

Learning Analytics

Courses

Each course has its code_module code, which is a unique identifier and a code_presentation presentation. The combination of the two makes a single entry. This table also contains information on the duration of each of the presentations.

	code_module	code_presentation	module_presentation_length
0	AAA	2013J	268
1	AAA	2014J	269
2	BBB	2013J	268
3	BBB	2014J	262
4	BBB	2013B	240
5	BBB	2014B	234
6	CCC	2014J	269
7	CCC	2014B	241
8	DDD	2013J	261
9	DDD	2014J	262
10	DDD	2013B	240

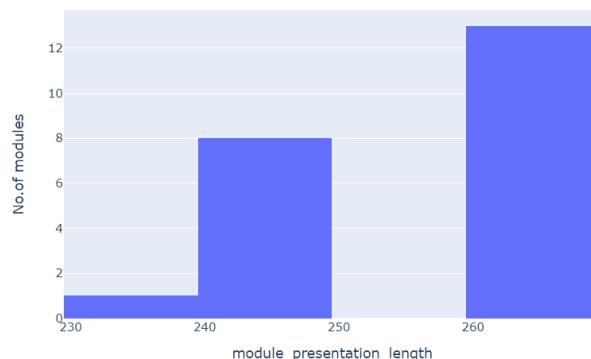


Each course may have 2 presentations in a year i.e. B and J. B presentations are usually different from J presentations. It is good to analyze the types separately.

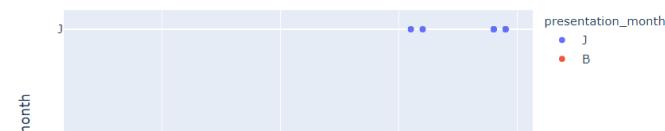
	code_module	code_presentation	module_presentation_le...	presentation_year
0	AAA	2013J	268	2013
1	AAA	2014J	269	2014
2	BBB	2013J	268	2013
3	BBB	2014J	262	2014
4	BBB	2013B	240	2013
5	BBB	2014B	234	2014
6	CCC	2014J	269	2014
7	CCC	2014B	241	2014
8	DDD	2013J	261	2013
9	DDD	2014J	262	2014
10	DDD	2013B	240	2013

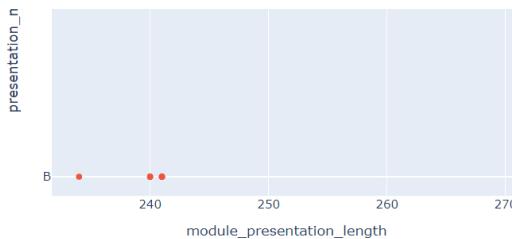


We have two groupings of days. This can be related to the year or the period in which each presentation takes place.



As we can see, the start month shows a clear separation between presentations that last up to ~ 245 days and those that last from ~ 260 days.





Assessment

The combination of code_module and code_presentation determines a single course and must also exist in the courses file. There should not be two or more unique id_assessments.

Only 3 types of assessment: TMA (assessed by monitor), CMA (assessed by computer) and Exam (final exam).

The evaluation date is counted in days from the beginning of the course.

	code_module	code_presentation	id_assessment	assessment_type	date	weight
0	AAA	2013J	1752	TMA	19	10
1	AAA	2013J	1753	TMA	54	20
2	AAA	2013J	1754	TMA	117	20
3	AAA	2013J	1755	TMA	166	20
4	AAA	2013J	1756	TMA	215	30



Sum of weights is 100% for all types of assessment except tests.

All courses are 100% by weight, except GGG courses, which may be a policy of the course itself.

		weight
AAA	2013J	100
AAA	2014J	100
BBB	2013B	100
BBB	2013J	100
BBB	2014B	100
BBB	2014J	100
CCC	2014B	100
CCC	2014J	100
DDD	2013B	100
DDD	2013J	100
DDD	2014B	100



How many final exams in each course ?

