EXPLORING STUDENT'S BEHAVIOUR: AN EXPLORATORY DATA ANALYSIS IN PYTHON

Mohammed Ali Arsalan M[1]

Department of Mathematics, School of Advanced Sciences

Vellore Institute of Technology, Vellore, Tamil Nadu, India

mohammed.ali2022@vitstudent.ac.in

Mohamed Affan M[2]

Department of Mathematics, School of Advanced Sciences

Vellore Institute of Technology, Vellore, Tamil Nadu, India

mohamedaffan.m2022@vitstudent.ac.in

Mohammed Saqui T[3]

Department of Mathematics, School of Advanced Sciences

Vellore Institute of Technology, Vellore, Tamil Nadu, India

mohammedsaqui.t2022@vitstudent.ac.in

Abstract

This project uses Python's Exploratory Data Analysis (EDA) tools to examine student behaviour. The goal is to gain understanding of numerous facets of student behaviours and activities, as well as how they affect academic performance. A dataset with data on student demographics, academic performance, and everyday activities is used in the study. The analysis focuses on several important topics of interest, including social media use, study habits, academic achievement, hobbies, and financial perceptions. The study uses EDA approaches to investigate the connections between these variables and identify recurring trends and patterns throughout the student body. The results show that, when compared to male students, female students tend to spend more time studying every day. Additionally, female student score better academically overall. A considerable majority of college students work part-time, according to the analysis. Additionally, it draws attention to a rising tendency in male students' video game usage. The initiative also investigates how much time students spend on social media, how much time they spend on their hobbies, how they perceive their financial conditions, and when they prefer to study. These observations give educators and organisations important knowledge that helps them better understand students and develop strategies to improve academic performance and student engagement. This study gives a thorough investigation of student behaviour using EDA approaches and delivers useful information for those involved in education. The findings lay the groundwork for future research in the field of student behaviour analysis as well as add to the body of existing knowledge in that area. In the end, the research findings can help in the creation of efficient interventions and regulations meant to advance student performance and wellbeing.

Keywords: Exploratory Data Analysis, Python, Student Behaviour, Academic, Educators.

Introduction

Understanding the dynamics of student activities, habits, and their effects on academic performance requires a thorough understanding of student behaviour. Learning institutions and educators can improve learning outcomes, increase student engagement, and foster a positive learning environment by acquiring insights into many facets of student behaviour. Exploratory Data Analysis (EDA) methods and Python's strength as a programming language combine to create a solid foundation for studying and interpreting student data. This project's goal is to use Python's EDA tools to do a thorough analysis of student behaviour. We want to identify significant patterns, correlations, and trends that can throw light on many aspects of student behaviour by utilising a comprehensive dataset that includes student demographics, academic performance, and everyday activities. An essential component of this analysis is comprehending students' study patterns. We can gauge students' dedication to academic goals by looking at how much time they spend studying each day. To find out whether female students spend more time studying than their male counterparts, we also investigate potential gender variations in study habits. This investigation sheds light on the gender dynamics in study behaviours, which may be used to better provide educational support. Another important area of attention is academic success. We can determine whether more study time results in better marks by looking at the relationship between study time and academic outcomes. To find out if female students typically outperform male students in academic endeavours, we also study potential gender variations in academic performance. These revelations can educate educators and organisations about gender differences in academic attainment and direct the creation of focused solutions. A comprehensive comprehension of student behaviour must also include an awareness of their interests. We can learn more about students' extracurricular activities and potential sources of motivation by analysing data pertaining to their interests and hobbies. With the use of this study, a well-rounded curriculum can be created that incorporates the interests of the students and increases their general involvement and pleasure. A crucial component of student life is the need to balance finances. We can better understand the financial difficulties that students are facing by looking at how they perceive their financial status. This analysis offers important insights into students' financial health and might guide activities and policies meant to address financial issues that might impede students' academic achievement.

Additional topics of relevance include daily travel time and social media usage. We can identify potential elements that might have an impact on students' time management, engagement, and academic achievement by examining the time they spend on social media and during their daily commutes. These perceptions aid in comprehending pupils' digital behaviours and the difficulties they encounter in today's technologically advanced society. By utilising Python EDA approaches, we hope to deliver a thorough analysis of student behaviour through this project. The results of this investigation can give policymakers, educators, and other stakeholders in education useful information. Institutions can put evidence-based measures in place to increase student engagement, boost academic performance, and provide a positive learning environment by developing a deeper understanding of student behaviour. We will provide the literature review, methods, findings, and conclusions from our study in the sections that follow. We'll also give a summary of the main conclusions and their ramifications in our conclusion. The goal of this project is to add to the body of knowledge already available on student behaviour analysis and act as a useful resource for upcoming studies and instructional design.

