Ashley Horton

Learning Analytics Final Project Part 1 - Data Prep

1: Overview

Purpose:

I will be analyzing the beginning and middle of the year iReady reading data for my building. The idea will be to look for the growth percentage between a few different categories: lower achieving students versus higher achieving students, students' race, and gender.

Key Questions:

1. How does the growth percentage compare between lower achieving students and higher achieving students?
2. How does the growth percentage compare among different races?
3. How does the growth percentage compare between males vs. females?

Audience:

The target audience for these analytics will be the school administrators, data coach, reading interventionist, and learning coach.

Student/Learning Relationship and Usefulness:

Analyzing this data can be proven useful when planning for instruction and future assessments. It may help identify targeted groups of students for intervention purposes.

2: Describe Collection, Acquisition, and Storage of Data

Collection:

Data was collected through the iReady website as students completed their middle-of-year (MOY) assessments. iReady creates reports at the school, class, student, and batch level.

Acquisition:

I acquired the data through my iReady administrator account which grants me access to data for my entire building. The data has not been cleaned.

Other Details:

- Data is anonymized having student name, grade level, and username removed
- Data is not confidential since all identifying factors have been removed
- Since all identifying factors have been removed, the data is being stored on my school-issued device, the iReady website, and Tableau.

Ashley Horton

3: Understand your Dataset

Variables:

Variable	**Definition**	**Signal or Noise**
Student ID	Student's ID number through eSchool	noise
Academic Year	School year data from which the data was collected	noise
Gender	Gender of which the student identifies (male or female)	signal
Hispanic or Latino	Does the student identify as Hispanic or Latino?	signal
Race	Student's identifying race	signal
English Language Learner	If the student is identified as an English Language Learner	signal
Special Education	Does the student have an IEP	signal
Economically Disadvantaged	Is the student registered with their building as economically disadvantaged	signal
Migrant	Is the student's living situation that of a migrant (moving frequently for work)	signal
Start Date	Date student began diagnostic assessment for the calendar year (may be more than one date per student)	signal
Completion Date	Date student finished the diagnostic assessment for	noise

	the calendar year (may be more than one date per student)	
Diagnostic Used to Establish Growth Data	Were the results of this diagnostic used for the growth monitoring report	noise
Most Recent Diagnostic	Is this diagnostic the most recent for the student	noise
Duration (min)	The amount of time, in minutes, the student spent on the assessment	noise
Overall Scale Score	The overall scale score the student received on the diagnostic	signal
Overall Placement	The grade level the student placed based on their overall scale score	noise
Overall Relative Placement	The number of grade levels the student placed at or below their current grade level	noise
Phonological Awareness Scale Score	Student's scale score (number) for phonological awareness.	signal
Phonological Awareness Placement	Student's placement (grade level) for phonological awareness	noise
Phonics Scale Score	Student's scale score (number) for phonics	signal
Phonics Scale Placement	Student's placement (grade level) for phonics	noise
High-Frequency Words Scale Score	Student's scale score (number) for high-frequency words	signal
High-Frequency Words Placement	Student's placement (grade level) for	noise

	high-frequency words	
Vocabulary Scale Score	Student's scale score (number) for vocabulary	signal
Vocabulary Placement	Student's placement (grade level) for vocabulary	noise
Reading Comprehension: Literature Scale Score	Student's scale score (number) for reading comprehension: literature	signal
Reading Comprehension: Literature Placement	Student's placement (grade level) for reading comprehension: literature	noise
Reading Comprehension: Informational Text Scale Score	Student's scale score (number) for reading comprehension: informational text	signal
Reading Comprehension: Informational Text Placement	Student's placement (grade level) for reading comprehension: informational text	noise

Filtered Variables:

- Gender
- Hispanic/Latino
- Race
- English Language Learner
- Special Education
- Economically Disadvantaged
- Migrant
- Start Date

Missing Data:

- Growth percentage between the beginning of year and middle of year diagnostics.