CCC-type courses have two final exams, which does not seem to be usual.

		assessment_type
AAA	2013J	1
AAA	2014J	1
BBB	2013B	1
BBB	2013J	1
BBB	2014B	1
BBB	2014J	1
CCC	2014B	2
CCC	2014J	2
DDD	2013B	1
DDD	2013J	1
DDD	2014B	1



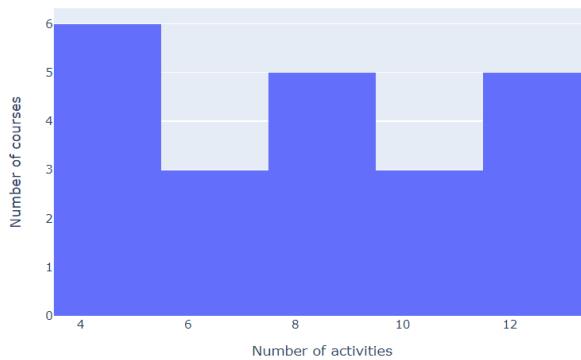
How is the distribution of the number of extra activities (CMA and TMA) per offer?

		assessment_type
AAA	2013J	Exam
AAA	2013J	TMA
AAA	2014J	Exam
AAA	2014J	TMA
BBB	2013B	CMA
BBB	2013B	Exam
BBB	2013B	TMA

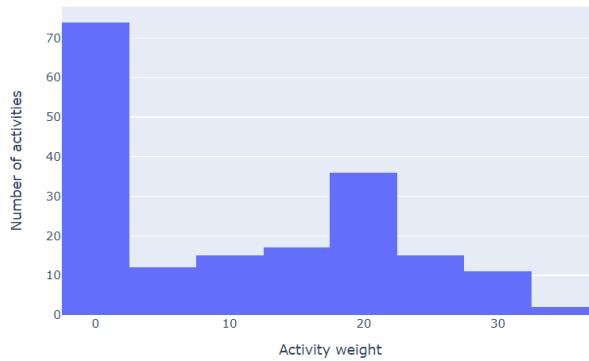
BBB	2013J	CMA	5
BBB	2013J	Exam	1
BBB	2013J	TMA	6
BBB	2014B	CMA	5

We can see that in the BBB and DDD courses there were changes in the number of activities applied in different course presentations.

Distribution of the number of extra activities (CMA and TMA) per offer

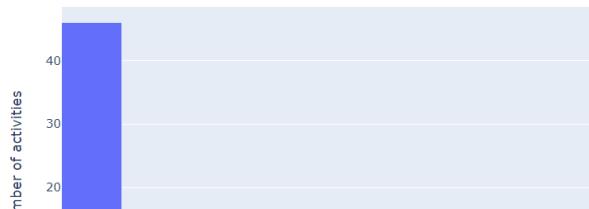


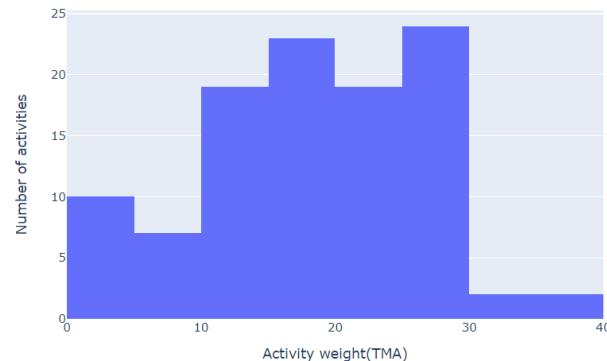
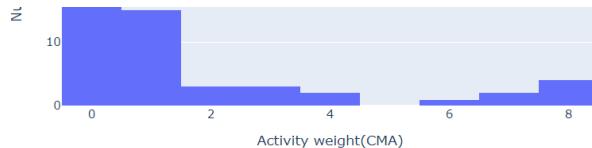
We see that few courses have more than 10 extra activities, and on average courses have 8.27 activities.



