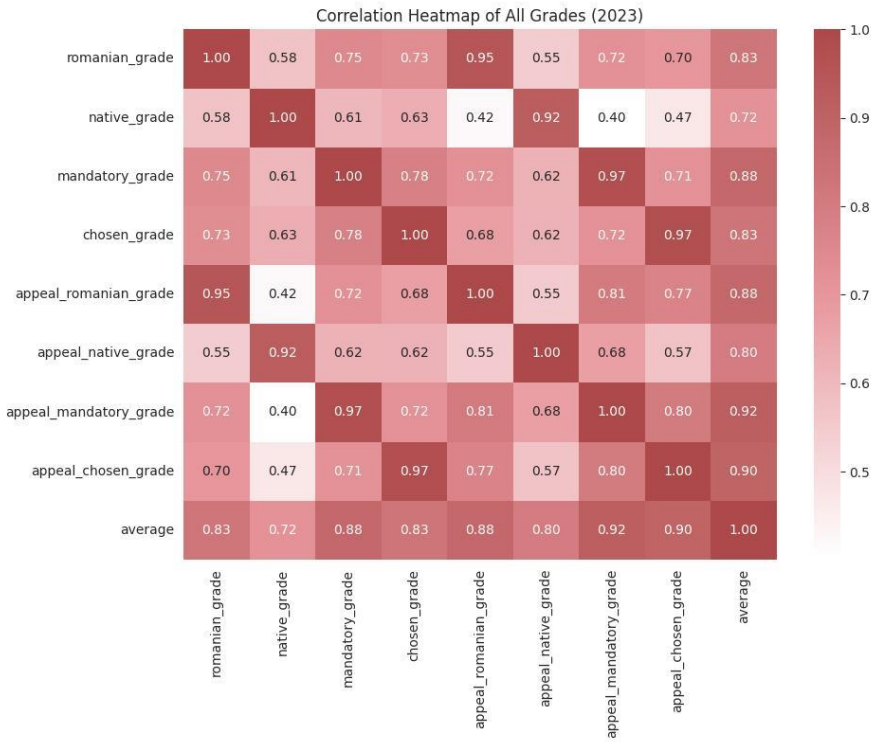


Figure 61: Correlation Heatmap of 2023 Exam Grades

As for the last year that we analyzed, the written and appealed grades still are consistently strongly correlated, and the final grade is most influenced yet again by the mandatory subject, just as how the mandatory appeal is the most influential amongst all appeal grades. This time, however, the native grade shows a stronger correlation to the final grade, compared to the 2022 results. All correlation coefficients are positive this time as well.

Once again, there is a strong linear relationship between written grades and appeal grades for the 2023 dataset as well, so we reject the null hypothesis and accept the alternative hypothesis.

T-Tests

T-tests are typically used for comparing the means of two numerical variables. They are commonly employed when you want to assess whether there is a significant difference between the means of two groups. We ran two t-tests, one for the final averages based on genders and the other for mandatory grades based on profiles.

Final Averages based on Genders

First, we examine the final averages for the two genders. It is important to mention that here we worked with only those average grades that are passing grades (greater than or equal 6). Here are the two hypotheses, the null and the alternative hypotheses (with a chosen significance level of 0.05).

H0: There is no significant difference in final averages between genders.

H1: The difference in final averages between genders is statistically significant.

```
pv = "{:.20f}".format(p_value)

if pv != 0:
  print(f"P-value: {pv}")
else:
  print(f"P-value: 0")
```

```
P-value: 0.00000000000000000000
```

Figure 62: p-value for T-test between Gender and Average

As a result, we can conclude that the p-value is less than the chosen significance level, so we have enough evidence to reject the null hypothesis, we accept the alternative hypothesis. The difference in final averages between genders is statistically significant. This t-statistic suggests a substantial difference between the means of the two gender groups. The negative sign indicates the direction of the difference (males have the lower mean).

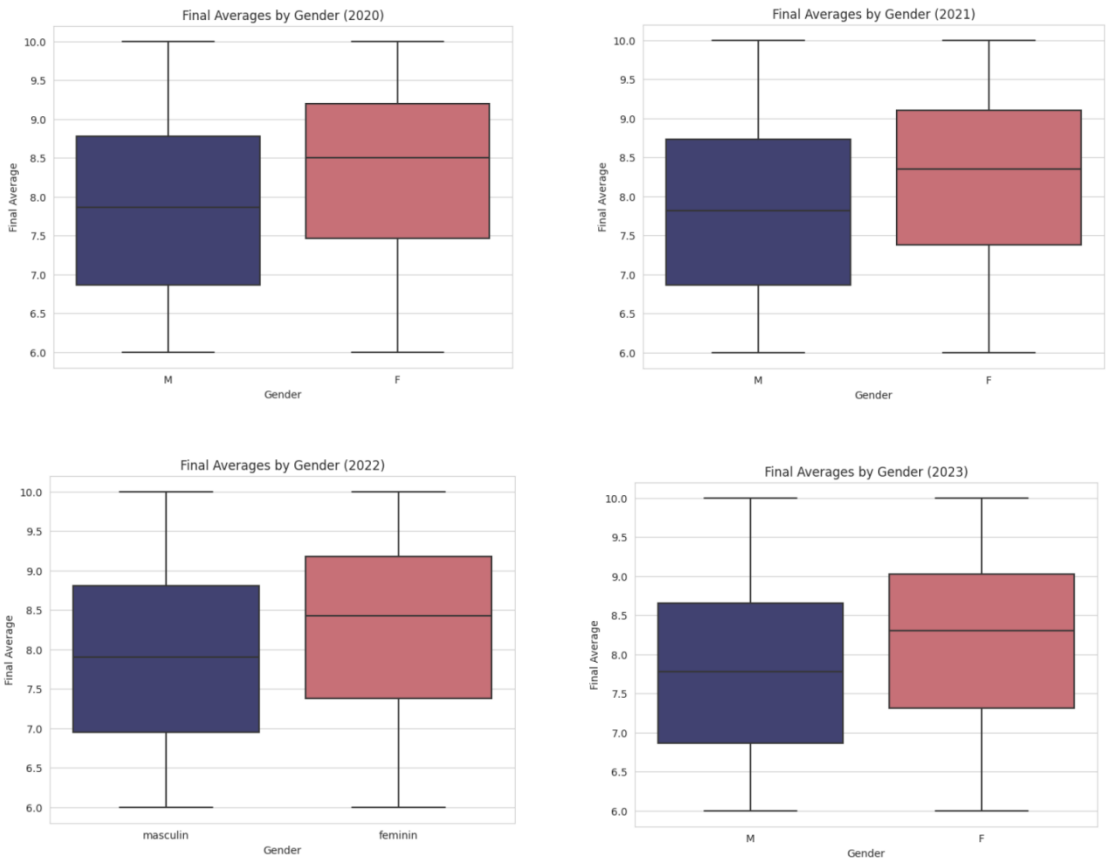


Figure 63: T-test boxplots (2020, 2021, 2022, 2023) for Gender and Final Average

The box represents the interquartile range (IQR), which is the range between the first quartile (Q1) and the third quartile (Q3). It contains the middle 50% of the data. The line inside the box represents the median, which is the middle value of the dataset when it is sorted in ascending order. The whiskers extend from the box to the minimum and maximum values within a certain range. The range is often determined by a multiplier (usually 1.5 times the IQR). Whiskers can provide information about the spread and potential outliers in the data. Individual data points beyond the whiskers are considered outliers. They are plotted as individual points and may suggest extreme values in the dataset.

