International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Examining the Impact of Project-Based Learning on Student Performance Across Demographic Factors

A. Poompavai[1], A. Poongothai[2], R. Divya[3]

[1]Assistant Professor, Department of Statistics, SDNB Vaishnav College, Chrompet, Chennai
[2]Research Scholar, Department of Statistics, Annamalai University, Chidambaram.
[3]Assistant Professor, Department of Statistics, SDNB Vaishnav College, Chrompet, Chennai

Abstract

In the 21st century, student performance is crucial for critical thinking, teamwork, creativity, digital literacy, cultural competence, self-direction, resilience, adaptability, and ethics. Project-based learning helps students acquire these skills and prepare them for a technologically sophisticated future. A study analysed performance differences between male and female students, racial/ethnicity trends, family education, lunch availability, test preparation course completion, math scores, reading scores, and writing scores. No significant differences were found between gender versus test preparation course and race ethnicity, race ethnicity and test preparation course, and race ethnicity and parental level of education. The study emphasizes the importance of continuous professional development in project-based learning for students to prepare them for a technologically sophisticated future.

Keywords: Student Performance, Race Ethnicity, scores, Chi Square test, t test, ANOVA, Professional Development, project based learning.

INTRODUCTION

In the 21st century, a student's progress is assessed beyond grades and test results, determining their ability to exhibit various skills required for societal success. These skills include critical thinking, teamwork, creativity, digital literacy, cultural competence, self-direction, resilience, adaptability, and ethics and global awareness. Critical thinking involves studying facts, evaluating perspectives, and expressing thoughts effectively. Collaborative abilities involve working as a team, generating new ideas, and applying creative thinking in various areas. Digital literacy involves using computers and digital technologies for information search and sharing. Cultural competence involves learning and enjoying different cultures, interacting with people from all backgrounds, and taking responsibility for one's education. Resilience and adaptability involve handling change, overcoming challenges, and learning from mistakes. Ethics and global awareness involve understanding global operations and moral issues. In the 21st century, student performance encompasses more than just academic success, it includes skills and knowledge that help students thrive in a rapidly changing world. Students possess a range of abilities, knowledge, and qualities that enable them to prosper in a knowledge-based society. Talent testing evaluates students based on more than just academic ability, including digital literacy, critical thinking, and social skills. Research on continuous professional development in project-based learning shows that interventions can help students acquire 21st-century abilities, raising grades and acquiring necessary knowledge for success in contemporary culture. However, objectively evaluating and grading these abilities can be challenging. Continuous study is crucial for assessing and comparing students' performance, even if these abilities are highlighted. Including 21st-century competencies in the classroom helps students prepare for an internationally technologically sophisticated future. Declining student performance outside conventional tests can negatively impact society and the economy. Overall, student performance encompasses a wider spectrum of qualities and talents than just intellectual ability, emphasizing the need for personal development to meet current challenges.

REVIEW OF LITERATURE

In 2015, Fiseha Berhanu and Addisalem Abera used educational data mining to predict students' academic performance using a decision tree algorithm. Data from Dilla University's College of Agriculture was collected over five years. After processing 49 attributes, 27 rules were generated to improve students' performance. The model determined students' performance based on courses, sex, and academic status. The algorithm achieved an accuracy of 84.95%.

Wan Maziah Wan Ab Razak's 2019 research focuses on the relationship between academic performance and the teaching and learning process among degree students in a higher learning institution. The study used questionnaires distributed to students from semester 4 and 5, with a sample size of 70. The results showed that the teaching and learning process had the highest beta value. The research recommends strategies for improving teaching methods, addressing students' learning needs, and addressing poor students to improve academic performance among these students.

Mesfin Tadese et al.'s 2022 study found that 66% of university students in Southern Ethiopia, particularly those aged 20-24 and in medical/health science, had good academic performance. Smoking cessation was found to be crucial for these groups. The study also suggested that involving older students in the academic environment could improve academic outcomes. The findings underscore the importance of education in producing qualified human power and addressing community issues.

