ASSISTments Dataset Description

IMPORTANT NOTE: Every value in this dataset is only a measure of the activity in ASSISTments within the date range specified. Only activity from assignments started within the date range is recorded within this dataset. For example, the student_answer_count column in the probelm_details table is only counting answers within the specified date range.

The ASSISTments standard dataset format consists of ten tables, documented in the sections below. IDs are shared between tables to enable cross referencing between tables. For example, if one wanted to know the total number of multiple choice problems completed by a student, one could use the problem_id field in the problem_details table and problem_logs table to combine the two tables and determine the answer to the above query.

ddets

The district_details table contains one entry for each district in the dataset.

district_id

A unique identifier for school districts.

location

A categorical value representing the location of the district. Locations are found through researching schools and school website domain owners. As such there is no guarantee that each location is accurate, however ASSISTments has found this method to be accurate enough for statistical purposes.

The location will be a state or possession of the United States, a list of which can be found here: https://pe.usps.com/text/pub28/28apb.htm, or “Non-US / Unknown” for districts that are either outside the United States or districts in an unknown location.

opportunity_zone

A categorical value representing whether or not a district is located in an opportunity zone as defined here: https://opportunityzones.hud.gov/.

This feature can be one of three values:

- Yes: The district is in an opportunity zone.
- No: The district is not in an opportunity zone.

- Contiguous: The district is adjacent to an opportunity zone, and thus may teach many students from an opportunity zone.

locale_description

This feature, when available, is a more specific description of the district's location. This measurement is only available sometimes, and only when the district is in an opportunity zone.

tdets

The teacher details `tdets` table contains one entry for each teacher in the dataset.

teacher_id

A unique identifier for teachers.

district_id

A unique identifier for school districts. ASSISTments uses teachers' school email addresses to identify their school district. If a teacher used a personal email address, they will not have a district id.

account_creation_date

A timestamp of when the teacher's account was created in the form YYYY-MM-DD HH:MM:SS.nnnnnnTZD.

assignment_reports_viewed_fraction

A measure of the number of assignment reports viewed by the teacher divided by the total number of assignment reports the teacher could have viewed. An assignment report is a summary of all students' performance on an assignment.

student_reports_viewed_fraction

A measure of the number of student reports viewed by the teacher divided by the total number of student reports the teacher could have viewed. A student report is a detailed report of a student's performance on an assignment.

graded_open_response_fraction

A measure of the number of open response questions graded by the teacher divided by the total number of open response questions the teacher could have graded. In very rare cases, problems are removed from ASSISTments, or have their type changed from an open response. In these cases, the teacher grading these problems is still tracked in the teacher logs table, but these problems are removed from this statistic because not enough information is available.

Not all teachers have open response questions available to grade. If the teacher has no gradable open response questions, this feature will be empty.

open_response_comment_fraction

A measure of the number of open response questions the teacher left a comment on divided by the total number of open response questions the teacher could have left a comment on. In very rare cases, problems are removed from ASSISTments, or have their type changed from an open response. In these cases, the teacher leaving a comment on these problems is still tracked in the teacher logs table, but these problems are removed from this statistic because not enough information is available.

Not all teachers have open response questions available to leave a comment on. If the teacher has no commentable open response questions, this feature will be empty.

cdets

The class details table contains one entry for each class in the dataset.

class_id

A unique identifier for classes of students.

teacher_id

A unique identifier for teachers.

class_creation_date

A timestamp of when the class was created by the teacher, in the form YYYY-MM-DD HH:MM:SS.nnnnnnTZD.

student_count

A measure of the number of students in the class.

problem_sets_assigned

A measure of the number of problem sets assigned to the class.

skill_builders_assigned

A measure of the number of skill builders assigned to the class.

adets

The assignment details table contains one entry for each assignment in the dataset.

assignment_id

A unique identifier for assignments.

class_id

A unique identifier for classes of students.

release_date

A timestamp of when the assignment became available for completion by students, in the form YYYY-MM-DD HH:MM:SS.nnnnnnTZD.

due_date

A timestamp of when the assignment was due to be completed by students, in the form YYYY-MM-DD HH:MM:SS.nnnnnnTZD.

assignment_type

A categorical value representing which type of assignment the assignment is.

There are two types of assignment:

- problem_set: Problem sets are assignments in which the student must complete every problem.