We see that there are a large number of activities with a weight of 0, corresponding to the activities of the GGG course, which gives weight 0 to everything except the final test. In addition, we have a large number of activities with a weight of 20% to 35%, and no activity with a weight between 36 percent and 100%. Which indicates that many courses have a large number of extra activities, agreeing with the average number of activities that we saw earlier.

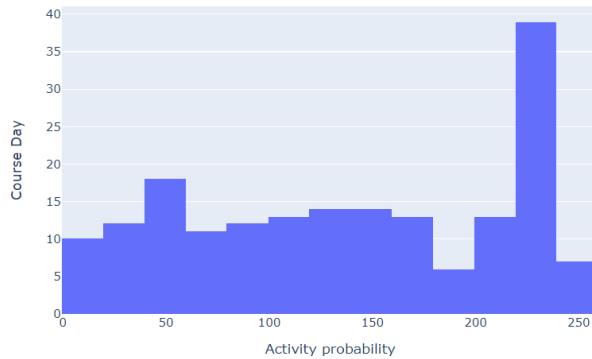
Distribution of weights when we group by type of activity





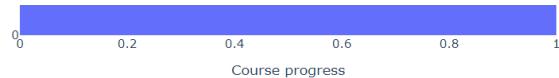
We see that activities corrected by monitors (TMA) carry more weight. Thus, it is possible to interpret CMAs as minor activities, as practical work, and TMAs as course projects, for example

Chances are of having an activity on each day of any course



We see that there are peaks along the course in which activities are most likely. These peaks can coincide with the end of modules within a course, for example. The fastest activity of the courses expires on the 12th, which may correspond to an activity of introduction to the content. In addition, a hiatus draws attention around the day 160 ~ 170 when there are no activities. This can be explained as a period that students have to prepare for the final activities of the course. This graph could be improved by normalizing the day of each activity using the duration of each offer. Thus, we would no longer have absolute days and would have a percentage of course completion. This could even show characteristics that are independent of the month of offer





In this new version, which takes into account the relative progress of the course, we see that less activities are due by 100% of the course, leaving this date exclusively for tests. In addition, we see more pronounced peaks.

VLE

The csv file contains information about the available materials in the VLE. Typically these are html pages, pdf files, etc. Students have access to these materials online and their interactions with the materials are recorded. The vle.

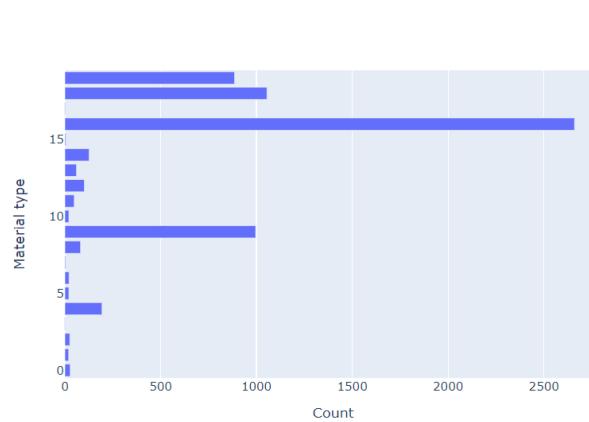
	id_site	code_module	code_presentation	activity_type	week_from	week_to
0	546943	AAA	2013J	resource	NaN	NaN
1	546712	AAA	2013J	oucontent	NaN	NaN
2	546998	AAA	2013J	resource	NaN	NaN
3	546888	AAA	2013J	url	NaN	NaN
4	547035	AAA	2013J	resource	NaN	NaN

Can a material be associated with more than one course

	code_module
1	6364
0	1

We see that there is only a value of 1 in the count of distinct modules associated with material ids. Thus, a material is unique.

