**Data Story by Karan Ashar ( kashar@syr.edu )**

I started of with some basic pre-processing steps. The column names were fixed first. A new column called ‘total’ was also created. ‘Total’ was the sum of all the students in each section.

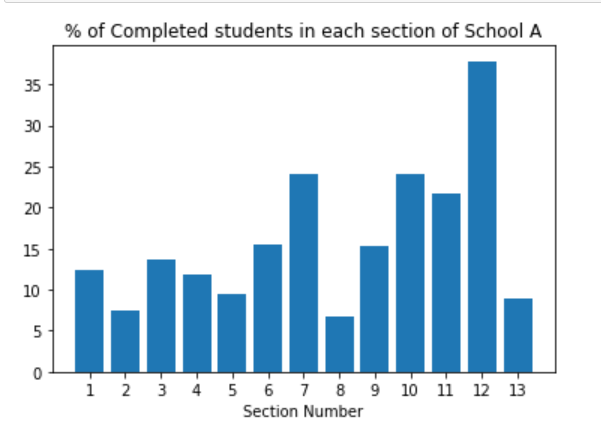
I started with specific school-wise analysis.

**School A**

**Note-** All the numbers of the students preset in each columns were converted to percentages so that they could be compared to each other.

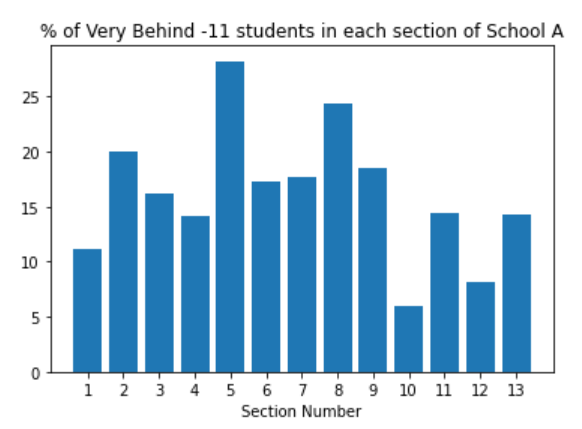
School A has 13 sections and a total of 932 students.

Now we see a distribution of the students section-wise who lie in the Completed category



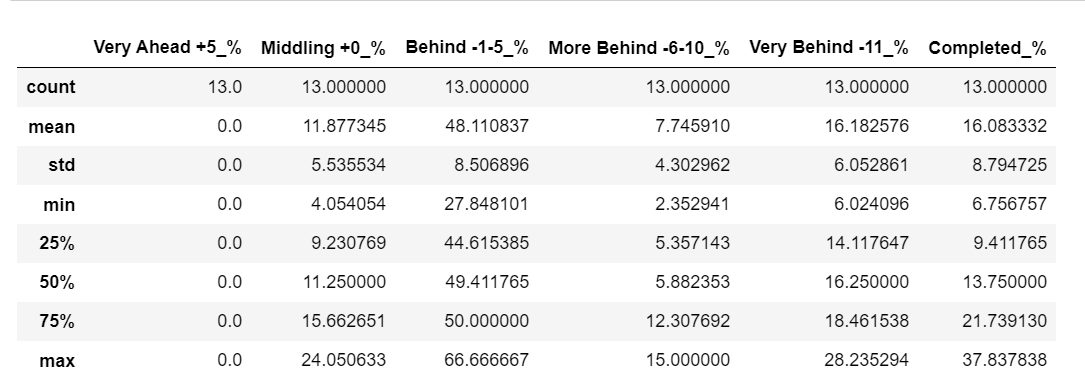
If we see Section 12 has the highest percentage of students in the Completed category with section 8 having the least percentage of students in this category.

Below is the distribution based on the Very-Behind -11 category.



Here we can see section 5 having the highest % of students in the Very Behind -11 category. This is putting a bad impression of section 5.  
Here we can see section 10 is doing the best. They have the least % of people in the Very Behind category.

Below we see a summary of the columns which are converted to percentages. This is to make sure we are doing a correct comparisons and drawing correct conclusions.  
This data is for the entire School\_A



Here we can say that, for School A - on average 48% of the students lie in the Behind -1-5.  
Similarly we can say that for School A – on an average 16% of the students have completed the course.