Literature Review

[1] Now-a-days the amount of data stored in educational databases is increasing rapidly. These databases contain hidden information for improvement of students' performance. Educational data mining is used to study the data available in the educational field and bring out hidden knowledge from it. The results provide steps to improve the performance of the students who were predicted to fail or promoted. After the declaration of the results in the final examination the marks obtained by the students are fed into the system and the results are analyzed for the next session. The comparative analysis of the results states that the prediction has helped the weaker students to improve and brought out betterment in the result.

[2] It represents an important step for the university and its further performance using non-traditional learning methods, since most of the lectures carried out at the university are still done in a traditional way (lecturer-centered). In the virtual class, more efforts should be directed towards decreasing frustrations by improving motivation and interactivity. Motivation could be strengthened by creating a sense of community and by building trust between students.

[3] This paper discusses the use of data mining techniques to improve student retention. The authors used clustering algorithms to group students based on their behavior and identify factors that contribute to student retention.

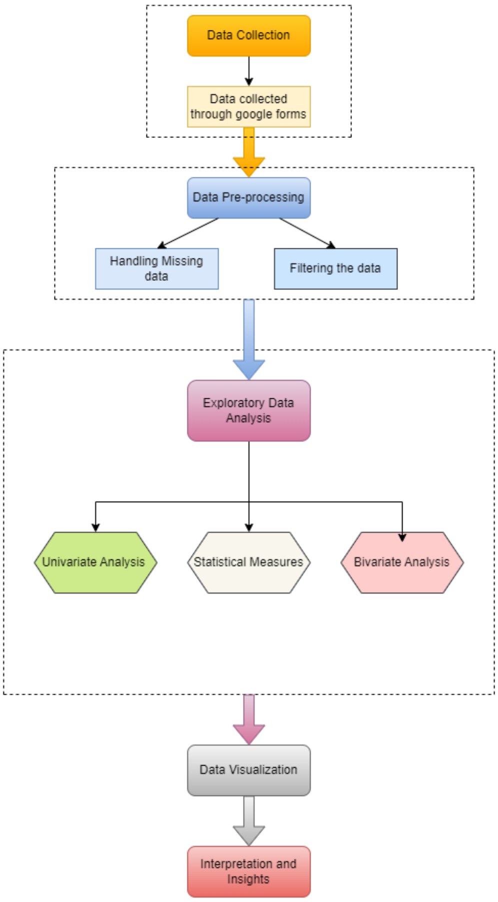
[4] This study used EDA techniques to analyze college students' time use and found that students spend most of their time on academic activities and leisure activities, with little time spent on employment, volunteering, or household activities.

[5] This paper discusses the EDA in school Exploratory data analysis (EDA) is an iterative, open-ended data analysis technique that enables practitioners to study data without bias in order to provide improvement procedures and make wise conclusions. Education is a data-rich industry that is ready to make the switch to a more in-depth, strategic use of data. By locating it within the literature linked to data-driven decision-making, this paper introduces the idea of EDA as an essential structure to be embedded in educational activities.

[6] Tom W. Archibald and David H. Feldman investigated that there aren't many studies that have predicted student achievement using data from LMSs. The frequency of a student's involvement with each LMS module, LMS log data, counts of hits, forum post information, and counts of exams seen and filed on the LMS are among the attributes.

[7] Diamond, M., & Mattia, A. discuss explains the use of visualization in business. Data analytics depend heavily on data visualisation, which enables analysts to build tools like dashboards for top executives. These tools can be created by corporations using a variety of software applications in order to alter data and make wise business decisions.

Methodology:



The methodology employed in this project involves several steps to analyze student behaviour using Exploratory Data Analysis (EDA) techniques in Python. The process includes data preprocessing, exploratory data analysis, and visualization of the key variables of interest.

Data Collection

The first step in completing an EDA process is to gather and prepare the necessary data for analysis.In this paper we collected the data through survey using google form and we examine the behaviour of students, the dataset include all available data related of each student such as 'Gender', 'Department', Height(CM), Weight(KG), 10th Mark,12th Mark, college mark, hobbies, daily studing time, prefer to study in, salary expectation, Do you like your degree?, possibility of choosing their career based on their degree , social media & video games spending Time, Travelling Time , Financial Status, Are you doing a part-time job right now?.

