

Student Academic Performance Project Proposal

Analysis of Factors that contribute to Student Academic Performance

Lab Group - Rikki Kendall, Tolulemi Gbile, Geary Stonesifer

Packages

Data

```
# A tibble: 480 x 17
  gender Nationality PlaceofBirth StageID GradeID SectionID Topic Semester
  <chr>   <chr>         <chr>      <chr>   <chr>   <chr>   <chr> <chr>
1 M      KW          KuwaIT     lowerl~ G-04    A       IT    F
2 M      KW          KuwaIT     lowerl~ G-04    A       IT    F
3 M      KW          KuwaIT     lowerl~ G-04    A       IT    F
4 M      KW          KuwaIT     lowerl~ G-04    A       IT    F
5 M      KW          KuwaIT     lowerl~ G-04    A       IT    F
6 F      KW          KuwaIT     lowerl~ G-04    A       IT    F
7 M      KW          KuwaIT     Middle~ G-07    A       Math  F
8 M      KW          KuwaIT     Middle~ G-07    A       Math  F
9 F      KW          KuwaIT     Middle~ G-07    A       Math  F
10 F     KW          KuwaIT     Middle~ G-07    B       IT    F
# ... with 470 more rows, and 9 more variables: Relation <chr>,
#   raisedhands <dbl>, VisITedResources <dbl>, AnnouncementsView <dbl>,
#   Discussion <dbl>, ParentAnsweringSurvey <chr>,
#   ParentschoolSatisfaction <chr>, StudentAbsenceDays <chr>, Class <chr>
```

Set seed

Section 1 - Introduction

Our general research question is as follows: Is it possible to predict a child's class level based off of the information provided in the dataset? If so, which variables are of significance? The information we're going to be using to unearth the answer to this research question is from Kaggle.com. The information was collected from a Learning Management System called Kalboard 360. The data is collected from a learner activity tracker tool which tracks the students behavior, like if they're reading an article, watching an informative video, or doing their classwork. The variables used in this data set and their data types are as follows:

Variables

- Gender - Student's Gender (categorical)
- Nationality - The student's nationality (categorical)
- Place of birth - the student's place of birth (categorical)

- Educational Stage - Elementary, Middle, or High school? (categorical)
- Grade Level - Grade student is in (categorical)
- Section ID - Classroom the student belongs to (categorical)
- Topic - Course topic (categorical)
- Semester - School Semester (discrete numerical)
- Parent Responsible for Student (categorical)
- Raised Hand - How many times the student raises their hand in the classroom (continuous numerical)
- Visited Resources - How many times a student uses the course content (continuous numerical)
- Viewing Announcements - How many times the student checks the new announcements (continuous numerical)
- Discussion Groups - How many times a student participates in discussion groups (continuous numerical)
- Parent Answering Survey - Did the parent answer the surveys provided by the school? (categorical)
- Parent School Satisfaction - Is the parent satisfied with the school? (categorical)
- Student Absence days - How many times the student has missed school (categorical)
- Class - Low, Medium, or High, depending on their grades/marks at the end of the semester (categorical)

Section 2 - Data analysis plan

The aim of our project is to attempt to predict a child's class based on relevant variables provided in the data set. In order to do so we will be using two different variables: the 'outcome' (dependent, response, Y), which is the child's class (L,M,H) and the 'predictor' (independent, explanatory, X) which can be any combination of the aforementioned variables in section 1.

Predictor

Explanatory Variables: Any combination of the variables listed above.

