WPI

MS Thesis: Analytics projects for key stakeholders in large scale online learning systems

Rahul Jakhmola
Advisor : Prof Neil Heffernan
Reader : Prof Mohamed Eltabakh

Overview

As a computer-based learning platform, ASSISTments helps both educators and students across the country by providing a number of tools :

- to aid in providing immediate feedback
- to report meaningful data, and
- deliver instructional support.

ASSISTments

20,000+
Teachers

500,000+
Students

30,000,000+
Problems Solved

In 2020-2021 school year

Overview

At this scale, data has enormous potential for **day-to-day decision making** around the core learning product.

Outside of existing projects, stakeholders can benefit from **insights at different levels of data granularity**.

Administrators
What is the big picture?

Teachers
Better feedback should help my students with homework..

Students
Another assignment !?

Overview

Project 1: Descriptive Analytics
Data driven decision making for administrators

Project 2: A/B Testing
Influencing individual teacher behavior

Project 3: Predictive Analytics
Observing individual student behavior

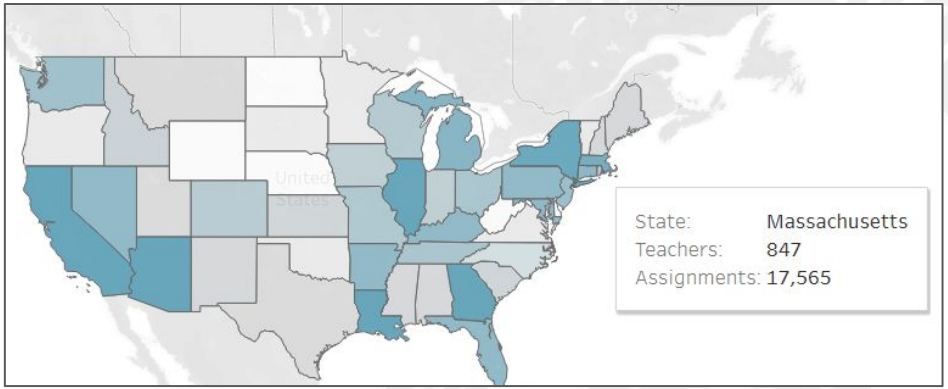
Administrators
What is the big picture?

Teachers
Better feedback should help my students with homework..

Students
Another assignment !?

Project 1:
Descriptive Analytics

Data driven decision making for administrators



Snippet of map visualization from a live Tableau dashboard

Project 1 : Descriptive Analytics

Admin team has multiple responsibilities like

- Teacher onboarding
- Teacher training
- Teacher retention
- Teacher support

....and a lot more

Goal : Track teacher and student activity.
Data is often needed to get answers to multiple questions.

Actual data requests
How are we doing as compared to last week/month/year?
How many teachers have signed up during covid?
What content do they use?
Can we observe EngageNY teachers who use assignment reports for grade 5?

Project 1 : Descriptive Analytics

Getting to insights faster from data request

Legacy Process

Data Need → IT team writes query → Admin team runs query → Data analysis in Excel → Insight + New Data Request

Day 0 → **Day 30**

New Process

Data Need → Tableau Dashboard → Insights + Deep Dive + Near real time Reports

Day 0 → **15 minutes**

Project 1 : Descriptive Analytics

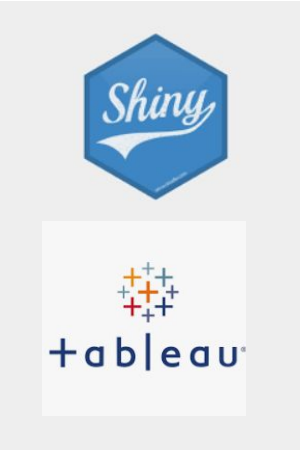
IT Infrastructure	Data Preparation	Dashboard Design	Mature Workflow

Project 1 : Descriptive Analytics

IT Infrastructure	Data Preparation	Dashboard Design	Mature Workflow

Database

Analysis



Dashboards

Hosting Server

Project 1 : Descriptive Analytics

IT Infrastructure	**Data Preparation**	Dashboard Design	Mature Workflow

Understand end user requirements

Prepare persistent data sources

Project 1 : Descriptive Analytics

IT Infrastructure	Data Preparation	Dashboard Design	Mature Workflow
Understand end user requirements		Prepare persistent data sources	

1. Track sign up activity.
2. Track assignment creation activity.
3. Track student responses on assignments.
4. Track assignment report views.

As lowest data granularity observed here is an assignment, the analysis dataset has one row per assignment.

