DOI: 10.36558/rsc.v12i3.7919

Data Analysis to Identify the Impact of the COVID-19 Pandemic on ENEM in 3 States of Northeast Brazil

Nelson Weber Neto
Universidade Federal do Maranhão (UFMA)
nelsonweberneto@gmail.com

Raimundo C. Soares
Universidade Federal do Maranhão (UFMA)

Luciano R. Coutinho
Universidade Federal do Maranhão (UFMA)

Ariel S. Teles
Universidade Federal do Maranhão (UFMA)
Instituto Federal do Maranhão (IFMA)
ariel.teles@ifma.edu.br

ABSTRACT

In Brazil, the main exam for evaluating the educational performance of basic education is the *Exame Nacional do Ensino Médio* (ENEM), which is also used for the admission of students to higher education. In 2020, with the arrival of the COVID-19 virus, basic education institutions needed to change their educational model with face-to-face teaching to the use of remote teaching methodologies. This work aims to identify the main impacts caused by the pandemic on ENEM in the states of Ceará, Maranhão and Piauí. Therefore, an analysis of ENEM data was carried out in 5 different years, using the 2017, 2018, 2019, 2020 and 2021 editions of the exam, in order to understand the pre-pandemic years and the first two years of the pandemic. The analysis aimed at making a comparison of the years and discover the impacts caused in the editions that occurred during the pandemic period. The results show that students with higher incomes perform better, private schools have a higher average performance than other types of schools, and the number of present and absent in the exam was lower in the years during the pandemic.

Keywords

Data Analysis, COVID-19, Pandemic, Education, Brazil, ENEM.

1. INTRODUCTION

On March 11, 2020, the World Health Organization (WHO) declared that the COVID-19 epidemic had become an international pandemic. Studies on the COVID-19 virus showed that the virus was highly contagious and can cause serious harm to humans [8]. According to Jill Robbin [23], the high social, educational and economic costs of school closures affect people in all communities, but disproportionately hit vulnerable groups and their families.

With the arrival of the COVID-19 pandemic, educational institutions began to create methods to avoid the education to be compromised, and the teaching process could continue [10]. The COVID-19 pandemic has directly affected the lives of students and teachers in different ways, regardless of educational level or course. The closing of schools has changed the structure of learning and teaching, as well as teaching and assessment methods [2]. By increasing the ability to teach remotely, schools have had to adapt to using remote and distance learning methods, so requiring students and teachers to adapt and learn about these methods quickly [11].

In Brazil, there are several education evaluation programs, from early childhood education to higher education. Programs such as the Basic Education Development Index (in Portuguese language, *Índice de Desenvolvimento da Educação Básica* - IDEB), Basic Education Assessment System (in Portuguese language, *Sistema de avaliação da Educação Básica* - SAEB), School Census, National High School Exam (in Portuguese language, *Exame Nacional do Ensino Médio* - ENEM), National Student Performance Exam (in Portuguese language, *Exame Nacional de Desempenho dos Estudantes* - ENADE), among others, serve as support for states and cities in the decision-making process. ENEM is an exam created and applied by the Ministry of Education of Brazil to measure the performance of students who have completed high school. It was originally conceived as a final exam to evaluation students at the end of basic education. However, in 2009, the exam was intended to replace the entrance exam of federal universities, and today it is already used as an entrance exam in several higher education institutions.

This paper is an extended version of our initial study published in [19], which presented a process of exploratory analysis of ENEM microdata in the years 2019 and 2020 in Maranhão, Ceará and Piauí, aiming to identify whether the pandemic impacted ENEM. The states of Ceará, Maranhão

and Piauí are located in the northeast region of Brazil, directly affecting the provision of quality education [20]. The studied states were also selected due to the synergy between them, which allowed the creation of the *Escola Regional de Computação Ceará, Maranhão e Piauí* (ERCEMAPI). This is an extended version of our study because we have deepened our article in the following points:

1. Inclusion of a background section covering topics related to our study;
2. Addition of 3 editions of the exam, thus analyzing 3 years before the pandemic (2017, 2018 and 2019) and 2 years after the pandemic (2020 and 2021)
3. Comparative and more in-depth analysis for each state.

The remaining of this paper is organized as follows. Section 2 provides a foundation on the topic addressed by this research. Section 3 presents the related work. Section 4 explains how the analysis was conducted, while Section 5 presents the results found for each state analyzed. Section 6 discusses the results found. Finally, Section 7 brings the final remarks.

2. BACKGROUND

2.1 Educational Data Mining

Educational data mining focuses on the development of methods that seek to extract insights using data collected in educational environments. Its main objective is to understand education, and then developing methods that help academic trajectory. According to Peña Ayala [21], the results of the application of data mining can provide immeasurable support for the decision-making process of school management.

The use of different data mining techniques can be seen as a potential basis for educational change and have a significant impact if seen as a tool that can help educational institutions at all levels. Certainly, data mining techniques can provide education policymakers with data-driven models that may serve to support their goals of improving the efficiency and quality of education. According to Sainath Aher and Louis Lobo [1], data mining can be used in educational systems to predict student dropout rates, relationship between students' college entrance exam scores and their success, predict student academic performance, find evidence about influencing factors in grades, and improving the management of education-oriented policies.

The application of data mining techniques to extract knowledge from educational systems has been treated by Rosa Arruabarrena et al. [3] as a formative evaluation technique with the purpose of continuously improving the evolution process of educational systems. All knowledge discovered through these techniques can be used by people involved, from students, teachers, educational management, and even those responsible for creating educational policies. Therefore, the knowledge generated can be seen differently by each point of view.