Outcome

Response Variable: Class

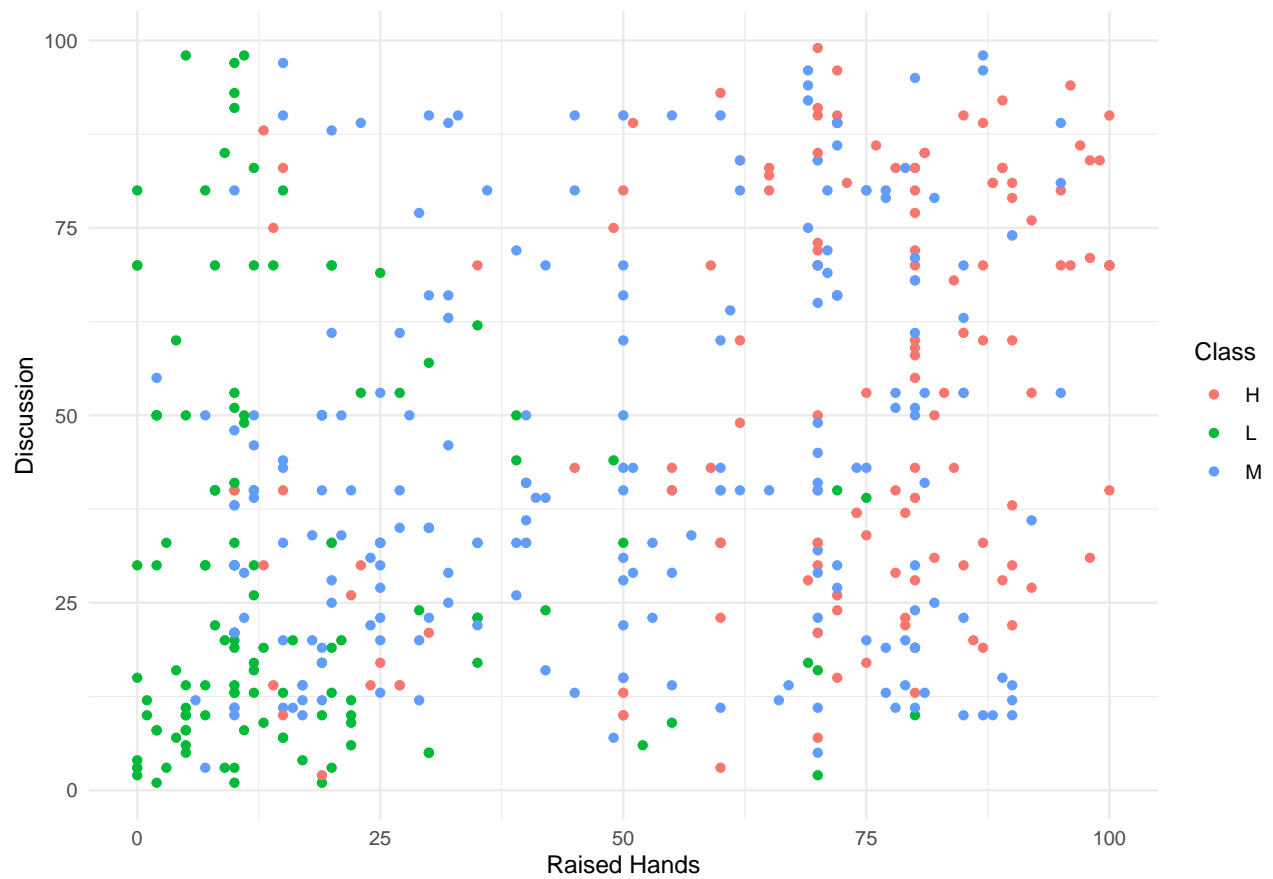
Summary Statistics/Visualizations

In order to attempt to prove that class is predictable, we will run several different linear regression models. With the given correlations expected, we should be able to extrapolate a child's class. However, what is crucial is to pick the most relevant of the variables. Initially, we will screen the variables based on intuition as to which are obviously not useful in determining class. Then we will those that are and run several linear regression models.

However, at first to portray the relevance of the variables we have provided several visualizations below to learn more about the data:

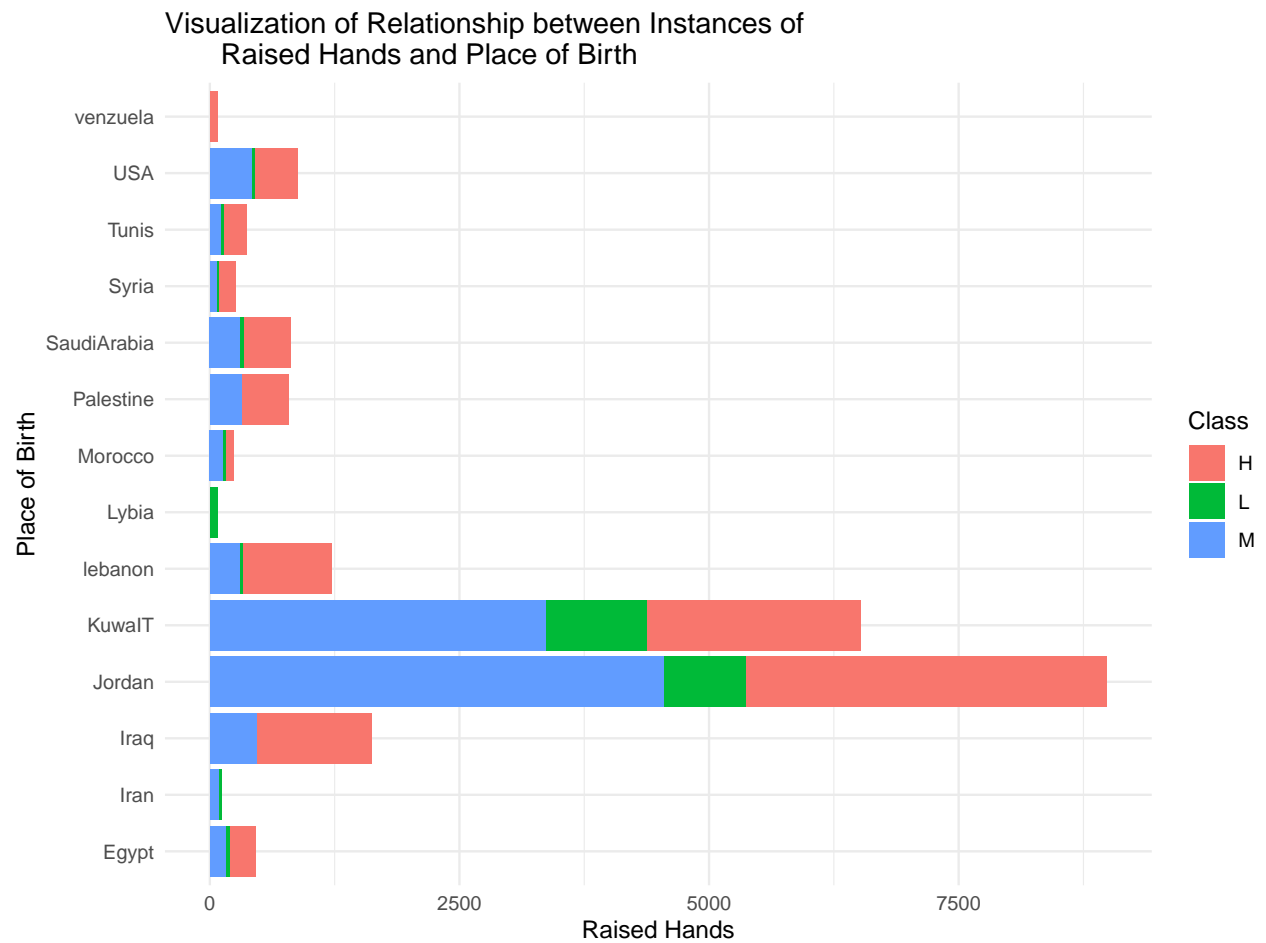
The first graph explores the relationship between the instances of raised hands and discussion. These two variables, although not necessarily dependent on one another, intuitively go hand in hand since both variables are related to class participation.

