



Research Subteam: Presentation 3



Maxwell Bach, Okubay Gebrelibanos,
Yili Hui, Arjun Patel, Manley Roberts



Review: Master Plan

- Aggregate/clean data
- Create a library of models
- Evaluate model success

Standard Approach

- Structuring queries with known variables
 - Examples: course_id, user_id, usage_key, earned_all, possible_all, earned_graded, possible_graded
 - Inner joining courses to create a single database of assignments
- Trial and error
 - Discerning if the assignments are the “same” across courses from term to term
 - How to find which variables are useful or complete

Problems Encountered

- Some of the variables are empty or simply placeholders
 - Continually retrieving outputs such as “0.0” “0.0” “0.0” for earned assignment scores or possible assignment scores
- Data is incomplete in several places
 - Sets like `courseware_studentmodule` are insufficient or contain only fragments of usable data
- Further issues
 - Clickstream data cannot be useful without adequate assignment data since our models depend on both

Possible Solutions

- More data science/SQL training
 - Possibly consulting with Integration subteam
- Using variables previously believed to be useless
 - Might be the case that actual useful data can be extracted from odd places
- Finding usable data in currently unknown places
 - Unsure of what every variable name in the clickstream data represents

To come next

- Interpreting meaningful results from data gathered
- Continuing MongoDB - Clickstream data
- Possibly adding Piazza data