The boxplot shows a higher median for females compared to males, it suggests that, on average, females have higher final averages. The whiskers for one gender are no longer than for the other, it indicates the same variability in final averages for both groups. There are no outliers.

Mandatory Grades based on Profiles (STEM and humanities)

Based on our previous correlation analysis, the mandatory grade is the one that influences most the final averages. That is why we chose this grade to see whether it is easier to perform well if you graduated from a humanities class or a stem profile.

Here we considered each grade value (from 1 to 10), and filtered for only the two above-mentioned profiles. These are the hypotheses (with a chosen significance level of 0.05):

H0: There is no significant difference in mandatory grades between profiles.

H1: The difference in mandatory grades between profiles is statistically significant.

```
pv = "{:.20f}".format(p_value)

if pv != 0:
  print(f"P-value: {pv}")
else:
  print(f"P-value: 0")
```

```
P-value: 0.00000000000000000000
```

Figure 63: p-value for T-test between Profile and Mandatory Grade (2020)

For 2020, the p-value is less than the chosen significance level, so we have enough evidence to reject the null hypothesis, we accept the alternative hypothesis. The difference in mandatory grades between profiles is statistically significant. This t-statistic suggests a substantial difference between the means of the two profile groups. The negative sign indicates the direction of the difference (real profilers have a lower mean).

Figure 64: T-test boxplot (2020) for Profile and Mandatory Grade

The boxplot shows a higher median for humanities profilers compared to stem profilers, it suggests that, on average, humanities profilers have higher final averages. The whiskers for the stem profile are longer than for the other, which indicates a higher variability in mandatory grades for this group. There are some outliers for the humanities grades.

```
pv = "{:.20f}".format(p_value)

if pv != 0:
  print(f"P-value: {pv}")
else:
  print(f"P-value: 0")
```

```
P-value: 0.11616766755168840553
```

Figure 65: p-value for T-test between Profile and Mandatory Grade (2021)

For 2021, the p-value is greater than the chosen significance level, so we do not have enough evidence to reject the null hypothesis, we accept it. The difference in mandatory grades between profiles is not statistically significant. This t-statistic suggests a non-significant difference between the means of the two profile groups.

Figure 66: T-test boxplot (2021) for Profile and Mandatory Grade

The boxplot shows a lower median for humanities profilers compared to stem profilers, it suggests that, on average, humanities profilers have lower final averages. The whiskers for the humanities profile are longer than for the other, which indicates a higher variability in mandatory grades for this group. There are some outliers for the stem grades.

```
pv = "{:.20f}".format(p_value)

if pv != 0:
  print(f"P-value: {pv}")
else:
  print(f"P-value: 0")
```

```
P-value: 0.00000000000000000000
```

Figure 67: p-value for T-test between Profile and Mandatory Grade (2022)

For 2022, The p-value is less than the chosen significance level, so we have enough evidence to reject the null hypothesis, we accept the alternative hypothesis. The difference in mandatory grades between profiles is statistically significant. This t-statistic suggests a substantial difference between the means of the two profile groups. The positive sign indicates the direction of the difference (real profilers have the higher mean).

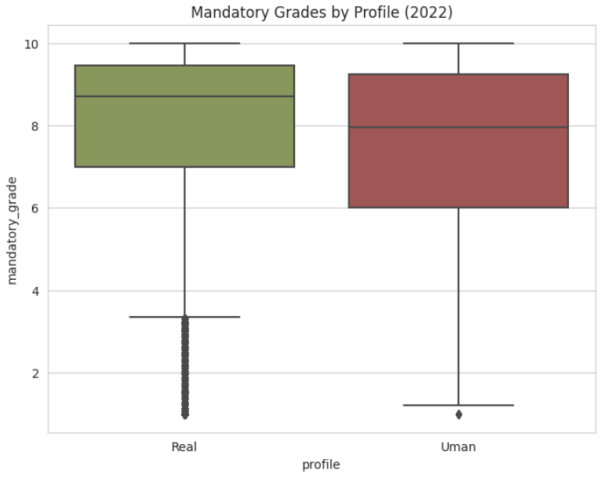


Figure 68: T-test boxplot (2022) for Profile and Mandatory Grade

The boxplot shows a lower median for humanities profilers compared to stem profilers, it suggests that, on average, humanities profilers have lower final averages. The whiskers for the humanities profile are longer than for the other, which indicates a higher variability in mandatory grades for this group. There are some outliers for both grades.

```
pv = "{:.20f}".format(p_value)

if pv != 0:
  print(f"P-value: {pv}")
else:
  print(f"P-value: 0")
```

```
P-value: 0.00000000003911131923
```

Figure 69: p-value for T-test between Profile and Mandatory Grade (2023)

For 2023, the p-value is less than the chosen significance level, so we have enough evidence to reject the null hypothesis, we accept the alternative hypothesis. The difference in mandatory grades between profiles is statistically significant. This t-statistic suggests a substantial difference between the means of the two profile groups. The positive sign indicates the direction of the difference (real profilers have the higher mean).

Figure 70: T-test boxplot (2023) for Profile and Mandatory Grade

The boxplot shows a lower median for humanities profilers compared to stem profilers, it suggests that, on average, humanities profilers have lower final averages. The whiskers for the humanities profile are longer than for the other, it indicates a higher variability in mandatory grades for this group. There are some outliers for both grades.

ANOVA

ANOVA (Analysis of Variance) is a statistical method used to analyze the differences among group means in a sample. It is very useful when comparing three or more groups. The main object of ANOVA is to determine whether there are any statistically significant differences between the means of the groups.

In the case of 2020's dataset:

- H0 (null hypothesis): There are no differences between average grades across specializations.
- H1 (alternative hypothesis): There are differences between average grades across specializations.