2.2 Open Data on Brazilian Basic Education

According to [14], ENEM is an assessment created and applied by the Ministry of Education of Brazil to measure the performance of students who have completed high school. The ENEM was originally conceived as a final exam to assess students at the end of basic education. However, in recent years, it has been revised as an admission test for public and private universities. Currently, the ENEM grade is composed through knowledge in different areas, namely: Human Sciences and their Technologies (History, Geography, Philosophy and Sociology); Natural Sciences and their Technologies (Chemistry, Physics and Biology); Languages, Codes and their Technologies (Portuguese Language, Literature, Foreign Language – English or Spanish, Arts, Physical Education and Information and Communication Technologies); Mathematics and its Technologies. However, not only the data related to the tests can be interesting to be analyzed. During the registration, students answer a questionnaire with 25 questions containing information about their socioeconomic level, family, education and work, in addition to the participants' data, including age, gender, year of completion of high school, type of school (public or private), among other information that can generate knowledge.

2.3 COVID-19 Pandemic

In December 2019, an outbreak of pneumonia of unknown cause broke out in Wuhan, Hubei Province, China. A few weeks later, in January 2020, the WHO declared the rapidly spreading COVID-19 outbreak a Public Health Emergency of International Concern. According to Ciotti et al. [7], the coronavirus pandemic spread rapidly and affected about 213 regions and countries around the world. The coronavirus disease 2019 (COVID-19) pandemic hit Latin America later than other continents. On February 25, 2020, Brazil registered its first case [26].

According to the Federal Government's COVID-19 Epidemiological Bulletin of Brazil, since the confirmation of the first case in February 2020 until the beginning of November 2022, Brazil has recorded about 34.9 million cases and 688.735 thousand deaths, one of the highest in absolute numbers of deaths [6]. The United Nations Educational, Scientific and Cultural Organization (UNESCO) has recognized that the outbreak of the coronavirus pandemic has affected the world's education systems [23]. According to Robbins [23], the high social, educational and economic costs of school closures affect people in all communities, but disproportionately affect vulnerable groups and their families. The Centers for Disease Control and Prevention (CDC) also expressed concern about the impact of school closures [30].

3. RELATED WORK

Data analysis has been used to study Brazilian basic education [28]. In Brazil, there are several studies that use data analysis techniques in the educational context, by analyzing the entire national territory. Alessandro Aparecido Barcellos et al. [5] used data mining techniques in the ENEM microdata of the year 2018 to correlate participants' financial data with their performance in the exam. Augusto Simon and Sílvio Cazella [27] used decision tree techniques to generate a predictive model with the results of ENEM data from the year 2015, and improve the average performance indicator of the participants in the knowledge area of Natural Sciences and its technologies. Eduardo Yoshinori Nakazone and Luis Marcelo Bortolotti [18] analyzed the ENEM microdata from the years 2015 to 2019 to understand the differences between tests over the years. The study carried out by Moraes et al. [9] comprised the factors and relationships of

the variables presented in the ENEM microdata 2017, then comparing with data during the pandemic and evaluating the performance of students in the Mathematics test and its technologies.

Several studies around the world have focused on the impact of the pandemic on education. For example, to emphasize the importance of the topic, the editorial [22] presented a proposal by the *United Nations International Children's Emergency Fund* (UNICEF) in response to the COVID-19 pandemic. Pereira Junior et al. [16] presented an analysis of student profiles during the pandemic. In the distance education modality, students with better financial infrastructure, that is, those who had access to the Internet and higher quality computers, has dealt better with this modality of education.

Our first effort was in [29], in which an analysis was made of ENEM data from the years 2019 and 2020 covering the entire Brazilian national territory. In this study, we used correlation analysis to understand which characteristics were most correlated with the participants' mean score. From the correlation, we noticed that the social class was the characteristic with the highest correlation with the grade. In addition, there was a disparity in grades between public and private schools.

We performed a more detailed analysis in the states of Ceará, Maranhão and Piauí [19]. We used data from 2019 and 2020, which were the data available by INEP [13]. In this study, we have expanded that first analysis, as explained previously.

There are several works using educational data analysis to understand educational issues [21, 28]. However, to the best of our knowledge, there is no article published in the scientific literature that has conducted a study to compare 5-year data of the states of Ceará, Maranhão and Piauí, in order to understand the impact of the COVID-19 pandemic on ENEM. This work therefore used 3 years before the pandemic and 2 years during the pandemic to improve the understanding of the educational situation of states during the pandemic period.

4. METHODOLOGY

This study aims to analyze ENEM data from the states of Ceará, Maranhão and Piauí for 2017, 2018, 2019, 2020 and 2021. ENEM microdata provides educational data for all states and territories in Brazil, by administrative dependency (Municipal, State, Federal and Private), candidates' average grades by year of academic completion, and other information about participants and their educational institutions.

For this study, the methodology Cross-Industry Standard Process for Data Mining (CRISP-DM) [4, 25] was used to measure the quality of results in a data mining project. The project life cycle in the CRISP-DM methodology is divided into six phases: business understanding, data understanding, data preparation, modeling, evaluation and distribution [4, 25]. CRISP-DM allows an exploratory analysis to be carried out following only the first three phases, based on this principle. We used this methodology in this study. Also, in this work, we used the Python programming language, Pandas, NumPy, and Matplotlib libraries, with Google Colaboratory as the development environment.

4.1 Business and Data Understanding

By following the CRISP-DM methodology, the first phase is to understand the business, in which the available resources, problems and objectives were identified. For this study, an analysis of the changes and impacts that occurred in the ENEM in the years 2017, 2018, 2019, 2020 and 2021 was performed.

The next phase of the methodology consists of understanding the data. For this purpose, additional documents (i.e., ENEM Reference Matrix, Technical Notes, Data Dictionary) in the ENEM database were read in order to understand each column of the dataset. In this study, we used five datasets: the ENEM microdata of the years 2017, 2018, 2019, 2020 and 2021, both provided by INEP in its open data portfolio. Each row corresponds to a participant, and the columns present various information about the participant, such as the state where he/she lives, year of high school completion, gender, race, age group, grade in each area of knowledge, socioeconomic data, among others. The ENEM microdata from 2020 has changed when compared to previous years due to the General Data Protection Law (in Portuguese language, *Lei Geral de Proteção de Dados* - LGPD). Information such as school identification code, requests for specialized care, among other data were removed from 2020.

