Analyzing the effect of certain factors on student performance in exam

Sara Shatnawi   
Data Science   
Jordan University of Science and TechnologyIrbid, Jordan  
[sashatnawi19@cit.just.edu.jo](mailto:sashatnawi19@cit.just.edu.jo)

Maram Aldiabat  
Data Science   
Jordan University of Science and TechnologyIrbid, Jordan

[mjaltheiabat19@cit.just.jo.jo](mailto:mjaltheiabat19@cit.just.jo.jo)

Lara Smadi

Data Science   
Jordan University of Science and TechnologyIrbid, Jordan   
[Lhalsmadi20@cit.just.edu.jo](mailto:Lhalsmadi20@cit.just.edu.jo)

*Abstract*

The observation strategy is most essential in the learning process, so observing the around factors for the student will affect their average at the end of the semester. Several factors affect student's performance or their average such as gender, race, lunch, test preparation, math score, writing score, reading score, and parental level of education. In this paper, these effects will be discussed by visualization processes to make the relation easy to discover, the student behaviors clear, and help in exploration for data and find the importance of the features. The result shows by map visualizations to show gender, education levels, and rate of lunch types depending on the area.

# ***Keywords—factors, education-levels, race, ethnicity.***

# Introduction

Students' production is an indispensable element in different institutions because it is one of the evaluation criteria of universities and schools, the criteria depend on the record of students' education level and achievements [1].

FARAH, I.S., MUHAMMAD, ABEERA, ABDUL studied the Influence of Examination Anxiety on students' performance. They apply their research among Medical Students in Lahore-Pakistan and used a cross-sectional descriptive survey at Central Park Medical College Lahore (CPMC). They did the statistical analysis using SPSS 20.00, their sample consist of (55%) female and (45%) male with ages between 20 to 25.

They made a comparison between these factors among male and female students and how anxiety affects them. The main result that they have got is more males doing physical exercise during the exam while more females were using anti-depressants than males during exams. So, they had more negative thoughts and self-criticism than males [2].

Dalia Alyahya, N.A has studied the effects of electronic exams on the student’s achievements in their exams. The sample consists of two groups: the experimental group and the non-experimental group. Their results show that there are no differences between writing, reading comprehension, and audio comprehension which refers means the electronic exams are better than traditional (face to face) exams. Their study is performed by interview 8 individuals and asks them a list of questions, then the data is analyzed using SPSS [3].

Azlina et al study the evaluation performances on Malaysian university’s facilities by data envelopment analysis. It is performed by a multi number of input data and output data. The input data such the number of employees in the facility, number of students, and the university budgets. The output such as the number of graduates and the number of published articles. Their result is that all facilities were efficient except one facility. Their methodology is using DEA is non-parametric technique is used for measuring the efficiency. The Azlina et al conduct this method to improve the performance of their university.

Their dataset is from 2009, their results were documented in the table which specifies the input and output for each facility, and so CF and FBA have the same efficiencies [4].

Balaram Duwal and Lalita Khonju conducted a study entitled Factors Affecting Student's performance: The case of Students at community college in Bhaktapur District in 2020. The aim of the study is to investigate the importance of teaching methods, student's study habits, and student's attendance on student's academic performance. Researchers used a survey method to collect data from 150 students to assess their opinions on their academic performance and the researchers used an analytical research design to explain the status of students. Regression analysis was used to identify the impact of independent variables on student's academic performance. Mann-Whitney test was used to identify the differences between perceived student's performance across gender, age, and faculty. Kruskal-Wallis test was conducted to test differences between perceived student’s performance across study year and guardian occupation. Teaching methods and student attendance have a positive significant impact on student's performance, student's habits have no significant impact, but socio-economic characteristics have no significant impact on student's academic performance. The study revealed that there is a low degree of a positive relationship between perceived teaching methodology and perceived student's academic performance. [5]