After finishing the analysis, we got F-statistic and P-value as a result. The F-statistic is a numerical value. This statistic is a ratio of the variance between groups to the variance within groups. A higher F-statistic suggests greater differences among group means. The F-statistic's value is 644.53, so there are greater differences among the groups.

```
ANOVA Result:
F_onewayResult(statistic=644.5269982844582, pvalue=0.0)
ANOVA is statistically significant. There are differences in average grades across specializations.
```

Figure 71: Result of the ANOVA about average grades across specializations in 2020

The other value, the P-value is the probability of observing such extreme results if there were no real differences among the groups. If the P-value is 0.0, that means that the observed differences are so extreme that they are highly unlikely to occur by chance alone. By the analysis, we got that the ANOVA is statistically significant. There are differences in average grades across specializations.

Based on the P-value, the null hypothesis is not true, there are differences between average grades across specializations.

We analyzed education form and average too.

- H0 (null hypothesis): There are no differences between average grades across education forms.
- H1 (alternative hypothesis): There are differences between average grades across education forms.

We got as a result, that the ANOVA is statistically significant, and there are differences in average grades across education forms. The F-statistic in this case is 3569.54, and the P-value is 0.0.

```
ANOVA Result:
F_onewayResult(statistic=3569.5411475861924, pvalue=0.0)
ANOVA is statistically significant. There are differences in average grades across education_form.
```

Figure 72: Result of ANOVA about average grades across education form in 2020

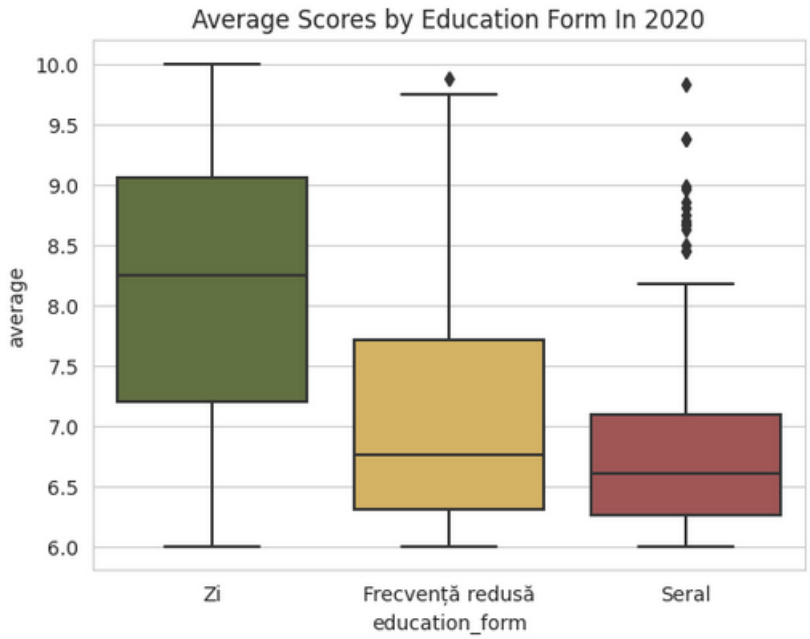


Figure 73: Average Scores by Education Form in 2020

We can see at the boxplot, that the 'Zi' education form has the highest average grade, and the 'Seral' education form has the lowest average grade. Also, the 'Zi' education form has the highest median value, between 8 and 8.5. The 'Seral' and 'Frecvență redusă' education forms have almost the same median, between 6.5 and 7.

In the 2021's dataset, at the first ANOVA test (between average grades across specialization), the null hypothesis is the same as we said before, there are no differences between average grades across specialization. The F-statistic is 485.15, so it suggests that there are differences, and the P-value is 0.0. Based on this we can reject the null hypothesis, there are differences.

```
F_onewayResult(statistic=485.15477305059795, pvalue=0.0)
ANOVA is statistically significant. There are differences in average grades across specializations.
```

Figure 74: Result of ANOVA about average grades across specialization in 2021

At the second ANOVA test (difference between average grades across education form), the value of F-statistic is 2460.27 and the P-value is 0.0, so we can reject the null hypothesis, there are differences between average across education form.

```
F_onewayResult(statistic=2460.2681859578624, pvalue=0.0)
ANOVA is statistically significant. There are differences in average grades across education_form.
```

Figure 75: Result of ANOVA about average grades across education form in 2021

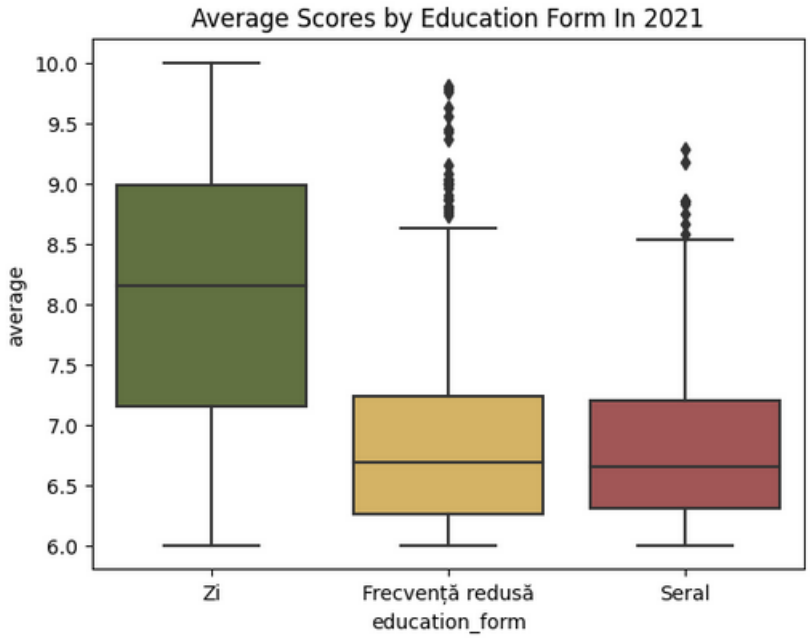


Figure 76: Average Scores by Education Form in 2021

This boxplot shows that the 'Zi' education form has the highest average again, but in this year, there is no as big difference between the average of 'Frecvență redusă' education form and 'Seral' education form as in the previous year. The values of median are almost the same as in the previous year, they are in the same interval.

We got the same result for the 2022's and the 2023's dataset. We rejected the null hypotheses, because the values of F-statistics were higher, and the P-values were 0.0, so there were differences between average grades across specialization and average grades across education form.

