AF DATA ANALYSIS

Chen(Cici) Chen cc4291@columbia.edu

1/2/2020

Goal

The goal of this analysis is to research the potential trend and results of the students from two AF middle schools: **Bushwick Middle School** and **Crown Heights Middle School**. In order to close the achievement gap for the students who have low performance within the neighborhoods, AF will utilize the results of this analysis as a reference to deliver equal educational opportunities for all the students regardless of their race and economic status. Beyond this analysis, AF as an engaged parter in the community will take this exercise as an example to discuss how to improve public education quality and enhance the students' academic performance in the larger conversation.

Introduction

The original dataset contains the information of students' **School Name**, **Grade level(5th, 6th)**, the student's beginning of year (**BOY**) and end of year (**EOY**) F&P reading scores.

To begin with, I deleted the dataset with missing values and the potential entry error of the dataset. It is of great importance to deal with those missing values and outliers, and after a careful check, I find that moving those values will not influence the big picture of our dataset. Take **BOY F&P Score** as an example(it is the column where most N/A comes from), the average of the score has a minor difference as you can be seen from the following result: With N/A mean=12 and without N/A= 12.9478, and the difference is 0.947761, which is small.

To better analyze the data, we need to clear up the dataset first before further analysis. I converted F&P scores to the corresponding F&P proficiency levels and rewrote the Grade levels/School names to a consistent format.

Now, the following shows how our dataset looks like:

Student.ID	School.Name	Grade.Level	BOY.F.P.Score	EOY.F.P.Score	BOY.PL	EOY.PL
10000001	Bushwick Middle School	5.0	11	16	Below Proficient	Advanced
10000002	Bushwick Middle School	5.0	11	16	Below Proficient	Advanced
10000003	Crown Heights Middle School	5.0	11	16	Below Proficient	Advanced
10000004	Bushwick Middle School	5.0	11	16	Below Proficient	Advanced
10000005	Bushwick Middle School	5.0	11	14	Below Proficient	Proficient
10000006	Bushwick Middle School	5.0	11	10	Below Proficient	Remedial

Additionally, this graph indicates the distribution of the number of students in each middle school. It is easy to see that most students in our analysis come from Bushwick Middle School and 5th grade of Crown Heights Middle School, and only a few students(15) come from 6th grade of Crown Heights Middle School.

As such, we need to consider the percentage of the students in the following analysis rather than numbers themselves to avoid the bias caused by the difference of total number of the students.