- skill_builder: Skill builders are assignments in which the student answers problems that require the same small set of skills to solve until the student answers a small number (usually 3) problems correctly in a row, or attempts to answer many problems (usually 20) without getting enough in a row. When a student answers enough problems in a row correctly, they are considered to have mastered the skill builder.

assigned_student_count

A measure of how many students have been assigned this assignment. This is based on the number of students in a class. It is possible that a teacher used google classroom to remove the assignment from specific student's dashboards, and if this is the case, ASSISTments has no way of knowing and therefore this number has the potential to be inaccurate.

started_student_count

A measure of how many students have started the assignment.

completed_or_mastered_student_count

A measure of how many students have completed the assignment if the assignment is a problem set or mastered the assignment if the assignment is a skill builder.

problem_count

A measure of how many problems are in an assignment. This is based on the maximum number of problems completed by a student during the assignment, therefore if no student completed the assignment, this value can be inaccurate.

mean_correct

A measure of the average correctness of students across all the problems in the assignment. This only takes into account students who have completed the assignment.

mean_time_on_task

A measure of the average time on task of students across all the problems in the assignment. This only takes into account students who have completed the assignment.

pdets

The problem details table contains one entry for each problem in the dataset, not including scaffolding problems, which are a type of tutoring.

problem_id

A unique identifier for problems.

content_source

A list of categorical values representing where the problem is sourced from. ASSISTments is currently improving our content tracking system, currently we have limited knowledge of a problem's source material.

The content sources are defined as follows:

- Undetermined: Unknown content source
- Certified Content: ASSISTments made math content
- Skill Builder: ASSISTments made math skill-builder content
- Textbook: Content from an indeterminable textbook
- Illustrative Mathematics: The Illustrative Mathematics Curriculum - https://www.illustrativemathematics.org/
- Open Up Resources: The Open Up Resources Math Curriculum - https://openupresources.org/
- Utah Math: The Utah Middle School Math Project Curriculum - http://utahmiddleschoolmath.org/
- Engage New York: The Engage NY Curriculum - https://www.engageny.org/
- Eureka Math: The Eureka Math Curriculum - https://greatminds.org/math
- State Tests: Content from released state tests

skills

A list of all the Common Core Mathematics Standards (http://www.corestandards.org/Math/) that apply to the problem.

Not all problems in ASSISTments have been tagged with their applicable skills, when a problem has no tagged skills, this feature will be empty.

problem_type

A categorical value representing what kind of answer the problem requires.

Examples or descriptions for each problem type are given below:

- Number: 12, -15.1, 14.5
- Numeric Expression: 5*3, 13.2/2 + 2, 5 - 5^2
- Exact Fraction: 3/5, -1/2, 17/19
- Algebraic Expression: 5x+2, 3y^2 + 2z, 3/x

- Multiple Choice: The student selects the only correct answer from multiple given answers
- Check All That Apply: The student selects all correct answers from multiple given answers
- Externally Run: A question that exists outside the ASSISTments platform, which connects to ASSISTments through the tutor to inform the tutor of the student's correctness
- Exact Match (ignore case): The student submits a written phrase that must match the text of the correct answer in every way except letter case.
- Exact Match (case sensitive): The student submits a written phrase that must match the text of the correct answer exactly.
- Ungraded Open Response: The student submits a written response which is graded after assignment completion by the teacher.
- Ordering: The student must order a set in the correct formation

tutoring_types

A list of categorical values representing what kinds of tutoring strategies are available for the problem,

Descriptions for each problem type are given below:

- Explanation: An explanation reveals the problem's answer to a student and often explains how to reach the answer.
- Hint: A hint provides a student with guidance on how to solve the problem without necessarily revealing the answer. The last hint available will reveal the answer.
- Scaffold: A scaffold is another problem with supplemental information that when solved, brings the student closer to the answer. A scaffold must be answered for the student to have another attempt at answering the original problem.

student_answer_count

A measure of how many students have answered this problem.

mean_correctness

A measure of the mean correctness of all students that have answered this problem.

mean_time_on_task

A measure of the mean time on task of all students that have answered this problem.

sdets

The student details table contains one entry for every student in the dataset.

student_id

A unique identifier for students.

class_id

A unique identifier for classes.

account_creation_date

A timestamp of when the student's account was created in the form YYYY-MM-DD HH:MM:SS.nnnnnnTZD.

started_problem_sets_count

A measure of how many problem sets have been started by the student.

completed_problem_sets_count

A measure of how many problem sets have been completed by the student.

started_skill_builders_count

A measure of how many skill builders have been started by the student.

mastered_skill_builders_count

A measure of how many skill builders have been mastered by the student.

answered_problems_count

A measure of the total number of problems completed by the student.

mean_problem_correctness

A measure of the average correctness of a student on every problem answered by the student aggregated from the problem_logs table.

If the student never completed a problem, this feature will be empty.

mean_problem_time_on_task

A measure of the average time on task of a student on every problem answered by the student aggregated from the problem_logs table.