Project 1 : Descriptive Analytics

Different metrics tracked for every assignment include:

1. **Class Level Metrics :**
   Class Size, Assignment Completion % etc
2. **Curriculum Details :**
   Curriculum, Grade, Chapter etc
3. **Successful report checking behavior (A Loop : True/False)**
   " A teacher assigns homework, then a student does it and then the teacher looks at the assignment report"

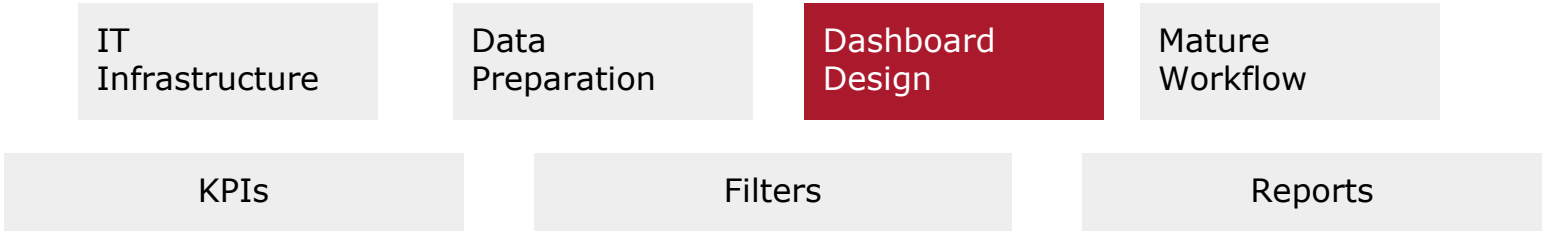
Project 1 : Descriptive Analytics

IT Infrastructure | **Data Preparation** | Dashboard Design | Mature Workflow

Understand end user requirements | **Prepare persistent data sources**

	Tasks	Solutions
1	Query across multiple DBs and tables	Cross DB joins with Foreign Data Wrapper
2	Optimize query speed for dashboards	SQL/Python/R scripts for data preparation and Tableau Extracts for analysis
3	Daily data refresh	SQL Views and scheduled Tableau extract refresh

Project 1 : Descriptive Analytics



The dashboards need to provide:

- Intuitive visualization to analyze data
- Easy controls to subset and filter data
- Downloadable reports for analysis results

Project 1 : Descriptive Analytics

IT Infrastructure	Data Preparation	Dashboard Design	Mature Workflow

KPIs	Filters	Reports

Key performance indicators created for assignments, which can then roll up to different levels.

These include:

- Total Teachers in a group (like country/state/experiment)
- Total Assignments created
- Assignment Completion rates
- Report checking rates (Loops) for teachers

Project 1 : Descriptive Analytics

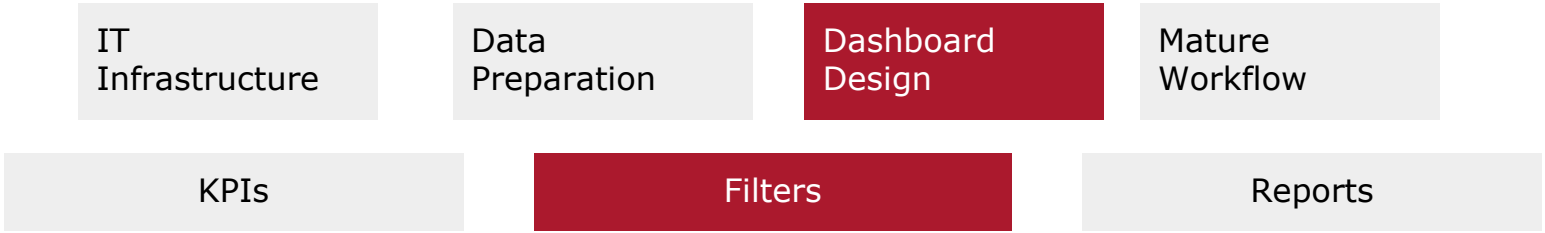


Tableau filters provide subset and deep dive capabilities in easy to use UX elements.
Following filters turned out to be extremely useful, and can be used together as needed:

- Map filter by state, and other teacher groups
- Usage dates
- Assignment responses
- Curriculum
- Teacher Name..and more

Project 1 : Descriptive Analytics

IT Infrastructure | Data Preparation | **Dashboard Design** | Mature Workflow

KPIs | Filters | **Reports**

The KPI reports can obtained starting from a bird's eye view of the entire country, to individual assignment level.

Links to assignment reports provide the provision to navigate to individual student responses on a particular assignment as well.