Students Distribution of Two AF Schools
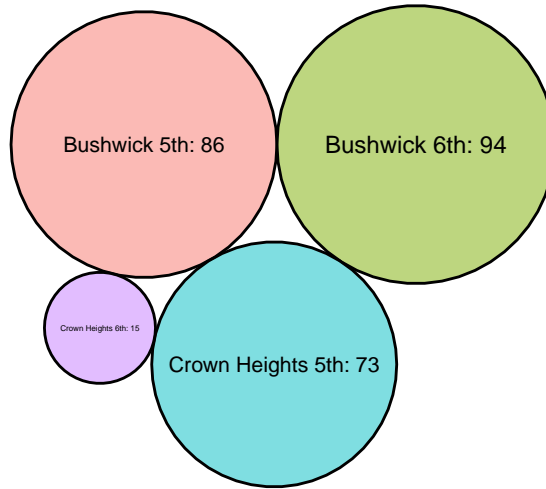
Let's get more detailed information at some statistics from our dataset, which shows that the mean of the F&P scores is generally improved after one-year education, no matter which school and grade the student is.
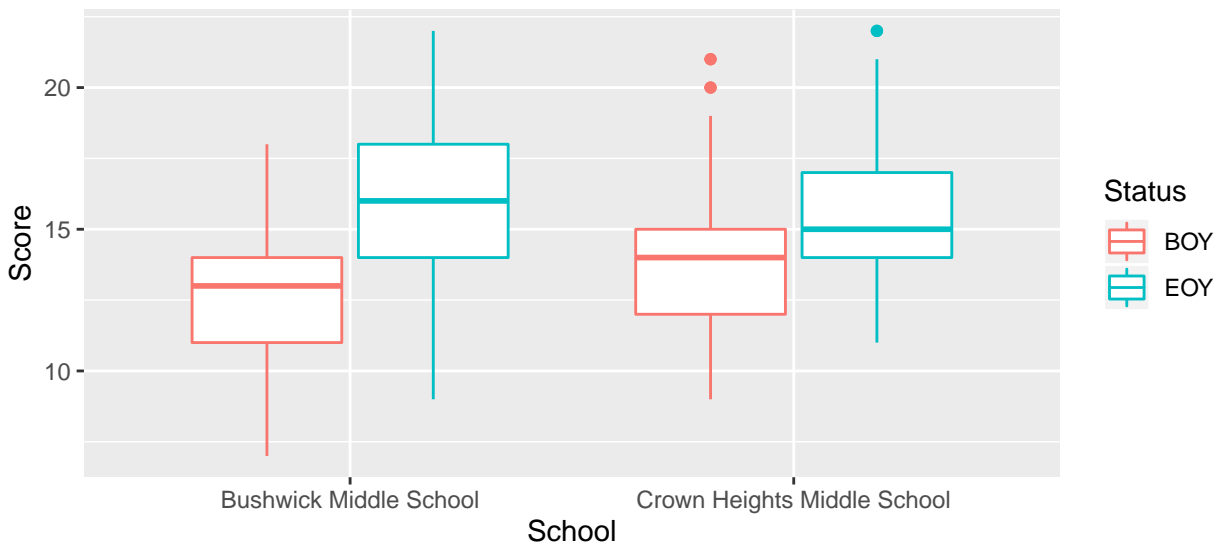
School Name	Grade	Count	BOY Mean	EOY Mean	BOY Max	EOY Max	BOY Min	EOY Min
Bushwick Middle School	5.0	86	11.01	13.53	16	19	7	9
Bushwick Middle School	6.0	94	13.74	17.88	18	22	9	12
Crown Heights Middle School	5.0	73	13.15	14.89	17	20	9	11
Crown Heights Middle School	6.0	15	18.07	18.60	21	22	15	15

Analysis

