

edX Dashboard

The dataset is provided by edX (<https://www.kaggle.com/edx/course-study>). It includes information about student performance and interaction with the course along with statistics relating to number of participants and their respective completion rates. The data covers four academic years from 2012 to 2016 and lists all courses offered by edX, split into four course subjects.

The dashboard helps answer the following question:

Question: Are there specific indicators that appear to be predictors for higher course completion rates for edX online courses?

Answer: The courses with the highest completion rates ($> 20\%$) are all short courses from the Humanities, History, Design, Religion and Education subject area. They vary in popularity. However, none of these courses is overly popular, with the most popular course recording 12,856 participants. 17.42% of participants posted in the course forums for these courses, compared to the average for all edX courses (9.09%). 78.85% of participants hold a bachelor's degree (compared to the average for all edX courses - 72.12%) and 45.43% are female (compared to the average for all edX courses - 33.08%). These statistics might shed light into the nature of these courses and their participants and assist course administrators in their efforts of increasing student completion rates. Our dashboard seems to indicate that short humanities courses have the highest completion rates and long STEM courses have the lowest completion rates.

How: The first chart is a scatter plot that shows the relationship between course completion and number of participants. It helps us see which courses have a high completion rate, and their respective popularity. The second chart is a treemap plot that links the high completion rate courses that we chose and demonstrates the courses' median completion times. Finally, the five pie charts add vital statistics about student grade, video plating, forum posting, previous education and gender.

The scatter plot and treemap plot have consistent colors, representing the four subject areas as classified by edX. They are aligned next to each other to ease comparisons when investigating a course in both plots. The pie charts are added below as they add to our course investigation. They are colored blue/red, with red being the percentage that is linked to course incompleteness (except in the gender case where color is meaningless and was chosen at random).

The first chart is the scatter plot on the left. This was placed in the top left as it is the first thing that a dashboard user would look at. The scatterplot utilizes position very well, showing clusters of courses with similar combinations of participant numbers and completion rates. Color in the scatter plot adds to positional information, further emphasizing clusters of subject areas, such as red courses with higher completion rates or a few blue courses with high popularity. The y-axis is the course completion percentage, which is the outcome variable we are investigating. The x-

axis is the log transformed number of participants. Two introductory computer science courses are especially popular, skewing the x-axis, and a log transformation was needed to normalize the variables.

The second chart is the treemap plot on the right. The treemap plot acts as a third dimension to the first chart, showing median course completion times (in hours) for each course. The chart also uses the same color scheme as the first chart and utilizes size by showing courses with higher completion times as larger rectangles. Adding these sizes for each data point in the first chart would produce an extremely messy visual and limit interpretation.

Selecting a course (or group of courses) in the first chart highlights the same courses in the second chart, showing its size (median hours) compared to other courses. Likewise, highlighting courses in the second chart will highlight courses in the first chart showing the course's position in terms of completion rate and popularity. Selecting a course from either the first or second chart will show relevant statistics in the pie charts.

Hovering over a course in either the first or second chart will display a tool tip with details about the course including the course number/title, institution and launch date. The dashboard user might also choose a particular academic year to see courses offered and their performance each year. Also, the dashboard provides a search function to look up a particular course and see its performance in the chosen year. Finally, the dashboard user can filter courses by course subject area, helping in comparisons between different areas.