4: Data Cleaning Subcycle - Refining Your Questions and Plan

Going through the data cleaning process showed me that I had a lot of noise in the data set and some variables that could potentially be signals depending on how precise I want my analysis to be. Many variables were removed (as seen in section 5 of this document) and some variables, which are currently labeled as noise, were kept in the set because they could be useful in answering the key questions.

5: Data Cleaning Subcycle - Steps for Wrangling

Removed Variables:

The following variables were removed due to the privacy of the students and/or they were not necessary in answering the key questions. These were removed prior to completing section 3 of this document.

- Last name
- First name
- Student Grade
- School
- User Name
- Class(es)
- Class Teacher
- Report Group(s)
- Percentile
- Grouping
- Lexile Measure
- Lexile Range
- Mid on Grade Level Scale Score

Removed Variables (noise):

The following variables were removed after completing section 3 of this document due to them being considered “noise”.

- Academic Year
- Completion Date
- Diagnostic Used to Establish Growth Data
- Most Recent Diagnostic
- Duration
- Overall Relative Placement

Filtered Out:

The data set did not contain the calculated growth percentage from the beginning of year to middle of year assessment for each student. When calculating this, I noticed that many students had more than two assessments listed - either because they took the BOY or MOY more than once, or they had already taken the Growth Monitoring diagnostic in March. I only kept the scores that were from the assessments taken in

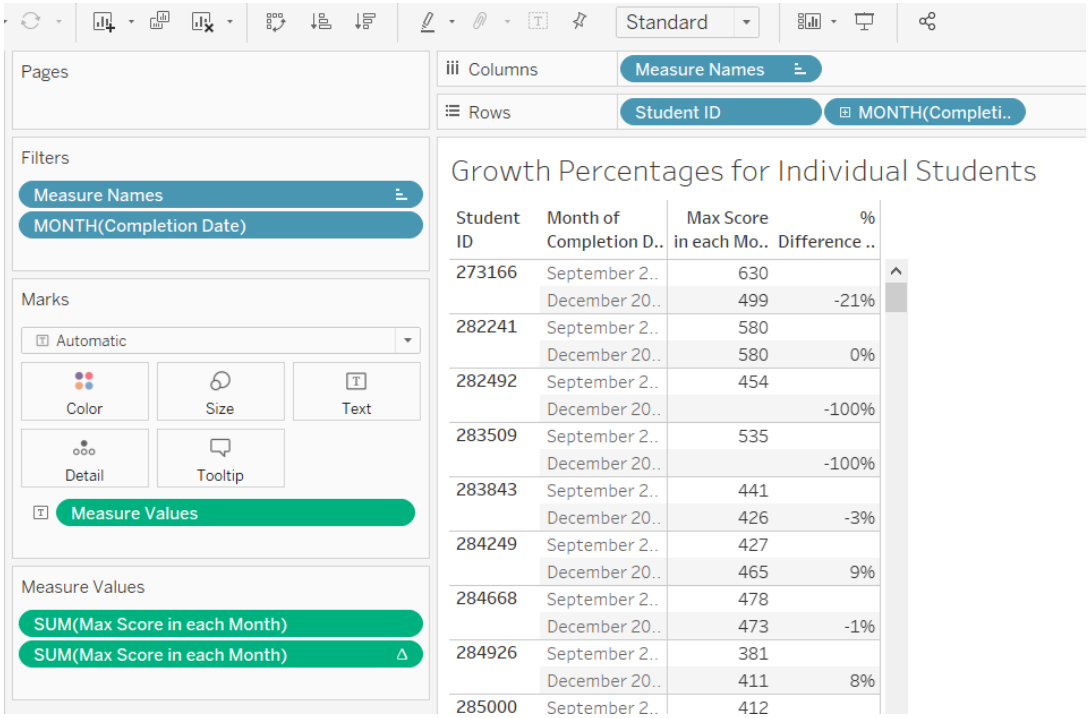
September and December. All others were filtered out. If a student took more than one test in the same month, the highest score was kept.