## Part 1: Two Schools Has Both Improved in Readings

Based on the F&P reading scores and the following boxplots, the medians of Bushwick Middle School and Crown Heights Middle School were improved, which indicates that the students have improved their readings.

We can have the similar result from the boxplot for two schools of 5th and 6th grade: The median of Bushwick Middle School increased more compared to the median of Crown Heights Middle School within each grade level, which shows that the students of Bushwick Middle School have improved their reading abilities faster than the students of Crown Heights Middle School.

Boxplots for Two Schools of 5th and 6th Grade
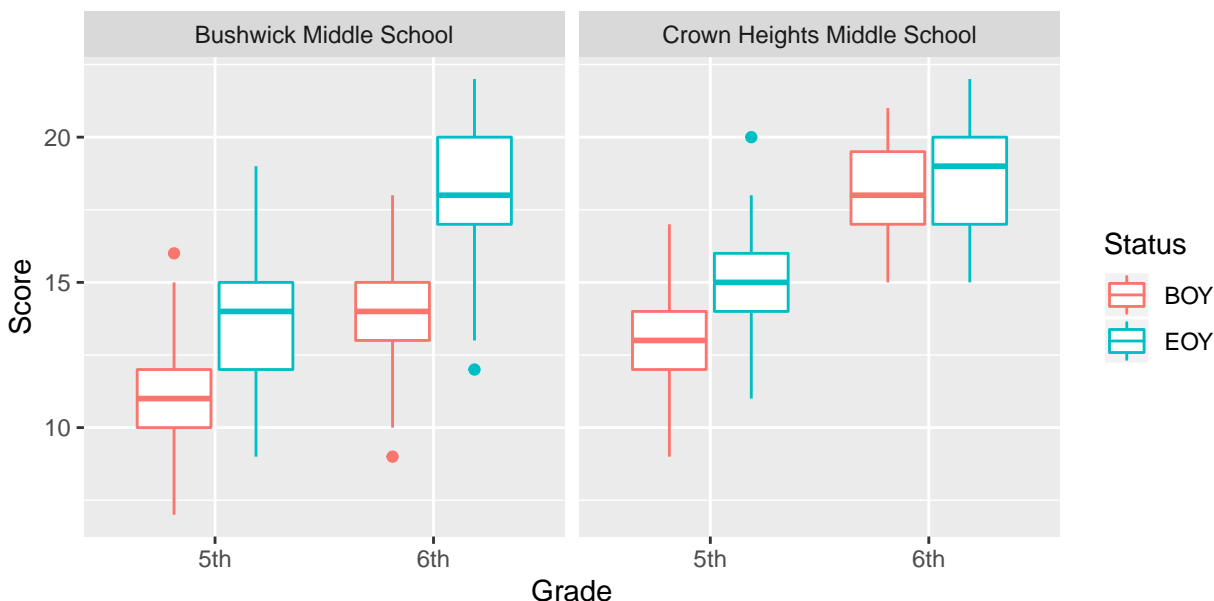
Part 2: Performance Compare Schools and Grades Indicates the Difference in Readings

