Cyber Security MOOC Data Analytics

Rushi Girdharbhai Vasoya

2026-01-15

CRISP-DM Cycle 1: Understanding Learn Engagement Across Course Runs

Business Understanding

Online learning platform, like Newcastle University, Collectively produce hundreds of thousands if not millions of record of learners interaction across multiple run of course. Course teams and the platform analysis seek insight into how learners engage with the course content as a part of course effectiveness and future design decision. Specially, While environment figures may indicate the reach of a course they do not necessarily reflect the way in which learner actively participate once enrolled.

The first investigation here in explores general trends in the pattern of learner engagement across multiple runs of course. Engagement is considered at the foundations level using enrollment data and the step activity records, which together provide insight into both learner participants and interaction with the course content. This foundation understanding must be established before more complex behavior indicators can be introduced in later analysis.

This study will try to establish whether engagement pattern are the same between run and explorer to what extent enrollment is covered into actual participation. The study is deemed to be successful if through the use of descriptive Statistics and visualization pattern of engagement can be described clearly and these findings set a meaningful foundation for future investigation in cycle 2.

Data Understandig and Preoaration

To data analysis relies on two main data sources provided from the Newcastle University: Enrollment data and step activity data. The enrolment data tracks when learner enrolled for each run of our course, while step activity record how learners have engaged with specific steps of Course.

Data from serveral runs of course where aggregated into structured data set. This data set represent a number of rows with several thousand entries from learner enrollment and intractions at a step level, with columns having learner IDS, Run IDS, steps and activity value the data column includes both category and numerical variables.

Initial discovery yielded a number of issues that impact data quality. A portion of student are represented in enrol data without being reflected in activity data, indicating that they do not participate after enrolling. Also, activity levels are highly variable from run to run, which is in a variability in activity distributions. This have been maintained since these are real activity patterns in learners.

Only enrollment and step activity data were chosen for this cycle in order to keep the focus of the analysis clear and easy to interpret. Both of these datasets are the most basic form of indicators for engagement and can be easily compared at highly level without adding any complexity of activity. The video and response data will be saved for investigation two.

For the diagnosis the environment file as well the step activity file from all the runs were combined in a consolidated form. The identifier for the running of the courses were maintained in order to allow further comparison of data, while aggregation was done in order to obtain the total of the enrollments and the step activities for every run.

Data Analysis

This section addresses the research question: what are overall patterns of the learner engagement across all course runs? The analysis is bounded by descriptive Comparisons of learner enrollment and step activity across multiple runs of the course in order to find broad engagement trends.
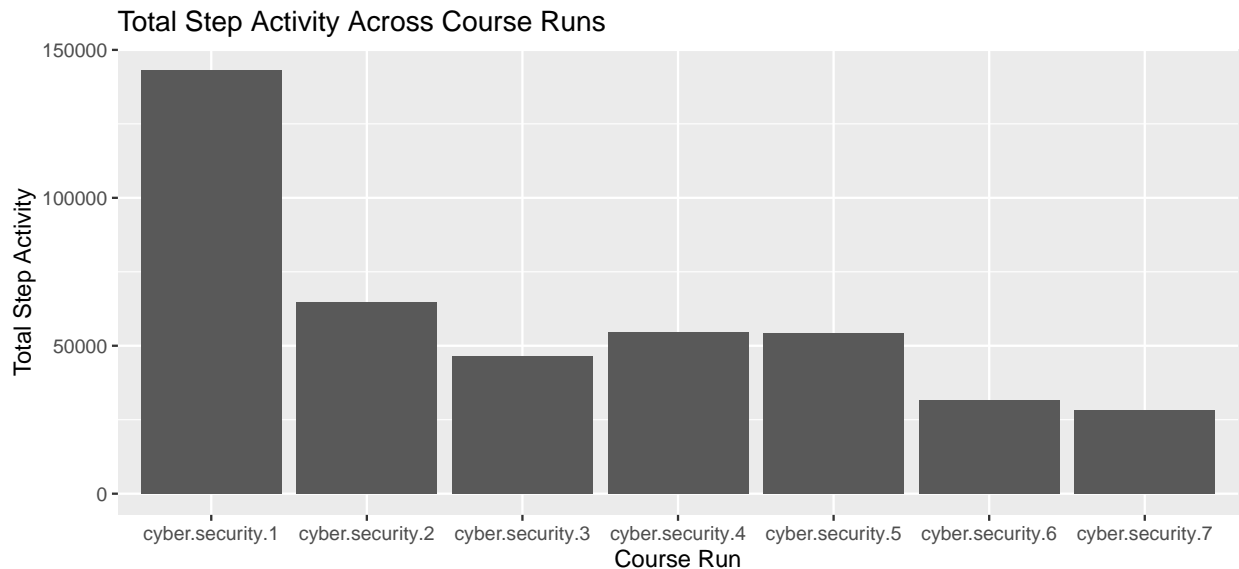
Learner engagement is considered from two complementary viewpoints. First, enrollment data is used to indicate the level of learner participation At a point of registration. Second step activity data is used to represent observable engagement with course content after enrollment. Comparing these measure occurs runs allow the variation in engagement to the be assessed and whether enrollment Translates into active Participation.

Descriptive statistics of total enrollments and total step activity were generated for each course run. The result have allowed the comparisons of engagement pattern across different iterations of course at a high level.