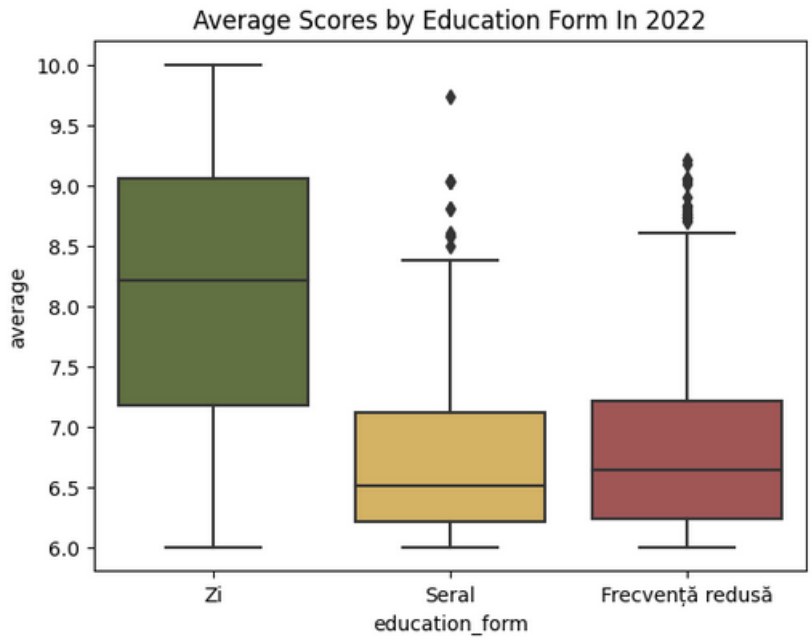


Figure 77: Average Scores by Education Form in 2022

The difference between 2022 and the previous years is that the average of 'Zi' education form is higher, it's a bit bigger than 9, the median of 'Seral' education form is smaller. It is between 6 and 6.5.

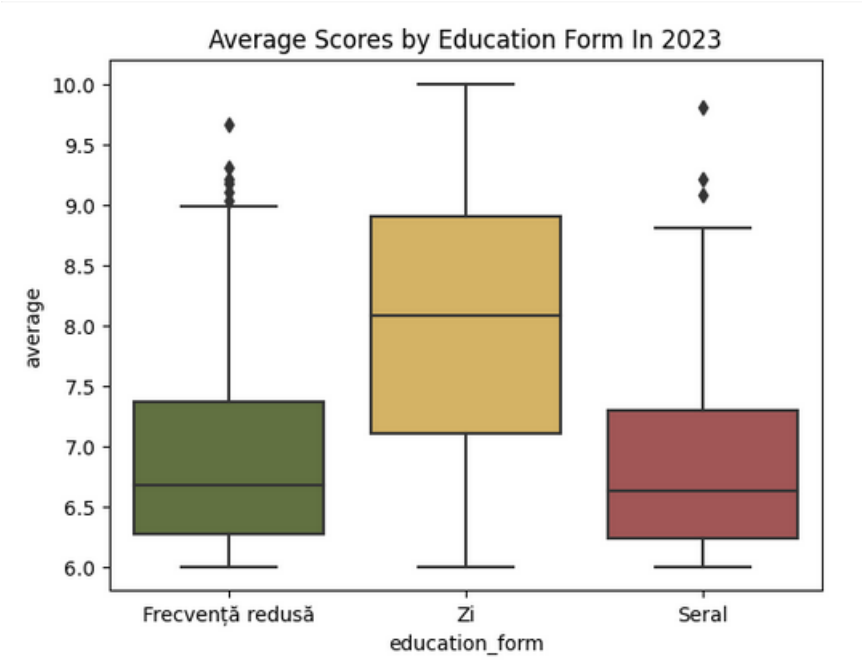


Figure 78: Average Scores by Education Form in 2023

In 2023 the highest averages of ‘Seral’ and ‘Frecvență redusă education forms are between 7 and 7.5.

Regression Analysis

We analyzed the regression between the final score and independent subject scores.

In statistics, regression refers to a set of techniques used to model the relationship between a dependent variable and one or more independent variables. The primary goal of regression analysis is to understand and quantify the relationship between variables.

For the regression we used the OLS model. The OLS model (Ordinary Least Squares) is a method to estimate the parameters of a linear regression model. The OLS model is the most appropriate for this regression analysis because it efficiently captures a linear relationship between independent grades and the final score.

First, we analyzed the grades without the native language and the average. In the result, the R square is 0.982, and this means that the independent variables in the model have a strong linear relationship with the final score. As we can see in the table below, the F-statistic is high and the P-value is 0.0, so, we can reject the null hypothesis, which states that there is no linear relationship between final score and independent variables.

It’s also visible, that the const (Intercept) is 0.13, so when all the grades are 0, the final scores value is 0. If the Romanian grade increases by one unit, then the final score increases by 0.0.3215 units. If the mandatory grade increases by one unit, then the final score increases by 0.3352 units. If the chosen grade increases by one unit, then the final score increases by 0.3315 units.

OLS Regression Results

Dep. Variable:	average	R-squared:	0.986
Model:	OLS	Adj. R-squared:	0.986
Method:	Least Squares	F-statistic:	2.273e+06
Date:	Sat, 13 Jan 2024	Prob (F-statistic):	0.00
Time:	11:42:27	Log-Likelihood:	58468.
No. Observations:	95475	AIC:	-1.169e+05
Df Residuals:	95471	BIC:	-1.169e+05
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.1300	0.003	39.877	0.000	0.124	0.136
romanian_grade	0.3215	0.000	828.662	0.000	0.321	0.322
mandatory_grade	0.3352	0.000	941.904	0.000	0.334	0.336
chosen_grade	0.3315	0.000	793.681	0.000	0.331	0.332

Figure 79: OLS Regression result without native grade 2020

We also analyze cases when there were native grades too. It is important to mention we only analyzed grades above 5, and the averages above 6. In this result, the R square is 0.992, so that the independent variables in the model have a strong linear relationship with the final score. We can reject the null hypothesis again because the P-value is 0.0. The const is 0.08, so when all the grades are 0, the average is also 0. If there is a one-unit increase in the Romanian grade, the final score increases by 0.2466 units. If the mandatory grade increases by a unit, the final score increases by 0.2502 units. If the chosen grade increases by a unit, then the final score increases by 0.2491 units. And if the native grade increases by a unit, then the final score increases by 0.2458 units.

