CSC8631 Report

Morgan Frodsham

08/11/2021

Learning Analytics for Cyber Security: Safety at Home, Online, in Life

Introduction

This report details the process and findings of exploratory data analysis for Newcastle University's FutureLearn course called "Cyber Security: Safety at Home, Online, in Life". Data analysis of educational platforms, such as FutureLearn, is a key part of learning analytics, which 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 environments in which it occurs."* (Add reference - 1st International Conference on Learning Analytics and Knowledge, Banff, Alberta, February 27–March 1, 2011, https://tekri.athabascau.ca/analytics/.) Learning analytics has the potential to help educational institutions undertake data-driven decision-making (reference - https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education) that better support students through their educational journey (reference - from bricks to clicks)

FutureLearn is a massive open online course (MOOC) provider (reference - https://www.futurelearn.com/info/blog/what-is-a-mooc-futurelearn). It promotes itself as *"a powerful new way to learn online"* (Reference - https://www.futurelearn.com/using-futurelearn/why-it-works) that is designed with the principles of effective learning. Learning effectiveness can be indicated by student engagement, their success rate, and completion time (reference - https://reader.elsevier.com/reader/sd/pii/S0747563218302590?token=A32E75AFAA1095275B9E0B7D3FA82CF29E5B168C4C92AE4B3D48A26C6456C9A142B81B117C93C71602CF948FB87D8F originRegion=eu-west-1&originCreation=20211120115140). The objective of this data analysis is to better understand the learning effectiveness of Newcastle University's Future Learn Cyber Security course.

(The success criteria of this project will include course completion as success rate, provision of personal identifiers such as demographic information as student engagement, and completion time?. Data analysis will be focused on the enrollments data, and use CRISP-DM as the methodology.) To become familiar with the data, the enrollments data sets for all 7 runs were combined into one full dataset.

Course completion

To understand the FutureLearn students' success rate (course completion) for Newcastle University's Cyber Security course, a variable was created from filtering the students by those who had data under fully_participated_at.

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
```

```
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

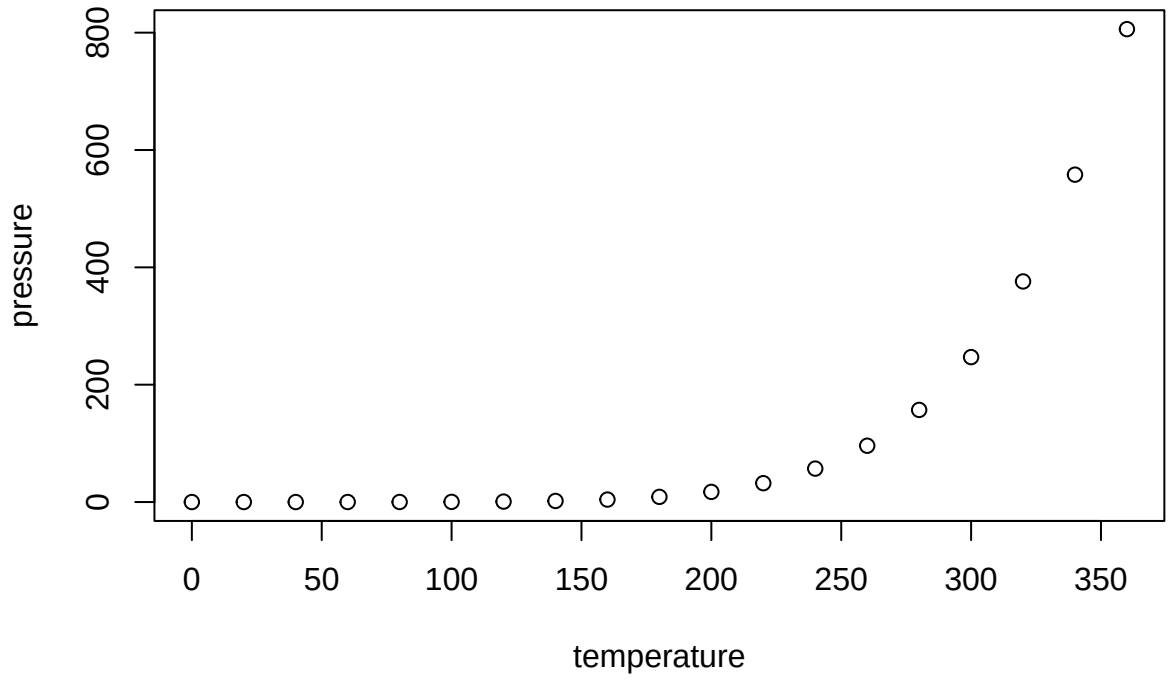
This showed that 2154 of 37296 students completed the course.

$How can I tell if all the student ID sare unique? Should I do that? How do I check that all the ID sare of the role "learner"?$

Gender

Of those who had completed the course, gender identty.

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Country

Age Range

Highest Education Leave

Employment Status

Employmenmt Area

Course Completed and Amount of Information Provided

Correlation between different types of information and course completion