Visualization of Relationship between Instances of Raised Hands and Discussion

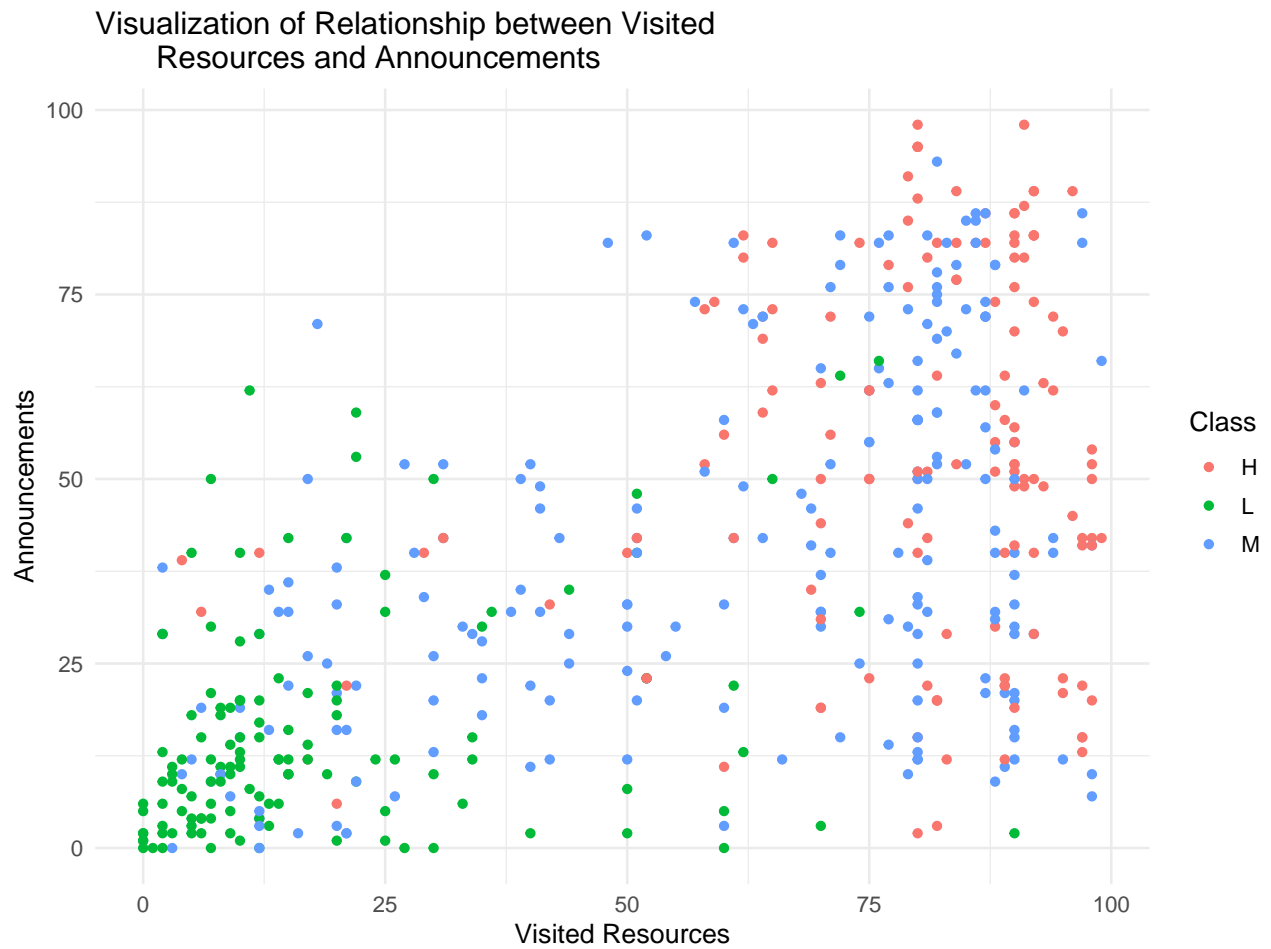


Given the disparity in results, whereby lower class is visibly on the lower spectrum of both Discussion and Raisedhands these variables will be important.

The following visualization corroborates the data in the prior scatter-plot for each country. This is to say that there is a clear trend among class in various countries given the variable, 'raisedhands.'