Data Preprocessing:

Data preprocessing is performed to ensure the dataset is clean, consistent, and ready for analysis. This includes handling missing values, removing outliers, and transforming variables if necessary.

Filtering the data:

For our research paper, we focused on analyzing the subset of data that pertains to students who scored above 35 marks in their 10th standard, 12th standard, and college examinations.By filtering the dataset based on this criterion, we aimed to narrow down our analysis to a specific group of high-scoring students. This subset of data includes individuals who achieved academic success by surpassing the minimum threshold of 35 marks in their respective educational stages.

Droping least necessary variables:

In our dataset 'Are you doing a part-time job right now?' column doesn't provide any useful information so we are deleting that column.By eliminating this column, we aimed to streamline our analysis and focus on the relevant variables that contribute to the research objectives. This step was taken to ensure the accuracy and validity of our findings in the project.

Handling missing values:

In our dataset, we encountered approximately 53 missing records for weight and height. To address this issue, we employed the mean imputation method to fill in the missing values for height and weight. By calculating the average values from the available data, we were able to estimate and substitute the missing measurements. Furthermore, we encountered missing values in the 'Do you like your degree' column, which is a categorical variable. To handle this situation, we decided to fill in the missing values with the category that had the highest frequency of occurrences in the dataset. By doing so, we aimed to ensure that the imputed values reflected the most common response in the dataset.These steps were necessary to address the missing data and ensure the completeness of our dataset for further analysis in our research paper.

Exploratory Data Analysis (EDA):

The EDA phase involves gaining a comprehensive understanding of the dataset and its variables.

Statistical measures such as mean, median, standard deviation, and correlation are calculated to assess the central tendencies and relationships between variables[8].

Univariate analysis is conducted to examine the distribution, range, and summary statistics of individual variables.

Bivariate analysis is performed to explore relationships between pairs of variables, such as studying time and academic performance or gender and study habits.

Visualization:

Data visualization techniques are employed to provide a visual representation of the findings.

Various types of plots, such as scatter plots, histograms, bar charts, and box plots, are used to illustrate the relationships and distributions of the variables.

Gender-based comparisons and trends are depicted through visualizations to identify any disparities or patterns in student behaviour.

Interpretation and Insights:

The results obtained from the EDA and visualization are interpreted and analyzed to derive meaningful insights about student behaviour. Patterns, trends, and relationships identified in the data are discussed in the context of the research objectives. The findings are compared with existing literature and theories to provide a comprehensive understanding of student behaviour.

Result and Discussion

To gain a comprehensive understanding of the dataset and provide meaningful insights for our research paper, we utilized the describe() method. This method allows us to generate descriptive statistics that include percentiles, mean, standard deviation, maximum value, and count for each

	Height(CM)	Weight(KG)	10th Mark	12th Mark	college mark
count	251.000000	251.000000	251.000000	251.000000	251.000000
mean	157.316372	60.844690	78.643147	70.331355	73.990159
std	20.654842	14.083744	12.166327	11.603714	11.986259
min	4.500000	20.000000	45.000000	45.000000	50.000000
25%	153.000000	50.000000	70.000000	61.000000	65.000000
50%	158.000000	60.844690	80.000000	70.000000	75.000000
75%	170.000000	70.000000	88.700000	79.750000	84.000000
max	192.000000	106.000000	100.000000	98.500000	100.000000

Table 1 Descriptive statistics

variable in the dataset.Table.1 shows the descriptive statistics of our dataset.

Gender-based comparisons and trends are depicted through visualizations to identify any disparities or patterns in student behaviour. Based on the provided information, we can deduce that the total count of male students in the dataset is 149, while the total count of female students is 76 as shown in figure.1.This finding indicates a gender imbalance within the dataset, with a higher representation of male students compared to female students. Such insights are valuable for understanding the demographic composition of the sample population under study.