The most common types of materials



6364

We see that two of the most used resources seem to be generic names: oucontent and resource. We interpret oucontent as external content and resource as an attached file. In addition to these two, we have subpages as one of the most referenced items, which We believe are pages for other content within the same course, and URLs as links to external sites.

StudentInfo

This file contains demographic information about the students together with their results.

	code_module	code_presentation	id_student	gender	region
0	AAA	2013J	11391	M	East Anglian Region
1	AAA	2013J	28400	F	Scotland
2	AAA	2013J	30268	F	North Western Region
3	AAA	2013J	31604	F	South East Region
4	AAA	2013J	32885	F	West Midlands Region

This table contains student data for the various course offerings.

A single entry is defined by a triple
` (id_student,code_module,code_presentation)`

The students appear to be all from England and neighboring countries, the regions all appear to be in the UK. That is why we have the IMD measure, which is a "poverty index" used in these countries. The graduation levels also seem to be the ones used there. We interpret it as follows:
Post Graduate Qualification: Post Graduate HE Qualification: Graduation / Bachelor A Level or Equivalent: High School or Equivalent Lower Than A Level: Incomplete High School No Formal Quals: No training. We have information about a student's previous attempts in the same module and the number of credits completed. We don't know how many credits each module is worth, but that same table can show that information. There are students with some type of disability



Number of students in the data

28785



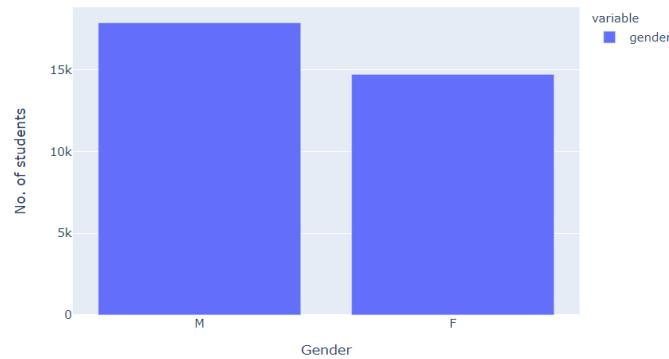
Average number of modules taken by students

	code_module
count	28785
mean	1.0868
std	0.2840
min	1
25%	1
50%	1
75%	1
max	3



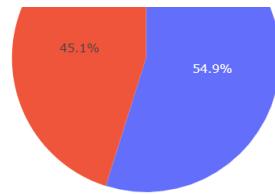
The average number of different modules is 1,086, which indicates that the vast majority of students take only one module. However, the maximum number of modules is 3, but few students reach that number. When we look at the number of entries in the table and the number of students, we have about 10,000 more entries than the number of students. Since few students take more than one module, We attribute that number of entries to more than multiple attempts. We will investigate this later.

Gender Distribution



Gender Distribution



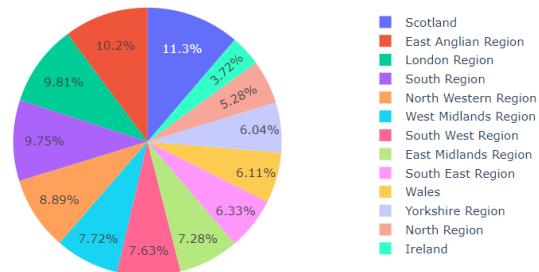


We have more men than women among students, but the proportion is not so distant: 55% men and 45% women.



Where do the students from the courses come from?

Regionwise Distribution



Mobility of students throughout their training

	region
1	0.8771
2	0.1144
3	0.0077
4	0.0008
5	0.0000

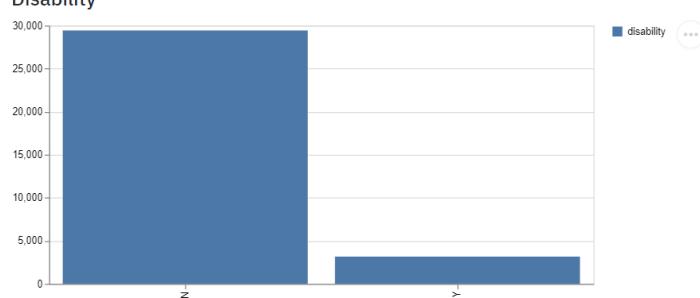
Most of the students remained in the same region throughout all modules. 12 percent of students lived in more than 1 region.

