4160 Learning Analytics Notes

Learning Analytics data is – as a rule – considerably more messy than *performance* data. In our previous section we worked with the snapshots of test/quiz (performance) data, which can be analyzed using a variety of psychometric methods. In this section, we are looking at behavior, changes over time, and telling a story.

To tell the story, as I mentioned in the section overview, we have to define a behavior of interest both as a concept (words) and as a formula (math/data manipulation steps). Then we can look at that behavior – as it changes over time – and compare it to outcomes like test/quiz scores or course-level results (pass, fail, withdraw).

* <https://www.kaggle.com/datasets/rocki37/open-university-learning-analytics-dataset/data>
* <https://www.kaggle.com/code/jennamatthews/open-learning-analytics-jam/edit>

I have added three csv files to this module. Throughout all the files, some columns are consistent. For example, code\_module and code\_presentation identify the course. The id\_student code identifies the same student across all three files. For the purpose of this assignment, I have limited the data set to a single course.

1. studentVleDeets.csv: This has activity data for student activity. The date is a simple counter, relative to the first day of the course. sum\_click gives a count of user clicks – which can be used to determine the volume of student activity. Finally, activity\_type has a variety of categories, including forumng (forums) and homepage and oucontent.
2. studentAssessmentsOfInterest.csv: This has assessment activity for students. The data of interest here includes the score (student grade on the assessment) and date\_submitted.
3. studentOutsSingleFile.csv: This has one row per course-student with their final result (Pass, Fail, Withdrawn) and date\_unregistration (withdrawal date) where applicable.