CSC8631 Project: Learning Analytics

Semester 1: 2023

This coursework is collectively worth 100% of the mark for the module. The ProjectTemplate directory is worth 80% of the module, and the oral presentation worth 20%. The deadline for submission is `16:30 Friday November 17th`.

If you have any questions regarding the coursework, please ask either during a practical session, or by emailing `joe.matthews@ncl.ac.uk`

Context

Learning Analytics, a rapidly-growing application area for Data Science, is defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environment in which it occurs".

Existing mechanisms to record student engagement (e.g. atten- dance monitoring) fail to capture the extent and quality of engage- ment outside of the classroom environment. Further complementary sources of data are routinely collected about our learners (e.g. use of on-campus facilities, Virtual Learning Environment (VLE) and Re- Cap access, and student wellbeing referrals); however, these currently reside in a number of silos.

Learning Analytics seeks to aggregate these sources of data to de- rive shared insights, and provide effective measures of engagement. Insights may inform learning design, inform intervention processes for at-risk students, and improve student attainment.

The Task

We have data from 7 runs of a massive open online course (MOOC) entitled "Cyber Security: Safety At Home, Online, and in Life" made by Newcastle University and made available to the public by the online skills provider FutureLearn

We have raw data collected by FutureLearn on leaners as they progressed through the course, along with some characteristic information collected from their profiles. Learner IDs allow for combining of information between different data sheets.

Project Deliverables

The task for this project is to investigate aspects of the FutureLearn data which you feel would be of interest to a provider of a course such at this one. There are no restrictions on what you can investigate, and you will not be judged by the "success'' of your analysis (i.e. it does not matter whether your analysis actually produces the interesting thing you were looking for). Assessment here is based around the approach to the analysis you take, and how well you observed best practice in your approach.

For the **ProjectTemplate Directory** submission you should submit a single zip file containing:

- **Your complete ProjectTemplate directory**. This is the directory that is created when you run `create.project` at the start of your work, and will contain your code and output from your research into the FutureLearn data. Typically the main subdirectories used in this project will be:
  - **Data** – Containing the FutureLearn data files

  - **Munge** – Containing the R scripts which handle data preprocessing
  - **Cache** – Containing the cached preprocessed data
  - **Reports** – Containing the analysis report and reflective log described below.
  - **(Optional) src** – Containing scripts for producing outputs for the report
- **Analysis report** – A report describing the 2 cycles of CRISP-DM which make up your investigation. This report should be written in R Markdown, and has a maximum page limit of 20 pages. (Note this is an upper limit and not an expectation, it is perfectly possible to write a very high scoring report in under 10 pages.)
- **A reflective log** – A maximum 2 page reflective summary from working on the project. Within this you should discuss your experience of working with the different tools and techniques from the module and completing your project. You should make sure to consider your approach to tackling the module, and any thoughts on how you will approach similar tasks in the future.
- **README file** – A plaintext file (called README) which outlines steps needed to run your analysis (any extra packages required etc), along with the location of the deliverables in your submitted directory (i.e. where is your git log, where is your reflective log etc)
- **Git log** – A plaintext file containing the contents of your local git repository for the project.

To produce your Git log, you can run

```
git log > GitLog.txt
```

in the terminal. The `GitLog.txt` file should be included within your ProjectTemplate directory.

For the **Oral Presentation** submission, you should submit a zip file containing:

- A 5 minute recorded presentation where you highlight the analysis you carried out, and your main findings. This should focus on the aspects of the data you chose to investigate and the results you found, you do not need to discuss use of the tools from the module here (ProjectTemplate, RMarkdown etc)
- The slides and/or other materials used in your presentation (these can be created using whichever software you like)

Project Guidance

- You will not be judged on how sophisticated your investigation is – you should choose an area to investigate which you think would be interesting to the relevant business, but also that is achievable within the timeframe of this project
- We are primarily assessing how well you follow the best practice principles outlined in the module, i.e. using CRISP-DM to structure the investigation, using ProjectTemplate to structure the analysis and ensure it is reproducible, using dplyr/ggplot to carry out the analysis, and R Markdown to produce the analysis report
- The data files you are given are the raw data files from FutureLearn, it is entirely possible you will encounter issues around data quality during your analysis. How you choose to account for this is entirely up to you, we're just looking to see that you've checked for data quality issues and acknowledged any that are there.
- In the presentation you do not need to mention any of the tools/techniques (CRISP-DM, ProjectTemplate etc), instead you should focus on what you chose to investigate, the data you used, what you found, what that would mean for a business stakeholder etc