

Learning Analytics

Exploratory Data Analysis on data set from future learn cybersecurity course data set

Thamizhiniyan Pugazhenth

17/11/2021

INTRODUCTION

We generate data on practically every aspect of our lives, which can be harnessed and analysed to reveal powerful insights about our behavior, preferences, and future actions, and the amount, velocity, and diversity of data we generate is fast increasing. Students in higher education leave a digital footprint during their studies that tells us about their learning and experiences at university. Universities can use this information to better understand how students learn and improve the student experience. This is referred to as “Learning Analytics.”

OVERVIEW

Learning analytics has the potential to be a huge help in improving the university student experience. Universities should use analytics systems that are designed in consultation with students, supported by an ethical framework or policy, driven by the improvement of learning and teaching processes and student engagement, tailored to the specific needs of each institution, and embedded in an institution’s strategic plan to ensure students get the most benefit from analytics. Higher education institutions that want to serve their students in a data-driven environment must have good data management. This entails assigning data management ownership and leadership to one person or area, ensuring data accuracy and consistency, keeping data secure, ensuring appropriate access controls are in place so that people who need the data can access it and those who don’t can’t, creating a culture where data is everyone’s asset and responsibility, and developing good data capability among their employees.

The statistical analysis of past and present data gathered from learners and the learning process to develop prediction models that improve the learning environment. Most analytics systems in use in the UK – including those covered so far in this research – now focus on dashboards that visualize previous and current learner data. This is referred to as “post-mortem analytics,” in which historical data is used to guide and target present interventions. While this is useful for tracking learner progress and identifying and supporting disengaged or struggling learners, big data analytics has the potential to go even further in higher education. Predictive learning analytics is defined as “the statistical analysis of historical and current data derived from learners and the learning process to create models that allow for predictions that improve the learning environment within which it occurs,” according to a recent ECAR Working Group Paper. This gives learners, tutors, and institutions access to a wealth of information that can help them with their learning, tutoring, and overall strategic decision-making¹.

Learning analytics data could act as a measure of the learner’s commitment to learning, training and employment of staff and teaching intensity².

DATASET DESCRIPTION

In Future Learn MOOC data set₃, we have different sets of data files from seven different runs of cyber security course on different time frames from future learn website. Each set contains enrollments, question response, step activity, team members, video stats, weekly sentiment survey, leaving survey and archetypes.

Enrollment data file contains the data about each person who enrolled in each run of the course. The column of the data file reads the designed id to each enrolled, the time stamp which they enrolled and unenrolled, role of each enrolled, fully participated at, purchased statement at, gender, country, age range, highest education level, employment status, employment area and country from which they are accessing from. Question response data file contains the data about the questions answered by each learner in the quiz they had attended. The columns are learner id, number of the quiz question, type of the question, week number, step number, response they had given, submitted at and whether it is correct or not. Step number data file explains the data about the step each learner had visited. The data are learner id, step number, week number, first time they had visited the step and the time they had completed the step at. Weekly survey sentiment data file contains the data about the rating and review after each week. It reads the columns id, responded at, week number, experience rating and reason behind the rating.

Leaving survey data file contains the data behind leaving. It explains the data about the learner id, time they left at, leaving reason, last completed at, last completed step number and week number. Video stats contains the data which deals with the video associated with the course. It has everything about the videos like the question it associated with, title, duration, total views, downloads, caption views, transcript views, HD view, percentage of the video viewed – five percent, ten, twenty-five, fifty, seventy-five, ninety-five, hundred, console in which it had been viewed – device, desktop, mobile, tv, tablet, unknown data and the continent from which it had been viewed – Europe, Oceania, Asia, North America, South America, Africa and South America.

Learning Archetypes represent user behavior and general characteristics, and an archetype can be a general guideline about end user's behavior and are general guidelines on how a user behaves₄. The types of archetypes are Fixers, Vitalizers, Advancers, Preparers, Explorers, Flourishers, Hobbyists₅. Learning Archetypes data file segregates learners with different archetypes. It contains the data which has columns – learner id, responded at and the type of archetype the learner is.

From the data set, we will draw different analysis in terms of similarities, differences, positives, negatives, and the improvements they can make along with the recommendations we can tell the future learn team to enhance the user experience.

First, we will consider enrolled people from all seven runs of cyber security course as a single data frame. We will classify the enrollments into four categories – age, gender, highest education level and employment status and plot the data to get clear view about the trend of enrollments.

When we see all four plots, one factor that is very evident is the rate of “Unknown” entries across all four columns of enrollment data. The team at Future Learn need to get all details of an enrolled learner without any compromise. They are able to understand the trend of learners and their background if they get those details. Moreover, it will improve the team to concentrate on areas in which they are lacking and improve even more on areas in which they are good at.

From the below plot regarding age range, we cannot see any discrete differences in the block of age range across ages. Leave “Unknown” group aside. If they set the personal details entry as mandatory, we can fetch even more clear picture about the age range. They will create even more courses or tweak cyber security course according to the learners who represent majority of the age range.

