DATABASE

The secondary dataset, collected from Kaggle, includes data on gender, race/ethnicity, parental level of education, lunch, test preparation course completion, math score, reading score, and writing score. It allows for analysis of performance differences between male and female students, racial or ethnic groups, and socioeconomic factors. The dataset also provides total and average scores based on these scores, allowing for comparisons across various demographics. The data also provides insights into literacy and comprehension levels among students. The dataset also includes a total score and average score based on these scores.

METHODOLOGY

EXPLANATORY DATA VISUALIZATIONS

Explanatory Data Visualizations are charts or graphs used to share data analysis findings. They simplify charts by excluding irrelevant data and using styling features to highlight key messages. Carefully selected data is displayed in a way that clearly and concisely conveys the key message. These visualizations are typically used in the data analysis and interpretation phase, ensuring the key message is accessible to all users.

BAR GRAPH

A bar graph is a graphical representation of data using rectangular bars, with the length of each bar proportional to the value it represents. It is commonly used in business data and is drawn on a two-dimensional plane with the x-axis representing categories and the y-axis representing frequencies. Bar graphs are unique in having equal width and space between bars, can be drawn horizontally or vertically, and have a height equivalent to the data they represent. They are widely used in mathematics and statistics for easy comparisons between different categories and are the most widely used method of data representation in various industries. They help in studying patterns over long periods of time.

PIE CHART

A pie chart is a circular graph that represents data in a fraction or proportionate form, making it easier for average readers to understand. It can be represented with percentages and has limited use-case applications as it can only represent parts of a whole. Common uses include identifying growth areas in businesses, identifying market trends, and comparing the contributions of different marketing strategies or tools. Pie charts are useful for displaying the parts-to-whole relationship, demonstrating each element's contribution to the whole, and comparing the contributions of different categories.

HISTOGRAM

A histogram is a visual representation of a grouped frequency distribution with continuous classes. It is an area diagram consisting of rectangles with bases and intervals between class boundaries, with areas proportional to frequencies in the corresponding classes. The heights of rectangles are proportional to corresponding frequencies of similar classes and different classes. To plot a histogram, mark class intervals on the X-axis and frequencies on the Y-axis. The height of each rectangle is proportional to the corresponding class frequency if the intervals are equal, or the area of every individual rectangle is proportional to the corresponding class frequency if the intervals are unequal. Histograms are used for various purposes, including numerical data analysis, process changes, customer requirements analysis, and comparison of bar graphs and histograms.

DENSITY PLOT

A density plot is a smooth curve that represents the distribution of data in each range, rather than the frequency of values. It shows the proportion of data in each range, rather than the frequency of values. Histograms show the counts of values in each range, while density plots show the proportion of values in each range. Both histograms and density plots are useful for visualizing data distribution, but they show slightly different things. Histograms show counts, while density plots show proportions. Both can help understand patterns in data.

DOT PLOT

A dot plot is a data visualization technique that uses data points as dots on a graph to represent trends or groupings. It is similar to a histogram, displaying the number of data points in each category or value on the axis. There are two types: Cleveland and Wilkinson dot plots, commonly used by the Federal Open Market Committee (FOMC). Dot plots visually group data points based on their value, providing a quick analysis of central tendency, dispersion, skewness, and modality. They are typically used for smaller data sets and can be vertically or horizontally stacked for easy comparison. The spread of dots can be calculated by subtracting the minimum value from the maximum value on the chart.

QQPLOT

The QQ plot is a graphical tool used to assess if a set of data is plausible from a theoretical distribution like normal or exponential. It is a visual check, not an exact proof, but allows for an immediate evaluation of the assumption's validity. A normal QQ plot is a scatterplot created by plotting two sets of quantiles against each other. If both sets come from the same distribution, the points should form a straight line.

PAIRPLOT

A pair plot, also known as a scatterplot matrix, is a graph that visualizes the relationship between each pair of variables in a dataset. It combines histograms and scatter plots, providing a unique overview of the dataset's distributions and correlations. A pair plot is a statistical analysis technique that utilizes scatterplots, histograms, and boxplots to visually represent the relationships between different variables. The primary purpose of a pair plot is to simplify the initial stages of data analysis by offering a comprehensive snapshot of potential relationships within the data. Pair plots are crucial in Exploratory Data Analysis (EDA) as they enable data scientists to visually understand the distribution of single variables, identify relationships, and detect anomalies. The key elements of a pair plot are histograms, which show the distribution of a single variable, and scatter plots, which show the relationship between two variables.