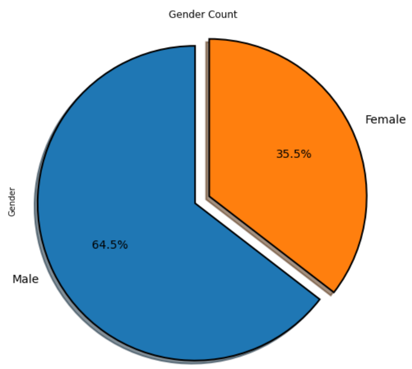


Fig. 1 Gender count

In our project, we calculated the average student marks for the 10th standard, 12th standard, and college levels. This analysis allows us to gain

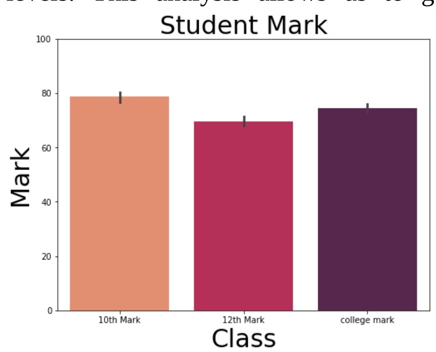


Fig.2 Student Marks

insights into the overall academic performance of the students in our dataset as shown in figure.2.

Certification plays an important role in getting selected for corporate jobs.So let's see how many students have completed any certification courses.

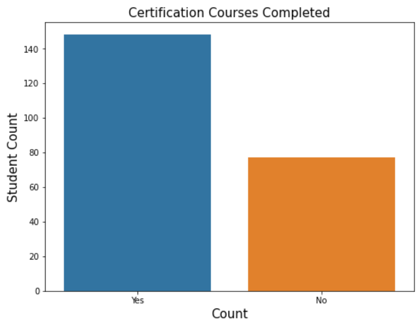


Fig. 1 Certification Count

When analyzing the preferred study times among students, we found a diverse range of responses. A significant portion of the participants mentioned flexible study schedules, with no specific preference for a particular time of day. Additionally, around 30% of students reported being early birds, favoring morning study sessions before their regular classes or commitments.

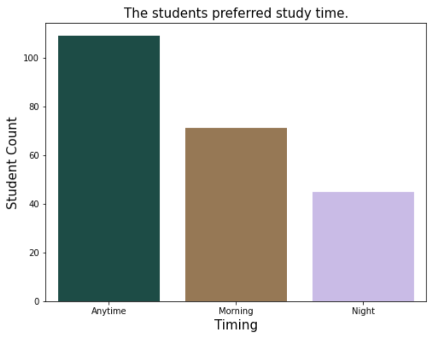


Fig. 2 Preferred Study time

During our investigation into student hobbies, we discovered a wide range of interests and activities as in fig.5. The most common hobbies among the participants included sports, with approximately 40% of students engaging in various physical activities. Additionally, around 35% of students expressed a passion for cinema and watching movies. Another hobby among students was reading, with approximately 20% of participants indicating their enjoyment of literature and spending time with books. From our dataset, we could arrange the favourite hobbies of students in a certain order. such as

Male students :- sports > watching movies > video games > reading books.

Female Students :- watching movies > sports > reading books > video game.

Here Video games are stated as a hobby by the fewest female students.

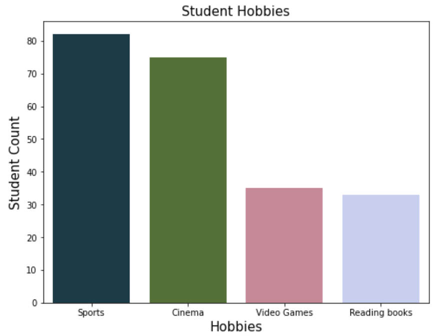


Fig. 5 Student Hobbies

Now lets perform some bivariate analysis for our dataset and find the useful insights

In terms of daily social media usage, our research indicated that students spend an average of 1 to 2 hours per day on various social media platforms.

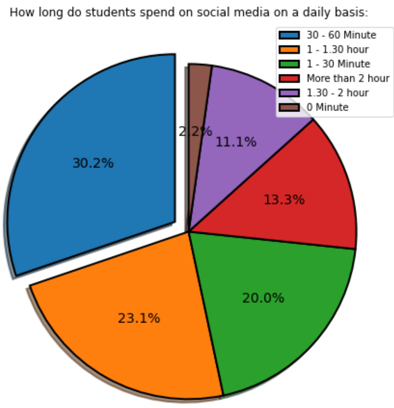


Fig.6 Social media usage

This finding aligns with the broader trend of increased digital connectivity and the pervasive role of social media in modern society. However, it is crucial to note that usage patterns varied among individuals, On comparing social media usage

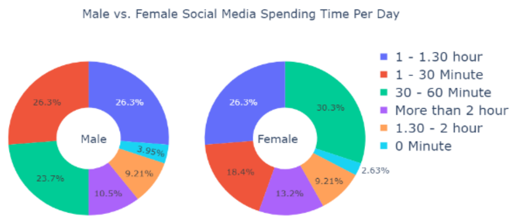


Fig.7 Social media usage

between male and female students we could conclude that there is no significant difference in how much time female and male students spend on social media each day as shown in fig.7.

