Students' Academic Performance Dataset
(visualization with Seaborn, and Matplotlib)

Team Members:

Ruba Alkhattabi

Amani

Almarzoog

Abdullah

Alothman Razan

Alsallumi

## Instructor:

Mikio Harman

Introduction:

A Students' Academic Performance Dataset is an educational data set which is collected from a learning management system (LMS). Such a system provides users with synchronous access to educational resources from any device with an Internet connection. In our project, we have visualized the dataset in order to facilitate the extraction of information that helps in deducing results that may contribute to the development of education.

Exploratory Data Analysis:

Here we will show several of the plots that we made using the matplotlib and seaborn libraries. Which are two of the libraries used in the Python programming language to do data visualization.

- Plot of Raised Hands In Each Topic


Raisedhands In Each Topic
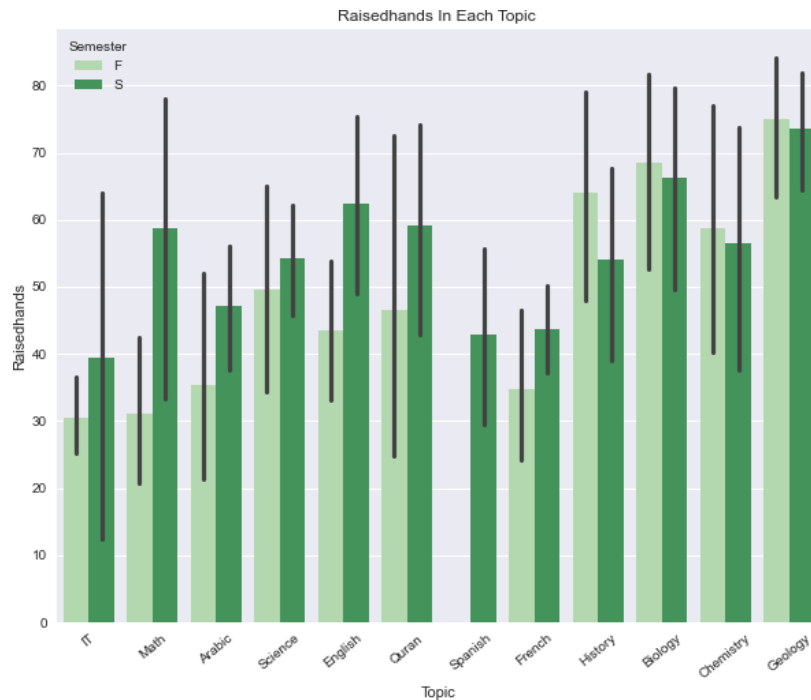
In this plot, we show the number of students raising their hands on each topic. The plot contains two bars for each topic. And each bar represents a semester. The Dark green means the first semester. And light green means in the second semester.

- Plot for Number of Students in each Stage ID

Count Of Stage ID
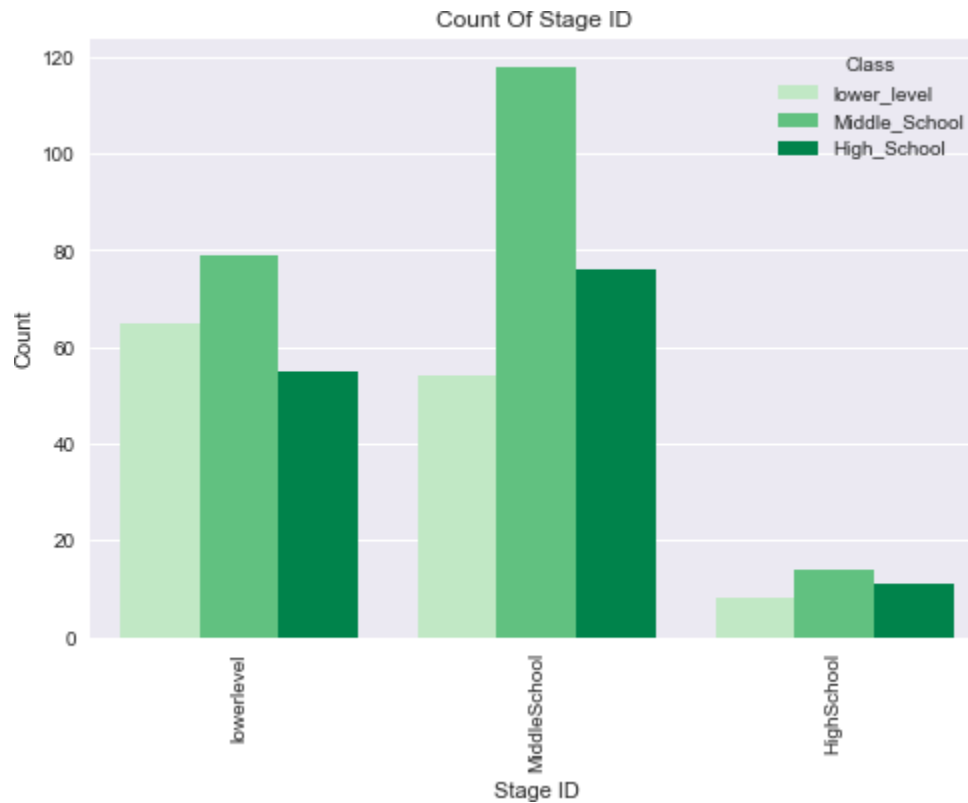
This plot shows the count in each Stage ID(Lower Level, Middle School, and High School). Also, shows us 3 bars, a bar for the low level, the second for the Middle School, and the third for the High School.