In this section, I will discuss more within 5th and 6th grades in two middle schools based on their F&P Proficiency Levels. As two middle schools have different numbers of students in each grade, I will consider the corresponding percentage rather than numbers.

As the primary metric for performance, we will consider the students who are **Proficient** or **Advanced** as **Good Performance**, and the students who are **Below Proficient** or **Remedial** as **Bad Performance**.

Here is a summary of our proficiency levels in two middle schools:

|	BOY of Bushwick	BOY of Crown Heights	EOY of Bushwick	EOY of Crown Heights
Advanced	25	45	71	37
Proficient	62	32	58	32
Below Proficient	68	8	32	16
Remedial	25	3	19	3

According to the F&P proficiency levels, I created the bar plots to compare the performance of the students.

For Bushwick Middle School, good performance students rate increase from **13.89%+34.44%=48.33%** to **39.44%+32.22%=71.67%**, which also reflects the numbers in the "Good Performance Rate" table after.

For Crown Heights Middle School, there is **51.14%+36.36%=87.50%** of the students perform well at the beginning of the year, and after one year, there is only **42.05%+36.36%=78.41%** of the students performs well.

As such, the students of Bushwick Middle School performs better than the students of Crowns Heights Middle

School after one-year training of readings, because the area of blue and light blue in the barplot of Bushwick Middle School becomes larger compared to the blue and lightblue area change of Crown Heights Middle School.

Performance of Two Middle Schools
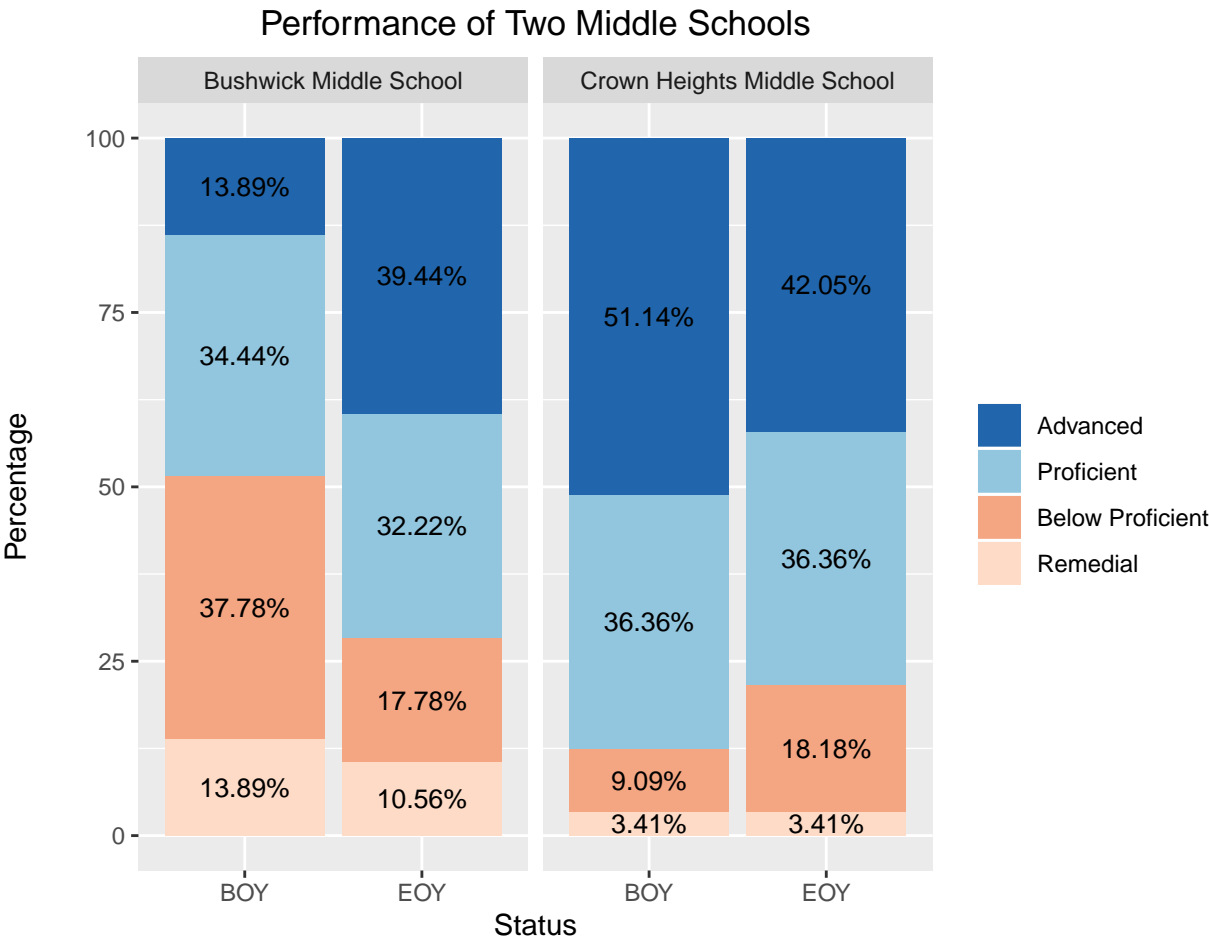
The good performance rate table shows the same result as the barplot above.