Next investigation into the daily travel time for students yielded intriguing results. On average, participants reported spending approximately 30 minutes traveling to and from their educational institution. However, it is important to consider that the range varied significantly among students, with some commuting as little as 10 minutes and others requiring up to an hour or more each way. This disparity could be attributed to factors such as the geographical location of the institution, transportation availability, and the student's choice of residence.

On comparing travel time of male and female

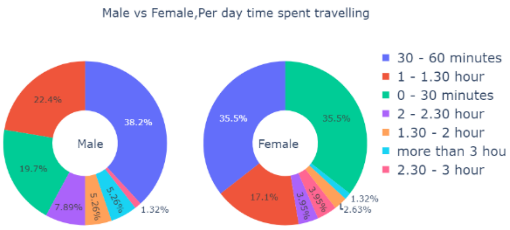


Fig.8 travelling time of male vs female students

student we could conclude that female students spend less time travelling which is around 0-30 mins and male students travel around 30-60 mins as in fig.8.

Regarding students perspectives on their financial situation, our findings exhibited a mixed sentiment as shown in fig 10. While some students (around 55%) reported feeling relatively comfortable with their financial circumstances, and remaining(approximately 40%) expressed concerns and acknowledged financial challenges. The reasons cited for these concerns varied, including the rising costs of tuition fees, textbooks, and living expenses.

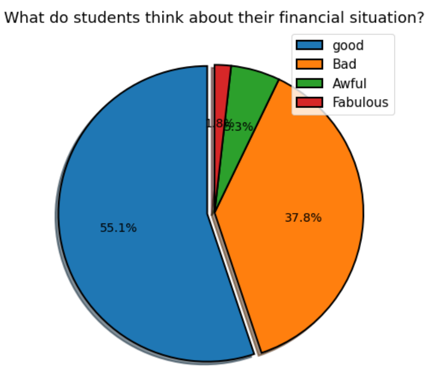


Fig. 9 financial situation

On comparing male and female students we could conclude that most female students feel their financial status is better than male students as in fig.10.

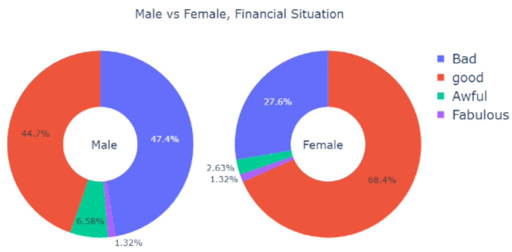


Fig. 10 financial situation of male vs female students

Next to investigate the height distribution among students and provide an understanding of the height characteristics within the student population. The study collected height data from a representative sample of students, allowing for a comprehensive analysis of their height distribution.

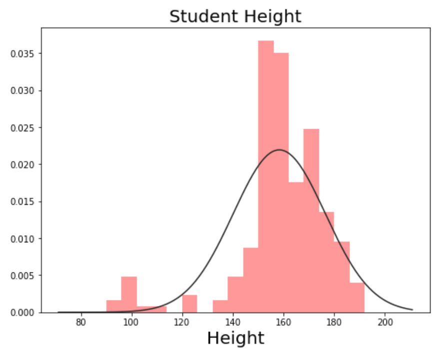


Fig. 11 height distribution of students

From fig.11, we could come a conclusions that 90% of the student heights fall between 140 cm and 180 cm.

On comparing height distribution of male and

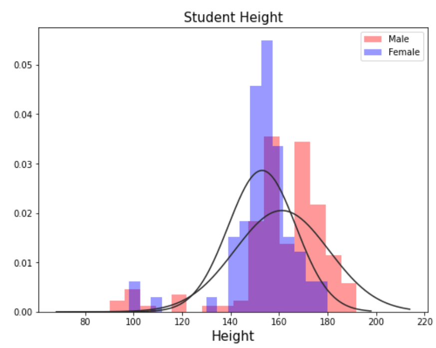


Fig. 12 Height distribution of male vs female

female students as in fig.12 we can understand that 90% of the female student heights falls between 140 cm and 165 cm. and 90% of the male student heights fall between 147 cm and 180 cm.

Next we investigated the weight distribution among students and gain insights into their overall weight patterns.The analysis revealed a diverse range of weights among the student population. The weight distribution followed a normal distribution curve, where student weights fall between 40 kg and 80 kg as shown in fig.13.

