



10 Academy Batch 3: Week 10

Moodle Database: Educational Data Log Analysis

Overview

This week you will analyse the 2019 10 Academy learners activity in the Moodle Learning Management System. The moodle LMS is a highly scalable framework, and all students activities are stored in a highly structured database.

Business Need

Many educational facilities such as colleges, universities, bootcamps rely on scalable and versatile Learning Management Systems.

The Moodle LMS is a free and open-source learning management system written in PHP and distributed under the GNU General Public License. It is used for blended learning, distance education, flipped classroom and other e-learning projects in schools, universities, workplaces and other sectors. With customizable management features, it is used to create private websites with online courses for educators and trainers to achieve learning goals. Moodle allows for extending and tailoring learning environments using community-sourced plugins¹.

In 2019, 10 Academy used the Moodle LMS to manage about 1000 students in their 6 months data science training. Learners, course instructors, and all admins interacted with

¹ <https://en.wikipedia.org/wiki/Moodle>

the Moodle system for almost all the training activities. All events from these activities are logged in the moodle postgres database.

10 Academy, like any other educational facility, is interested to understand the learners skill and knowledge development, and is interested to build models that are able to predict possible dropouts as well as classify learners into doing, well, doing ok, and struggling groups. 10 Academy is also interested in automating the process of reminding learners approaching deadlines, providing quick feedback based on their overall community engagement and performance. Moreover, given the main goal of 10 Academy training is to make students job ready, it wants to measure students' performance across many relevant metrics.

Your challenge: assuming a role-model of a freelancer hired by 10 Academy, you are to explore the 10 Academy Moodle logs stored in the database together with many other relevant tables. By the end of your analysis, you are expected to build a Tableau dashboard that illustrates the progress of students across time.

Data:

You can download the anonymised 10 Academy moodle postgres database copy [here](#).

Check [this reference](#) for commands used to restore a postgres database from pg_dump generated .sql file - which is what you download with the link given above. In general it is a three step process

1. Download and install postgres database if it is not already installed in your computer
2. In the command line create a database named *moodle* with psql for the user

your_user_name

```
$ sudo -u {your_user_name} createdb moodle
```

3. Add the tables from the dump to the local database you just created

```
$ psql -U tenac -d moodle -f devmoodle_anonymised.sql
```

Moodle logs

The first place to start to understand moodle logs is to understand the model of events in moodle [here](#). In the link, the description of the log column names are also explained.

For dashboard creation illustrating users engagement can be extracted using moodle log. For more information on how to do analytics with moodle log entries, see [this](#) . Important columns to know:

Events: Events are atomic pieces of information describing something that happened in Moodle. Events are primarily the result of user actions, but could also be the result of the [cron](#) process or administration actions [undertaken via the command line](#).

Most Important Tables (MIT)

Moodle database is complex - with more than 400 connected tables! In this project we are interested only in the subset of the tables. The most important tables we will consider in this challenge are (tables in bold are VIP)

- **mdl_logstore_standard_log**
- **mdl_context**
- **mdl_user**
- **mdl_course**
- **mdl_modules**
- **mdl_course_modules**
- **mdl_course_modules_completion**
- mdl_grade_items
- **mdl_grade_grades**
- mdl_grade_categories
- mdl_grade_items_history
- mdl_grade_grades_history
- mdl_grade_categories_history
- mdl_forum
- mdl_forum_discussions
- mdl_forum_posts

WARNING: You may not have time to understand all tables, so be selective and focus only on what is essential for this challenge.

Learning Outcomes

- Understanding complex database schema

- Understanding how user interactions data is modeled using events
- Working with postgres database
- Educational logs data exploration
- Education data mining
- Complex Tableau dashboard building

Team

- Yabebal Fantaye
- Jean henock viayinon
- Usman Abdul Ganiy Babatunde

Key Dates

- Discussion on the case - 1130 Rwanda time on Monday 21 September 2020.
Use #all-week10 to pre-ask questions.
- Interim Solution - 2000 Rwanda time on Tuesday 22 September 2020.
- Final Submission - 2000 Rwanda time on Saturday 26 September 2020

Grading for the week

There are 100 points available for the week.

20 points - community growth and peer support. This includes supporting other learners by answering questions (Slack), asking good questions (Slack), participating (not only attending) daily standups (GMeet), and sharing links and other learning resources with other learners.

25 points - presentation and reporting.