Good Performance Rate	BOY	EOY	Comparison
Crown Heights Middle School	87.50%	78.41%	Decrease
Bushwick Middle School	48.33%	71.67%	Increase

Then we look into the grades to see which grade of Bushwick Middle School performs better(the percentage of the students who are advanced and proficient becomes larger).

From the graph, the 6th grade of Bushwick Middle School has the largest improvement from **15.96%** to **61.7%**, and the 5th grader of Bushwick Middle School has the least improvement from **11.63%** to **15.12%**. However, Crown Heights Middle School has a minor decrease in good performance rate in both the 5th and 6th grades.

There is one point we need to notice: in both 5th and 6th grade of Crown Heights Middle School, the persentage of the students who are bad performance becomes larger after one year, and it will be essential for use to find out the reasons after.

Performance of Two Middle Schools of 5th and 6th Grade
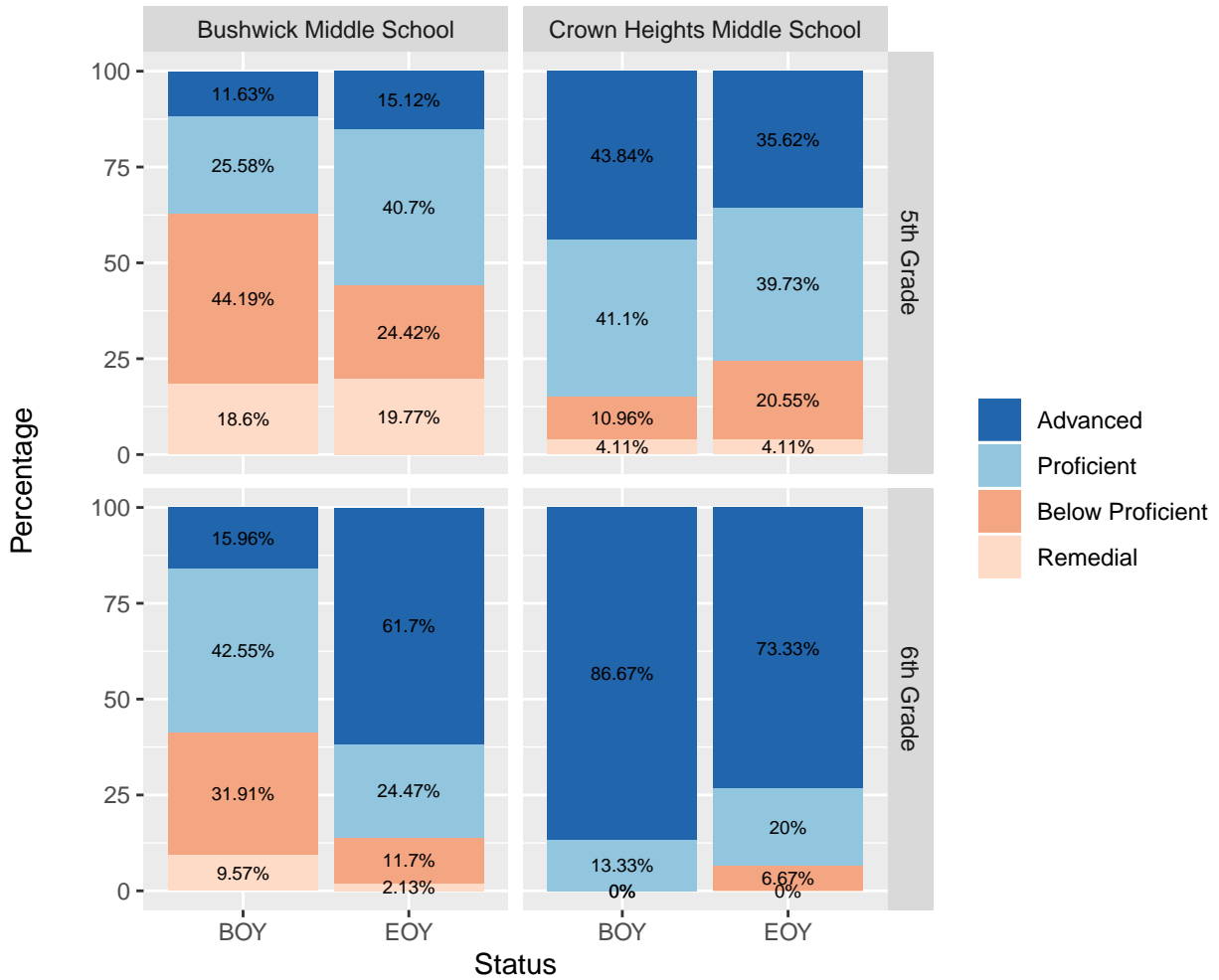
Part 3: Achievement Gaps Becomes Closer

In the following, I defined the **Score Difference/Gap** as the score difference between **BOY F&P score** and **EOY F&P score**: the larger the gap is, the better the improvement is. In the following four density plots, if the position of the line of the mean gap is on the right more, the improvement is more siginficant, and the **Achievement Gaps** becomes closer.

School Name	Grade Level	Mean of Score Difference/Gaps
Bushwick Middle School	5th	2.523256
	6th	4.138298
Crown Heights Middle School	5th	1.739726
	6th	0.533333