M.S. Farooq, A.H. Chaudhry, M. Shafiq, and G. Berhanu conducted a study entitled factors affecting student’s quality of academic performance (a case of secondary school level in 2011). The study aims to investigate the effects of socioeconomic status and parents' education on student's academic achievement in Math and English. Researchers used a survey method to conduct this descriptive study. They conducted a study on 300 males and 300 females of 10th Grade in Pakistan. Researchers design a questionnaire to collect data about variables of parent's education, parent's occupation, socio-economic status (SES), urban/ rural belongingness, and students’ gender. They applied descriptive and inferential statistical measures to analyze collected data. Multiple comparisons by applying ANOVA using SPSS 16 were used to explore the effects of various factors on student's achievement. Researchers concluded that there is a significant effect between socio-economic status and parents' education on student's achievement although girls' performance is better than males. They recommended conducting more studies on other samples and include more and new variables .[6]

# Research Question

Exploring and analyzing the influencing specific factors among student performance in different exams.

The students’ performance in the test should be discovered to enhance the infrastructure for their exams or improve their abilities, then improve the students’ scores by studying the factors that affect them such as gender, lunch, test preparation course, parental level of education, and their race. All these factors will affect positively or negatively the fulfillments of the students. If there is a factor effect negatively on the achievements such as the test preparation course, the school will work to complete the courses for their students or make them free to enhance their accomplishment.