- Plot for Show Nationality VS Stage ID
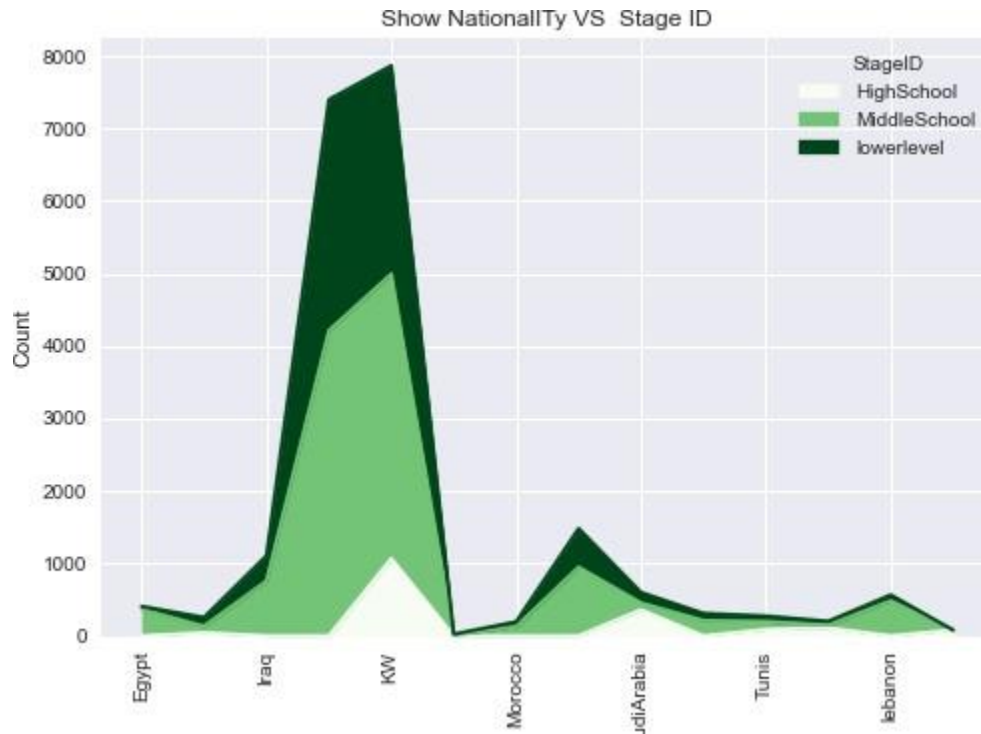
In this plot, we look at The Total Discussion sum over the Nationality in each school level. As you can see, Kuwait has the highest discussion count of all.

4

- Plot for show Grade ID VS Discussion

Boxplot GradeID VS Discussion
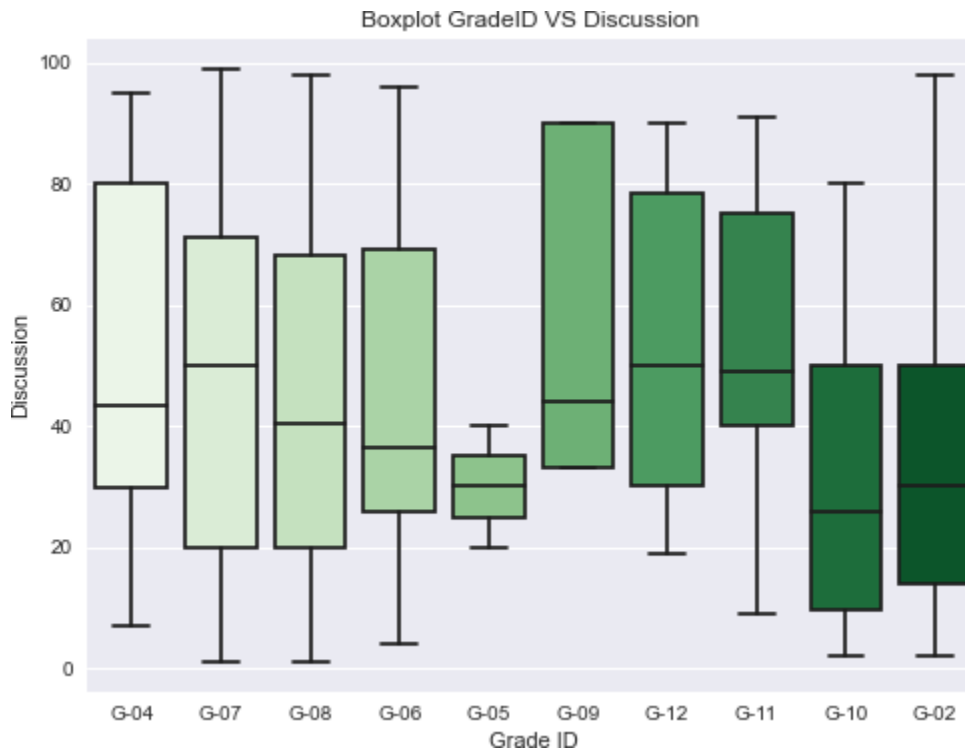
In this plot, show the grade Id as they represent from G1 to G12 in the dataset and their relationship with Discussions. This grade Id data means grade levels. The boxplot also represents the lowest value, the highest value, and the median value.