STATISTICAL ANALYSIS

CHI SQUARE TEST

The Chi-Square test is a statistical method used to compare observed and expected data and determine if it correlates with categorical variables. It helps determine if a difference between two variables is due to chance or a relationship. This test is necessary to test a hypothesis about the distribution of categorical variables, which can be nominal or ordinal and cannot have a normal distribution due to their limited values.

$$\chi^2 = \frac{\sum_{i=1}^{n}(O_i - E_i)^2}{E_i}$$

The degrees of freedom in statistical calculations represent the number of variables that can vary, ensuring the validity of chi-square tests. These tests compare observed data with expected data based on a null hypothesis. Observed values (O_i) are collected by the observer, while expected values (E_i) are anticipated frequencies based on the null hypothesis.

Chi-Square tests consist of two main types: independence and goodness-of-fit. The independence test examines the likelihood of two sets of variables being related using counts of values for nominal or categorical variables. It requires a large sample size and independence of obseravations. The goodness-of-fit test determines if a variable is likely to come from a given distribution using data values and the distribution idea. Both tests are essential in statistical hypothesis testing.

t TEST

A t-test is a statistical tool used to evaluate the means of one or two populations using hypothesis testing. It can be used to determine if a single group differs from a known value, whether two groups differ from each other, or if there is a significant difference in paired measurements. To use a t-test, one defines the hypothesis and sets an acceptable risk of drawing a faulty conclusion. Then, a test statistic is calculated from the data and compared to a theoretical value from a t-distribution. The outcome determines whether to reject or fail to reject the null hypothesis.

The one-sample t-test is a statistical method commonly employed for small sample sizes or when the population standard deviation is unknown.

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

$\bar{X}$ is the sample mean, μ is the hypothesized population mean, s is the sample standard deviation and n is the observations in the sample.

The t-test statistic measures the standard errors of the sample mean ($\bar{x}$) from the hypothesized population mean (μ). A larger t-value indicates a larger difference between the sample mean and the population mean, as the sample mean is further from the peak of the distribution, suggesting a more significant difference.

ANALYSIS OF VARIANCE

ANOVA is a statistical test used to analyze the difference between the means of more than two groups, typically when data is collected about one categorical independent variable and one quantitative dependent variable. It determines if the dependent variable changes according to the level of the independent variable. ANOVA calculates the means of treatment levels from the overall mean of the dependent variable, rejecting the null hypothesis if any group means are significantly different from the overall mean. The F test compares the variance in each group mean from the overall group variance, finding a higher F value if the variance within groups is smaller than the variance between groups. ANOVA's assumptions are the same as general assumptions for any parametric test, including independence of observations, a normally-distributed response variable, and homogeneity of variance.

RESULT AND DISCUSSION

A bar chart is a pictorial representation of data, presenting categorical data with rectangular bars proportional to their values. This data set represents the gender of the dataset, with 518 females and 482 males.

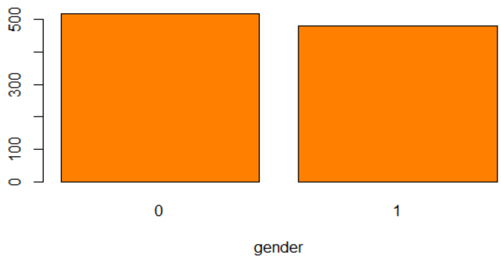

Figure1. Bar chart

A bar plot displays the number of students who received lunch and those who did not, with the height of the bar representing the data in each category, with 355 students not receiving lunch and 645 receiving lunch.

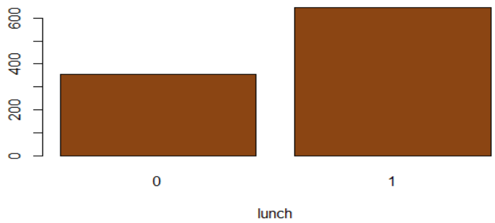

Figure2. Bar plot of receiving lunch

A bar plot displays the number of students who have prepared for and have not, with the height of the bar representing the data in each category, indicating 642 students who have prepared and 358 who have not.