4.2 Data Preparation

The preparation phase includes selecting, cleaning, and transforming the data to be used in the project. At this stage, it is decided whether to use the entire dataset or a subset to include certain columns and rows [24]. The first process is to create a new column containing the average grades achieved by participants.

Then, an average is calculated based on the five knowledge area grades, as the dataset only had the grade averages for each of the knowledge areas. After the column with the grade averages of all areas of knowledge and essay was created, the rows with empty data in the grade average were removed, in which these rows represented absent participants, either on one exam day only, or both days. Subsequently, numerical data were transformed into categorical data, which improves understanding. In this step, ENEM microdata dictionary was also used, which is a tool made available by INEP to assist in understanding the columns to be explored. It is worth mentioning that the data preparation was the same for all the years analyzed.

4.3 Exploratory Analysis

In this step, preliminary data investigation was carried out to discover patterns, detect anomalies, test and validate hypotheses with the help of descriptive statistics and graphical representations [17]. The purpose of this study step was to explore data that could be translated into important information to understand the impact of the pandemic on ENEM. In addition to exploratory data analysis, the main focus of this step was to generate comparative information and analyze whether there was an impact on ENEM. During this phase, the analysis was done by filtering one step at a time, but keeping the other variables, so we used the same data predefined in the preparation phase in all states. To filter only the columns of each state, the Boolean indexing available in the Pandas library was used. This type of indexing is used to filter values from a column in the dataset.

Selected columns in the dataset have been defined to allow for a more comprehensive analysis. For this purpose, columns on socioeconomic issues, information about the participants' high school and their performance were selected. For all analyses, we considered only present participants, who are characterized by their presence in the two days of the test, which is the way INEP determines to be considered present. After selecting the variables, we generated an exploratory analysis using notebook in Google Colaboratory, using graphs and tables from the three states.

5. RESULTS

The results of the analyses are presented individually for each state in the following subsections. From these results, we will discuss, and we will compare the results of the states.

5.1 Ceará

Figure 1 shows the number of absent and present participants in Ceará state. In the years 2017, 2018 and 2019, the number of participants present was higher than the number of absentees. In the first year of the pandemic, this pattern changed, and the number of absentees exceeded the number of present participants. In 2021, it is clear that the number of present candidates was higher than the number of absentees, a pattern seen in the pre-pandemic years that changed only in the first year of the pandemic. However, the number of subscribers to take the test was the lowest in the 5 years analyzed. Therefore, this leads us to believe that students who signed up in 2021 felt safer to go to the test location to take it.

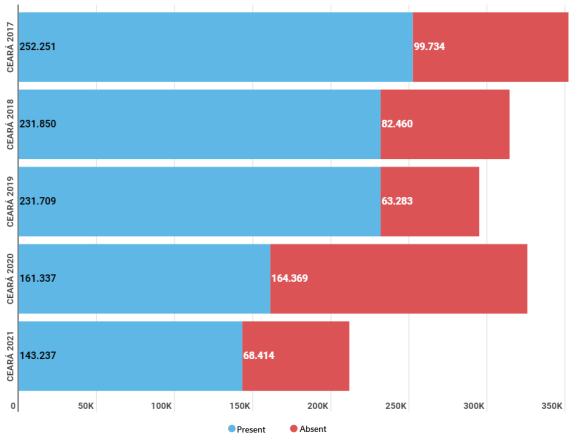

Figure 1: Number of candidates registered in Ceará state: present (blue) and absent (red).

Figure 2 shows different violin plots with the grade of participants from Ceará state, in relation to declared family income. The order of family income information presented in the graph goes from 'A' to 'Q', where 'A' represents the participants who declare having no family income, and 'Q' are the participants who declare a monthly family income above R$19,000.00 (nineteen thousand Brazilian Reais) - the highest range. We can note that the average of students who reported higher family income have a higher grade average than students from other social classes. In the year 2021, 22,347 participants declared that they did not have any type of family income. In contrast, 1,405 participants declared that their income exceeded R$22,000.00 (twenty-two thousand Brazilian Reais) - the highest range. In Figure 2, the behavior of the graph shows that participants with higher social class have higher average grades than the other candidates, regardless of the year.

Figure 3 displays the average grade in relation to the type of administrative dependency in box plots. In the graph, we can see the quartiles and analyze the performance of the students. We can also analyze the difference between the grades of private schools and public ones. As previously mentioned, participants with higher social classes perform better, when we analyze the graphs in Figures 3 and 2. We consider that students from higher family income are in private schools. In Ceará state, schools with the best performance in ENEM are private, followed by federal, state, and municipal.

In 2019, the state had the top two positions with private schools. We could only find one public school in the 680th position, which was a federal one. In the dataset for the years 2020 and 2021, this analysis could not be performed because INEP removed school identifiers from the dataset due to containing data that may not comply with the LGPD [15].

5.2 Maranhão

Figure 4 presents a graph with the number of present and absent candidates in Maranhão state. We can note that Maranhão followed a pattern similar to Ceará state. The year 2017 was the one with the highest number of registered students, and the year 2021 was the year with the least number of registered students. In 2020, the high number of absentees was also evident: about 68,771 participants more than in 2019. In 2021, the number of presents decreased by 34,965 participants. When comparing the year 2017 with 2021, we can note that the difference in the number of registered students is 153,237 participants.

Figure 5 shows the average grade of the participants in relation to the type of declared family income. From 2017 to 2021, we can note that the average grade of students with the highest family income is always higher than the other family incomes. The graph also shows that there is no difference in the years presented: in all years, students with higher income have the highest averages. The state of Maranhão also had an increase of 0.5 points in average grade, from 493.81 in 2019 to 494.31 in 2020, in the first year of the pandemic. It is worth mentioning that, by considering the total number of people registered to take the exam in 2021, 14.713 declared they do not have any type of income, and only 690 participants declared they have an income greater than 22 thousand Brazilian Reais. This shows that, despite the higher grades, the number of participants is lower than for other types of family income.