Proportion of students with some type of disability

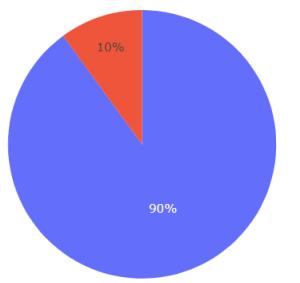
	disability
N	0.9029
Y	0.0971

9.7 percent of students have some type of disability.

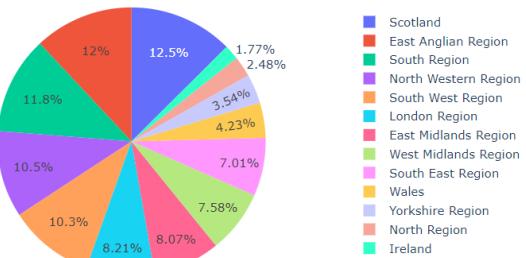
Disability



Disability Distribution



Regionwise Disability Distribution

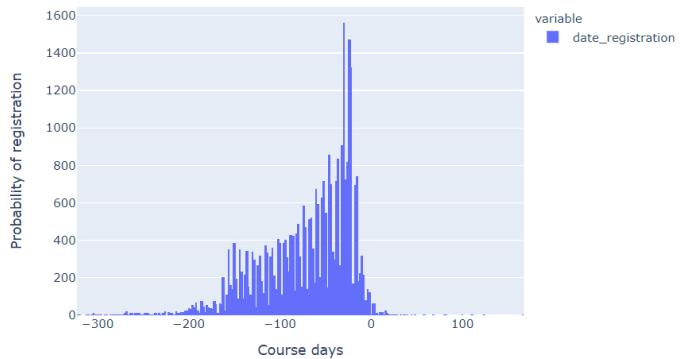


Student Registration

This file contains information about the time when the student registered for the module presentation. For students who unregistered the date of unregistration is also recorded.

	code_module	code_presentation	id_student	date_registration	date_unregistration
0	AAA	2013J	11391	-159	
1	AAA	2013J	28400	-53	
2	AAA	2013J	30268	-92	
3	AAA	2013J	31604	-52	
4	AAA	2013J	32885	-176	

When do students register the most?



date_registration
count
mean
std
min

Count: 32548
Mean: -69.4113
Std: 49.2665
Min: -322

25%	-100
50%	-57
75%	-29
max	167

Usually, students register before the course starts. The registration peak starts 50 days before the start, but some students even register later. The last time someone registered was on 167.



When do more students drop out of courses?

date_unregistration	
count	10072
mean	49.7576
std	82.4669
min	-365
25%	-2
50%	27
75%	109
max	444

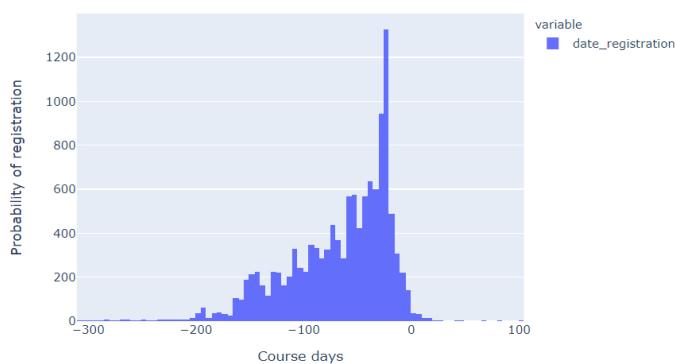
About 1/4 of the students drop out of the course before it even starts, which is interesting, because then 1/4 of the student records in courses are cases in which one person did not even attend a class. Peak withdrawals occur in the first 50 days.