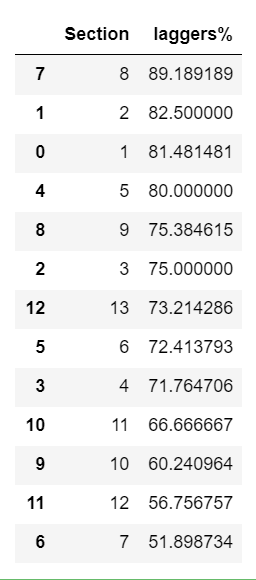
Next, I divided the categories into two classes – gainers and laggers.  
The categories in gainers were – [ Middling +0, Very Ahead +5, ‘Completed’]   
The categories were laggers were basically the ones which are not in the gainers.  
laggers - [ Behind -1-5, Very Behind -11, More Behind]

Using these 2 lists I created 2 new variables – gainers% and laggers%

gainers% is the percentage of total students that lie in one of the categories defined by the gainers list.

laggers% is the percentage of total students that lie one of the categories defined by the laggers list.

Belowwe can see the laggers% in descending order.

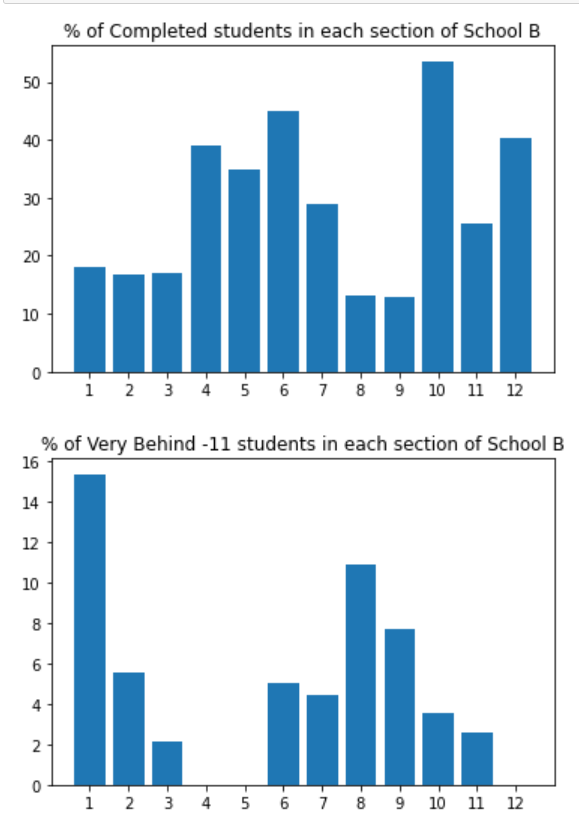


As we can see section 8 performs the worst. 89% of its students are laggers.  
Whereas section 7 is performing the best. 51% of its students are laggers. This also means that around 49% of its students are gainers which is the highest for School A

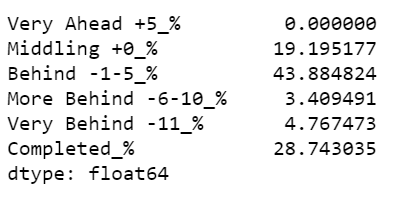
**School B**

Similar analysis was done for School B as well.

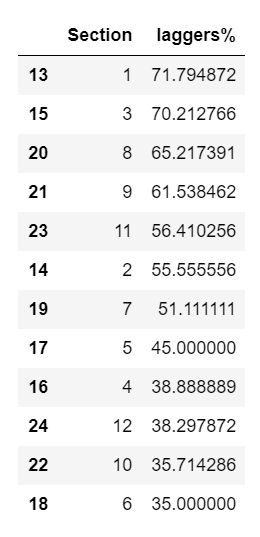
The plots below show that:  
The section with the highest % of Completed students is section 10.  
The section with the highest % of Very Behind students is section 1.  
One interesting thing to note here is that section 4, 5, and 12 do not have any students in the Very Behind -12 category. That’s impressive.



On an average 43.8% of the students lie in the Behind -1-5 category for every section



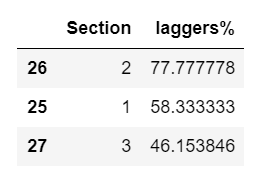
We can see below that section 1 has the highest% of laggers. While section 6 is performing the best with only 35% of laggers.