OLS Regression Results

Dep. Variable:	average	R-squared:	0.992
Model:	OLS	Adj. R-squared:	0.992
Method:	Least Squares	F-statistic:	1.536e+05
Date:	Sat, 13 Jan 2024	Prob (F-statistic):	0.00
Time:	11:46:00	Log-Likelihood:	4878.2
No. Observations:	4783	AIC:	-9746.
Df Residuals:	4778	BIC:	-9714.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0850	0.011	7.806	0.000	0.064	0.106
romanian_grade	0.2466	0.001	234.938	0.000	0.245	0.249
mandatory_grade	0.2502	0.001	235.008	0.000	0.248	0.252
chosen_grade	0.2491	0.001	190.188	0.000	0.247	0.252
native_grade	0.2458	0.001	193.901	0.000	0.243	0.248

Figure 80: OLS Regression result with native grade 2020

We got the same result in 2021 from the point of view of hypothesis, the independent variables in the model have a strong linear relationship with the final score. We can see that the mandatory grade has the biggest effect on the final score, if this grade increases by a unit,

the final score increases by 0.3391 units. After this comes the chosen grade, if this increases by a nit, the final score increases by 0.3353 units, and if the Romanian grade increases by a unit, then the final score increases by 0.3109 units.

OLS Regression Results

Dep. Variable:	average	R-squared:	0.983
Model:	OLS	Adj. R-squared:	0.983
Method:	Least Squares	F-statistic:	1.743e+06
Date:	Sat, 13 Jan 2024	Prob (F-statistic):	0.00
Time:	12:16:44	Log-Likelihood:	48203.
No. Observations:	88590	AIC:	-9.640e+04
Df Residuals:	88586	BIC:	-9.636e+04
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.1579	0.004	42.806	0.000	0.151	0.165
romanian_grade	0.3109	0.000	727.260	0.000	0.310	0.312
mandatory_grade	0.3391	0.000	845.343	0.000	0.338	0.340
chosen_grade	0.3353	0.000	712.154	0.000	0.334	0.336

Figure 81: OLS Regression result without native grade 2021

If we analyze the grades where are native grades too, the final score increases by 0.2391 units if the Romanian grade increases by a unit. If the mandatory grade increases by a unit, the final score increases by 0.2515 units. If the chosen score increases by a unit, then a final score increases by 0.2528 units, and if the native grade increases by a unit, then the final score increases by 0.2467 units.

OLS Regression Results

Dep. Variable:	average	R-squared:	0.990
Model:	OLS	Adj. R-squared:	0.990
Method:	Least Squares	F-statistic:	9.896e+04
Date:	Sat, 13 Jan 2024	Prob (F-statistic):	0.00
Time:	12:17:14	Log-Likelihood:	3887.0
No. Observations:	4129	AIC:	-7764.
Df Residuals:	4124	BIC:	-7732.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0912	0.014	6.719	0.000	0.065	0.118
romanian_grade	0.2391	0.001	207.469	0.000	0.237	0.241
mandatory_grade	0.2515	0.001	190.358	0.000	0.249	0.254
chosen_grade	0.2528	0.002	158.521	0.000	0.250	0.256
native_grade	0.2467	0.001	165.085	0.000	0.244	0.250

Figure 82: OLS Regression result with native grade 2021

Results for 2022 and 2023:

OLS Regression Results

Dep. Variable:	average	R-squared:	0.987
Model:	OLS	Adj. R-squared:	0.987
Method:	Least Squares	F-statistic:	2.258e+06
Date:	Sat, 13 Jan 2024	Prob (F-statistic):	0.00
Time:	12:46:05	Log-Likelihood:	58440.
No. Observations:	91331	AIC:	-1.169e+05
Df Residuals:	91327	BIC:	-1.168e+05
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.1052	0.003	32.011	0.000	0.099	0.112
romanian_grade	0.3196	0.000	828.659	0.000	0.319	0.320
mandatory_grade	0.3361	0.000	891.982	0.000	0.335	0.337
chosen_grade	0.3343	0.000	757.795	0.000	0.333	0.335

Figure 83: OLS Regression result without native grade 2022

OLS Regression Results

Dep. Variable:	average	R-squared:	0.992
Model:	OLS	Adj. R-squared:	0.992
Method:	Least Squares	F-statistic:	1.436e+05
Date:	Sat, 13 Jan 2024	Prob (F-statistic):	0.00
Time:	12:46:51	Log-Likelihood:	4711.6
No. Observations:	4391	AIC:	-9413.
Df Residuals:	4386	BIC:	-9381.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0728	0.011	6.507	0.000	0.051	0.095
romanian_grade	0.2413	0.001	223.880	0.000	0.239	0.243
mandatory_grade	0.2513	0.001	221.101	0.000	0.249	0.254
chosen_grade	0.2517	0.001	183.012	0.000	0.249	0.254
native_grade	0.2474	0.001	189.569	0.000	0.245	0.250

Figure 84: OLS Regression result with native grade 2022

OLS Regression Results

Dep. Variable:	average	R-squared:	0.983
Model:	OLS	Adj. R-squared:	0.983
Method:	Least Squares	F-statistic:	1.843e+06
Date:	Sat, 13 Jan 2024	Prob (F-statistic):	0.00
Time:	13:29:14	Log-Likelihood:	52883.
No. Observations:	94004	AIC:	-1.058e+05
Df Residuals:	94000	BIC:	-1.057e+05
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.1469	0.003	42.189	0.000	0.140	0.154
romanian_grade	0.3158	0.000	754.366	0.000	0.315	0.317
mandatory_grade	0.3370	0.000	838.026	0.000	0.336	0.338
chosen_grade	0.3332	0.000	731.681	0.000	0.332	0.334

Figure 85: OLS Regression result without native grade 2023

OLS Regression Results

Dep. Variable:	average	R-squared:	0.992
Model:	OLS	Adj. R-squared:	0.992
Method:	Least Squares	F-statistic:	1.330e+05
Date:	Sat, 13 Jan 2024	Prob (F-statistic):	0.00
Time:	13:29:43	Log-Likelihood:	4782.2
No. Observations:	4448	AIC:	-9554.
Df Residuals:	4443	BIC:	-9522.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0995	0.011	8.970	0.000	0.078	0.121
romanian_grade	0.2416	0.001	222.271	0.000	0.240	0.244
mandatory_grade	0.2506	0.001	218.196	0.000	0.248	0.253
chosen_grade	0.2511	0.001	192.763	0.000	0.249	0.254
native_grade	0.2455	0.001	208.509	0.000	0.243	0.248

Figure 86: OLS Regression result with native grade 2023

Chi-Square Test

Besides using numerical variables in models to create inferences, we considered it would be useful and insightful to also compare some of the categorical variables as well. Since our main focus is to examine what factors influence the outcome of the final status variable, we experimented mainly with that. We modelled three chi-square tests, looking for associations between gender and status, profile and status, and also candidate environment and status.

So, a chi-square test can be used to determine if there is a significant association between two categorical variables. The expected frequencies represent what you would expect to observe in each cell if there was no association between the variables. The chi-square statistic measures the discrepancy between the observed and expected frequencies. A larger chi-square value indicates a larger difference. There are two hypotheses that we can consider for each case, the null and the alternative hypotheses. In addition, we calculated Cramér' V indicators for each case and year. Cramér's V is a measure of association for nominal (categorical) variables. It's based on the chi-square statistic and provides a normalized measure of association, ranging from 0 to 1. Interpretation of Cramér's V: 0: no association between the variables; 1: perfect association between the variables. The higher the value of Cramér's V, the stronger the association between the variables.

