

Private Datasets

Person-Course-Survey Dataset

The Person-Course-Survey dataset is at the level of one row per-person, per-course. So, for example, if one individual enrolled in three MITx or HarvardX courses during the period covered by the dataset (for this release, Fall 2012, Spring 2013, Summer 2013, Fall 2013, Spring 2014 and Summer 2014), that person would have three rows associated with their user ID.

The following is a list of variables for the “Person Course Survey” dataset, which will comply with FERPA constrained datasets:

List of Variables

course_id

administrative; course ID in the standard format as org/number/semester

registered

administrative; Boolean flag, indicating that the user registered in the course (obsolete: should always be 1 or true)

viewed

administrative; Boolean flag, indicating that the user visited the course at least once

explored

administrative; Boolean flag, indicating that the user viewed at least half of the chapters of the course

certified

administrative; Boolean flag, indicating that the user earned a certificate in the course

completed

administrative; Boolean flag, indicated that the user completed the course by achieving a final grade above the overall passing threshold for the course ($grade \geq passing_grade$)

cc_by_ip

administrative; Two-letter country code corresponding to the modal IP

address

countryLabel

administrative; Full name of country corresponding to modal IP address

city

administrative; Full name of city corresponding to modal IP address

region

administrative; Two-Letter code of region or U.S. State corresponding to modal IP address

subdivision

administrative; Full name of region/subdivision or U.S. State corresponding to modal IP address

postalCode

administrative; Postal code corresponding to modal IP address

un_major_region

administrative; UN defined sub-continent major geographic regions corresponding to modal IP address

un_economic_group

administrative; UN defined major geographical economic groups (developing and developed nations) corresponding to modal IP address

un_developing_nation

administrative; UN defined developing nation groups files corresponding to modal IP address

un_special_region

administrative; Special Regions Set 1 files: Latin America and the Caribbean.csv, Sub-Saharan-Africa.csv corresponding to modal IP address

LoE

user-provided; [Level of education](#) specified when registering for an edX user account

'p': Doctorate

'm': Master's or professional degree

'b': Bachelor's degree

'a': Associate degree

'hs': Secondary/high school

'jhs': Junior secondary/junior high/middle school

'el': Elementary/primary school

'none': No formal education
'other': Other education
(blank): User did not specify level of education
'p_se': Doctorate in science or engineering (no longer used)
'p_oth': Doctorate in another field (no longer used)
NULL: This student signed up before this information was collected

YoB

user-provided; Year of birth

gender

user-provided; gender, one of 'm', 'f', or 'o' (or NULL)

grade

administrative; Final grade earned in the course, on a scale of 0 to 1

passing_grade

administrative; Passing grade for this course, on a scale of 0 to 1. This value should be the same within a course; Note: This value may be null if the course has course content hosted-off platform, but uses edX for registrations

start_time

administrative; Date of enrollment in the course

first_event

administrative; Timestamp of user's first event in the course from tracking logs

last_event

administrative; Timestamp of users' last event in the course from tracking logs

nevents

administrative; Number of tracking log events

ndays_act

administrative; Number of days with activity, from tracking logs

nplay_video

administrative; Number of video play events from tracking logs

nchapters

administrative; Number of chapters visited by the user

nforum_posts

administrative; Number of forum posts and comments made by user. This represents the sum of forum threads and comments.

nforum_votes

administrative; Number of votes received by user's forum posts

nforum_endorsed

administrative; Number of endorsed forum posts by user

nforum_threads

administrative; Number of forum post threads by user (defined by an [edX comment thread](#))

nforum_comments

administrative; Number of forum post comments by user (defined by an [edX comment](#))

nforum_pinned

administrative; Number of forum posts by user which were pinned (good indicator of staff)

roles

administrative; Roles played by the user in the course, 'Student' or 'Staff'; 'Staff' when any of the following roles are assigned: *roles_isBetaTester* or *roles_isStaff* or *roles_isInstructor* or *roles_isCCX* or *roles_isFinance* or *roles_isLibrary* or *roles_isSales* or *forumRoles_isAdmin* or *forumRoles_isCommunityTA* or *forumRoles_isModerator*; 'Student' otherwise; NULL indicates missing data from both edX roles tables: [student courseaccessrole](#) and [django comment client role users](#)