- plot for Count of Topics

Count of Visited Resources, Viewing Announcements and Discussion by Topic
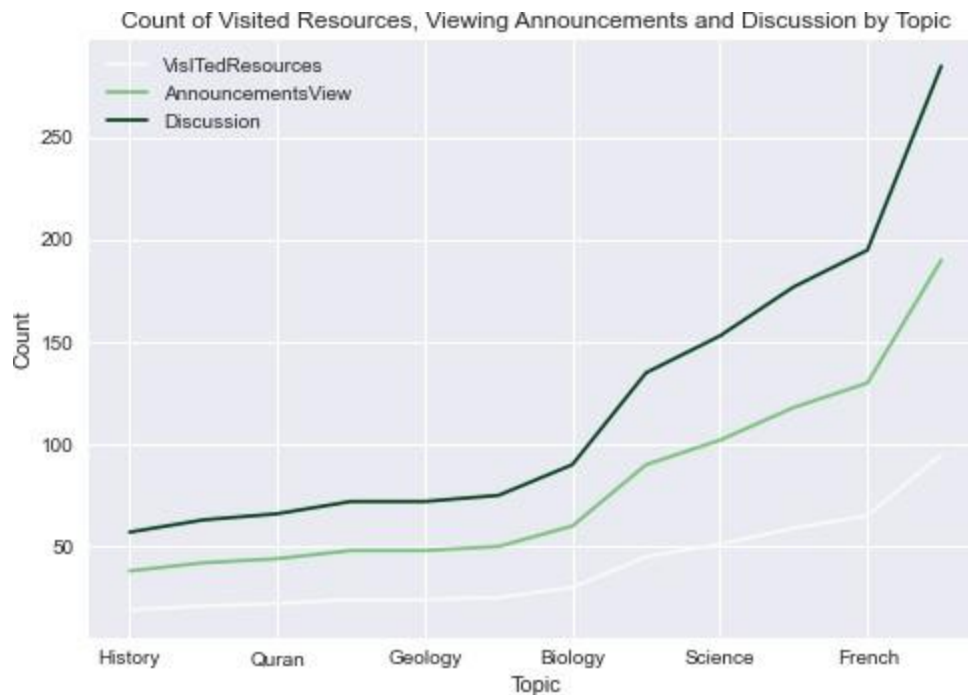
This plot describes the count of three features for each topic, as we see the discussion has the highest count for topics, and the visited resources have the lowest count for topics.

Also, French is the highest topic with a count of discussion around 200 times, a count of viewing announcements about 140 times, and a count of visited resources above 50 times.

And history is the lowest topic with a count of the discussion above 50 times, a count of viewing announcements about 50 times, and a count of visited resources below 25 times.

- Figure to Show Students Performance

Figure to Show Students Performance

In this figure, It contains a set of 4 subplots that were previously shown. There is a relationship between them that helps to extract results for the development of education.