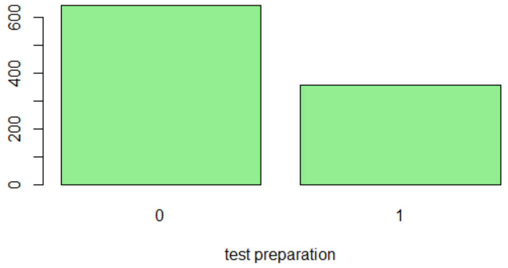

Figure3. Bar plot for test preparation

A pie-chart represents values as colored circles with corresponding numbers, representing race ethnicity. Group A represents 9%, B 19%, C 32%, D 26%, and E 14%.

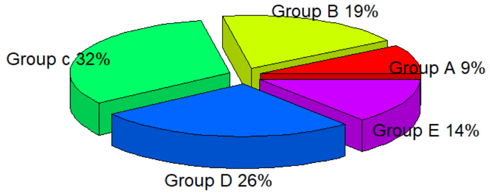

Figure4. Pie Chart for Race Ethnicity

A pie-chart is a visual representation of values as colored slices of a circle, with each slice labeled and corresponding to a number. This chart shows the percentages of parental education, with high school being 20%, some high school 18%, bachelor's degree 12%, some college 23%, associate's degree 22%, and master's degree 6%.

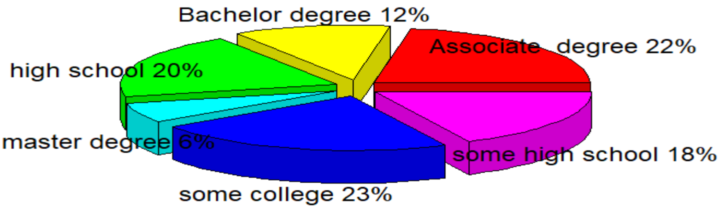

Figure5. Parental Education Pie chart

Histograms visually represent data distribution by displaying the frequency of values in a set, identifying common and least common values. They help visualize data types and bins, revealing that over 600 students scored over 150, while others had significantly lower scores.

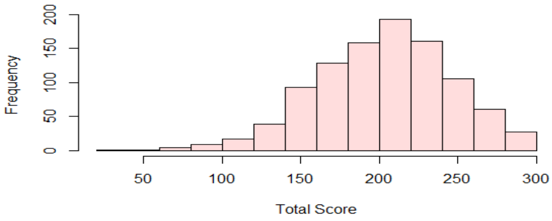

Figure6. Histogram for total score of the students

The visual representation of a histogram reveals that most students achieve over 50%, as depicted in the figure7.

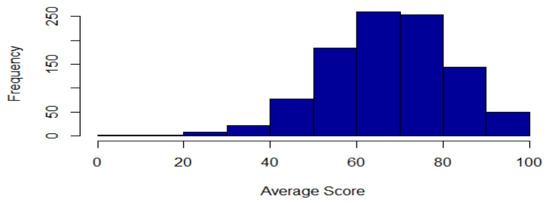

Figure7. Histogram for Average Score of the Students

A dot plot is a graph that uses dots to represent data points along a number line, useful for smaller data sets. It has a centre, middle, and spread calculated by subtracting minimum and maximum data values. In the figure 8, race ethnicity is visualized with mean and spread values for math, reading, and writing scores.

