course\_id: administrative, string, identifies institution (HarvardX or MITx), course name, and semester, e.g. “HarvardX/CB22x/2013\_Spring”   
   
userid\_DI: administrative, string, first portion identifies dataset (MHxPC13 corresponds to MITx HarvardX Person-­‐Course AY13), second portion is a random ID number. Example ID: “MHxPC130442623”.   
   
registered: administrative, 0/1; registered for course, =1 for all records in person-­‐course.   
   
viewed: administrative, 0/1; anyone who accessed the ‘Courseware’ tab (the home of the videos, problem sets, and exams) within the edX platform for the course. Note that there exist course materials outside of the ‘Courseware’ tab, such as the Syllabus or the Discussion forums.   
   
explored: administrative, 0/1; anyone who accessed at least half of the chapters in the courseware (chapters are the highest level on the “courseware” menu housing course content).   
   
certified: administrative, 0/1; anyone who earned a certificate. Certificates are based on course grades, and depending on the course, the cutoff for a certificate varies from 50% -­‐ 80%.   
   
final\_cc\_cname\_DI: mix of administrative (computed from IP address) and user-­‐ provided (filled in from student address if available when IP was indeterminate); during de-­‐identification, some country names were replaced with the corresponding continent/region name. Examples: “Other South Asia” or “Russian Federation”.   
   
LoE: user-­‐provided, highest level of education completed. Possible values: “Less   
than Secondary,” “Secondary,” “Bachelor’s,” “Master’s,” and “Doctorate.”   
   
YoB: user-­‐provided, year of birth. Example: “1980”.   
   
  
gender: user-­‐provided. Possible values: m (male), f (female) and o (other).   
   
grade: administrative, final grade in the course, ranges from 0 to 1. Example: “0.87”.   
   
start\_time\_DI: administrative, date of course registration. Example: “12/19/12”.   
   
last\_event\_DI: administrative, date of last interaction with course, blank if no   
interactions beyond registration. Example “11/17/13”.   
   
nevents: administrative, number of interactions with the course, recorded in the   
tracking logs; blank if no interactions beyond registration. Example: “502”.   
   
ndays\_act: administrative, number of unique days student interacted with course.   
Example: “16”.   
   
nplay\_video: administrative, number of play video events within the course.   
Example: “52”.   
   
nchapters: administrative, number of chapters (within the Courseware) with which   
the student interacted. Example: “12”.   
   
nforum\_posts: administrative, number of posts to the Discussion Forum. Example: “8”.   
   
roles: administrative, identifies staff and instructors, but blank as staff and   
instructors were removed from this release.   
   
inconsistent\_flag: administrative, identifies records that are internally inconsistent. Due to a variety of data issues, including missing tracking logs and one course (CS50x) which has virtually no logs because most of the course content is hosted outside of the edX platform, a portion of the records have null values for nevents but have non-­‐null values for ndays\_act, nforum\_posts, or nchapters. The source   
for nevents and for last\_event\_DI is the tracking logs, whereas ndays\_act, nforum\_posts, and nchapters come from a data source known as the “Courseware Student Module”1. Due to the two different sources, if something is wrong with the Tracking Logs2 for a class or a student, then records in Person Course can be internally inconsistent and have a value of ‘1’ in this column.

age- derived column from Year Of Birth

total\_activity- derived from start and end date of activity