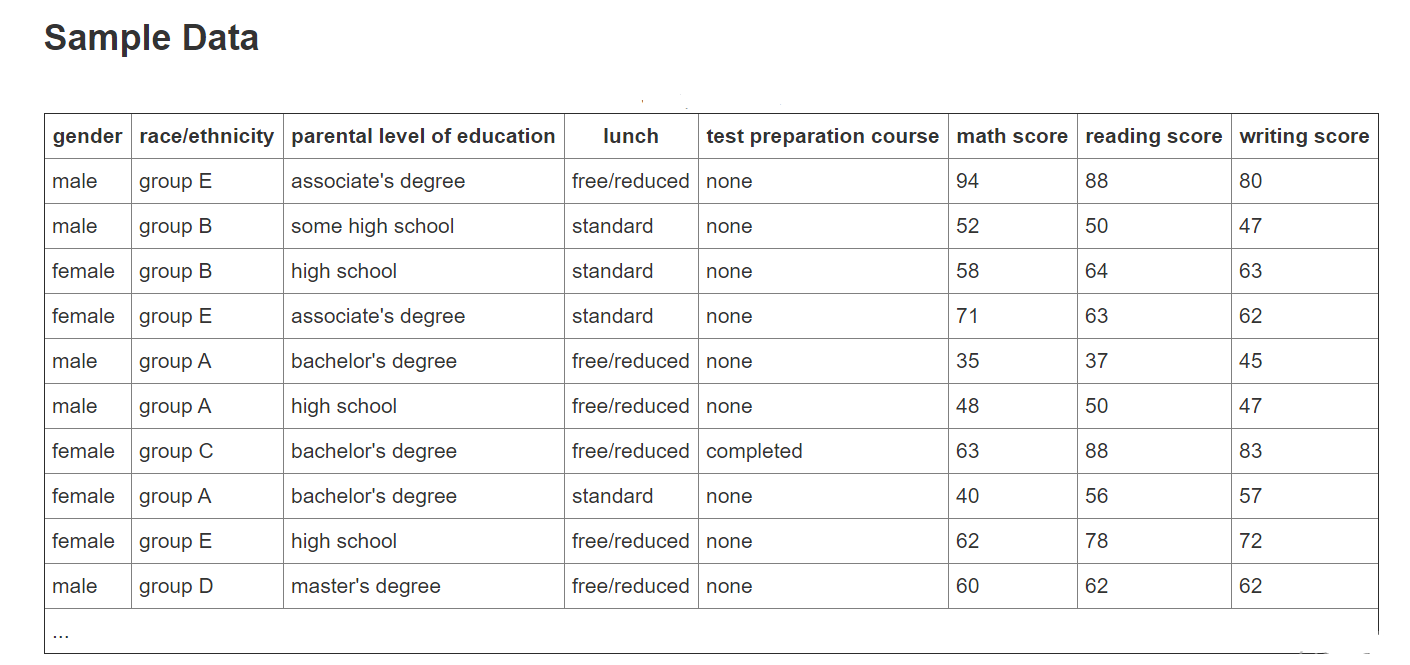
# Methodology

DATASET

The dataset has been used from Kaggle, consists of 1000 rows and 8 columns there are no null values, five parameters are categorical variables, and three are numeric variables.

The columns identifying as follow:

* Gender which represents the sex of the student consists of t2values (male, female)
* Race/ethnicity which means the race of the student has been divided into four groups (group A, group B, group C, group D).
* Parental level of education consists of six values: bachelor's degree, some college, master's degree, associate degree, high school, some high school.
* Lunch consists of two values: standard and free/reduced.
* Test preparation course: if the student completes the preparation course or not, has two values (none or completed).
* Math score: the score which is numbers between 0-100 that the student got in the math exam.
* Reading score: the score which is numbers between 0-100 that the student got in the math exam.
* Writing score: the score which is numbers between 0-100 that the student got in the math exam.



METHODOLOGY

After cleaning the data, Grouping is applied over the gender by males, females and find the mean and stander division in math, reading, writing scores to display the highest grades in both groups. Then apply the same steps on the race groups to show which is the highest groups get the high mean and sd. Also, studying the relationships between these variables and different factors.

To make it easier to understand the data, different visualization has been applied such as, the pairplot, relpolt, distplot, lineplot, the density and boxplot to show the relationship between two variables.

# Expoluration

There are set of graphs that illustrate the general information about paper dataset.

Such the percentage for each group of grades that between (A-D).

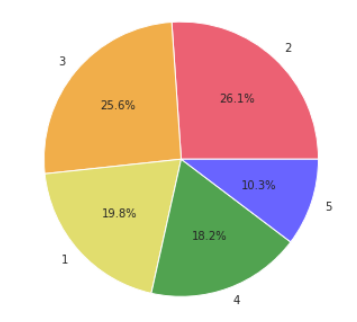


Fig.1 Total number of each grade

The graph shows that the total number of males and females is almost the same. there are 482 males and 518 females.

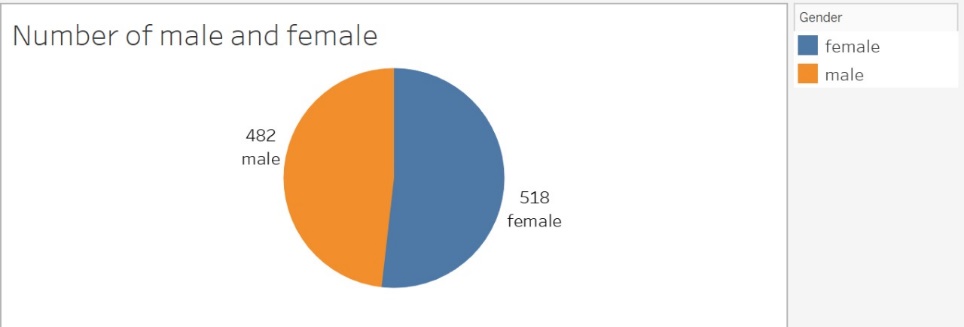


Fig.2 Total number of males and females

If we go deeper to investigate the number of males and females in each level of education, it found that almost there is a balance in their count unless in the master’s degree where the number of females exceeds the number of males.

Chart

Description automatically generated

Fig.3 Total number of males and females in education levels

Regarding the number of participants in each group, group C has the biggest number and group A is the smallest.