And on comparing comparing weight pattern of

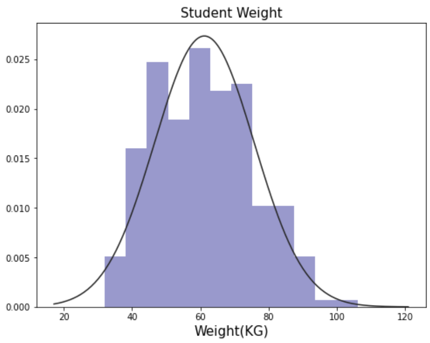


Fig.13 weight pattern of students

male and female students as in fig.14.we could conclude that 90% of the female students weight falls between 40 kg and 65 kg. and 90% of the male students weight falls between 50 kgs and 85 kgs.

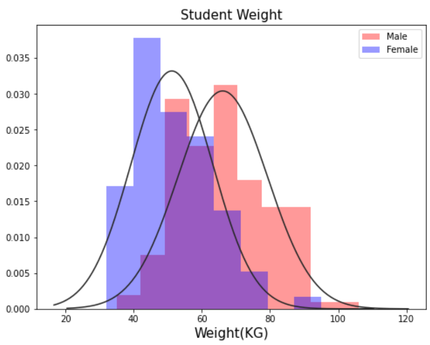


Fig 14 weight pattern of male vs female students

We further analysed the weight distribution, weight categories were established based on standard body mass index (BMI) ranges. These categories included underweight, normal weight, overweight, and 3 categories of obese. The percentage of students falling within each weight category was determined through BMI calculations using the height and weight data.

The pie charts fig. 15 shows us the BMI levels of students and around 45.8% are under weight and 42% are weighted normal. There are very few students falls under obese category.

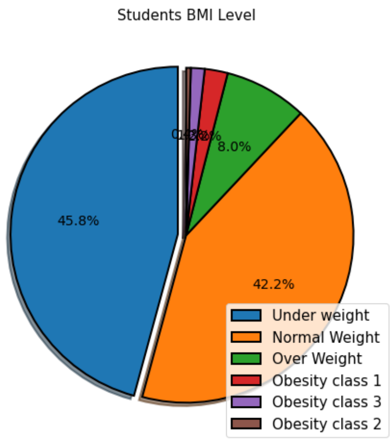


Fig.15 BMI Level of students

Conclusions

Python EDA techniques were used in this research to examine student behaviour. The research showed that there are gender inequalities in academic performance and study habits, and more time spent playing video games by male students. Insights were also gathered into the preferred study periods, pastimes, financial attitudes, commute time, and everyday social media usage of the students. Like Female students spend more time studying on a daily basis than male students. The hobbies of male and female students are different. These findings advance our knowledge of student behaviour and can help educators and institutions devise plans to boost academic achievement and student engagement. These results can be built upon by more study and analysis to improve the efficacy of instructional strategies and encourage student achievement.

References

[1] "An Exploratory Data Analysis of College Student Time Use" by Tom W. Archibald and David H. Feldman (2008)

[2] "Data mining and student retention: a case study" by M. Romero and E. Zafra.

[3] Diamond, M., & Mattia, A. (n.d.). Data visualization: An exploratory study into the software tools used by businesses. Journal of Instruction Pedagogies, 17, 1–7

[4] EDA On Student Behavior - https://www.kaggle.com/code/gunapro/eda-on-student-behavior/notebook#Data-Visualization

[5] Martin, W.W., & Bridgmon, K.D. (2012). Quantitative and statistical research methods. San Francisco, CA: Jossey-Bass.

[6] Rajeev Gandhi, B. (2012). Anticipation for Performance Improvement of Students using Machine Learning Approach. Artificial Intelligent Systems and Machine Learning, 4(5), 296-301. Retrieved from https://www.ciitresearch.org/dl/index.php/aiml/article/view/AIML052012007

[7] Purwoningsih, Tuti & Santoso, Harry & Hasibuan, Zainal. (2019). Online Learners' Behaviors Detection Using Exploratory Data Analysis and Machine Learning Approach. 1-8. 10.1109/ICIC47613.2019.8985918.

[8] Rebeka Lukman, and Majda Krajnc. “Exploring Non-Traditional Learning Methods in Virtual and Real-World Environments.” Journal of Educational Technology & Society, vol. 15, no. 1, 2012, pp. 237–47. JSTOR, http://www.jstor.org/stable/jeductechsoci.15.1.237. Accessed 22 Mar. 2023.