Do students who pass the course register earlier?

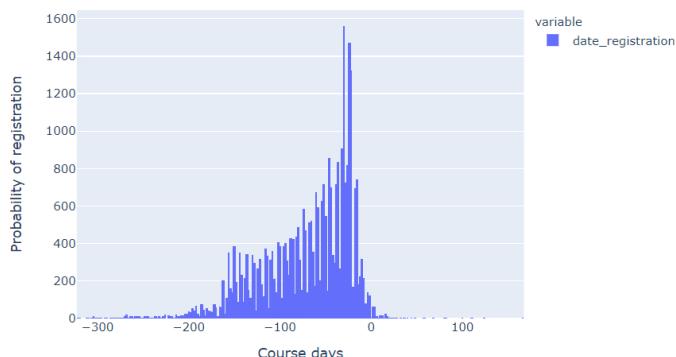
Registration date of approved students vs. all students



Approved



All



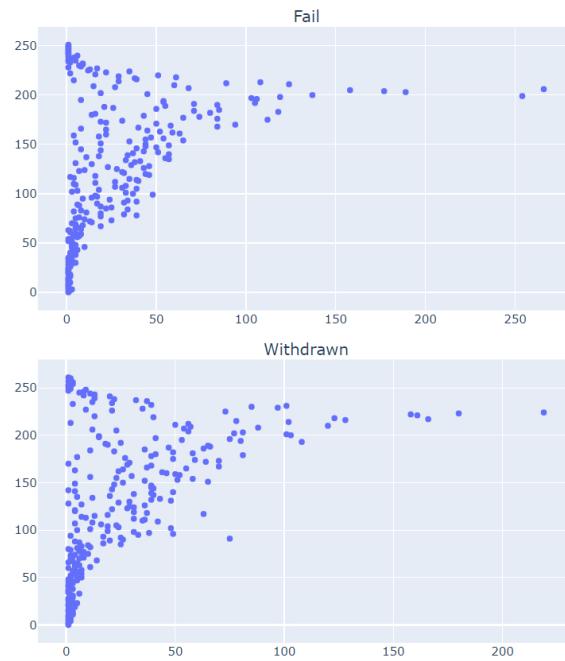
date_registration	
count	12360
mean	-66.7878

std	47.3743
min	-310
25%	-95
50%	-53
75%	-29
max	101

	code_module	code_presentation	id_student	gender	region
0	AAA	2013J	11391	M	East Anglian Region
1	AAA	2013J	28400	F	Scotland
2	AAA	2013J	30268	F	North Western Region
3	AAA	2013J	31604	F	South East Region
4	AAA	2013J	32885	F	West Midlands Region
5	AAA	2013J	38053	M	Wales
6	AAA	2013J	45462	M	Scotland
7	AAA	2013J	45642	F	North Western Region
8	AAA	2013J	52130	F	East Anglian Region
9	AAA	2013J	53025	M	North Region
10	AAA	2013J	57506	M	South Region

Although different, the distribution of registration dates for approved students does not differ much from the distribution of all students.

Do students who pass the course register earlier?



We chose to study the number of students who dropped out or failed modules vs. the day they registered. We expected to see a larger number of students with positive days, indicating that they registered after classes started. However, the abnormality we noticed was the peak of students who registered around 150 days earlier and dropped out of the course.

Student Assessment

This file contains the results of students' assessments. If the student does not submit the assessment, no result is recorded. The final exam submissions is missing, if the result of the assessments is not stored in the system.