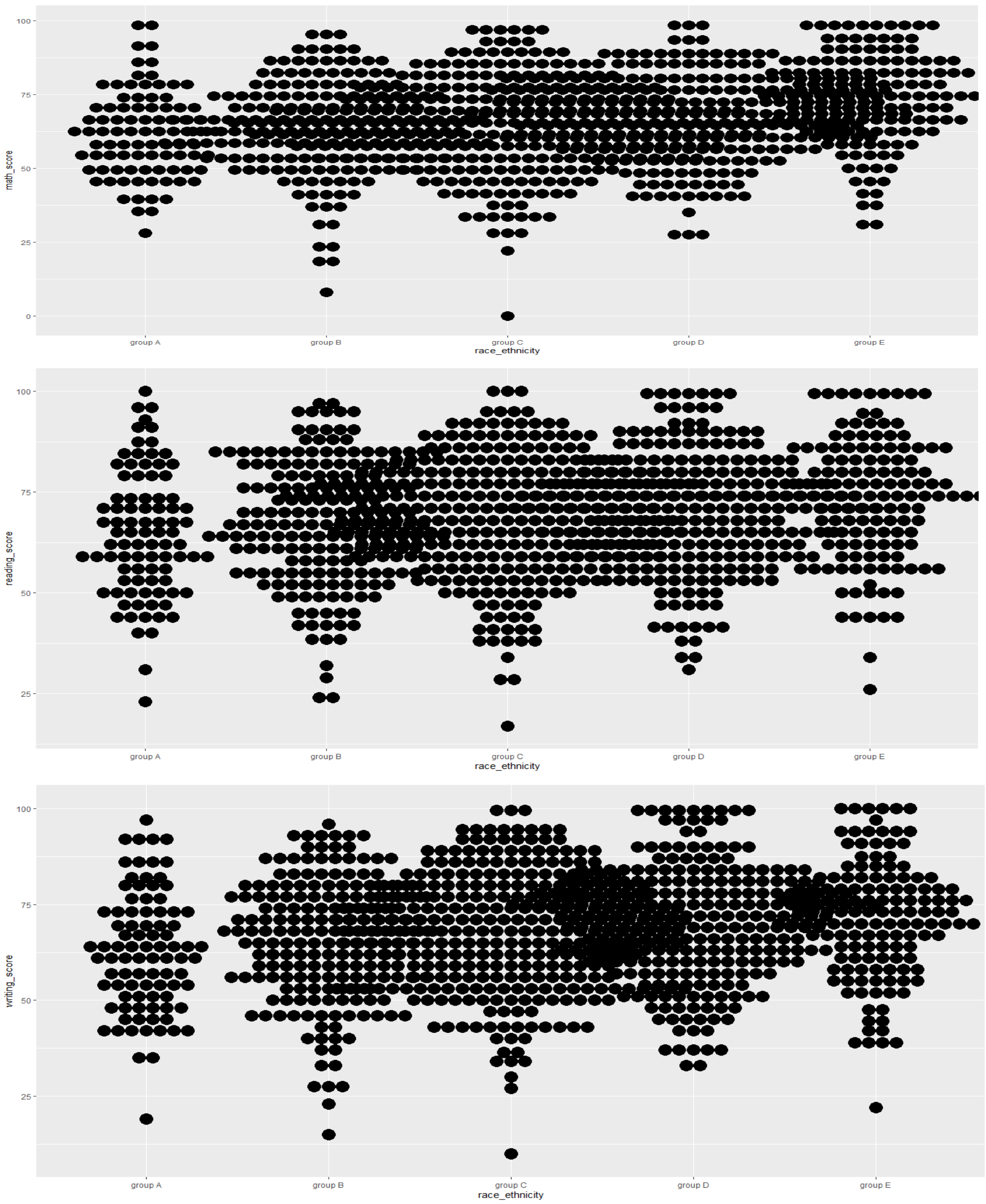

Figure8. Dot plot for Race Ethnicity

Figure 9 QQ plot is a graphical tool that compares two sets of scores, reading and writing, to determine if the data follows a theoretical distribution.

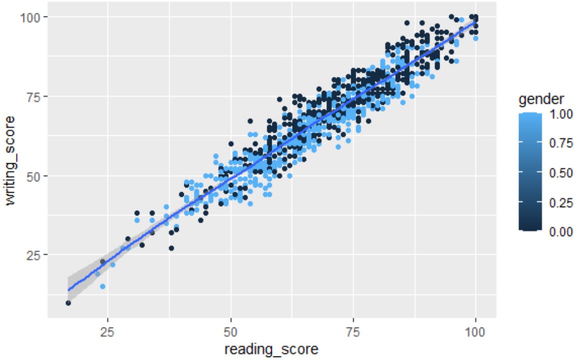

Figure9. qq Plot for Reading and Writing Scores of the Students

Pair plots are matrix-based visualizations that display the relationship between variables in a dataset, providing a sense of distribution and identifying potential issues. They combine histograms and scatter plots to simplify data analysis by offering a comprehensive snapshot.