We tested the models for each year, the results came out almost identical. That is why the interpretation will be written only once because it applies to each case.

Association between Gender and Final Status

First, these are the two hypotheses from which we will have to choose based on the significance level (p = 0.05):

H0: There is no association between Status and Gender.

H1: There is an association between Status and Gender.

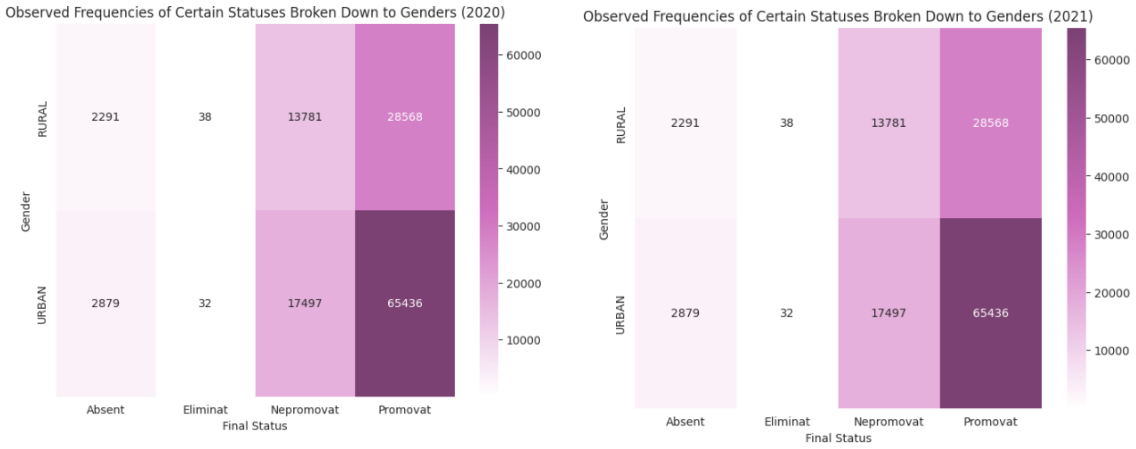
```
pv = "{:.20f}".format(p)

if pv != 0:
  print(f"P-value: {pv}")
else:
  print(f"P-value: 0")
```

```
P-value: 0.00000000000000000000
```

Figure 87: p-value for Association between Gender and Status

Based on what we can see in Figure x, we can say that the result is that the p-value is less than the chosen significance level, so we have enough evidence to reject the null hypothesis, we accept the alternative hypothesis. There is a statistically significant association between Status and Gender.



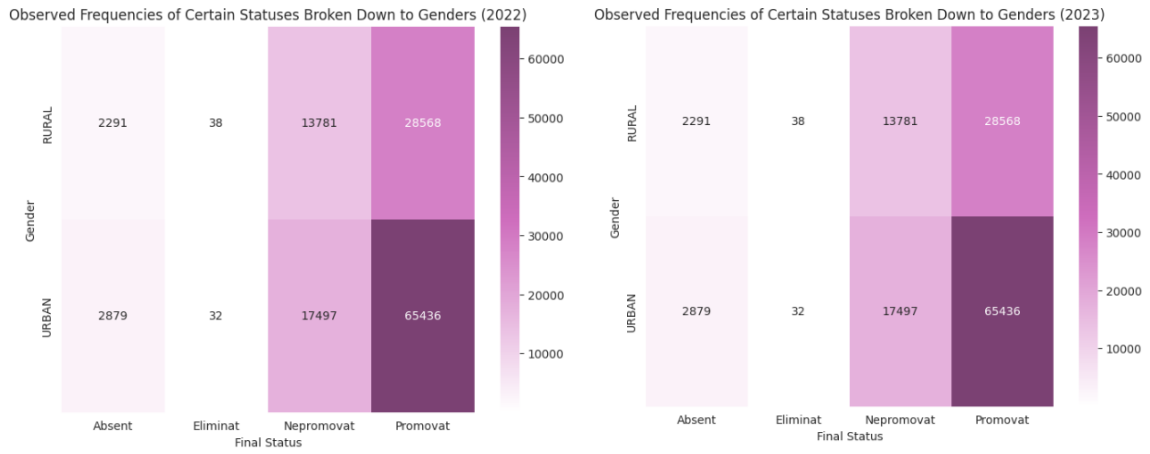


Figure 88: Chi-Square Contingency Table Heatmap (2020, 2021, 2022, 2023) for Gender and Status

We can see that most candidates passed the exam, and there are more girls in that group than boys. Based on this plot, we can speculate that male candidates perform more poorly than females, but the chi-square test does not say this exactly, only that there is an association between gender and final status.

As we mentioned before, we calculated the Cramér's V measure as well (Figure x).

```
Cramer's V (2020): 0.12829022693122932

Cramer's V (2021): 0.11053527480667767

Cramer's V (2022): 0.09312394894925773

Cramer's V (2023): 0.10255867438713534
```

Figure 89: Cramér's V measure for Chi-Square Test between Gender and Status

In this case, we can see that the association between the two variables in pretty weak, existent, but weak.

Association between Profile and Final Status

These are the two hypotheses from which we will have to choose based on the significance level (p = 0.05):

H0: There is no association between Status and Profile.

H1: There is an association between Status and Profile.

```
pv = "{:.20f}".format(p)

if pv != 0:
  print(f"P-value: {pv}")
else:
  print(f"P-value: 0")
```

```
P-value: 0.00000000000000000000
```

Figure 90: p-value for Association between Profile and Status

Based on what we can see in Figure x, we can say that the result is that the p-value is less than the chosen significance level, so we have enough evidence to reject the null hypothesis, we accept the alternative hypothesis. There is a statistically significant association between Status and Profile.