5 points - interim submission

5 - Requirements met, clear presentation

3 - Most requirements met, presentation acceptable

1 - Some effort made

20 points for the final submission. This is measured through:

- Clarity of graphs (5 points)
- Clarity of message (5 points)

- Professionalism/production value (free of spelling errors, use of same font, well-produced) (5 points)
- Balance between being 'full of information' and 'easy to understand' (5 points)

55 points - data analysis and coding

10 points - interim submission

The validity of recommendations made (5 points)

Quality of code (including readability) (5 points)

45 points - final submission

The validity of recommendations made (25 points)

Quality of code (20 points)

Badges

Each week, one user will be awarded one of the badges below for the best performance in the category below.

In addition to being the badge holder for that badge, each badge winner will get +20 points to the overall score.

Visualization - the quality of visualizations, understandability, skimmability, choice of visualization

Quality of code - reliability, maintainability, efficiency, commenting - in future this will be [CICD](#)

An innovative approach to analysis -using the latest algorithms, adding in research paper content and other innovative approaches

Writing and presentation - clarity of written outputs, clarity of slides, overall production value

Most supportive in the community - helping others, adding links, tutoring those struggling

The goal of this approach is to support and reward expertise in different parts of the Data Scientist toolbox.

Late Submission Policy

Our goal is to prepare successful learners for the work and submitting late when given enough notice, shouldn't be necessary.

For interim submissions, those submitted 1-6 hours late will receive a maximum of 50% of the total possible grade. Those submitted >6 hours late may receive feedback, but will not receive a grade.

For final submissions, those submitted 1-24 hours late, will receive a maximum of 50% of the total possible grade. Those submitted >24 hours late may receive feedback, but will not receive a grade.

When calculating the leaderboard score:

- Your two lowest weeks' scores will not be considered.

Instructions

The tasks are divided into four sub-tasks

- Moodle database schema understanding
- Data Extraction Transformation and Loading (ETL)
- Dashboard making with Tableau
- Automation and Deployment

Task 1 - Moodle database schema understanding

Your first task is to have a clear understanding of the moodle database schema - start by studying [this page](#). In parallel do the following

- Download the database and write an SQL script to count
 - the number of tables
 - the number of records in each of the tables given in the MIT section
 - Number of quiz submissions by hour of day
 - Monthly usage time of learners who have confirmed and are not deleted
 - Count of log events per user for the following verbs: ['loggedin', 'viewed', 'started', 'submitted', 'uploaded', 'updated', 'searched', 'resumed', 'answered', 'attempted', 'abandoned']
- Write a python class to pull
 - Overall grade of learners for
 - Number of forum posts

Task 2 - Data Extraction Transformation and Loading (ETL)

Based on task 1, write a python class to perform the following

- Compute dedication time (see note below)
- Compute login and activity counts.
- Based on the following metrics, group students as top 1%, 5%, 10%, 25%
 - Login count
 - Activity count
 - Dedication time

Dedication time of a learner is the time between login and last activity for that login session, which is the last activity before the next login event for a given learner. For more information, check these²³ references.

Moodle estimates time based in the concepts of Session and Session duration applied to Moodle's log entries:

- **Click:** every time that a user accesses a page in Moodle a log entry is stored.
- **Session:** set of two or more consecutive clicks in which the elapsed time between every pair of consecutive clicks does not overcome an established maximum time.
- **Session duration:** elapsed time between the first and the last click of the session.

Task 3 - Dashboard making with Tableau

Make a Tableau dashboard to illustrate

- Learners by country of origin & gender ordered by their grade

Your Tableau dashboard should be something similar to the one shown in the Example dashboard section below. This example dashboard is just a suggestion you can make yours with a different look and feel. Your dashboard could give more information than what is shown in the example.

Task 4 - Automation and Deployment

- Connect the database with Tableau to continuously monitor learners performance.
- Prepare a 7-10 slide that illustrates your dashboard functionality, and provides your deployment strategy.

² <https://moodle.org/mod/forum/discuss.php?d=316649>

³

https://docs.moodle.org/36/en/ad-hoc_contributed_reports#User.27s_accumulative_time_spent_in_course

Interim Submission (Due Tuesday 22.09 20hr Rwanda time)

Submit a github link to your jupyter notebook you used to address the tasks defined in Task 1. Make sure to place at least one submission by **Tuesday 22.09 20hr Rwanda time**.

Feedback

You may not receive detailed comments on your interim submission, but will receive a grade.