6: Reflect on Ethics

Correlation vs. Causation - When analyzing data comparing sub-groups of students, it may be easy to blame their “label” as the cause to their performance (whether positive or negative). For example, if we were to compare homeless students versus those with stable homes, we potentially could see a correlation between homelessness and low growth percentages. Further analysis would need to be done to confirm if homelessness is a cause for the low growth percentages.

7: Simple Design Plan

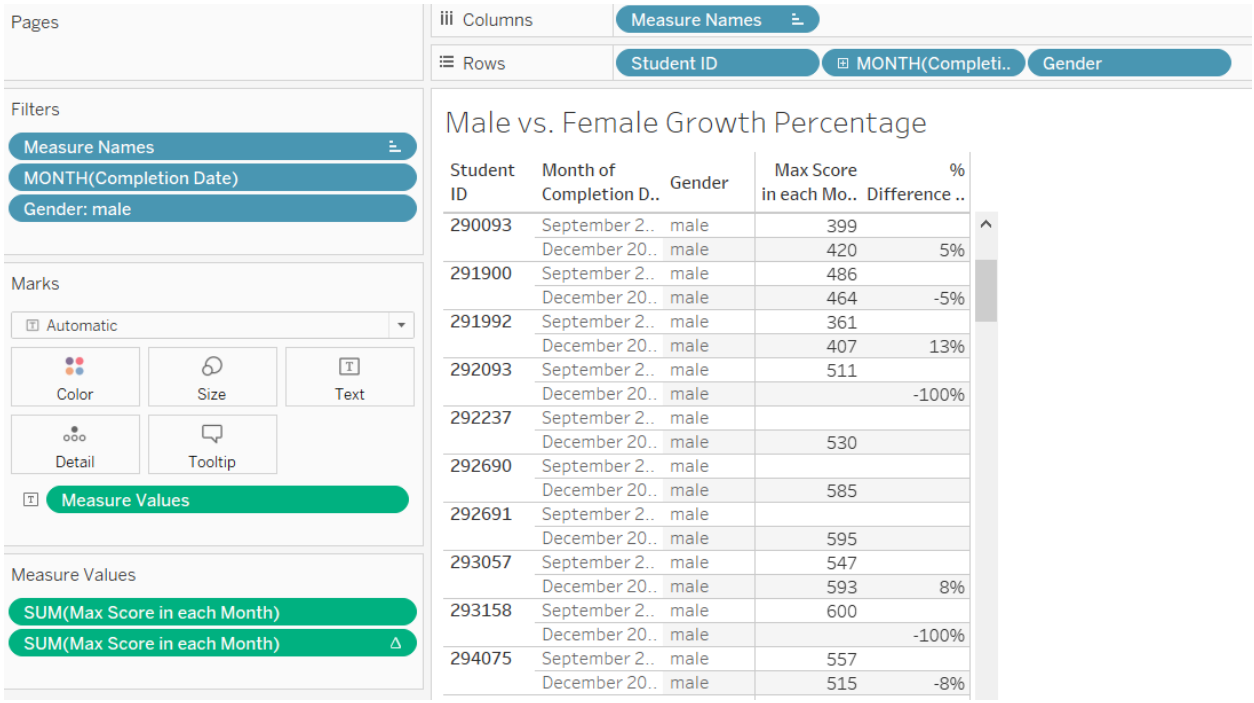
Example 1: This image shows the calculated growth percentage for each student for their beginning of year and middle of year assessments.