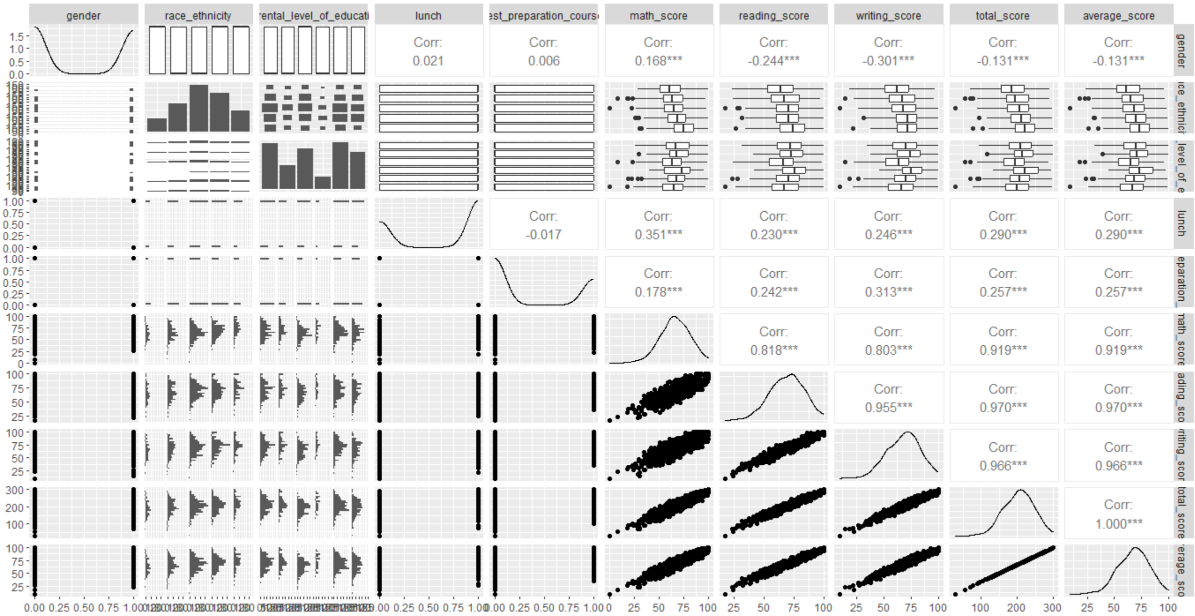

Figure10. Pair plots for correlation

Table 1 displays the number of students with different levels of parental education among the 1000 students.

Parental level of Education	Gender of the Students	
	Female	Male
Associate's Degree	116	106
Bachelor's Degree	63	55
High School	94	102
Master's Degree	36	23
Some College	118	108
Some High School	91	88
Total	518	482

Table1. Count of Students with their Parental Level of Education

Chi Square Test

Gender versus Test Preparation Course

H0: There is no significant difference between the gender and test preparation course.

H1: There is significant difference between the gender and test preparation course.

	0	1
Female	334	308
Male	184	174

Table2. Count of Students for their Test Preparation Course

The Pearson's Chi Square test results 0.015529 show no significant difference between gender and test preparation course, with a p value of 0.9008, which is greater than 0.05.

Gender versus Race Ethnicity

H0: There is no significant difference between the gender and race ethnicity.

H1: There is significant difference between the gender and race ethnicity.

	Female	Male
Group A	36	53
Group B	104	86
Group C	180	139
Group D	129	133
Group E	69	71

Table3. Count of Students for Race Ethnicity

The study found no significant difference between gender and race or ethnicity, as indicated by a Pearson Chi Square value of 9.0274 and a p value of 0.06042.

Race Ethnicity versus Test Preparation Course

H0: There is no significant difference between the race ethnicity and test preparation course.

H1: There is significant difference between the race ethnicity and test preparation course.

	0	1
Group A	58	31
Group B	122	68
Group C	202	117
Group D	180	82
Group E	80	60

Table4. Count of Students for their Test Preparation Course and Race Ethnicity

The Pearson's Chi Square (5.4875) test results show no significant difference between race ethnicity and test preparation course, with a value (0.2408) greater than 0.05.

