

Analyzing student grades in the Open University

Introduction

Business context: The Open University (<https://www.openuniversity.edu/>) is one of the largest universities in the world by number of enrolled students. In recent years, it has been working on developing an analytics platform called OU Analyse (<https://analyse.kmi.open.ac.uk/>) to track student performance across its many programs, with the aim of reducing dropout and increasing student engagement and learning. Their analytics team is growing, and they have been actively looking to hire strong candidates. You have been shortlisted as one of them.

Business problem: As part of the interviewing process, you have been given an anonymized dataset and asked to **compute the GPA (Grade Point Average) for all students in the October 2014 cohort and create a Tableau dashboard that includes their GPAs and demographic data.**

Analytical context: You will be working with a subset of the OULA dataset (https://analyse.kmi.open.ac.uk/open_dataset), which contains demographic and assessment data for a number of UK-based students from the October 2014 cohort (coded as 2014J). The data is in the OULA.xlsm (data/OULA.xlsm) file (click on the file name to download the file). You will find four worksheets in this Excel file:

- `student_info` contains demographic information about students as well as which "module" (course) they were enrolled in and whether they received a pass or a fail, or dropped out.
- `modules` lists the different modules that the students could enroll in. The actual names have been replaced with three-letter codes for privacy reasons.
- `assessments` has a table with all the assessments for each module, indicating the weight that each assessment had in the final grade.
- `student_assessment` contains the actual grades that each student received in each assessment on a 0-100 scale. Most students in this sample only enrolled in one course, but there are some who enrolled in more.

The full documentation can be found here
(https://analyse.kmi.open.ac.uk/open_dataset#description).

Deliverable: In this extended case, you will be submitting 1 excel file that contains all of the exercise answers and your Tableau Dashboard link.

Preparing the dataset

Before you create the dashboard, you need to prepare the dataset and compute the GPAs.

Important Note: *You may notice missing values.* This is common in large datasets. The tools we're using will generally ignore missing values, but it is a good idea to start thinking about how to handle them. For instance, in this dataset, at one point Excel converted all 10-20 under imd_band to October 20 , which is a typical behavior for Excel. Those anomalous values were later removed, but this leaves us with missing values. You are not expected to do any data cleaning here.

Exercise 1

Instructions: Go to the `student_assessment` worksheet and complete Column F with the **code_module** to which each assessment corresponds. The modules associated to each assessment are in the `assessments` worksheet.

For example, the **Code_Module** for **Student #569505** with **Assessment ID 1758** is "AAA".

Hint: You will need to use INDEX & MATCH functions.

Answer.

Exercise 2

Instructions: Go to the `student_assessment` worksheet and complete Column G with the weight that is associated with each assessment. Name Column G `weight` .

Remember that the weights are in the `assessments` worksheet.

Answer.

After solving the previous exercises, your worksheet should look like this:

A1		fx id_assessment						
	A	B	C	D	E	F	G	H
1	id_assessment	id_student	date_submitted	is_banked	score	code_module	weight	
2	1758	569505	19	0	85	AAA	10	
3	1758	570317	19	0	79	AAA	10	
4	1758	574505	19	0	43	AAA	10	
5	1758	588477	19	0	88	AAA	10	
6	1758	593846	17	0	67	AAA	10	
7	1758	593987	17	0	74	AAA	10	
8	1758	596288	18	0	85	AAA	10	
9	1758	603834	19	0	64	AAA	10	
10	1758	603861	-1	1	61	AAA	10	
11	1758	623840	19	0	80	AAA	10	
12	1758	626240	26	0	63	AAA	10	
13	1758	628088	25	0	80	AAA	10	
14	1758	630142	19	0	88	AAA	10	
15	1758	637665	19	0	68	AAA	10	
16	1758	639053	16	0	60	AAA	10	
17	1758	642878	19	0	52	AAA	10	
18	1758	648130	17	0	53	AAA	10	
19	1758	676546	19	0	77	AAA	10	
20	1758	677186	17	0	43	AAA	10	

Calculating the GPAs

Now we are ready to compute the GPA for each student-module pair. Since every assessment has a different weight, this will need to be a *weighted* GPA; i.e. a weighted average. The formula for a weighted average is this:

$$\text{WeightedAverage} = \frac{\text{score}_1 \times \text{weight}_1 + \text{score}_2 \times \text{weight}_2 + \dots + \text{score}_n \times \text{weight}_n}{\text{sum of weight}}$$

You can see that the points earned on each assignment are multiplied by the total possible points for each assignment. Those values are all added together, and then divided by the sum of *weight*, which is calculated by adding together the total possible points for all assignments.

The calculations we are going to use for GPA will look different than those on a typical 0 - 4.0 scale. Be sure to follow the instructions for each exercise in calculating GPA as it is used in this dataset.

Exercise 3

Before we can compute the GPAs, we need to know the sums of the weights for each module (the sum of *weight* terms in the formula above). In theory, the assessments in each module should add up to a total of 200 points (100 points for ordinary assessments plus 100 points for the final exam). However, this isn't true for all modules in this sample.

Instructions: Using the `assessments` worksheet, calculate the sum of the weights for each `code_module` using a **pivot table**. Place it in the worksheet called `Weight_Sums` in Cell A1 .