Figure 6 shows box plots with the average grade of participants from Maranhão state in relation to the type of administrative dependency of schools. The graph shows that this state did not suffer a significant change in student performance due to the type of administrative dependency of the school. The graph shows that private schools in Maranhão state obtained better results in the exams compared to the 3 other types of administrative dependency. Schools at the federal dependency have the highest averages when compared to public schools. Between the years 2017 and 2021, little change can be noticed between students' performance

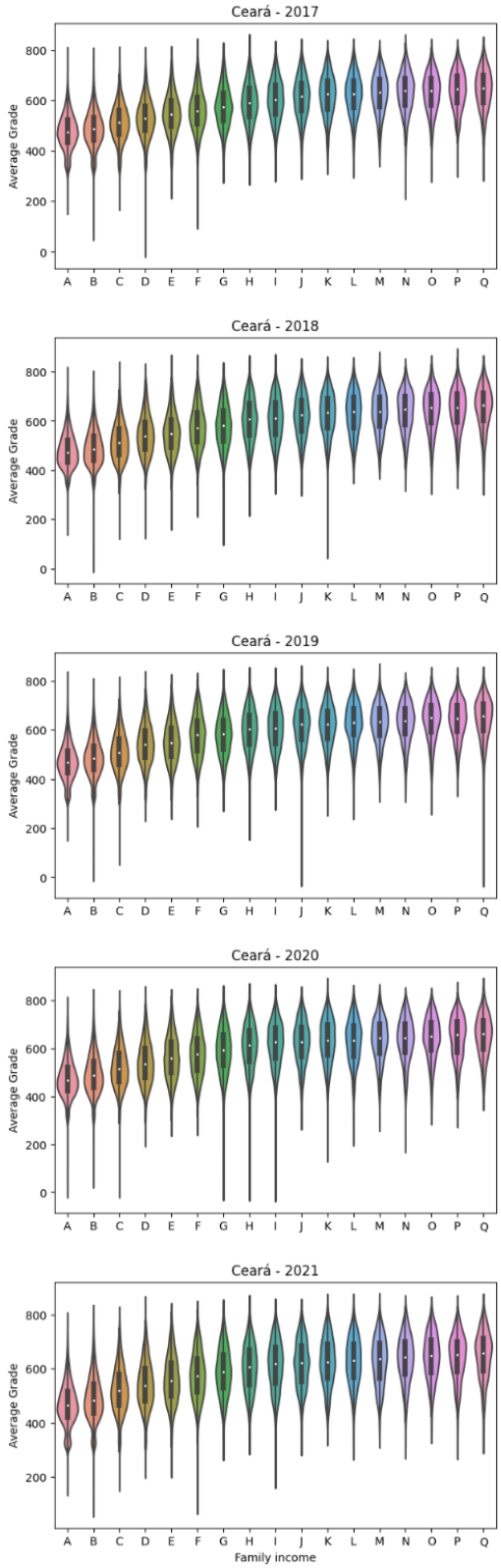

Figure 2: Average grade of students in Ceará state in relation to family income.

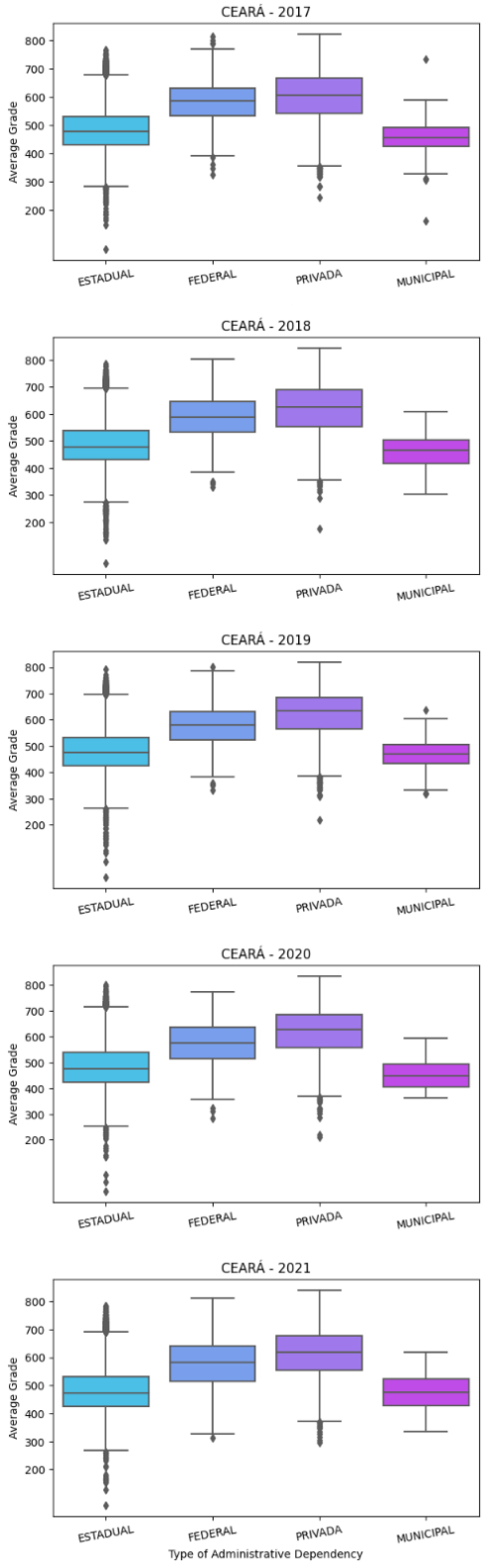

Figure 3: Average grade of students in relation to the type of administrative dependency of schools in Ceará state.