Race Ethnicity versus Parental Level of Education

H_0: There is no significant difference between the race ethnicity and parental level of education.

H_1: There is significant difference between the race ethnicity and parental level of education.

	Associate Degree	Bachelor Degree	High School	Masater Degree	Some College	Some High School
Group A	14	12	18	3	18	24
Group B	41	20	48	6	37	38
Group C	78	40	64	19	69	49

Group D	50	28	44	23	67	50
Group E	39	18	22	8	35	18

Table5. Count of Students for Race Ethnicity and Parental Level of Education

The Pearson's Chi-squared test (29.459 with df = 20) results show no significant difference between Race Ethnicity and Parental Level of Education, with a p-value (0.07911) greater than 0.05.

t TEST

Hypothesis:

H_0: There is no significant difference between the average scores of the students.

H_1: There is significant difference between the average scores of the students.

Test Results:

The t-test results show a significant difference between average score of the students, with a t-value of 150.32. The p-value is less than 2.2e-16, indicating a lower probability of obtaining such a large difference. The alternative hypothesis is set to check if the true difference in means is not zero. The 95% confidence interval indicates that the true population means difference lies within the range of 66.88593 and 68.65540. The sample estimates indicate a significant difference between the averages, with a sample mean of 67.77067.

One Way ANOVA

Hypothesis

H_0: All sample means are equal or non-significant for gender of the student

H_1: At least one sample means differs from the rest for gender of the student

	df	Sum of square	Mean sum of square	F ratio	P value
Math score	1	7.05	7.05	64.73	2.43e-15
Reading score	1	109.68	109.68	100.73	<2e-16
Writing score	1	24.54	24.54	225.45	<2e-16
Residuals	996	108.41	0.11		

The p-value is less than 0.05, indicating significant differences between the scores in the model summary.

Hypothesis

H_0: All sample means are equal or non-significant for test preparation course of the student

H_1: At least one sample means differs from the rest for test preparation course of the student

	df	Sum of square	Mean sum of square	F ratio	P value
Math score	1	7.26	7.258	36.55	2.10e-09
Reading score	1	6.45	6.454	32.51	1.57e-08
Writing score	1	18.36	18.356	92.44	<2e-16
Residuals	996	197.77	0.199		

The p-value is less than 0.05, indicating significant differences between the scores in the model summary for test preparation course.

Hypothesis
H_0: All sample means are equal or non-significant for race ethnicity of the student
H_1: At least one sample means differs from the rest for race ethnicity of the student

	df	Sum of square	Mean sum of square	F ratio	P value
Math score	1	62.7	62.65	49.347	3.97e-12
Reading score	1	4.1	4.05	3.190	0.0744
Writing score	1	6.4	6.44	5.074	0.0245
Residuals	996	1264.6	1.27		

The model summary shows significant differences in math and writing scores for race ethnicity, with a p-value less than 0.05, and no significant difference in reading score and race ethnicity, with a p-value greater than 0.05.

Hypothesis
H_0: All sample means are equal or non-significant for parental level of education of the student
H_1: At least one sample means differs from the rest for parental level of education of the student

	df	Sum of square	Mean sum of square	F ratio	P value
Math score	1	16	15.589	4.680	0.0307
Reading score	1	3	2.786	0.836	0.3607
Writing score	1	8	8.101	2.432	0.1192
Residuals	996	3317	3.331		

The model summary reveals significant differences in math scores for parental level of education, with a p-value less than 0.05, and no significant difference in reading and writing scores.

CONCLUSION

In the 21st century, student performance is crucial for critical thinking, teamwork, creativity, digital literacy, cultural competence, self-direction, resilience, adaptability, and ethics. Project-based learning helps students acquire these skills and prepare them for a technologically sophisticated future. A dataset was used to analyze performance differences between male and female students, racial/ethnicity trends, family education, lunch availability, test preparation course completion, math scores, reading scores, and writing scores. The study found no significant difference between gender versus test preparation course and race ethnicity, race ethnicity and test preparation course, and race ethnicity and parental level of education. However, there were significant differences in reading and writing scores for parental education levels, reading score, and race ethnicity, but no significant differences for test preparation courses and gender. Overall, the study highlights the importance of continuous professional development in project-based learning for students to prepare them for a technologically sophisticated future.