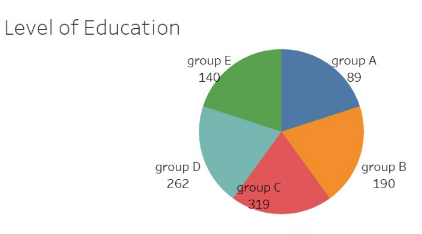


Fig.4 Total number of race groups/ ethnicity groups

The following graph gives more details for each education level. The changing in the size refers the number of a participant in each race.

Chart, pie chart

Description automatically generated

Fig.5 Total number of race groups/ ethnicity groups for education levels

The variance in color degree indicates the number of participants in each level of education, the higher hue means the higher number of participants.Chart, treemap chart

Description automatically generated

Fig.6 Total number for every education level

For more details, we can use this graph to analyze the number of males and females at each level.

Chart, treemap chart

Description automatically generated

Fig.7 Total number for every education level for both gender males/ females

Since the test preparation is one of the most important factors, it has been visualized to see how many students complete the test and how many did not. It clears the most students have not completed the test.

Chart, bubble chart

Description automatically generated

Fig.8 Total number for test preparation status in general

From the graph, it has been noticed that most students did not complete the test in all levels of education, maybe because they do not have time or faced some problems in the test.

Chart

Description automatically generated

Fig.9 Total number for test preparation status for each education level

The fig.9, fig.10, and fig.11 help in analyzing for each the minimum and maximum result in Math, reading, writing test in each level of education. where the values between 0 and 100.

Chart

Description automatically generated with medium confidence

Fig.10 Max and Min of math score for each education levels

Chart, scatter chart

Description automatically generated with medium confidence

Fig.11 Max and Min of reading score for each education levels

Chart, scatter chart

Description automatically generated

Fig.12 Max and Min of writing score for each education levels

The following visualizations show the differences between the average score of males and females in each race. The female average score is higher than the male average score, regardless of race.

The part of whole figures shows the percentage of pass and fail in each exam math, writing, and reading. The percentage of pass in reading is the higher percentage equals 91%, and percentage of fail in reading is lower percentage equals 9%. However, the lower percentage of pass is in math, and higher percentage of fail is in math.

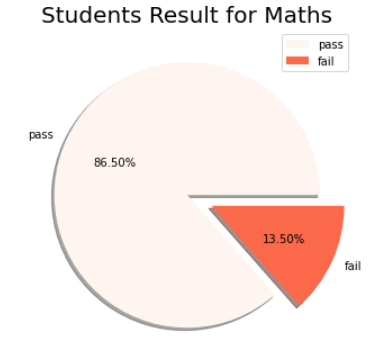


Fig.13 the student result for math

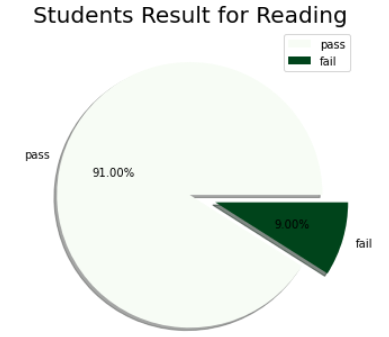


Fig.14 the student result for reading

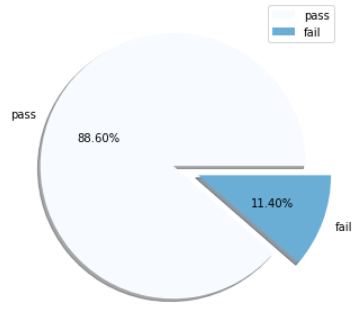


Fig.15 the student result for writing

The frequency of each grade (A, B, C, D, and F) in each education level is important to measure the failure and pass in each level. The result, in bachelor’s degree has the B grade is most frequent, in some collage the D is most frequent grade, in the master’s degree the A grade is most frequent, the associate degree has B is most frequent grade, in the high school has D is most frequent, and in some high school the B grade is most frequent.