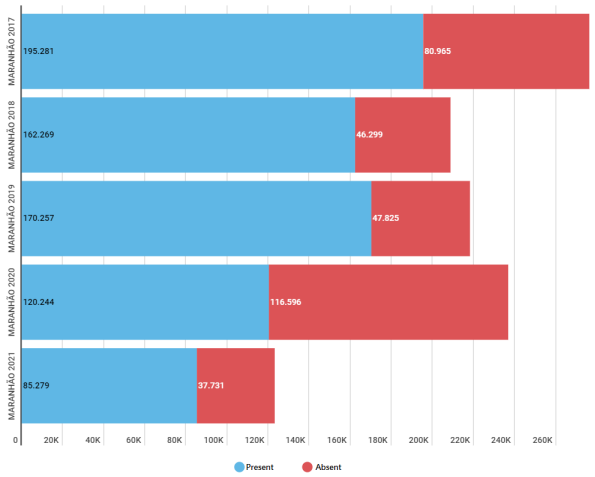

Figure 4: Number of participants registered in Maranhão state: present (blue) and absent (red).

on the exam. The dependency in public schools remained the same in 2020 and 2021, with federal schools with the best performance, followed by state and municipal schools.

Among the schools with the best performance in ENEM in the year 2019, a school from Maranhão only appears in the 102nd position, a private one. The first public school to appear in the ranking is in the 2396th position, which is a federal one.

5.3 Piauí

The state of Piauí also shows the impacts on the number of absentees and presents as seen in Figure 7. In pre-pandemic years, the state maintained an average of 98.990 attendees, while in 2020 and 2021, the average of presents at 63.293. The year 2021 had the lowest number of absentees. However, it also had the only 77.299 participants - the number of enrolled in the test does not even reach the number of presents in the year 2019. 2020 was the year with the highest percentage of absentees, i.e., about 3 times more absentees than the previous year. This is still below the national average, which is about 52% abstention [29]. This high percentage of abstentions in the state indicates that, in the first year of the pandemic, those enrolled chose not to take the exam due to the risks of the pandemic.

The state of Piauí, similarly to the states of Maranhão and Ceará, maintained the same behavior in both graphs: from 2017 to 2021, participants with a "Q" family income had a higher grade than the other participants, as shown in Figure 8. Piauí, in turn, had about 10.135 participants with income type "A" and 485 with income type "Q". The grade averages have not changed much over the years. However, if we compare the years 2017 and 2021, the number of participants who declare to be in the income classes "A", "B" or "C" is 108.411 in 2017, and it is 55.474 in 2021.

The state of Piauí shows a disparity in performance between state, federal and municipal participants. Figure 9 shows that the top participants came from private dependency schools. When analyzing outliers, we can see that public schools are rarely close to reaching the maximum grade. This type of behavior is repeated in all five years ana-

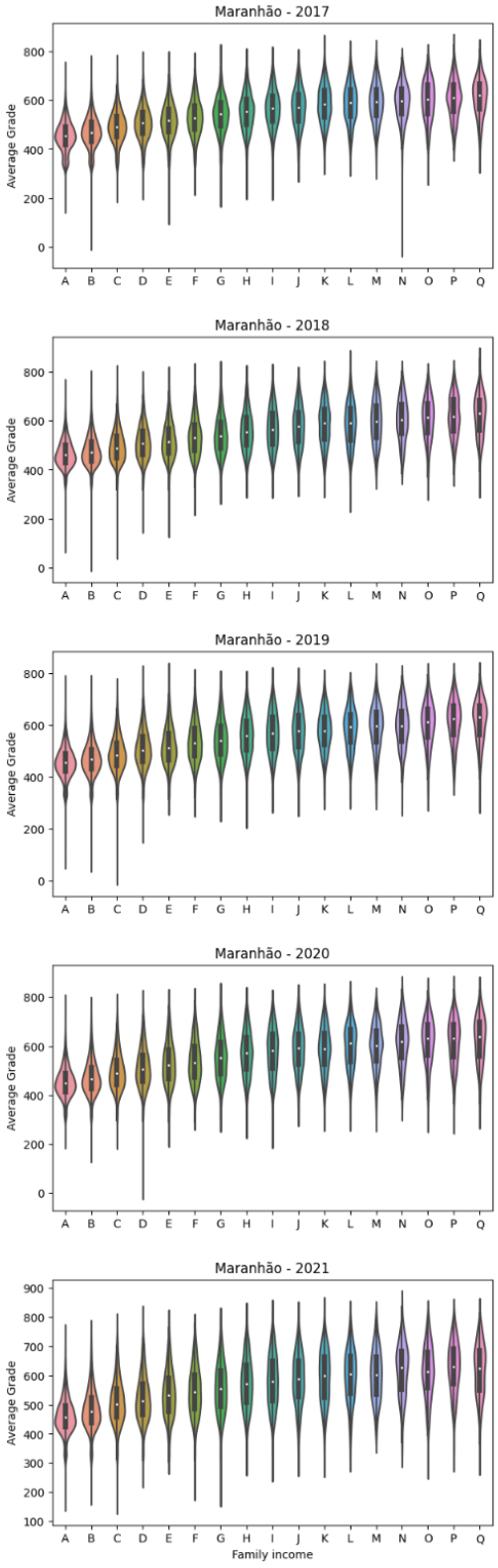

Figure 5: Average grade of students in Maranhão state in relation to family income.