This is the table that relates students to assessments.

	id_assessment	id_student	date_submitted	is_banked	score
0	1752	11391	18	0	78

1	1752	28400	22	0	70
2	1752	31604	17	0	72
3	1752	32885	26	0	69
4	1752	38053	19	0	79

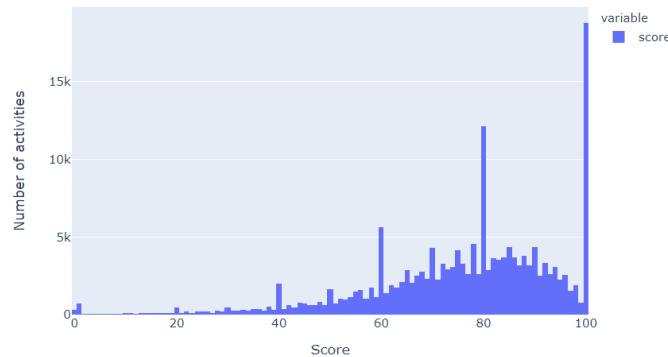


We have 173 lines with invalid values. We will investigate what those values are.

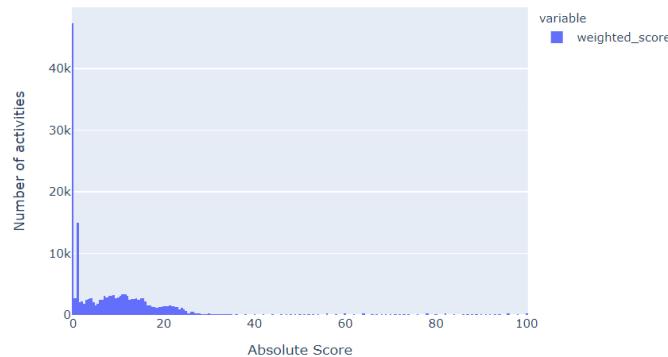
	id_assessment	id_student	date_submitted	is_banked	score
104964	34860	544930	51	0	NaN
122892	34877	581307	175	0	NaN
20497	15000	476278	173	0	NaN
104977	34860	545791	80	0	NaN
159251	37415	610738	87	0	NaN
33812	15013	487543	200	0	NaN
104980	34860	545944	49	0	NaN
104983	34860	546164	22	0	NaN
108572	34864	185496	230	0	NaN
98336	30712	578725	235	0	NaN
69153	25339	560260	201	0	NaN

Invalid values are in the score column. Therefore, the note of these activities can possibly be ignored.

Average grade of activities

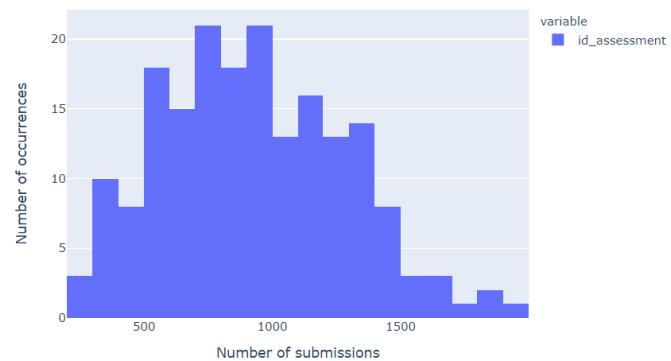


We have a lot of 100%, but this is not very informative of the weight that these 100 percent carry. We will incorporate weight data from the activities to get a better idea of what the grades are like.



We see that when taking weight into account, the distribution changes a lot. This is because of the many activities that have a weight of 0. In addition, we see that activities that are not tests count towards scores below 40.