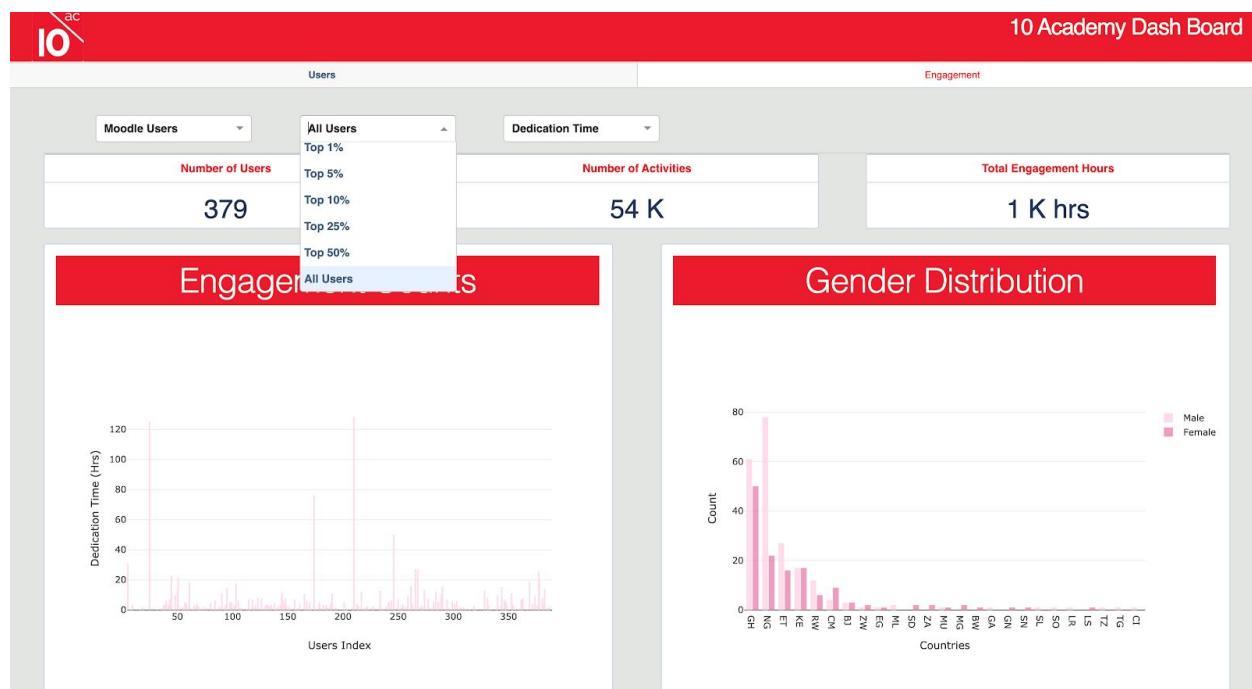
Final Submission (Due Sat 26.09 20hr Rwanda time)

1. Github link to your code
2. Link to your public Tableau dashboard
Due **Sat 26.09 20hr Rwanda time**.

Feedback

You will receive comments/feedback in addition to a grade.

Example dashboard



Users

Engagement

Moodle Users

All Users

Dedication Time

Activity Count

Login Count

Dedication Time

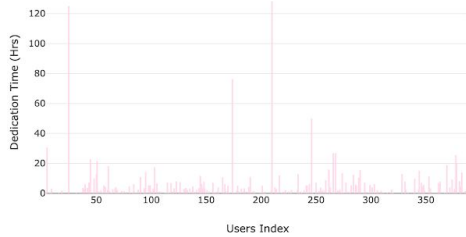
Number of Users

379

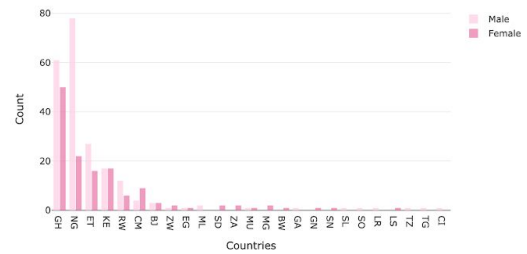
Total Engagement Hours

1 K hrs

Engagement Counts



Gender Distribution



References

- https://docs.moodle.org/39/en/Custom_SQL_queries_report
- https://docs.moodle.org/39/en/ad-hoc_contributed_reports
- <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.331.667&rep=rep1&type=pdf>
- http://informatics.ue-varna.bg/conference19/Conf.proceedings_Informatics-50.years%20177-187.pdf