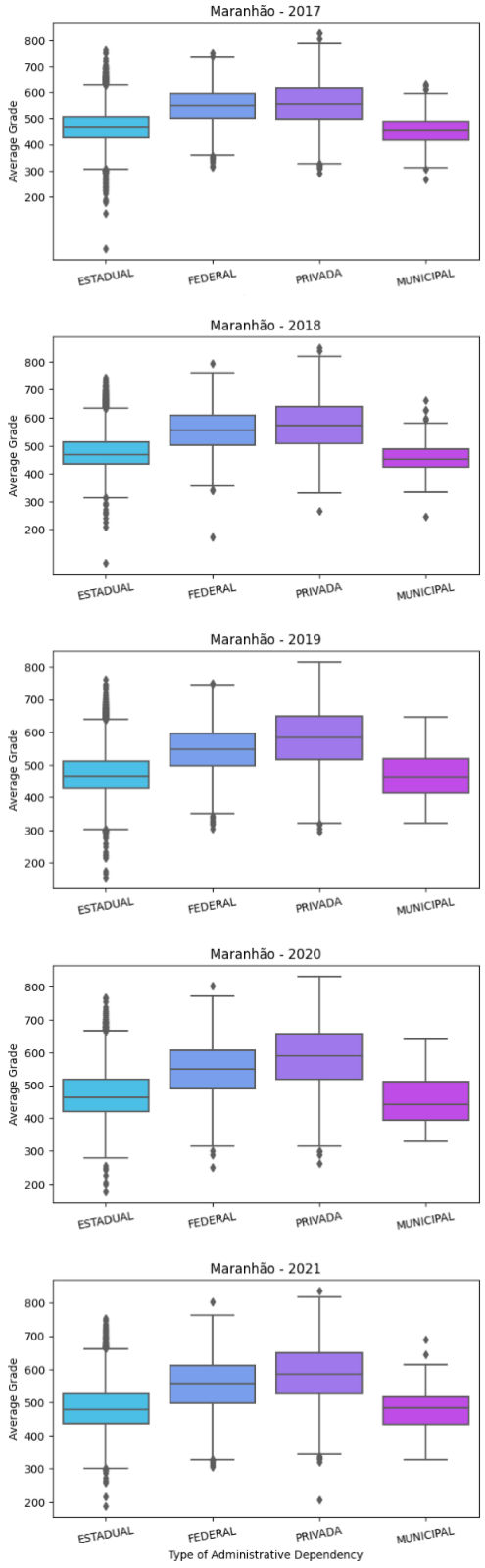

Figure 6: Average grade of students in relation to the type of administrative dependency of schools in Maranhão state.

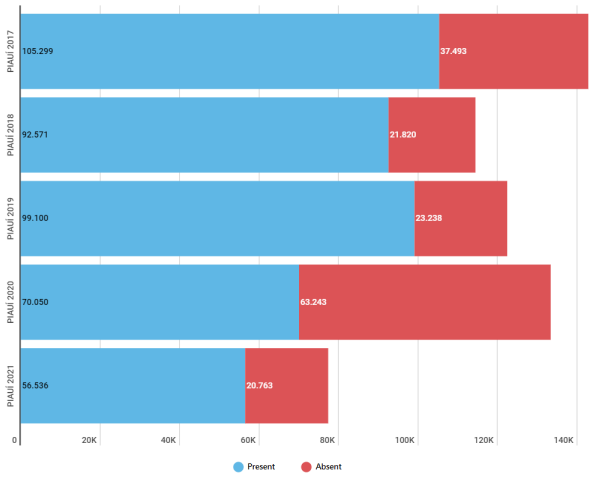

Figure 7: Number of participants registered in Piauí state: present (blue) and absent (red).

lyzed. When analyzing the graph, we can also note that the general performance of the students did not change during the pandemic, that is, participants, regardless of dependency school they studied, kept the grade averages very similar to previous years. Because the exam uses the Item Response Theory (IRT) [12], the test can be leveled in comparison to the performance of all other participants who also took the exam.

6. DISCUSSION

We found that the presence of participants in the ENEM in 2020 was greatly affected by the pandemic in both states. The number of participants in 2021 has dropped sharply in all three states. Taking into account the 3 states, all of them had the highest number of abstentions in 2020, which was the first year of the pandemic. Knowledge about the disease, high transmission rates, economic concerns and insecurities about the difficulty of accessing the necessary knowledge may have been the factors that led participants not to take the exam. Another reason could be the change in the date that the exam usually took place, which are October and November. Due to the pandemic, the exam of 2020 was applied in January 2021.

In 2021, states had the lowest number of participants among the 5 years analyzed. We believe that, in 2021, students had already been taking remote classes for a long time, so students preferred not to register for the exam. Therefore, in 2020, high abstention in all states leads us to believe that students who took the test were better prepared, and participants who enrolled in 2021 would also feel more comfortable taking the test, as they were well-prepared. The increase in student performance during the pandemic period leads us to believe that students who took the exams in the pandemic could be studying for the test in advance, making their performance better.

All states had many absentees in 2020 and a very low number of participants in 2021. This leads to the hypothesis that absentees and those who did not participate (people who did not register for the exam) may be from more vulnerable social groups and are more likely to be absent in the

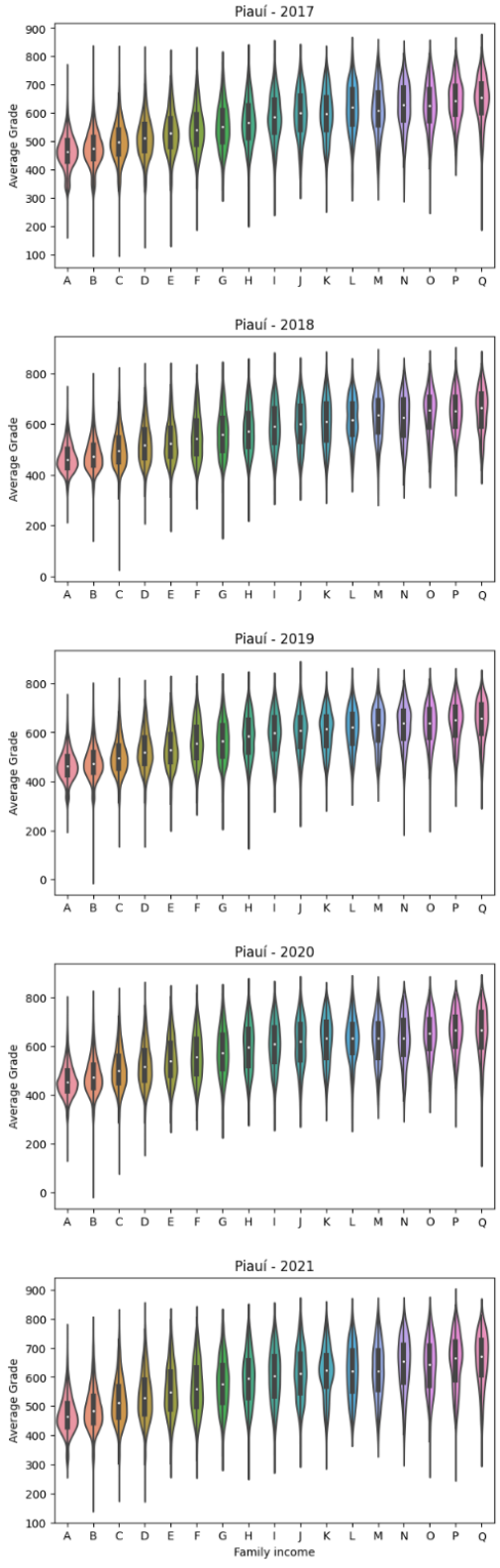

Figure 8: Average grade of students in Maranhão state in relation to family income.