Number of submissions per activity

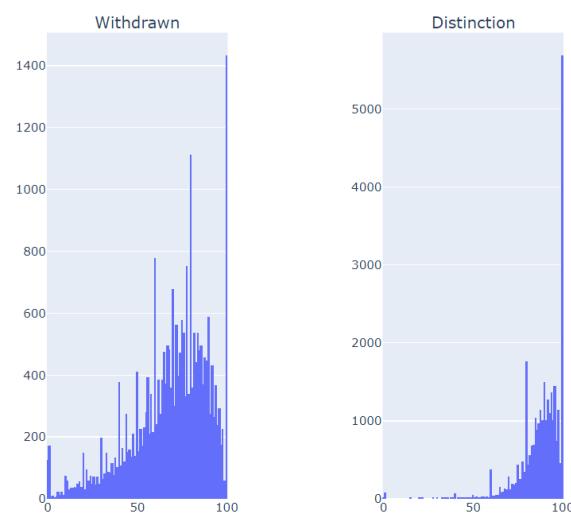
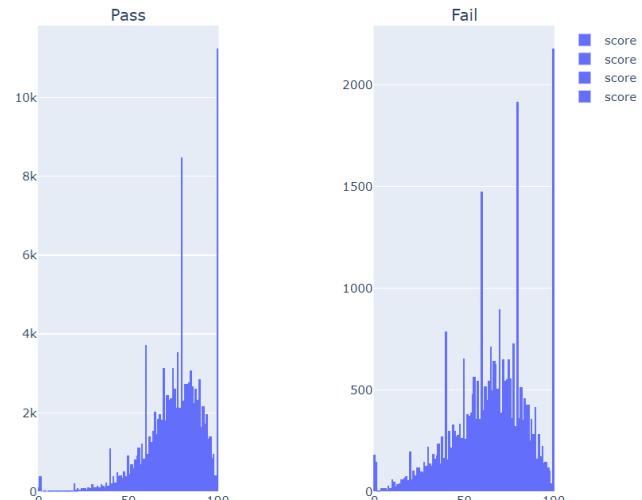


id_assessment	
count	188
mean	925.0638
std	360.0015
min	278
25%	652.7500
50%	909
75%	1,192.2500
max	1917



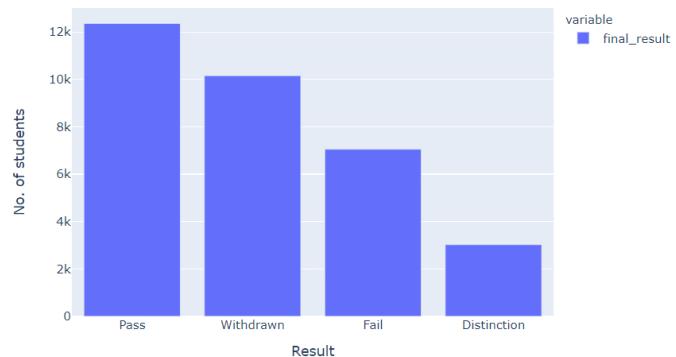
On average, activities receive 925 submissions.

Result Analysis

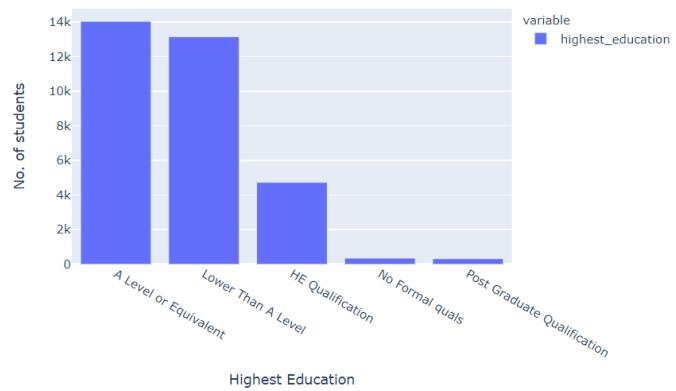


We see some interesting behaviors in the distribution of students' grades. Firstly, students who have passed the distinction have better grades than the rowers. Then, the passers have a higher proportion of grades greater than 70 in relation to those who fail, and less grades less than 60. When analyzing failing and dropping students, we noticed that dropouts have a higher proportion of better grades than failing ones. This may indicate one of the reasons why a student gives up on a subject, which is poor performance in some key activity.

Final Result in given Data



Highest Education



Studied Credits