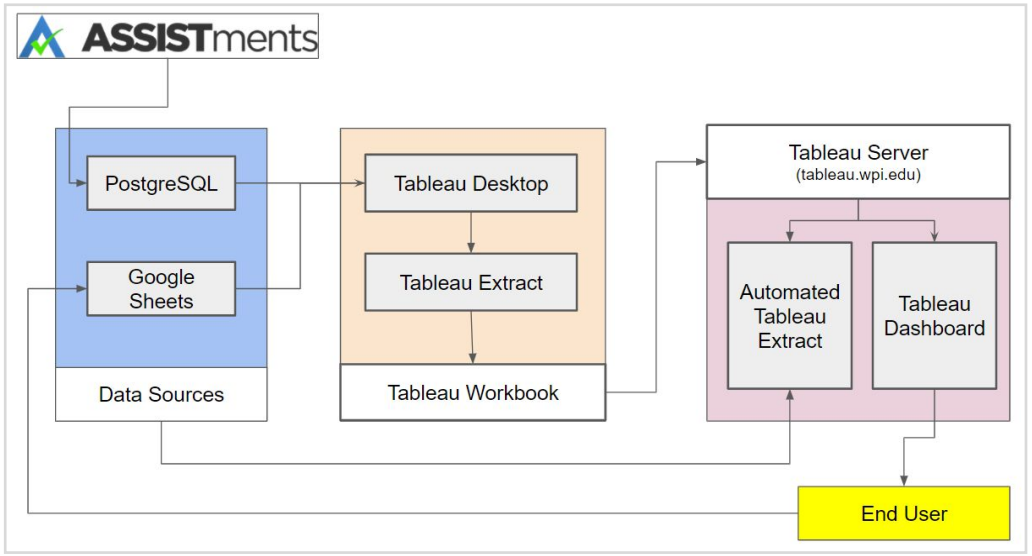
Tableau Dashboard → Data Filter → csv report

Project 1 : Descriptive Analytics

Demo

Project 1 : Descriptive Analytics

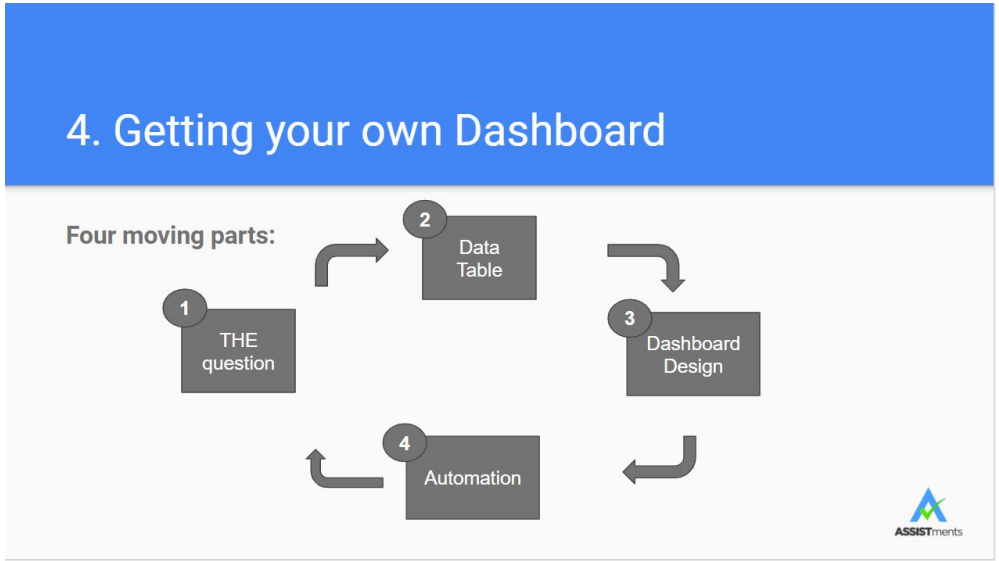
IT Infrastructure | Data Preparation | Dashboard Design | **Mature Workflow**



Project Components

Project 1 : Descriptive Analytics

IT Infrastructure	Data Preparation	Dashboard Design	Mature Workflow



New Dashboard Requests:

Four step iterative process to getting a dashboard up and running from concept to production.

Project 1 : Descriptive Analytics

IT Infrastructure	Data Preparation	Dashboard Design	Mature Workflow

User feedback examples

Cristina 4:32 PM

@Rahul Jakhmola I really like your new open response dashboard. Can you allow me to put in a Problem ID and get to that spot. So I can ask myself Are there any examples of OR answers for this problem. Also can it use the PR ID not the number. Also how about the text answers from students could you load both and then have a filter to filter out text and just look at uploaded.

One more thing if I sort by Correct can they stay clumped by problem ID?