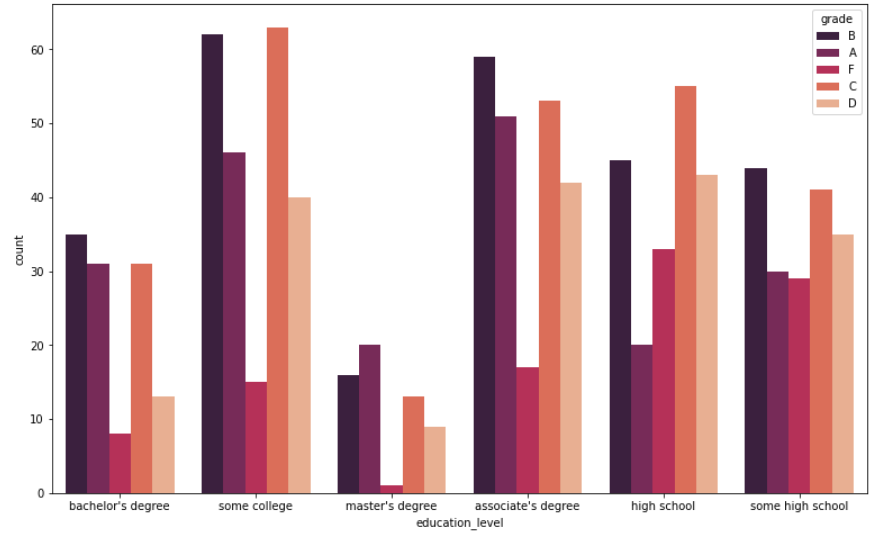


Fig.16 the number of each grade in each education levels

# Feature Importance

# Data visualization helps in determining the importance of each feature, the average score has the highest importance and the gender has the lowest importance.

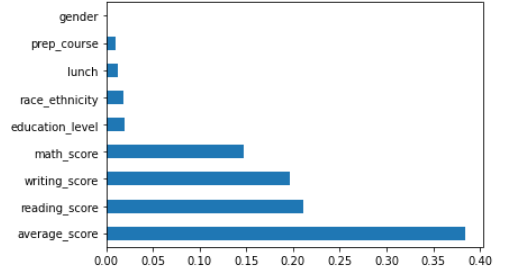


Fig.17 The importance of writing score, reading score, and math score features

# Results

This picture shows the general average of the student's total marks in the geographical regions, divided according to the level of education. When the education level was a master’s or bachelor’s degree, the general average of the students’ scores was high in most of the geographical areas specified on the map.

When education level was colleges, overall average student grades were average in most geography indicated on the map.

But when the education level was high school, the overall average of the student’s total scores decreased in most of the geographical areas specified on the map.

Map

Description automatically generated

Fig.18 The education levels depending on area

The picture below similar the previous one, but it also shows the percentage of males and females according to each geographical region and the size of the circle shows the number of students in each geographical region divided by the level of education.

.Map

Description automatically generated

Fig.19 Percentage of males and females according to geographical regions divided in education levels

The picture below shows the number of students in each geographical region and the number of males and females in each region, the largest number of students was in Russia, and the total number was 319, the number of females is 180 and males 139.

Map

Description automatically generated

Fig.20 The number of students in each geographical region and the number of males and females in each region

For more analysis of the factors, the below picture shows the rates of free and standard lunch in each geographical region, and does that affect the general average of student’s grades in these regions?

The lunch does not have that great effect on grades, in America the highest overall average for students ’grades, and the percentage of free and ideal lunches were equal, and in Canada, it was also equal, but the overall rate for student’s grades was low.

Map

Description automatically generated

Fig.21 The rates of free and standard lunch in each geographical region

The figure below shows the average score for males and females for each race group. The average scores for females are higher than males in all race groups.

A picture containing text

Description automatically generated

Fig.22 Average score for Male and Female among races.