If the student never completed a problem, this feature will be empty.

tlogs

The teacher logs table contains one entry for each action level interaction between a teacher and the ASSISTments platform.

teacher_id

A unique identifier for teachers.

assignment_id

A unique identifier for assignments.

student_id

A unique identifier for students.

For entries where the action isn't related to a specific student, this feature will be empty.

problem_id

A unique identifier for problems.

For entries where the action isn't related to a specific problem, this feature will be empty.

action

A categorical value representing the interaction between the teacher and ASSISTments.

The actions are defined as follows:

- assignment_report_viewed: the teacher viewed a summary of all students' performance on an assignment.

- student_report_viewed: the teacher viewed a detailed report of a student's performance on an assignment.
- comment_written: the teacher writes a comment for a student on their open response answer
- open_response_graded: the teacher grades a student's open response problem

timestamp

A timestamp of when the action took place in the form YYYY-MM-DD HH:MM:SS.nnnnnnTZD.

alogs

The assignment logs table contains one entry for every instance of a student doing an assignment.

log_id

A unique identifier for an instance of a student doing an assignment.

student_id

A unique identifier for students.

assignment_id

A unique identifier for assignments.

start_time

A timestamp of when the assignment was started in the form YYYY-MM-DD HH:MM:SS.nnnnnnTZD.

mean_correct

A measure of the average correctness of a student on every problem in the assignment answered by the student aggregated from the problem_logs table.

If the student started, but never completed a problem in the assignment, this feature will be empty.

time_on_task

A measure of the total time on task of a student on the assignment aggregated from the problem_logs table.

assignment_completed

A measure of whether or not the student completed the assignment, either by answering all problems, in the case of a problem set, or reaching mastery, in the case of a skill builder. This feature is true if the assignment was completed, and false if not.

plogs

The problem logs table contains one entry for each instance of a student answering a problem, not including scaffolding problems, which are a type of tutoring.

log_id

A unique identifier for an instance of a student doing an assignment.

student_id

A unique identifier for students.

assignment_id

A unique identifier for assignments.

problem_id

A unique identifier for problems.

start_time

A timestamp of when the problem was started in the form YYYY-MM-DD HH:MM:SS.nnnnnnTZD.

time_on_task

An estimate of the number of seconds the student spent solving the problem based on an aggregation of student_log information.

answer_before_tutoring

A measure of whether or not the student attempted to answer the problem before requesting tutoring, which is true if the student attempted to answer before tutoring, and false if not.

For entries where the student never attempted to answer or requested tutoring, this feature will be empty.

fraction_of_hints_used

A measure of what percent of hints were viewed by the student when solving the problem.

For entries where the problem has no hints available, this feature will be empty. Some hints have been deleted from the database, in this case the fraction_of_hints_used can be greater than one, because students had more hints available to them when they completed the assignment than when we collected this data.

attempt_count

A measure of how many attempts the student made to answer the problem.

answer_given

A measure of whether or not the student was given the answer to the problem, which is true if the student was given the answer, and false if the student was not.

problem_completed

A measure of whether or not the student completed the problem, which is true if the problem was completed and false if not.

correct

A measure of whether or not the student's first answer to the problem was correct, which is one if the student answered correctly on their first attempt, and zero if not.

For entries where the student did not complete the problem and the teacher did not enter a grade for the incomplete problem, or the problem was an open response question that the teacher did not grade, this feature will be empty.

slogs

The student logs table contains one entry for each action level interaction between a student and the ASSISTments tutor.

log_id

A unique identifier for an instance of a student doing an assignment.

student_id

A unique identifier for students.

assignment_id

A unique identifier for assignments.

problem_id

A unique identifier for problems.

For entries where the action isn't related to a specific problem, this feature will be empty.

action

A categorical value representing the interaction between the student and the tutor.

The actions are defined as follows:

- assignment_started: The student has started the assignment.
- assignment_resumed: The student has resumed the assignment after leaving the tutor.
- assignment_finished: The student has completed the assignment.
- problem_started: The student has started a problem.
- problem_finished: The student has finished a problem.
- tutoring_requested: The student has requested tutoring for a problem.
- correct_response: The student submitted a correct answer.
- wrong_response: The student submitted a wrong answer.
- open_response: The student submitted an answer to an open response problem.
- work_submitted: The student submitted their work for review.
- answer_requested: The student received the answer to a problem.
- continue_selected: The student moved on to the next problem after completing the current problem.

timestamp

A timestamp of when the action took place in the form YYYY-MM-DD HH:MM:SS.nnnnnnTZD.