From: Tignor, DonnaLee
Sent: Tuesday, November 10, 2020 8:54 AM
To: Jakhmola, Rahul
Subject: Re: [EXT] Re: Tableau Data

Hi Rahul
It worked! I can now access this new report through the link that you provided. However, this new version doesn't provide important info needed for evaluating the mentee's use of ASSISTments as compared to the old version in individual reports. For example, the old version provided me with 1) individual assignment dates for each assignment, 2) how many times each individual report was accessed by the teacher, 3) individual assignments number of problems completed as well as 4) individual assignment completion rates. Is there any way this info can be accessed for individual reports without having to click in and out of each assignment? It is an important component in my evaluation. In the current report, I can't tell at a glance, whether or not a mentee has been recently assigning, checking reports, and if students are completing at a proficient rate since the last time we met. Thanks for your help.

Kind regards,
DonnaLee

Project 1 : Descriptive Analytics

Successes, Conclusion and Scope

- Dashboards have become integrated into the admin team workflow, and have been viewed 2000+ times till date. TAF team has also procured Tableau for nonprofits licenses for future use.

- Well defined workflow ensures that new dashboard requests are welcome and can be catered to as needed.

- Current development involves development of a standard data warehouse for ASSISTments student workers and employees to create their own Dashboards and reports.

Project 1 : Descriptive Analytics

Questions

Project 2:
A/B Testing

Influencing individual teacher behavior

Project 2 : A/B Testing

A sizeable number of assignment reports available were not viewed by the teachers in the previous school year (2019-2020).

Without access to student responses, teachers are unable to incorporate valuable information in instructional strategies.

36%
(4,930/13,606)
Teachers never checked their assignment reports.

41%
(141,102/343,997)
Assignments had student responses but the report was never checked.

Project 2 : A/B Testing

Idea

- E-mails have been used in marketing contexts to increase engagement.
- Explore use of e-mail prompts in encouraging teachers to attend to their students data in a K-12 remote classroom setting.

Project 2 : A/B Testing

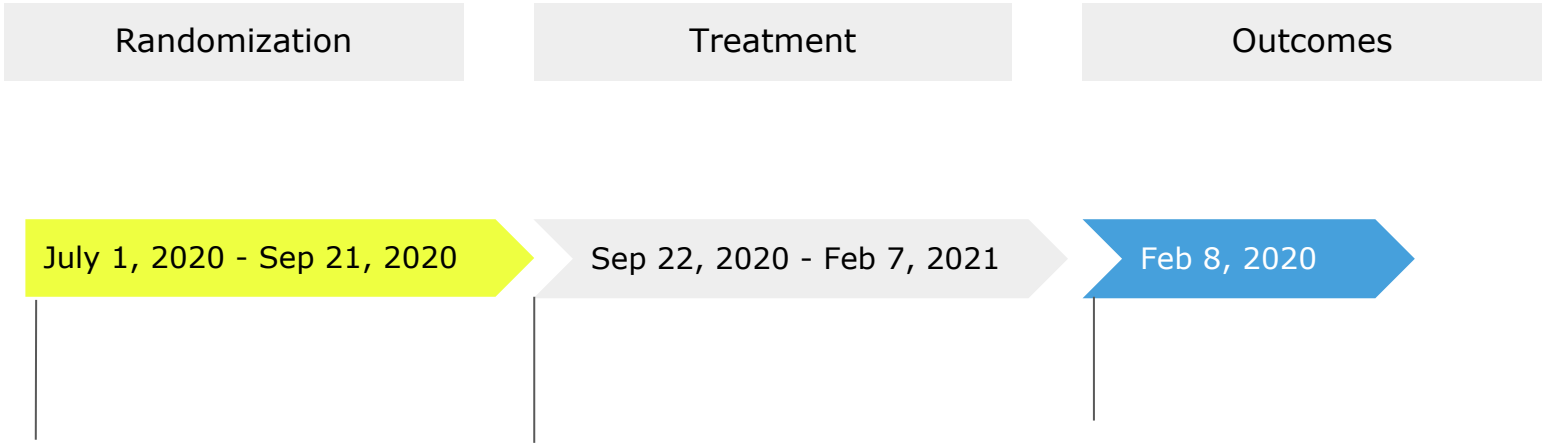
Research Questions

1. Does the sending of email prompts increase teacher engagement as measured through assigning and report-checking activity?