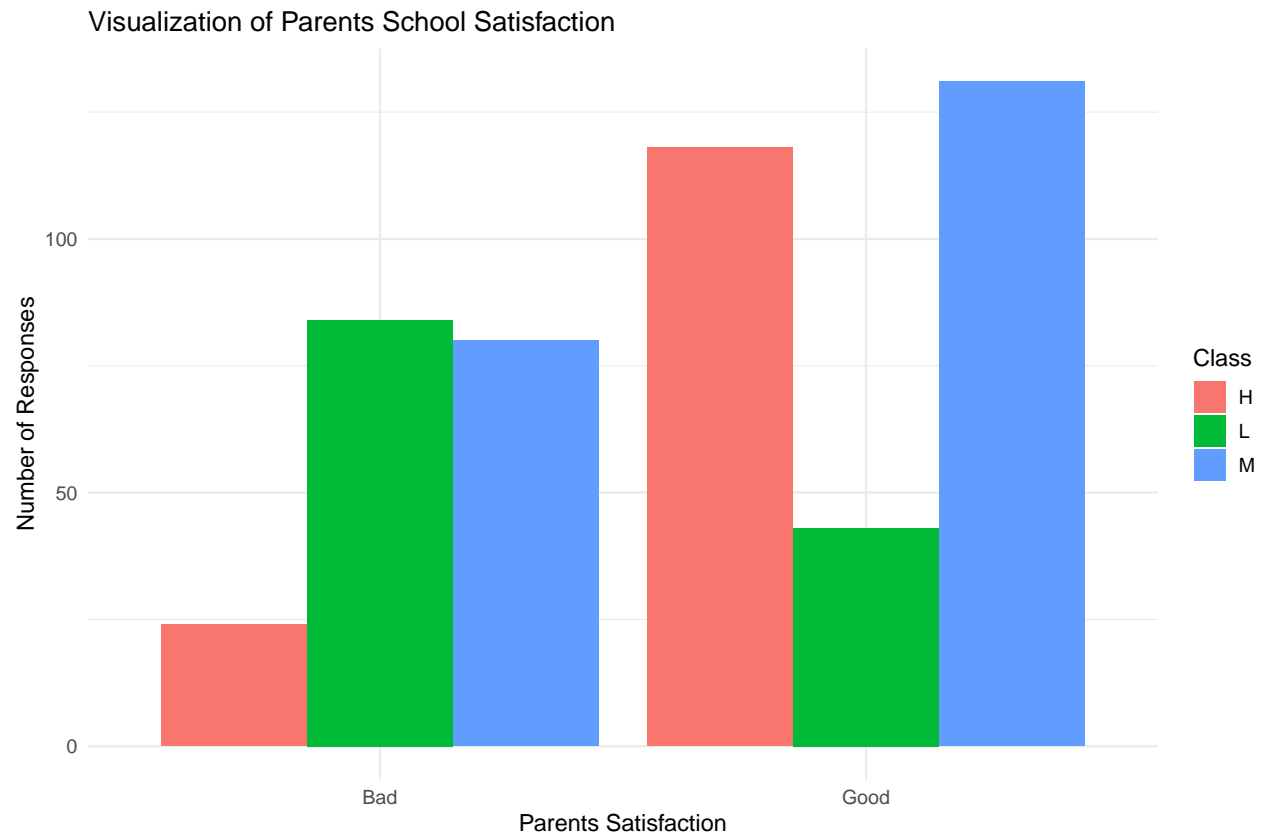


To continue to explore the variables, the next scatterplot takes two more data entries to examine a connection between class and Visited Resources / Announcements.



Again, we begin to see some trends when plotting relevant data side by side. Even more so in this visualization than the first scatter-plot, there is clear distinction between the classes of students.

Lastly, I have attempted to draw a connection to class using Parents satisfaction to examine a potential relationship with class.



This data provides interesting insight to the fact that in the Higher class bracket, there is a significant reduction in instances of negative feedback to any given school. Conversely, it appears the lower class bracket more often has negative responses to their school.

Given the visualizations of the variables above, it appears there are significant trends in the data that we aim to use to answer our research question: Is it possible to predict a child's class level based off of the information provided in the dataset?

Section 3 - Data

Observations: 480

Variables: 17

```
$ gender           <chr> "M", "M", "M", "M", "M", "F", "M", "M", "F...
$ Nationality      <chr> "KW", "KW", "KW", "KW", "KW", "KW", "KW", ...
$ PlaceofBirth     <chr> "KuwaIT", "KuwaIT", "KuwaIT", "KuwaIT", "K...
$ StageID          <chr> "lowerlevel", "lowerlevel", "lowerlevel", ...
$ GradeID          <chr> "G-04", "G-04", "G-04", "G-04", "G-04", "G...
$ SectionID        <chr> "A", "A", "A", "A", "A", "A", "A", "A", "A...
$ Topic            <chr> "IT", "IT", "IT", "IT", "IT", "IT", "IT", "Math"...
$ Semester         <chr> "F", "F", "F", "F", "F", "F", "F", "F", "F...
$ Relation         <chr> "Father", "Father", "Father", "Father", "F...
$ raisedhands      <dbl> 15, 20, 10, 30, 40, 42, 35, 50, 12, 70, 50...
$ VisITEDResources <dbl> 16, 20, 7, 25, 50, 30, 12, 10, 21, 80, 88,...
$ AnnouncementsView <dbl> 2, 3, 0, 5, 12, 13, 0, 15, 16, 25, 30, 19,...
$ Discussion       <dbl> 20, 25, 30, 35, 50, 70, 17, 22, 50, 70, 80...
$ ParentAnsweringSurvey <chr> "Yes", "Yes", "No", "No", "No", "Yes", "No..."
```

```
$ ParentschoolSatisfaction <chr> "Good", "Good", "Bad", "Bad", "Bad", "Bad"...  
$ StudentAbsenceDays      <chr> "Under-7", "Under-7", "Above-7", "Above-7"...  
$ Class                   <chr> "M", "M", "L", "L", "M", "M", "L", "M", "M..."
```