Score Difference/Gaps in Two Schools of 5th and 6th Grade

Generally speaking, all four lines are positive, which means that they have all improved their scores, while Bushwick Middle School has larger score difference compared to Crown Heights Middle School.
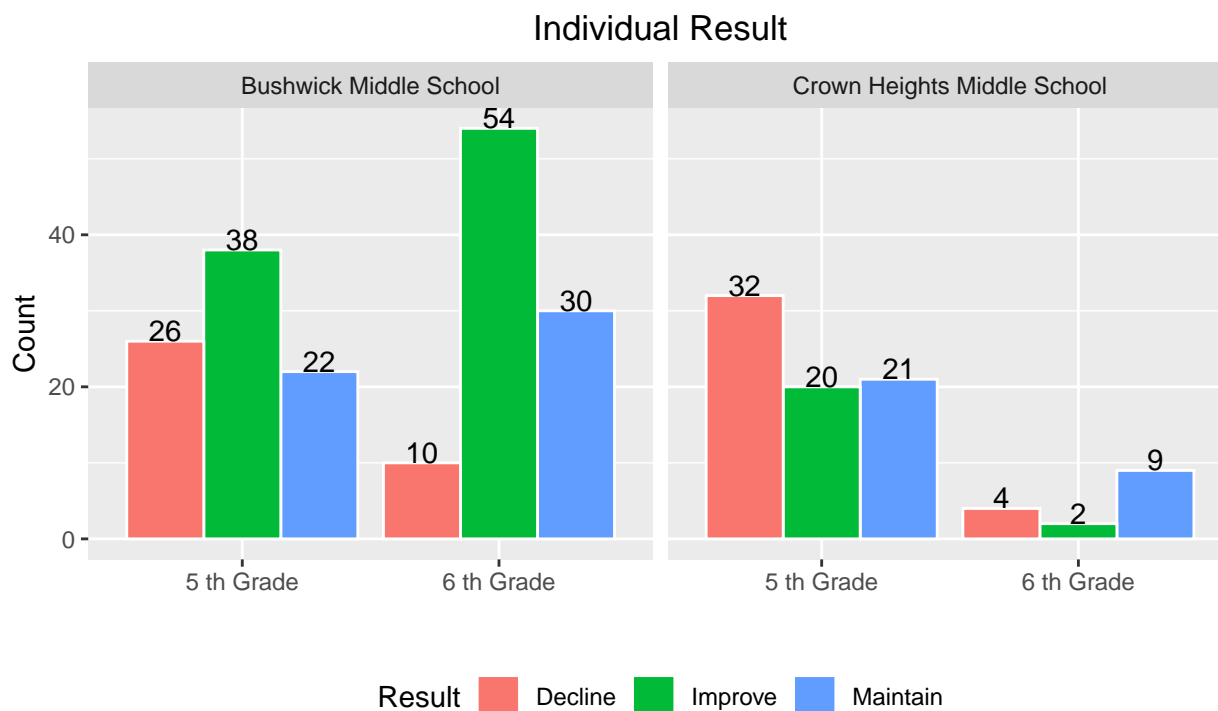
From the above density plots, it is easy to see that the 6th grade of Bushwick Middle School has the largest score difference, which is **4.138298**. Consequently, it has an average score improvement in the score of around 4, and the proficiency level will increase as well. Moreover, the 6th grade of Crown Heights Middle School has the smallest score difference, which is **0.533333**.

We need to pay particular attention to the 6th grade of Crown Heights Middle School. In the graph, it has two clear peaks, which means that most of its 6th-grade students have small score difference, and a small part of its 6th-grade students has large score improvements.

Part 4: Individual Performance Matters

In order to have a better understanding of the students' performance, I tracked the performance of each individual student in the dataset to specifically research whether the student improves the reading skills or no during the past year based on the F&P Proficiency Level and their own improvements.

I counted the number of students who **improve, decline and maintain** the proficiency level, and I had the following results:

Individual Result

Bushwick Middle School

54
38
30
26
22
20 21
10
9
32
4
2

Count
40
20
0

5 th Grade   6 th Grade

Crown Heights Middle School

5 th Grade   6 th Grade

Result   ■ Decline   ■ Improve   ■ Maintain

The total impoved students of Bushwick Middle School is **38+54=92** students, and Crown Heights Middle School has **20+2=22** students. We need to consider the persentage: $\frac{92}{180(total\ available\ BM\ students)} = 51.11\%$ of the Bushwick Middle School students improve themselves compared to their own proficiency levels one year ago. On the contrast, only $\frac{22}{88(total\ available\ CH\ students)} = 25\%$ of the Crown Heights Middle School students improve themselves compared to their own proficiency levels one year ago.

Conclusion and Notes

Bushwick Middle School and Crown Heights Middle School both improved their scores in readings after one year, while we need to pay more attention to the proficiency level as the increasing score might due to the nature improvement without the training. Crown Heights Middle School shows some decrease in the past year in the proficiency level but it shows a better ability to transform bad performance students to good performance students.

Besides this, we can try to find the reasons why some students of Crown Heights has relatively bad performance compared to their proficiency level one year before. Crown Heights Middle School seems to decline in the performance, it is because the students are generally good before they participate in this F&P assessment, so they do not have big room to improve their readings. However, **90.9%** of those students who have bad performance becomes good performance students after one year in Crown Heights Middle School. The same rate in Bushwick Middle School is **65.59%**. This result shows that Crown Heights is good at transforming bad performance students to good performance students.

The next stage of this analysis is to find better approaches to improve those students in Crown Heights Middle School after they achieve some degree of reading proficiency level. Otherwise, the students of Bushwick Middle School will face a similar situation(decrease in performance) when they have a big percentage of students of good performance at the beginning of the year.

Other things that we need to notice: The sample is still too small to have a big conclusion, and it will be better to analyze more data as needed. Also, the calculation not shown in the report could be seen at the end of the rmd file.