2. Does the inclusion of a direct link to a report within email prompts increase teacher engagement?

Project 2 : A/B Testing

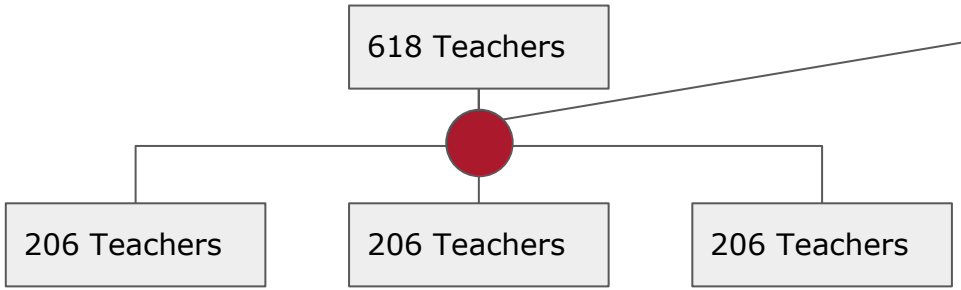
Experiment Design



Project 2 : A/B Testing

Experiment Design

Randomization | Treatment | Outcomes



Stratified Sampling.
14 strata created with k-means clustering.

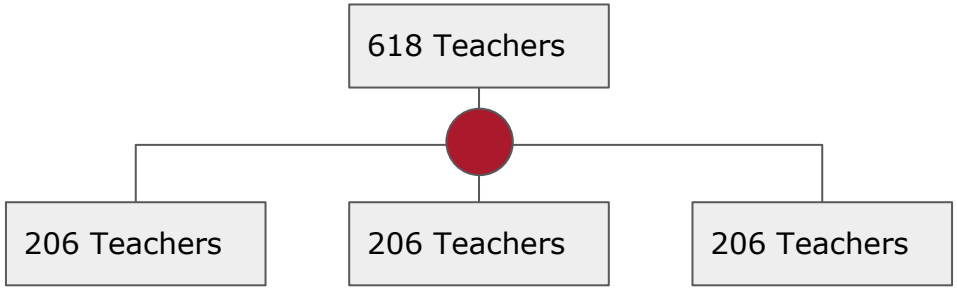
4 Clustering attributes

- Account Age (Days)
- Assignments Made
- Classes Added
- Students Added

Project 2 : A/B Testing

Experiment Design

Randomization | **Treatment** | Outcomes



One Control and Two treatment conditions.

Group 1
Receives **no email**

Control Group

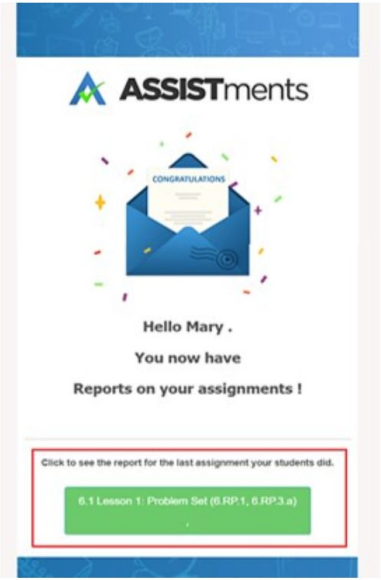
Group 2
Receives **email with a call to action** to click link to report.

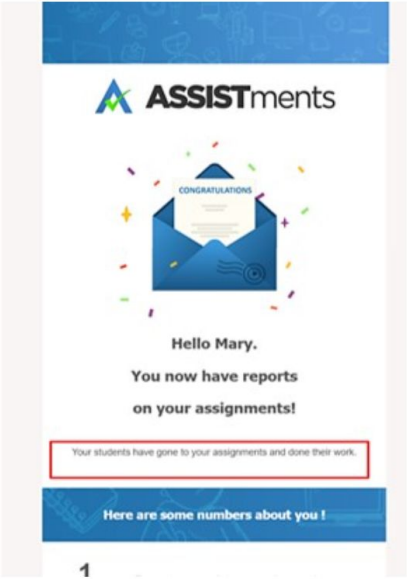
Group 3
Receives **email without call to action** to click link to report

Project 2 : A/B Testing

Experiment Design

Randomization	Treatment	Outcomes





Project 2 : A/B Testing

Experiment Design

Randomization	Treatment	Outcomes

Following dependent variables are observed

- Number of new complete assignment/report loops for each teacher
- Number of new assignments given by each teacher

Project 2 : A/B Testing

Observations

Group	Total Teachers	Assignment Creators	Loop Creators
No email (Control)	206	84	39
No Call to Action	206	81	46
Call to Action	206	75	42

Post experiment data shows no significant difference among the three groups.

Assignment Creators:
Teachers who made any new assignments after the email was sent.

Loop Creators:
Teachers who created new loops after the email was sent.

