Kyle\_Walter\_Homework2

Kyle Walter

1/27/2021

Assignment, review the data from the 5 Schools implementing the same math course over teh summer to the 30 sections, (6 at each school). The course is currently about 3/4 of the way through. What is the story the data is telling about the course so far.

First let’s load the data into R so we’re able to work with

require(data.table)

## Loading required package: data.table

DF <- fread("data-storyteller.csv")

Now that the data is loaded let’s impsect it by school

DF[, .(VeryAhead=sum(`Very Ahead +5`), Middling=sum(`Middling +0`), Behind=sum(`Behind -1-5`), MoreBehind=sum(`More Behind -6-10`), VeryBehind=sum(`Very Behind -11`), Completed=sum(Completed)), .(School)]

## School VeryAhead Middling Behind MoreBehind VeryBehind Completed  
## 1: A 0 113 450 73 154 142  
## 2: B 0 84 201 14 22 125  
## 3: C 0 11 39 4 12 19  
## 4: D 0 3 8 2 6 3  
## 5: E 0 11 56 7 15 27

As we can see from a quick glance of the data, most of the students are behind overall. With no students Very ahead. And a small portion already completed.

Since no students are Very Ahead, defined as more than 5 lessons a head. Let’s if this makes sense. Since we know there are 35 lessons how many lessons would we expect the students to have completed at 3 quarters through the course.

35\*.75

## [1] 26.25

Roughly 27 should be completed at this point in the course. In order to be very head the students would have to have 32 lessons done, so this makes sense. The numbers would likley quickly jump through this category into completion at this point in the course. However; our behind categories ar quite high. Let’s dig into them a little bit.

First we’ll turn section into a character variable. We won’t be doing math on this value that R imported as a number.

DF$Section <- as.character(DF$Section)

Next we’ll melt the data set down to make it easier to analyze, and add percentage columns by section and by school. This will help us see if there are other problems going on in particular classes or schools.

DFMelt <- melt.data.table(DF)

## Warning in melt.data.table(DF): id.vars and measure.vars are internally  
## guessed when both are 'NULL'. All non-numeric/integer/logical type columns  
## are considered id.vars, which in this case are columns [School, Section, ...].  
## Consider providing at least one of 'id' or 'measure' vars in future.

DFMelt[, ':='(SectionTotal=sum(value)), .(School, Section)]   
DFMelt[, ':='(SchoolTotal=sum(value)), .(School)]  
DFMelt[, ':='(SectionPercent=value/SectionTotal, SchoolPercent= value/SchoolTotal)]

First let’s see if there anything interesting by school

DFMelt[variable=="Behind -1-5"|variable=="More Behind -6-10"|variable=="Very Behind -11",  
 .(SchoolPercent=sum(value)/mean(SchoolTotal)), .(School)]

## School SchoolPercent  
## 1: A 0.7263948  
## 2: B 0.5313901  
## 3: C 0.6470588  
## 4: D 0.7272727  
## 5: E 0.6724138

Over all schools are showing more than half the enrolled students are behind. Over all it does tell us that course will need to be re-evaluated to figure out where the problems are.

Is it merely that the students overall are just slightly behind or very behnd?

Behind <- DFMelt[variable=="Behind -1-5"|variable=="More Behind -6-10"|variable=="Very Behind -11"]  
  
Behind <- Behind[, (SchoolPercent=sum(value)/mean(SchoolTotal)), .(School, variable)]  
dcast.data.table(Behind, School~variable)

## Using 'V1' as value column. Use 'value.var' to override

## School Behind -1-5 More Behind -6-10 Very Behind -11  
## 1: A 0.4828326 0.07832618 0.16523605  
## 2: B 0.4506726 0.03139013 0.04932735  
## 3: C 0.4588235 0.04705882 0.14117647  
## 4: D 0.3636364 0.09090909 0.27272727  
## 5: E 0.4827586 0.06034483 0.12931034

Overall the students classifed as behind the course are within 5 lessons of being on track. This is good news as they have a potential to catch up and complete the course on time with a little extra effort. That said, with this large of group, it would be nice to see if there was a specific lesson that most of the students were struggling with. The data however, does not contain such information.

From this view of the data, school D looks like a lost cause with 27% of their students very far behind. However; when looking at the total students of School D we can quickly see that percents alone are not very telling

DFMelt[School=="D", .(TotalStudents=sum(value)), .(School, Section)]

## School Section TotalStudents  
## 1: D 1 22

As there is only 1 section and 22 total students.

The percent of students in the “More Behind” classifcation runs from about 3 to 8 percent. School A has the 2nd highest percentage, but overall the most students of the schools.

Schoo A also has the 2nd highest average students per session, which puts in question for those students, who are more behind or very behind, will likely lack access to the resources to get caught up.

School E which has only 1 section will likely also show low finish rates amoung its students in the last two buckets as it has 122 students in its signle section being offered.

DFMelt[, .(SectionAverage=mean(SectionTotal)), .(School)]

## School SectionAverage  
## 1: A 71.69231  
## 2: B 37.16667  
## 3: C 28.33333  
## 4: D 22.00000  
## 5: E 116.00000

In conclusion, while the data tells us at this point in time large section of students are classified as behind, it does appear they are still within reach being complete by the end of the course as they are within 5 lessons of being on schedule. Additional data should be collected if availabe to see if a particular lesson caused the students to fall behind. Another consideration is what is the student’s math history prior to the course, is this one at too high a level or do we have any seperation of students based on prior learning or ability?

The other consideration that should be reviewed for schools A and E is given the high number of students in a session, is there only 1 teacher or do the students also have access to teaching assistants. If the former, one of the way to improve would be to add additional teachers so the students will have ample access to assistance needed for the course.