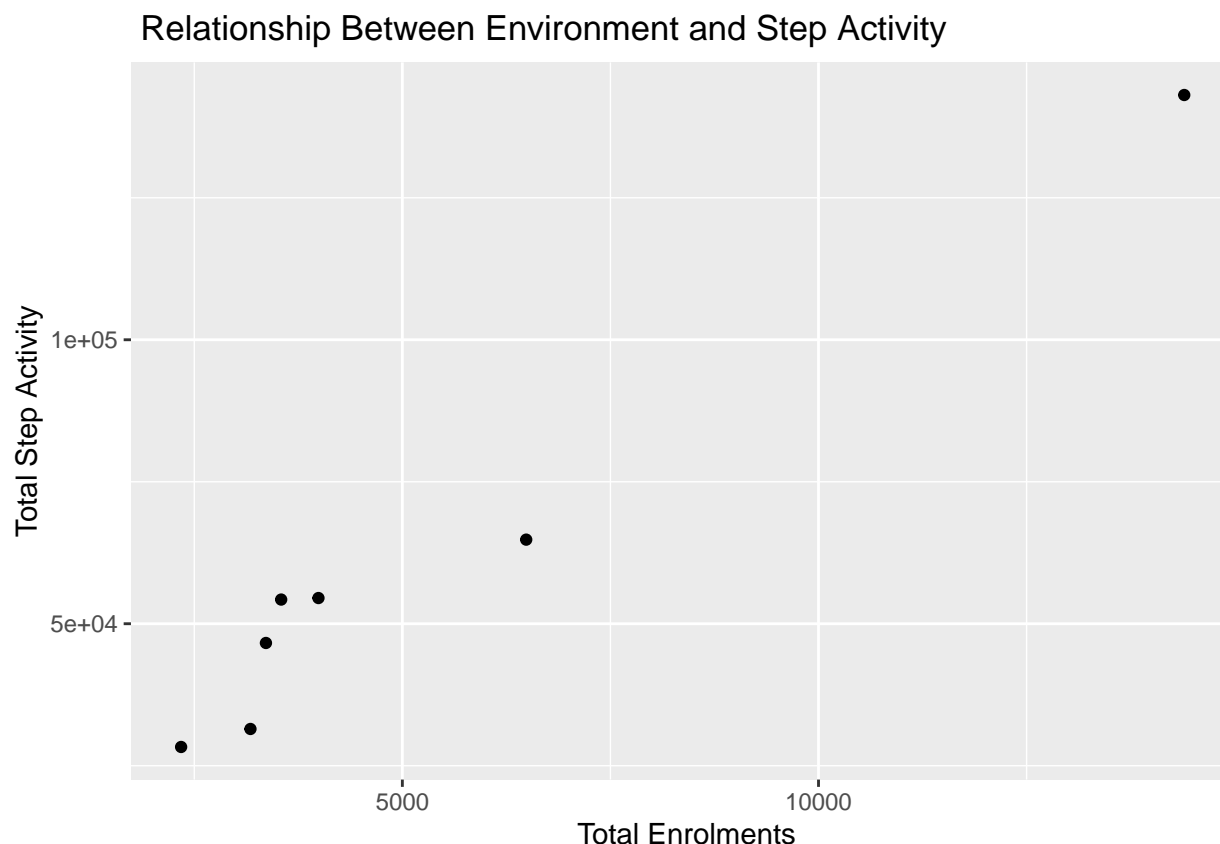Table 1: Summary of enrollments and step activity occurs course runs

run_id	total_enrolments	total_step_activity
cyber.security.1	14394	143092
cyber.security.2	6488	64809
cyber.security.3	3361	46614
cyber.security.4	3992	54524
cyber.security.5	3544	54257
cyber.security.6	3175	31472
cyber.security.7	2342	28304

To further support interpretation, graphical representation Were employed to illustrate contrast in the level of engagement for multiple instance of a course.

Analysis of the findings indicates that the level of engagement of learners differ significantly across the runs of the courses. Although some courses have high level of step engagement, other tend to have low levels of engagement, yet the number of enrolled learners in the two scenarios is similar. This implies that the behavior of learners in terms of their engagement with the course differs across the runs of the courses.

Further analysis of the relationship between enrollments and step activity data reveals that greater enrillments are not necessarily correlated to greater engagement levels.

Relationship Between Environment and Step Activity

In other runs, There may be many learners who enroll, but there may be fewer engagement with course steps. In this case It clearly show that there are learners who become disengaged after enrolling making it crucial to include behavioral engagement measures in addition to participation counts.

Overall, it can be seen that there is a descriptive analytical summary about engagement that proves engagement intensity is not consistent between runs and does not scale with enrollment.

Evaluation

Objective of cycle one was to build a founded fundamental understanding of learner engagement by leveraging enrollment data and data from step activities. Investigation cycle 1 was successfully in establishing pattern of learner engagement and the fact that learner participation and engagement differ from each other.

These results satisfy requirements for success that were identified in business understanding phase. Data from multiple course runs has been combined successfully, and engagement has been summarized as both numerical and graphical it is evident that data on environments is not adequate in describing engagement and the data on step activity is useful in Complementing data environments.

These outcomes highlights some critical Limitations in measure of high level engagement. It can be noted that step activity provides insight into learner engagement volume, Though it fails to provide an understanding

of learner activity in terms of different categorical of content. Consequently, these outcome generate further the concern in relation to learner engagement activity.

Therefore, findings from cycle 1 provide a direct motivation for what is explored cycle 2. In the subsequent study, a deeper understanding of learner engagement will be gained by including a far Sophisticated Set of data points related to learner engagement, like video based, interaction based, learning response, and architect data.