Student ID	Month of Completion D..	Max Score in each Mo..	% Difference ..
273166	September 2..	630	
	December 20..	499	-21%
282241	September 2..	580	
	December 20..	580	0%
282492	September 2..	454	
	December 20..		-100%
283509	September 2..	535	
	December 20..		-100%
283843	September 2..	441	
	December 20..	426	-3%
284249	September 2..	427	
	December 20..	465	9%
284668	September 2..	478	
	December 20..	473	-1%
284926	September 2..	381	
	December 20..	411	8%
285000	September 2..	412	

Examples 2-4: These visualizations will be simpler in that they will show an average growth percentage for the subgroups in a chart and then also displayed visually with a graph. I’ve begun the process of adding variables and filtering out subgroups as seen below with the male vs. female comparison.

Ashley Horton



Student ID	Month of Completion D..	Gender	Max Score in each Mo..	% Difference ..
290093	September 2..	male	399	
	December 20..	male	420	5%
291900	September 2..	male	486	
	December 20..	male	464	-5%
291992	September 2..	male	361	
	December 20..	male	407	13%
292093	September 2..	male	511	
	December 20..	male		-100%
292237	September 2..	male		
	December 20..	male	530	
292690	September 2..	male		
	December 20..	male	585	
292691	September 2..	male		
	December 20..	male	595	
293057	September 2..	male	547	
	December 20..	male	593	8%
293158	September 2..	male	600	
	December 20..	male		-100%
294075	September 2..	male	557	
	December 20..	male	515	-8%

What's left now is to calculate the average growth percentage for each subgroup.

Ashley Horton

Learning Analytics Final Project Part 2 - Analyze and Visualize

1: Overview

Purpose:

I will be analyzing the beginning and middle of the year iReady reading data for my building. The idea will be to look for the growth percentage between a few different categories: lower achieving students versus higher achieving students, students' race, and gender.

Key Questions:

How does the growth percentage of students' beginning of year and middle of year iReady tests compare among different groupings?

- Special education students vs. non-special education students
- White vs.Black/African American vs. Native Hawaiian vs. Asian vs. American Indian
- Males vs. females
- Kindergarten vs. 1st vs. 2nd vs. 3rd vs. 4th vs. 5th grade

Audience:

The target audience for these analytics will be the school administrators, data coach, reading interventionist, and learning coach.

Student/Learning Relationship and Usefulness:

Analyzing this data can be proven useful when planning for instruction and future assessments. It may help identify targeted groups of students for intervention purposes.

2: Exploration and Analysis

- Growth % All Worksheet
    - Added student ID, month completion, and measure names variables
    - Filtered month completion to September and December as to only include the beginning of year and middle of year assessments
    - Measure Values - max score for each month is visible; formula for % difference between both scores
    - Some students had only taken one of the assessments and had a % difference of 0 or -100%. These students were excluded from the data sets.
    - Column grand totals (average) were added to view the average growth percentages

- Remaining worksheets for each individual variable being analyzed - a column was added for that specific variable and filtered to include only one type (example: six worksheets were completed- one for each grade level)
- Race/Ethnicity - nine students were listed as “null” and were excluded from the visual.
- Race/Ethnicity - American Indian was included in the visual, but it is important to note that there are only two students who identify with that ethnicity
- Race/Ethnicity - Native Hawaiian was included in the visual, but it is also important to note that there is only one student who identifies with that ethnicity.
- Visuals were created using Piktochart by importing the variable names and

Patterns

I found it interesting that the growth rates were consistent across variables and comparisons. I also noticed that the growth rate for SPED students was higher than that of non-sped students. From what I know about IEPs, I would attribute this growth to the SPED students being provided with more assistance during testing (ie. having the test read to them, assistance with narrowing down answer choices, etc.). I would also be interested in seeing the growth percentage from beginning of year to the end of year, but unfortunately, students have not taken the end of year assessment yet.

I wasn’t expecting to find any specific patterns, however, I did think there would have been some larger growth rates between variables. I was surprised to see that most of the growth rates were positive, with the exception of Asian students, American Indian students, and 5th grade.