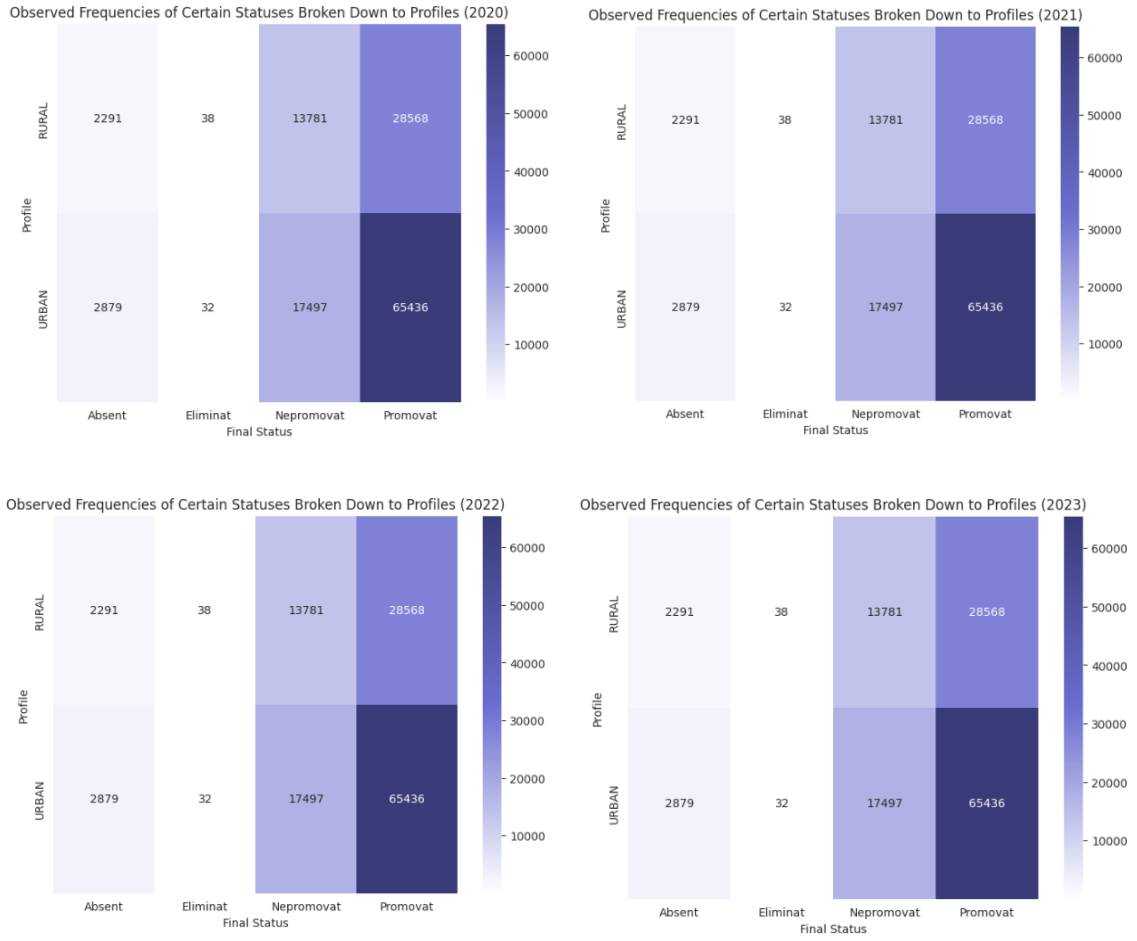


Figure 91: Chi-Square Contingency Table Heatmap (2020, 2021, 2022, 2023) for Profile and Status

We can see that most candidates that passed the exam are from the profiles 'Real' and 'Uman'. Many did not pass either but there are more who have not passed from the groups 'Servicii' and 'Tehnic'. This could be because less smart students choose these profiles, but the chi-square test does not prove it.

Here is the calculated Cramér's V measure (Figure x).

```
Cramer's V (2020): 0.2744067130923297

Cramer's V (2021): 0.26287733647431977

Cramer's V (2022): 0.25631768888807394

Cramer's V (2023): 0.25306861699296623
```

Figure 92: Cramér's V measure for Chi-Square Test between Profile and Status

Here, we still have a weak association, but it is slightly stronger than in the previous case.

Association between Candidate Environment and Final Status

These are the two hypotheses from which we will have to choose based on the significance level (p = 0.05):

H0: There is no association between Status and Profile.

H1: There is an association between Status and Profile.

```
pv = "{:.20f}".format(p)

if pv != 0:
  print(f"P-value: {pv}")
else:
  print(f"P-value: 0")
```

```
P-value: 0.00000000000000000000
```

Figure 93: p-value for Association between Candidate Environment and Status

Based on what we can see in Figure x, we can say that the result is that the p-value is less than the chosen significance level, so we have enough evidence to reject the null hypothesis, we accept the alternative hypothesis. There is a statistically significant association between Status and Candidate Environment.

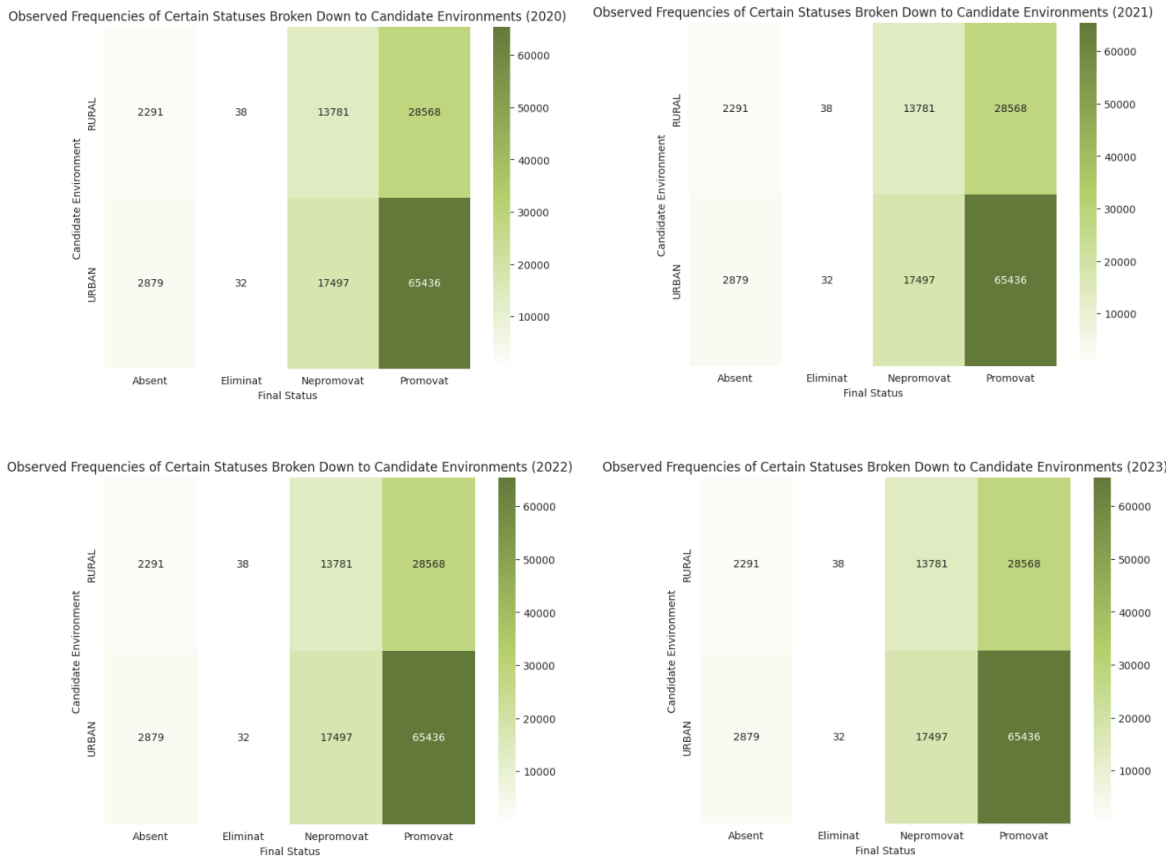


Figure 94: Chi-Square Contingency Table Heatmap (2020, 2021, 2022, 2023) for Candidate Environment and Status

