

# CSC8631- Data Management and Explanatory Data Analysis Project

SATISH PILLA

5/20/2022

## **BUSINESS UNDERSTANDING:**

Future Learn, an online learning platform, offers a course called Cyber Security: Safety at Home, Online. This three-week self-paced course focuses on critical cyber security issues. The courses are divided into three parts, each of which is studied over three weeks: internet privacy, payment security, and home security. From 2016 to 2018, the course was offered seven times. Throughout the course, a variety of data was collected. This project aims to evaluate the data and get some useful insights that can be utilized to improve the course and establish a viable business model for future courses.

## **DATA UNDERSTANDING:**

Each time the course is run, around 8 data files are generated.

Datasets:

1. cyber-security-enrolments
2. cyber-security-leaving-survey-response
3. cyber-security-step-activity
4. cyber-security-video-stats
5. cyber-security-step-activity
6. cyber-security-archtype-survey-responses
7. cyber-security-question-response
8. cyber-security-weekly sentiment-survey

The enrollments data set contains the learners' information, including their unique IDs, the course's enrolled and unenrolled dates and times, and additional fields such as gender, nation, age, highest education level, employment location, and current employment status. The vast majority of the students' data was not captured. A figure or action that symbolizes universal human nature patterns is known as an archetype. The information that categorizes the students is contained in the archetype data collection.

The details of the learners who left the course at what stage and for what reason can be found in the leaving survey response dataset. Each learner's responses to quiz questions conducted at a certain moment in each week are collected in the question response data collection. The quiz's results are also included. Learners who started and left the step at particular times are recorded in the step activity data collection. The video statistics data set includes videos of certain steps with titles and information such as time, views, downloads, viewed percentage, and learners' viewed continents. The weekly survey replies data collection contains the learner's input on the course.