

# DATA1001 Semester 1, 2019

## Project 3 data dictionary

This data set contains 10845 anonymised student grades from junior Mathematics units of study offered in 2018. The data provided is real data from the University's student system and has been provided by Institutional Analytics and Planning, the department of the University responsible for student data reporting and analysis.

Below is a list of definitions for each of the fields:

- **Fake Student Identifier:** A unique identifier for each student in the dataset. These are analogous to a student ID but have been anonymised. Each unique fake identifier can be considered to relate to exactly one unique student.
- **Fake UoS Identifier:** The “name” of the unit of study. To preserve anonymity, this is not the actual unit of study code for the unit but rather a made-up identifier. You can assume that each of these identifiers relate to a junior unit of study offered by the School of Mathematics and Statistics in 2018, though due to the low sample sizes of some units of study, only the top fifteen units of study by total 2018 enrolments have been included in this dataset.
- **Semester:** The semester that the unit of study was run in. Unit of study occurrences outside of the standard Semester 1 and 2 (for example, intensive or Summer/Winter school units) have been excluded from this dataset.
- **Domestic / international status:** Whether the student is a domestic or an international student.
- **Gender:** The gender of the student. “M” denotes male students and “F” denotes female students. To preserve anonymity, students who identify as neither male nor female have been coded as female. This is the same approach that is used by the Department of Education when reporting aggregated student statistics: see the notes on ‘Gender’ at <http://highereducationstatistics.education.gov.au/DataNotes.aspx>
- **Mode of Study:** Whether the student is full time or part time. Full time is defined as taking 18 or more credit points in the semester in which the student took the unit of study.
- **Age:** The age of the student as at the first day of semester for the unit of study. This is reported as one of four bands:
  1. 18 and under
  2. 19-21
  3. 22-25
  4. Over 25
- **Unit of Study Mark:** The final mark achieved by the student. Student enrolments with non-standard grades (for example, discontinuations, withdrawals, absent fails and the like) have been removed from the dataset to preserve anonymity.

- **UoS Changes:** The cumulative number of times a student changed units of study, recorded at the start of week 1, the start of week 2 and at census date (either March or August 31<sup>st</sup> in the relevant semester). This is intended to provide a quantitative variable to measure a student's indecision (or otherwise) with respect to their units of study.

A unit of study change is defined as an instance when a student either adds or withdraws from a unit of study within the semester in which they were enrolled in the relevant unit. This is calculated by:

1. adding the number of times a student added a unit in Sydney Student to the number of times a student withdrew from a unit, and then:
2. subtracting the total number of units of study which the student was enrolled in.

If a student adds and drops the same unit multiple times these will each be counted separately towards the total. Replacing one unit with another counts as two changes; a drop followed by an add.

- **Canvas access:** Fifteen columns are included, one for each week of semester plus the mid-semester break and STUVAC, which denote whether or not the student accessed the Canvas site for that unit of study at least once during the given week. A 1 denotes that the student did access the site, a 0 denotes that they did not. Students that did not record *any* access to Canvas over the whole of semester have been excluded from this dataset. This is intended to provide a time-series variable to measure a student's diligence (or otherwise) with regards to accessing course material.