Project 2 : A/B Testing

Analysis

Variable	Description
Independent Variables	
Account Age	Age of the teacher's account (in days) before condition assignment
Prior Assignments	Number of assignments made up to the time of condition assignment
Number of Classes	Number of classes created by the teacher at the time of condition assignment
Number of Students	Number of students enrolled across all of the teacher's classes
Assigned Condition	Experimental condition to which the teacher was assigned
Dependent Variables	
Number of New Assignments	Number of new assignments created after condition assignment
Number of Completed Loops	Number of assignment/report loops completed by the teacher after condition assignment

Project 2 : A/B Testing

Analysis

Dependent Variable:	Assignment/Report Loops			Log Number of Assignments		
Variable	B	Std. Error	p-value	B	Std. Error	p-value
Intercept	1.292	0.331	<0.001***	2.550	0.354	<0.001***
Email with No Call to Action (Treatment 1)	0.201	0.226	0.376	0.003	0.242	0.991
Email with Call to Action (Treatment 2)	0.243	0.227	0.285	-0.059	0.243	0.807
Log of Prior Teacher Assignments	0.062	0.111	0.575	0.580	0.119	<0.001***
Square Root of Age of Teacher Account	-0.100	0.057	0.079	-0.046	0.060	0.447
Log of Number of Students	0.112	0.083	0.179	-0.190	0.089	0.034*

Regression analysis results observing the log of assignment/report loops completed by teachers (left) and the log number of assignments created (right) as dependent variables.

Project 2 : A/B Testing

Inference

- The conducted study represents, in this context, a null finding. While this does not suggest that email prompts have no effect, it does mean that the effect is so small given the sample size of teachers
- Coefficients lean in favor of email campaigns.
- It seems questionable teachers would disagree that student reports are unhelpful.
- Several possible reasons for no effect:
  - Teachers already have a lot of information on students.
  - It’s likely teachers did not open the email - there was no tracking for that data in this study.

Project 2 : A/B Testing

Conclusion and Scope

- Large scale A/B testing for email campaign is being incorporated into ASSISTments.
- Recently procured Hubspot, a CRM tool which enables A/B testing and email tracking.
- API integration done with database. Running new campaigns to influence user behavior and iterate rapidly in the next school year.

Project 2 : A/B Testing

Observations

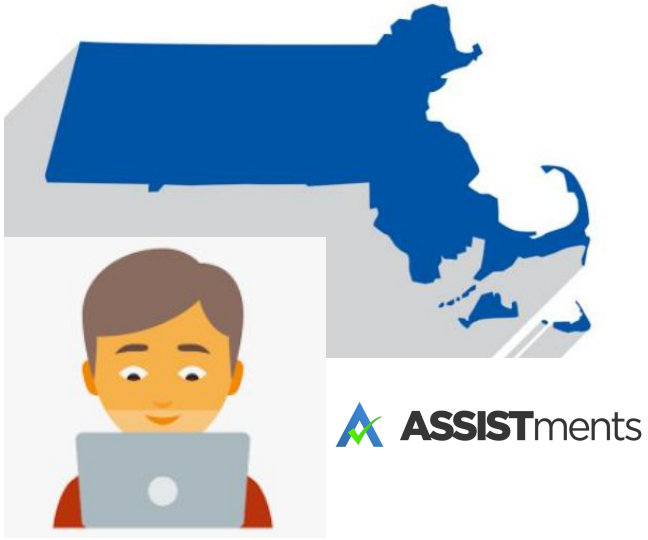
Project 3:
Predictive Analytics

Observing individual student behavior

Project 3 : Predictive Analytics

7918 students in MA did
203,268 homework assignments on ASSISTments in
2018-2019 school year.

However
Over **8%** assignments were left incomplete, as students left in the middle of the assignments (dropout).