- **Figure to Show Students Absence**

Figure to Show Students Absence Days
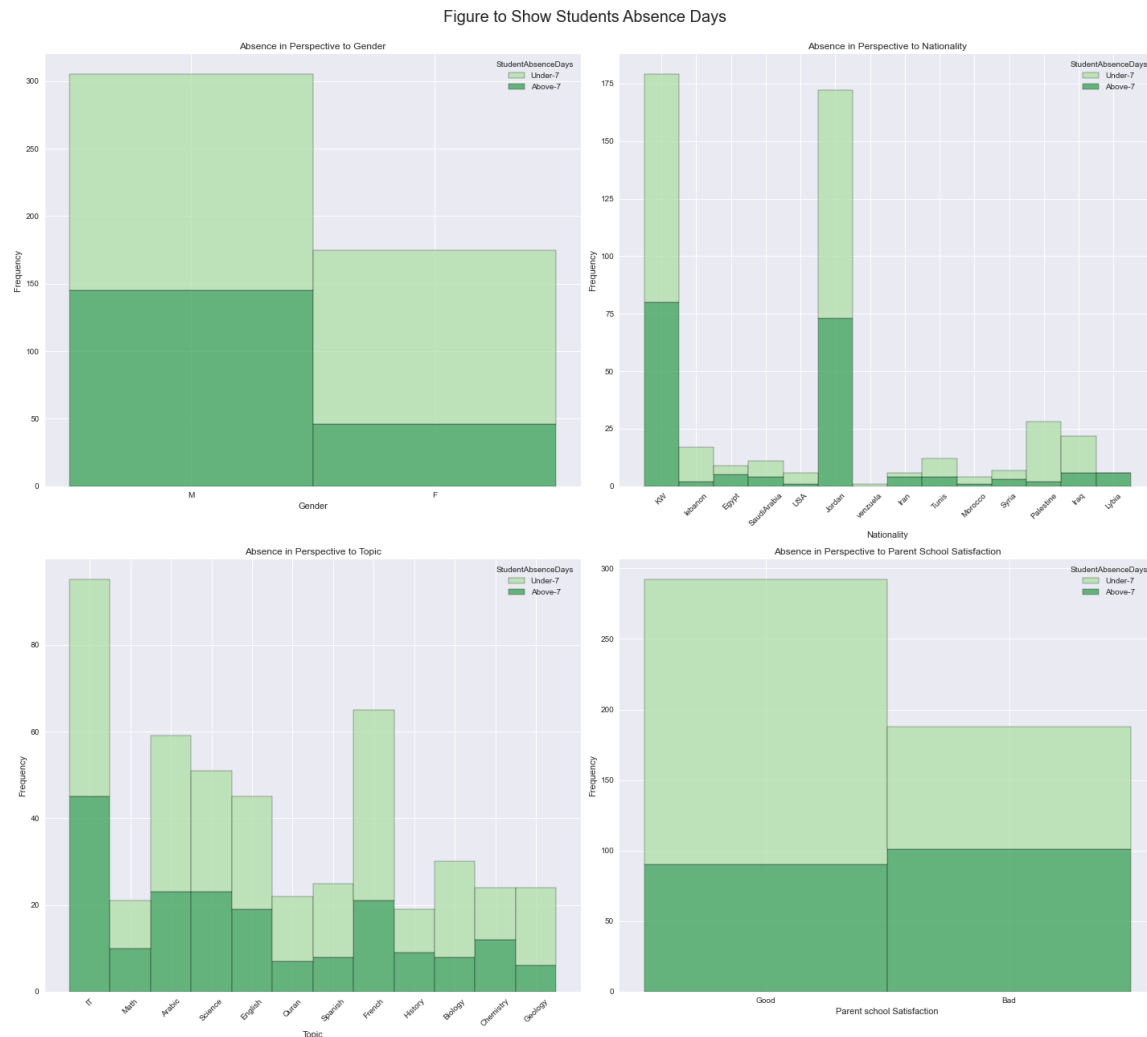
In this figure, we are exploring the effect of four variables on the student absence days from school. The first graph explores the role gender has to play in student absences, we can see that the percentage of Males absence above 7 days are higher than Females. Also, we have an almost 2 to 1 ratio of males to females in the data set.

The second graph looks at the nationality and whether they have an impact on absentees. We can see that most of the data are collected from Jordan and Kuwait, Palestine and Lebanon have a low absence percentage and Kuwait and Jordan have almost the same percentages.

The third graph explores the subject that the students are taught and whether it has an effect or not. We can say with confidence that the topic is neutral in terms of absence impact and most of the topics have identical rates.

The last graph is about parent satisfaction with the school and here we see a clue that might lead us to the correlation between the two variables. The higher your parents are satisfied with the school the lower your absence days you will have.

Summary of students' performance

When we come to the conclusion of students' performance in different grades and topics. We find that most of them have good performance. So, when focusing on the plot of Raised Hands In Each Topic, we see that students raise their hands in most subjects. Also, when we see in the plot for show Grade ID VS Discussion, we find that there is effectiveness between discussions and students' grades. Then, we see the extent to which parents are satisfied with their children. So, let see the extent to which parents are satisfied with their children in the figure to Show Students Absence. Then, we find that parent's satisfaction with the school is good, so this is a positive.

Conclusion

In this project, we've gone over several ways to plot a distribution plot using Seaborn and matplotlib with Python programming language.

Resources

[1]   Students' Academic Performance Dataset

[2]   Seaborn: Statistical Data Visualization