We can see that there are far more students from the urban area hence there are more who passed and also more who did not pass. But the ratios are very different, we can say that candidates from the rural areas perform worse, but the chi-square test does not prove this, only that there is an association.

Here is the calculated Cramér's V measure (Figure x).

```
Cramer's V (2020): 0.1599125505239331

Cramer's V (2021): 0.15139181827703865
```

```
Cramer's V (2022): 0.1422953590231465

Cramer's V (2023): 0.12994766843334327
```

Figure 95: Cramér's V measure for Chi-Square Test between Candidate Environment and Status

There is definitely a weak association between the two examined variables.

Conclusions

Descriptive Statistics

After completing data cleaning, we performed Exploratory Data Analysis (EDA) on four datasets, exploring central tendency and variability measures, frequency distributions to help detect typical scores and distribution patterns in each feature across the datasets.

The candidates between 2020 and 2023 are almost equally distributed between the two genders, theory-based teaching with school attendance by day is most common amongst institutions, most candidates are from urban environments. The most frequented specialization is Mathematics and Informatics and STEM ('Real') profiles are popular. There is an overwhelming majority of Hungarian-native students who take the native exams each year. When it comes to the mandatory and chosen subject, candidates frequent Mathematics and Biology respectively. The most popular foreign language to take is English.

Next, we covered the influence of important categorical features on the four written exam grades and the final grade (average). We looked at six important factors to consider when testing for inferences: gender, profile, specialization, form of study, environment of candidate and native language.

In terms of all written subject grades more female students achieved higher scores, and more males the passing grade for all the years we analyzed.

It is true for all years that the best performance was achieved by students who go to school by day and are from an urban environment. The highest-ranking mother tongue overall goes to the German-native students in all four years. The highest scorer specialization in 2020 & 2021 is Mathematics and Informatics in terms of median, and whilst the Natural Sciences specialization ranked last out of the five most popular specializations in 2020 & 2021, in 2022 they achieved the highest median final score. In 2023 the highest median was achieved by Social Sciences.

Overall, it is safe to say that gender, profile, specialization, form of study, environment of candidate and native language greatly influence the Romanian, native language, mandatory subject and chosen subject grades, thus having an influence on the final grade (average) as well.

With a final overlook on all four years' data, with higher emphasis on candidate's gender and profile, we concluded that the Romanian average grade experienced a slight linear decrease

between 2020 and 2023, the same with the native and chosen grades after 2021. In case of the mandatory grade and the overall final grade there is no linear trend.

Up until the recent year (2023) there has been a linear decline of attending candidates from both genders, having more male and ‘Real’ profiled candidates. In the most recent round of examinations, the number of candidates increased, although only slightly.

Inferences

Correlation Analysis

There is a high correlation between the four written exams and their appeal grades. Most likely, there were no significant increases or decreases in the scores after appealing. The appeal grade seems to follow the grade closely, showing a strong linear relationship. When it comes to the final score, the most influential written exam is usually the mandatory subject, as is their appeal.

T-Test

Looking at the T-test between Gender and Final Average grade, the evidence from both statistical tests and visual representations strongly supports the assertion that there is a significant and consistent difference in final averages between genders, with females generally achieving higher grades than males.

Next, moving on to the T-test statistics between Mandatory Grade and Profile, in summary, there is a consistent and statistically significant difference in mandatory grades between STEM and humanities profiles across the years, with STEM profilers generally having lower grades. Visual representations through boxplots further elucidate these distinctions, encompassing variations in medians, whisker lengths, and the presence of outliers. These findings contribute valuable insights into the academic performance disparities associated with different academic profiles. The only year is 2021 when no statistically significant difference was found in mandatory grades between STEM and humanities profiles.

ANOVA

Our analysis of data from 2020 to 2023 shows clear differences in average grades across specializations and education forms. The F-statistic values were consistently high, suggesting substantial variations among group means.

The boxplots visually confirmed these differences. 'Zi' education form had the highest average grades, while 'Seral' education form generally had lower averages.

In conclusion, there are significant and consistent variations in average across specializations and education forms throughout these years. Possible reasons for these differences could be how subjects are taught, the curriculum, or the types of students. Exploring these factors further could help us understand these trends better.

Regression

Our regression analysis using the Ordinary Least Squares (OLS) model revealed a strong linear relationship between subject scores and the final score. The mandatory grade stood out as the most influential factor, and the native the least influential factor. We can say that there is no significant difference between years.

Chi-square tests

All three chi-square tests reveal statistically significant associations between the respective categorical variables (Gender, Profile, Candidate Environment) and the final exam status. The visual representations (heatmaps) provide insights into the distribution of candidates across different categories of the variables. Cramér’s V measures indicate that the associations, while statistically significant, are generally weak. The strength of association is slightly stronger for Profile and Final Status compared to the other variables. These findings collectively contribute to a better understanding of how certain categorical variables may be associated with the outcome of the final exam status, allowing for further exploration and potential interventions in educational contexts.

SWOT

STRENGTH	WEAKNESS
• Use of multiple inferential statistics models, and various visualizations, useful for students to see how they compare to others	• Very long
OPPORTUNITY	**THREAT**
• Analyze further data in the future every year when the results come out, educate people, come up with even more models and tests	• Others do the same

References

[1] Relalitatea.net, ISTORIA BACALAUREATULUI Cine a introdus “bacul” şi prin ce schimbări extreme a trecut în timp, 2015: https://www.realitatea.net/stiri/ultimele-stiri/istoria-bacalaureatului-cine-a-introdus-bacul-si-prin-ce-schimbari-extreme-a-trecut-in-timp_5dcc9201406af85273d0b0b9

[2] BAC+: https://www.bacplus.ro/

[3] data.gov.ro: https://data.gov.ro/dataset

[4] edu.ro: https://edu.ro/