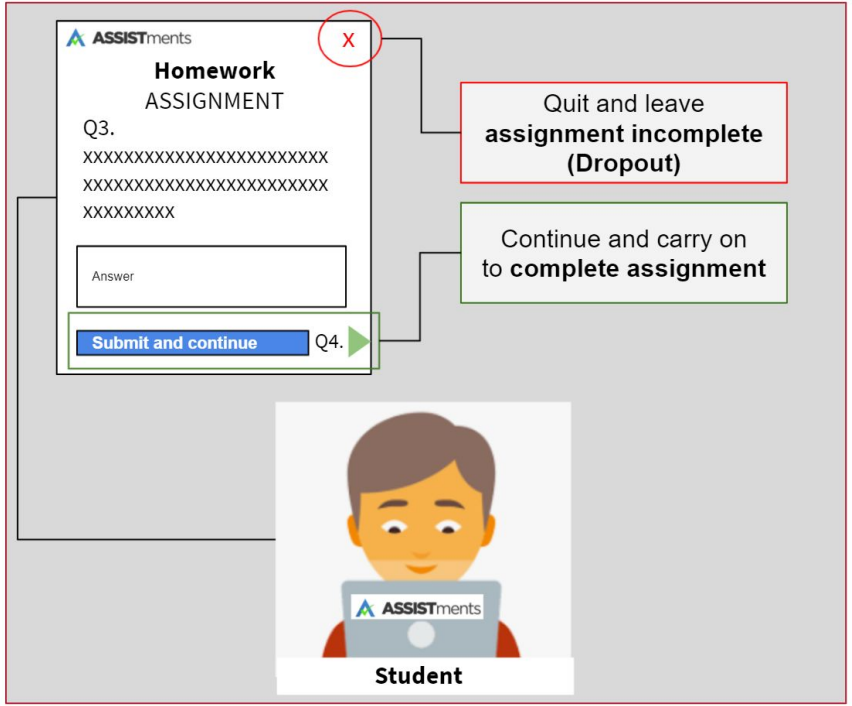
Project 3 : Predictive Analytics

Teachers assign homework, and students in their class solve the problems online.

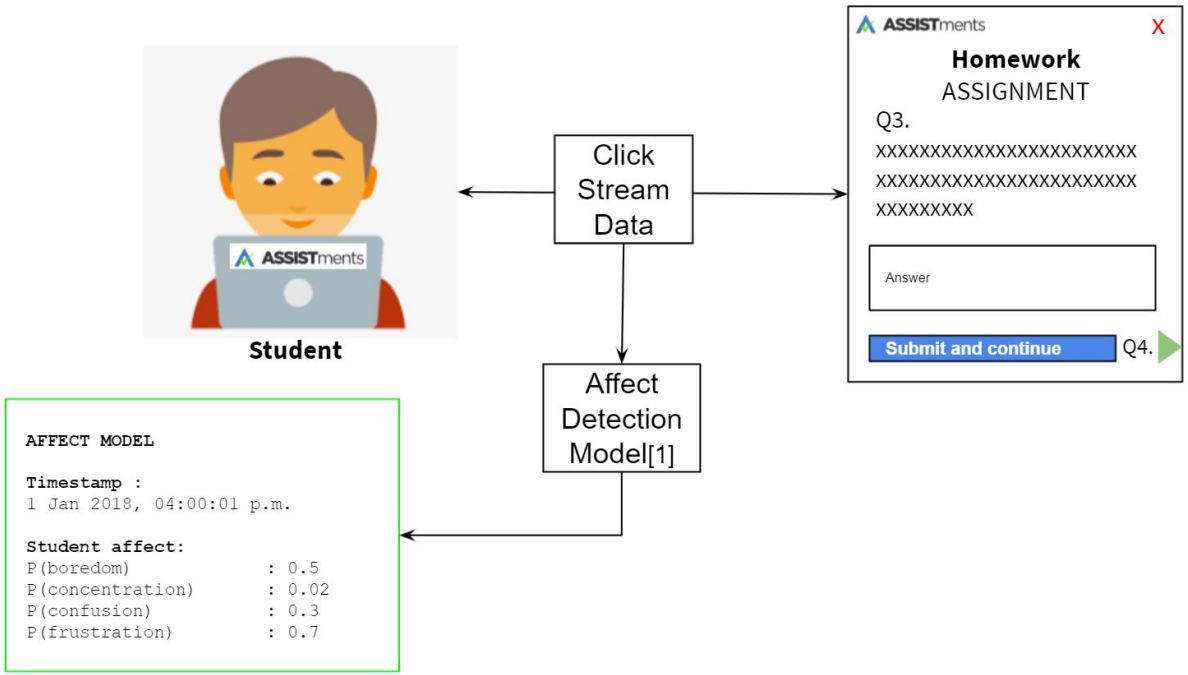
Students feel emotions like confusion, boredom, frustration, concentration can even stop doing the assignment in the midst of it.