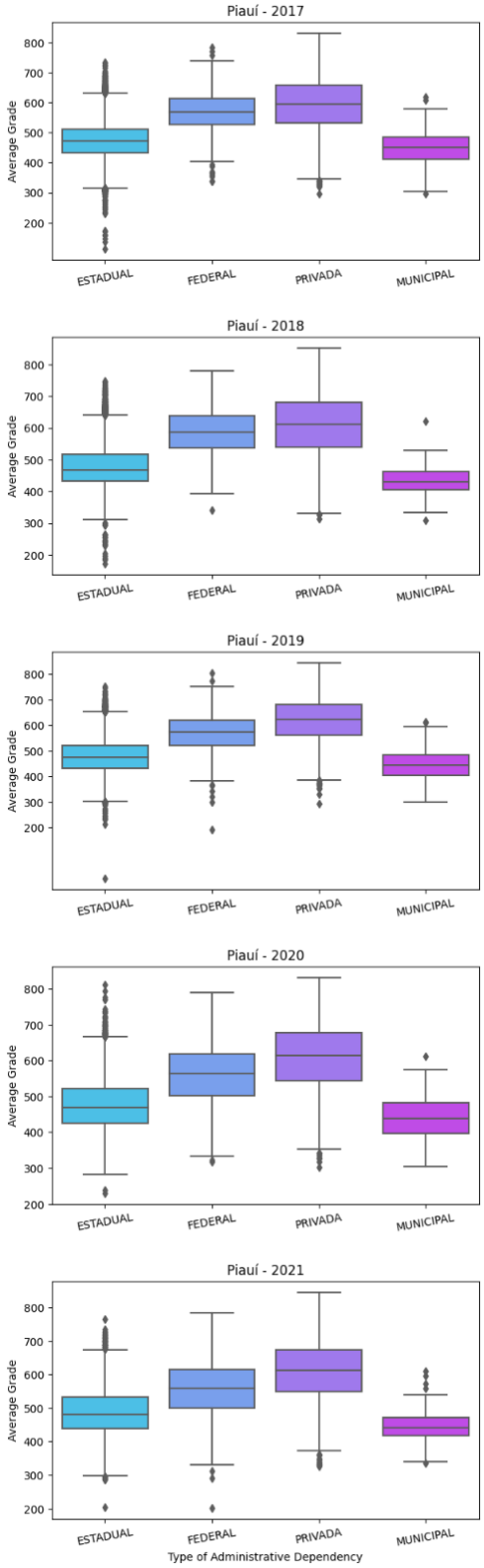

Figure 9: Average grade of students in relation to the type of administrative dependency of schools in Piauí state.

exam. Violin plots that show the average grade in relation to social class Figures 5, 2 and 8 highlight this hypothesis related to social inequality. We can note that, in both states, the number of participants from lower social classes decreased in the exam in 2021 when compared to the year 2017.

We can also note that, in the three states and in both years, private schools had the highest grades, followed by federal, state and municipal schools. The state of Ceará, in turn, showed that, when compared to Brazil, some private schools were among the best in the country. In the years of the pandemic, we were not able to analyze the schools individually. However, the private dependency schools and the federal ones remained with better performance. Furthermore, most private school students belong to a higher social class, and as we have seen, the higher the social class, the higher the average grade in all states analyzed.

This study has limitations to be acknowledged. First, the results presented here represent only three states in the northeast region of Brazil. Second, the information taken from the dataset due to LGPD limited us to exploring the pandemic years in some aspects. For a better understanding of the impact of the pandemic, as a future work, we plan: (i) to analyze all states of the Brazil; (ii) to use correlation analysis considering not only student grades, but also the success rate of exam questions; and (iii) to develop descriptive models.

7. FINAL REMARKS

By applying educational data analysis, this study identified relevant information to answer whether the COVID-19 pandemic impacted on ENEM in the states of Ceará, Maranhão and Piauí. For this purpose, the study presented results of an exploratory and comparative analysis of the ENEM microdata from 2017 to 2021 of the three states. The results of this study showed that the performance of the participants was not impacted by the pandemic, and the average grade of the states in the 5 years remained at similar values. In addition, the results showed that the number of abstentions in the first year (2020) of the pandemic took a huge reduction, with more than 50% of participants absent from the exam. Also, the number of subscribers was impacted in 2021, with the number of candidates being the lowest in the 5 years analyzed.

Acknowledgements

The authors would like to thank the Brazilian National Council for Scientific and Technological Development (CNPq) and the State Funding Agency of Maranhão (*Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão* - FAPEMA) for the support given to their research projects.

8. REFERENCES

[1] S. B. Aher and L. Lobo. Data mining in educational system using weka. In *International Conference on Emerging Technology Trends (ICETT)*, volume 3, pages 20–25, 2011.

[2] A. Y. Alqahtani and A. A. Rajkhan. E-learning critical success factors during the COVID-19 pandemic: A comprehensive analysis of e-learning managerial perspectives. *Education Sciences*, 10(9):216, Aug. 2020.

[3] R. Arruabarrena, T. A. Pérez, J. López-Cuadrado, J. Gutiérrez, and J. Vadillo. On evaluating adaptive systems for education. In *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 363–367. Springer, 2002.

[4] A. I. R. L. Azevedo and M. F. Santos. Kdd, semma and crisp-dm: a parallel overview. *IADS-DM*, 2008.