**School C**

Similar analysis was done for School C

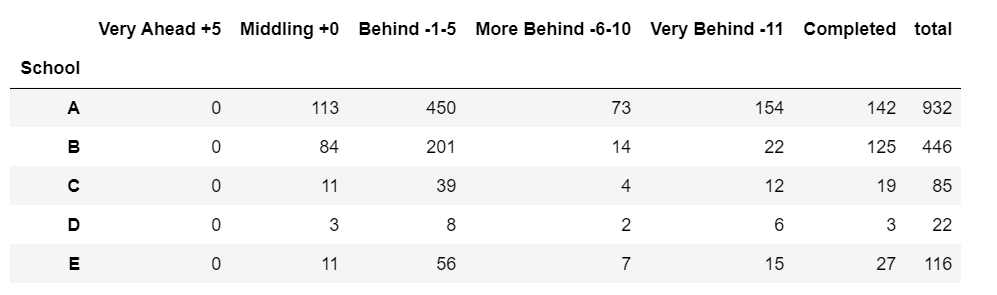
Here to note was that section 2 was performing the worst. Where as section 3 was the best performer amongst the 3 sections



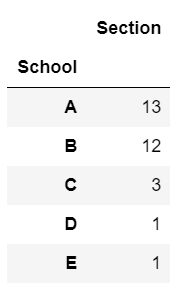
**All School Analysis**

Now I have done some analysis to compare School rather than sections as done before.

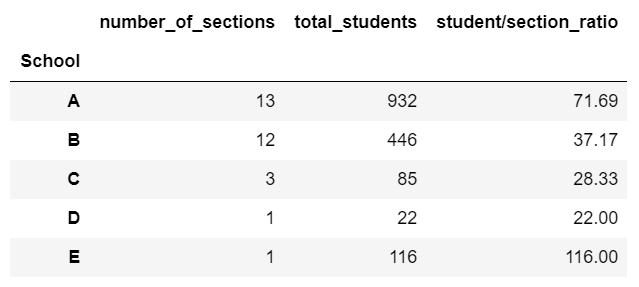
Below we can see the total number of students per school as well as the number of students in each of the categories per school.  
We can see School A has the highest number of students where as School D has the least number of students.



We now take a look at the number of sections in each school. School A has the highest number of sections

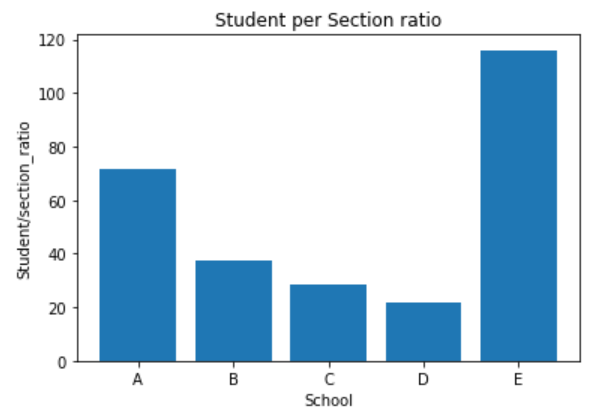


I then combined the above two numbers and created a new feature - ‘Students per section’. It is a ratio of the number of students to the number of sections in that school

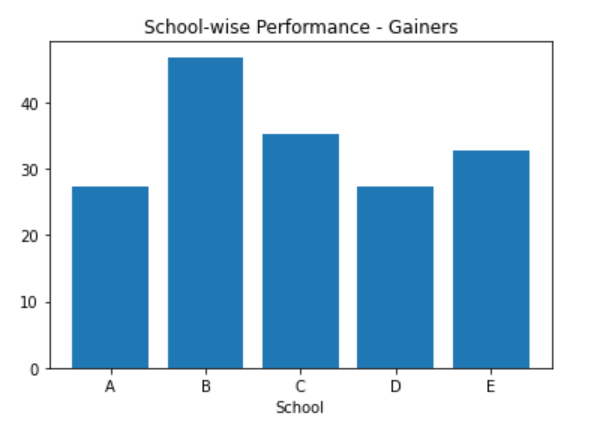


We can see that the class size of School D is the smallest. School E has the highest class size.

Below is a bar chart representing the same table:



Similar variables like before were created of the School Analysis. Gainers% and Laggers% were created.



Here we can see School B has the best performance. It is the highest% of gainers.   
School A and School D are at the bottom and share a similar distribution.

Below we can see School D is the worst, followed by School A by a very narrow margin.