High math frustration is associated with:

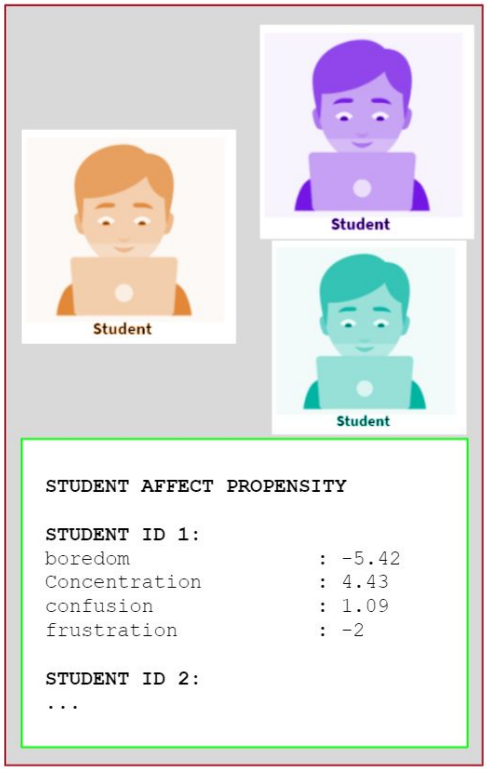
- Lower STEM outcomes
- Reduced College attendance



Project 3 : Predictive Analytics



Project 3 : Predictive Analytics



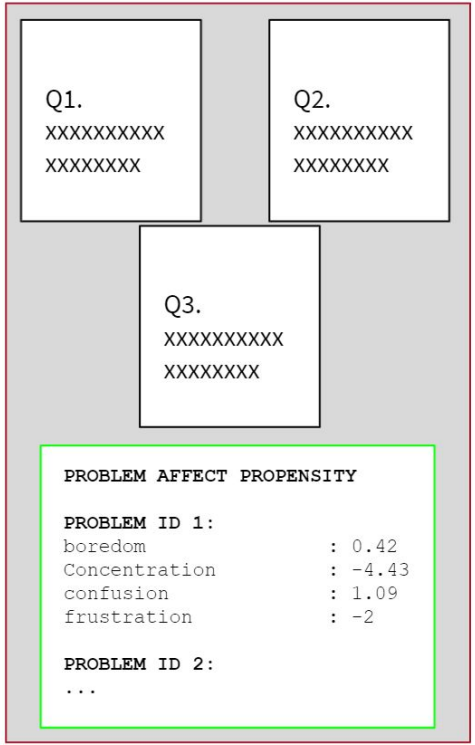
What causes dropout?

Is it student level (trait) or problem level (state).

For Example:

- Student Boredom Propensity v/s
- Problem Boredom Propensity

Affect propensity scores are calculated for students and problems by IRT model (Rasche models : which give us Z- scores).



Project 3 : Predictive Analytics

1. **1 Row per assignment :**
   For every assignment, we observe the last problem solved.

2. **Dropout variable :**
   If the student drops out, dropout is assigned the value ‘1’, otherwise ‘0’.

3. **Prior Dropout :**
   Prior dropout rates of students are noted before current assignment.

4. **Student and Problem Affect propensities :**
   For the student interaction with the last problem, we get student and problem affect propensities as trait and state variables.

```
LOGISTIC MODEL

Dropout ~
      Prior_dropout +

      student_concentration_propensity +
      student_boredom_propensity +
      student_confusion_propensity +
      student_frustration_propensity +

      question_concentration_propensity +
      question_boredom_propensity +
      question_confusion_propensity +
      question_frustration_propensity
```

Project 3 : Predictive Analytics

LOGISTIC MODEL

```
Dropout ~
      Prior_dropout +
      student_concentration_propensity + student_boredom_propensity + student_confusion_propensity +
      student_frustration_propensity +
      question_concentration_propensity + question_boredom_propensity + question_confusion_propensity +
      question_frustration_propensity
```

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.8776825	0.0103125	-279.047	< 2e-16	***
prior_dropout	3.9012357	0.0456599	85.441	< 2e-16	***
student_concentration_propensity	-0.0029178	0.0022828	-1.278	0.20118	
student_boredom_propensity	0.0006981	0.0022136	0.315	0.75249	
student_confusion_propensity	-0.0031933	0.0023846	-1.339	0.18052	
student_frustration_propensity	-0.0076202	0.0025617	-2.975	0.00293	**
question_concentration_propensity	-0.0052919	0.0017588	-3.009	0.00262	**
question_boredom_propensity	0.0077022	0.0017155	4.490	7.13e-06	***
question_confusion_propensity	-0.0016415	0.0015136	-1.085	0.27814	
question_frustration_propensity	-0.0016226	0.0014365	-1.130	0.25868	

Project 3 : Predictive Analytics

- Prior dropout is the strongest predictor of future dropout in assignments.

- Affect propensity values for students and assignments appear to be significant, but show small coefficients.
    - Lower student frustration is significantly associated with dropping out.
    - Lower question concentration propensity and higher question boredom propensities are associated with dropouts.

Design systemic interventions to address students who have dropped out in the past.

Project 3 : Predictive Analytics

Questions