[5] A. A. Barcellos, S. Isotani, C. Diego, and N. Damasceno. Mineração de dados abertos - enem 2018. *Anais dos Trabalhos de Conclusão de Curso da Pós-Graduação em Computação Aplicada à Educação*, 2018.

[6] S. E. Brasil. Epidemiológico 17. *Boletim Epidemiológico*, 50(17), 2019.

[7] M. Ciotti, M. Ciccozzi, A. Terrinoni, W.-C. Jiang, C.-B. Wang, and S. Bernardini. The COVID-19 pandemic. *Critical Reviews in Clinical Laboratory Sciences*, 57(6):365–388, 2020.

[8] S. J. Daniel. Education and the COVID-19 pandemic. *PROSPECTS*, 49(1-2):91–96, Apr. 2020.

[9] C. P. de Moraes, R. T. Peres, and C. E. Pedreira. Eficácia escolar e variáveis familiares em tempos de pandemia: um estudo a partir de dados do enem. *Interfaces da educação*, 12(35):635–658, nov. 2021.

[10] T. P. Educação. Ensino a distância na educação básica frente à pandemia da covid-19. *Nota Técnica*, 2020.

[11] F. N. Gularte, F. K. V. do Nascimento, and A. V. de Carvalho. ENEM em tempos de pandemia: desafios enfrentados pelos docentes frente o ensino remoto da rede estadual de presidente kennedy/TO. *Research, Society and Development*, 10(14):e571101422339, Nov. 2021.

[12] INEP. Teoria de resposta ao item. `https://download.inep.gov.br/educacao_basica/enem/nota_tecnica/2011/nota_tecnica_tri_enem_18012012.pdf`, 2011. Online; Acessado em 04 de junho de 2022.

[13] INEP. Enem - microdados. `https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem`, November 2020. Online; Acess in 07 april 2022.

[14] INEP. Exame nacional do ensino médio (enem). `https:https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem`, Fevereiro 2022. Online; Acessado em 10 de maio de 2022.

[15] INEP. Nota de esclarecimento | divulgação dos microdados. `https://www.gov.br/inep/pt-br/assuntos/noticias/institucional/nota-de-esclarecimento-divulgacao-dos-microdados`, Fevereiro 2022. Online; Acessado em 07 de maio de 2022.

[16] L. P. Junior, S. N. Matos, and H. B. Borges. Análise dos perfis de alunos do ensino superior sobre a realização de aulas na modalidade a distância durante pandemia da covid-19 usando algoritmos de aprendizagem de máquina. *RENOTE*, 18(2):336–345, Jan. 2021.

[17] M. Komorowski, D. C. Marshall, J. D. Salciccioli, and Y. Crutain. Exploratory data analysis. *Secondary Analysis of Electronic Health Records*, pages 185–203, 2016.

[18] E. Nakazone and L. M. Bortolotti. Análise de dados históricos do enem entre 2015 à 2019. In *Congresso de Tecnologia-Fatec Mococa*, volume 4, 2021.

[19] N. W. Neto, R. Soares, L. Coutinho, and A. Teles. Análise exploratória de dados para identificar o impacto da pandemia da covid-19 no enem dos estados do ceará, maranhão e piauí. In *Anais da X Escola Regional de Computação do Ceará, Maranhão e Piauí*, pages 31–40. SBC, 2022.

[20] D. A. Oliveira and A. M. Clementino. As políticas de avaliação e responsabilização no brasil: uma análise da educação básica nos estados da região nordeste. *Revista Iberoamericana de Educación*, 83(1):143–162, May 2020.

[21] A. Peña-Ayala. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, 41(4, Part 1):1432–1462, 2014.

[22] N. Reuge, R. Jenkins, M. Brossard, B. Soobrayan, S. Mizunoya, J. Ackers, L. Jones, and W. G. Taulo. Education response to COVID 19 pandemic, a special issue proposed by UNICEF: Editorial review. *International Journal of Educational Development*, 87:102485, Nov. 2021.

[23] J. Robbins. Unesco: 290 million students stay home due to coronavirus. *VOA Learning English*, 2020.

[24] J. S. Saltz. Crisp-dm for data science: Strengths, weaknesses and potential next steps. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 2337–2344, 2021.

[25] C. Shearer. The crisp-dm model: the new blueprint for data mining. *Journal of data warehousing*, 5(4):13–22, 2000.

[26] A. W. C. Silva, A. A. CUNHA, G. C. Alves, R. A. Corona, C. A. G. d. M. DIAS, R. NASSIRI, S. VEDOVELLI, T. R. F. VILHENA, V. B. C. FAVACHO, J. F. SOUSA, et al. Perfil epidemiológico e determinante social do covid-19 em macapá, amapá, amazônia, brasil. *Revista Científica Multidisciplinar Núcleo do Conhecimento*, 2020.

[27] A. Simon and S. Cazella. Mineração de dados educacionais nos resultados do ENEM de 2015. In *Anais dos Workshops do VI Congresso Brasileiro de Informática na Educação (CBIE 2017)*. Brazilian Computer Society (Sociedade Brasileira de Computação - SBC), Oct. 2017.

[28] R. D. C. Soares, N. W. Neto, L. R. Coutinho, F. J. D. S. e Silva, D. V. dos Santos, and A. S. Teles. Mineracão de dados da educação básica brasileira usando as bases do INEP: Uma revisão sistemática da literatura. *RENOTE*, 19(1):361–370, July 2021.

[29] N. Weber Neto, R. C. Soares, L. Reis Coutinho, and A. Soares Teles. A pandemia da covid-19 impactou o enem? uma análise comparativa de dados dos anos de 2019 e 2020. *RENOTE*, 20(1):223–232, 2022.

[30] Z. Wu and J. M. McGoogan. Characteristics of and important lessons from the coronavirus disease 2019 (covid-19) outbreak in china: summary of a report of 72 314 cases from the chinese center for disease control and prevention. *jama*, 323(13):1239–1242, 2020.