The figure below shows the differences between males' and females' scores in reading, writing, and mathematics, broken down by race. The scores for females are higher than males in reading and writing, while the scores for males are higher in mathematics in all cases.

A picture containing graphical user interface

Description automatically generated

Fig.23 Average score for each score among races.

This visualization shows the average scores broken down by parental learning scores. Note that the completion of the test preparation leads to higher scores in every case.

Graphical user interface, application, Word

Description automatically generated

Fig.24 Average score for each education level.

This visualization illustrates the differences between males' and females' scores in reading, writing, and mathematics, broken down by parental education levels. The scores of females are higher in reading and writing than males, while the scores of males are higher in mathematics in all cases

.

Chart, bar chart

Description automatically generated

Fig.25 Average score for math, reading, and writing scores among education level.

The figure below shows the count for each grade among gender. Grade B is more frequent in males and grade F is less frequent in males. Nevertheless, grade C is more frequent in females and grade F is less frequent in females.

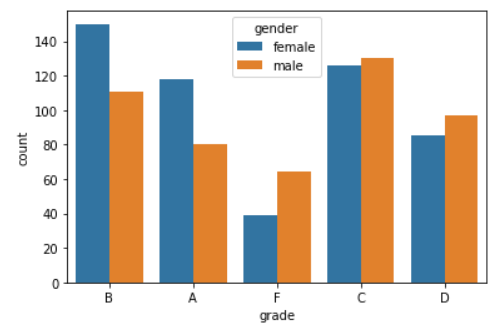


Fig.26 Total number of males and females in each group of grades

The race ethnicity consists of five groups each group has its properties and common features. The average score is calculated for each group and show the results in the next figure. The group E has the highest average score. Nevertheless, the group A has the lowest average score.

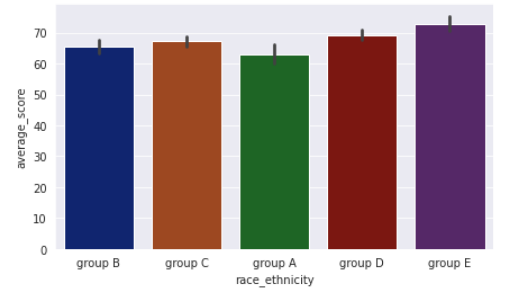


Fig.27 Average score for race groups

##### References

1. M. of Education Malaysia, National higher education strategic plan (2015). URL <http://www.moe.gov.my/v/pelan-pembangunan-pendidikan-malaysia-2013-2025>.
2. FARAH REHMAN , I.S., MUHAMMAD UMAR ZUBAIRI , MUHAMMAD UMAR, ABEERA SHAHZAD, ABDUL REHMAN, *Influencing Factors of Examination Anxiety among Medical Students in Lahore-Pakistan.* 2020. **Vol. 14**: p. 4
3. Dalia Alyahya, N.A., *The Impact of Electronic Tests on Students’ Performance Assessment.* International Education Studies, 2019. **Vol. 12**: p. 11.
4. Azlina Shaikh Awadz, R.R., Romiza Md Akhir, Chong Kim Loy, *Performance Evaluation of Faculties at a Private University A Data Envelopment Analysis Approach By Azlina.* Global Journal of Management and Business Research, 2012. **Vol. 12**: p. 9.
5. Balaram Duwal, a.L.K., *Factors Affecting Students' Academic Performance : The Case of Students at Community Colleges in Bhaktapur District.* The International Research Journal of Management Science, Dec 2020. **Vol. 5**: p. 16.
6. S. FARAH REHMAN , M.U.Z., MUHAMMAD UMAR, ABEERA SHAHZAD, ABDUL REHMAN, *FACTORS AFFECTING STUDENTS’ QUALITY OF ACADEMIC PERFORMANCE: A CASE OF SECONDARY SCHOOL LEVEL* Dec - 2011. **Volume 5**(2): p. 14.