The 21st-century education curriculum should emphasize critical thinking, imagination, teamwork, and digital literacy to improve student performance in a rapidly changing environment. Mixed learning environments can help develop research approaches and critical skills, but evaluating these abilities can be challenging due to their complexity. Professional liberty and innovative teaching strategies can help instructors be more effective and students excel in these disciplines. A good student's performance in the classroom should encompass a broader range of abilities, reflecting the capabilities of the twenty-first century. Performance-based examinations, tracking and evaluating student work over time, and consistent assessment criteria are essential for rubric development. Educational institutions should incorporate explicit instructions and opportunities for 21st-century skills in the curriculum, and teachers should receive continuous education and support to develop these skills in the classroom.

REFERENCE

1. Fiseha Berhanu, Addisalem Abera (2015), "Students' Performance Prediction based on their Academic Record", International Journal of Computer Applications, Volume 131, Issue 5, PP: 27 – 35.
2. An open repository and analysis tools for fine-grained, longitudinal learner data. K. Koedinger, K. Cunningham,A. Skogsholm, and B. Leber. Montreal, QC, Canada,: s.n., 2008. 1st Int. Conf. Educ. Data Mining,. pp. 157-166.
3. Selecting Optimal Subset of Features for Student Performance Model. Hany M. Harb, Malaka A. Moustafa. 2012, International Journal of Computer Science Issues, pp. 253-262.
4. Data Mining: A prediction for performance. Pal, B.K.Bharadwaj and S. No. 4, s.l. : International Journal of Computer Science and Information Security (IJCSIS), 2011, Vol. Vol. 9.
5. Educational Data Mining: A Review of the State of the Art. Romero, Cristobal. 2010, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, pp. 601-618.
6. Mesfin Tadese, Alex Yeshaneh and Getaneh Baye Mulu (2022), "Determinants of good academic performance among university students in Ethiopia: a cross – sectional study", BMC Medical Education, Volume 22, Article No: 395.
7. Regier J. Why is academic success important. Appl Sci Technol Scholarsh. 2011:1–2.
8. Karande S, Kulkarni M. Poor school performance. Indian J Pediatr. 2005;72(11):961–7.
9. Khan KW, Ramzan M, Zia Y, Zafar Y, Khan M, Saeed H. Factors affecting academic performance of medical students. Life Sci. 2020;1(1):4.
10. Mekonen T, Fekadu W, Mekonnen TC, Workie SB. Substance use as a strong predictor of poor academic achievement among university students. Psychiatry J. 2017;2017:7517450.
11. Ullah S, Sikander S, Abbasi MMJ, Rahim SA, Hayat B, Haq ZU, et al. Association between smoking and academic performance among under-graduate students of Pakistan, a cross-sectional study. 2019.
12. Trotter L, Wakefield M, Borland R. Socially cued smoking in bars, nightclubs, and gaming venues: a case for introducing smoke-free policies. Tob Control. 2002;11(4):300–4.
13. Wan Maziah Wan Ab Razak et al., (2019), "Academic Performance of University Students: A case in a Higher learning Institution", KnE Social Sciences, Volume 3, Issue 13, PP: 1294.
14. Cibarra, C. U, Hanseth, O (1998), Toward a contingency view of infrastructure and knowledge: an exploratory study. Paper presented at the Proceddings of the international conference on Information systems.
15. Hijazi, S and Naqvi, S (2006), "Factor Affecting Students Performance. A case of Private Colleges, Bangladesh", Journal of Sociology, Volume 3, Issue 1, PP: 12 – 17.

16. Khozaei, F., Ayub, N., Hassan, A. S., & Khozaei, Z. (2010). The factors predicting students satisfaction with iniversity hostels, case study, university sains Malaysia, Asia Culture and History, Volume 2, Issue 2, PP:148.
17. Singh, S Malik, S and Singh P, Research Paper Factors Affecting Academic Performance of Students.
18. Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. American psychologist, 35 (2), 151.
19. psychologist
20. students' satisfaction with university hostels, case study, universiti sains Malaysia.
21. Asian Culture and History, 2(2),