nprogcheck

administrative; Number of progress check events from tracking logs

nproblem_check

administrative; Number of problem check events from tracking logs

nforum_events

administrative; Number of all kinds of forum events, from tracking logs

mode

administrative; Mode of registrant, e.g. honor, audit, verified

is_active

administrative; 1 if enrollment active on date of this data, else 0

cert_created_date

administrative; date when certificate was generated (if applicable)

cert_modified_date

administrative; date when certificate was modified (if applicable)

cert_status

administrative; [certification status](#), specified as 'downloadable', 'notpassing', 'unavailable', 'audit_notpassing', 'audit_passing' or 'error'. (Note: 'audit_passing' and 'audit_notpassing' added 26 Jan 2016 for audit enrollments created after 1 Dec 2016)

verified_enroll_time

administrative; time when user initially had their first id-verified enrollment event

verified_unenroll_time

administrative; time when user initially had their last id-verified un-enrollment event, if applicable. NULL is never unenrolled

profile_country

administrative; two-letter country code as specified by user in their profile (only available after ~mid-2014)

y1_anomalous

administrative; if this user was anomalously missing and had to be side-loaded from mongo then this field = 1, else null

email_domain

administrative; domain of the email address of the user, i.e. what appears after the '@'

language

administrative; modal transcript language; null indicates either transcripts were not used by the person (if they existed for that course) or that no transcripts were available for that course

language_download

administrative; modal transcript language download count for any video

language_nevents

administrative; modal transcript language events

language_ndiff

administrative; count of different transcript languages used per person

enrolled in that respective course

ntranscript

administrative; Number of video show transcript events from tracking logs

nshow_answer

administrative; Number of show answer events from tracking logs

nvideo

administrative; Number of video events (of all kinds) from tracking logs

nseq_goto

administrative; Number of 'sequence goto' navigational events from tracking logs

nseek_video

administrative; Number of 'video seek' events from tracking logs

npause_video

administrative; Number of 'video pause' events from tracking logs

nvideos_unique_viewed

administrative; Number of unique videos watched

nvideos_total_watched

administrative; Percentage of all videos watched in the course

avg_dt

administrative; Average time difference in seconds between consecutive events from tracking logs

sdv_dt

administrative; Standard deviation of difference in seconds between consecutive events from tracking logs

max_dt

administrative; Maximum difference in seconds between consecutive events from tracking logs

n_dt

administrative; Number of consecutive events used in time difference computations, from tracking logs

sum_dt

administrative; Total elapsed time (in seconds) spent by user on this course, based on time difference of consecutive events, with 5 min max cutoff, from tracking logs

roles_isBetaTester

administrative; 1 if user is a course beta tester with 'role' = 'beta_tester' from edX table: [student courseaccessrole](#); otherwise NULL

roles_isInstructor

administrative; 1 if user is a course instructor with 'role' = 'instructor' from edX table: [student courseaccessrole](#); otherwise NULL

roles_isStaff

administrative; 1 if user is course staff (or other staff roles) with 'role' = 'staff' from edX table: [student courseaccessrole](#); otherwise NULL

roles_isCCX

administrative; 1 if user is ccx coach with 'role' = 'ccx_coach' from edX table: [student courseaccessrole](#); otherwise NULL

roles_isFinance

administrative; 1 if user is finance admin with 'role' = 'finance_admin' from edX table: [student courseaccessrole](#); otherwise NULL

roles_isLibrary

administrative; 1 if user is library user with 'role' = 'library_user' from edX table: [student courseaccessrole](#); otherwise NULL

roles_isSales

administrative; 1 if user is sales admin with 'role' = 'sales_admin' from edX Table: [student courseaccessrole](#); otherwise NULL

forumRoles_isAdmin

administrative; 1 if user is a forum admin with 'name' = 'Administrator' from edX table: [django comment client role users](#); otherwise NULL

forumRoles_isCommunityTA

administrative; 1 if user is a forum community TA with 'name' = 'Community' from edX table: [django comment client role users](#); otherwise NULL

forumRoles_isModerator

administrative; 1 if user is a forum moderator with 'name' = 'Moderator' from edX table: [django comment client role users](#); otherwise NULL

forumRoles_isStudent

administrative; 1 if user is a forum student (everyone in the course seems to

be assigned this) with 'name' = 'Student' from edX table:
[django comment client role users](#);; otherwise NULL;