After you have completed the Pivot Table in the `Weight_Sums` worksheet, go to the `ANSWER` worksheet and answer the following question: What is the `sum of weight` for **code_module** BBB ? Enter your response in `Cell X84` .

Answer.

Exercise 4

Instructions: Go to the `student_assessment` worksheet and complete `Column H` using the `VLOOKUP()` function with the sums of the weights you calculated in Exercise 3. Remember that the sum of weights is particular to each module.

Name the column `weight_sum` .

Answer.

Exercise 5

5.1

Instructions: Compute the product *score* × *weight* in `Column I` of the `student_assessment` worksheet for each student-assessment pair.

Name the column `score x weight` .

Answer.

5.2

Instructions: For each student and assessment pair, compute the quantity

$$\frac{\text{score} \times \text{weight}}{\text{sum of weight}}$$

Complete `Column J` in the `student_assessment` worksheet using the formula above.

Name the Column **(score x weight)/weight_sum**.

The `student_assessment` worksheet should look like this after you add this column:

<div> <div> <div>↶</div> <div>↷</div> <div>📄</div> <div>🔍</div> </div> <div> <div>Arial</div> <div>10</div> <div>B</div> <div>🔍</div> <div>🔍</div> <div>🔍</div> <div>...</div> </div> <div> <div>☰</div> <div>ab</div> <div>🔍</div> </div> <div> <div>General</div> <div>←0</div> <div>.00</div> <div>→0</div> <div>Σ</div> </div> </div>										
A1	fx id_assessment									
	A	B	C	D	E	F	G	H	I	J
1	id_assessment	id_student	date_submitted	is_banked	score	code_mod	weight	weight_su	score x weight	(score x weight) / weight_sum
2	1758	569505	19	0	85	AAA	10	200	850	4.25
3	1758	570317	19	0	79	AAA	10	200	790	3.95
4	1758	574505	19	0	43	AAA	10	200	430	2.15
5	1758	588477	19	0	88	AAA	10	200	880	4.4
6	1758	593846	17	0	67	AAA	10	200	670	3.35
7	1758	593987	17	0	74	AAA	10	200	740	3.7
8	1758	596288	18	0	85	AAA	10	200	850	4.25
9	1758	603834	19	0	64	AAA	10	200	640	3.2
10	1758	603861	-1	1	61	AAA	10	200	610	3.05
11	1758	623840	19	0	80	AAA	10	200	800	4
12	1758	626240	26	0	63	AAA	10	200	630	3.15
13	1758	628088	25	0	80	AAA	10	200	800	4
14	1758	630142	19	0	88	AAA	10	200	880	4.4
15	1758	637665	19	0	68	AAA	10	200	680	3.4
16	1758	639053	16	0	60	AAA	10	200	600	3
17	1758	642878	19	0	52	AAA	10	200	520	2.6
18	1758	648130	17	0	53	AAA	10	200	530	2.65
19	1758	676546	19	0	77	AAA	10	200	770	3.85
20	1758	677186	17	0	43	AAA	10	200	430	2.15

Answer.

Exercise 6

Calculate the GPAs for each student in each course, using a pivot table. Put it in a new worksheet called `gpas`. The rows should be the individual students' IDs, and the columns should be the modules. Each cell should correspond to the GPA that each particular student obtained in each module.

This is a sample of what your pivot table should look like:

Row Labels	AAA	BBB	CCC	DDD	EEE	FFF	GGG
6516	31.75						
8462				21.5			
23698			49.99				
24734	23.75						
25261			0.52				
25572				3.5			
26192	43.15						
26315		31.9					
26734		9.25					
27116			62.67333333				
27417				14.025			
28061	38.7						
28787			18.66				
28952			2.006666667				
29411			0.42				

Row Labels AAA BBB CCC DDD EEE FFF GGG

...

Instructions: Create a Pivot Table in the GPAS worksheet, position it in Cell A1 . Then, enter the value under column CCC for **student 29639** in Cell X235 in the ANSWER worksheet.

Hint: You can remove the "Grand Totals" row and column from your pivot table. To do so, after you have created it, right-click on it and select Show settings . Then uncheck the boxes below Show grand totals for (if you are using a Mac, after you right-click you will probably see the option to remove the grand total immediately).

Answer.

Exercise 7

Now reference the GPAs from the pivot table in the gpas worksheet to a new column in the student_info worksheet. Call the new column student_module_gpa . Your worksheet should end up looking like this:

Excel OULA - Saved															Search (Alt + Q)																							
File Home Insert Draw Page Layout Formulas Data Review View Automate Help															Editing															Share		Comments		More				
Undo Redo Font Color Background Color Text Color Paragraph Styles Styles															General															Sum		Average		Find		Replace		More
N2															fx																							
	A	B	C	D	E	F	G	H	I	J	K	L	M	N																								
1	code_module	code_presid	student	gender	region	highest education	imd_band	age_band	num_of_p	studied_or	disability	final_resul	student_module_gpa																									
2	AAA	2014J	6516	M	Scotland	HE Qualification	80-90%	55<=	0	60	N	Pass	31.75																									
3	AAA	2014J	24734	F	South Region	Lower Than A Level	60-70%	0-35	0	60	N	Pass	23.75																									
4	AAA	2014J	26192	F	East Anglian Region	A Level or Equivalent	60-70%	35-55	0	60	N	Distinction	43.15																									
5	AAA	2014J	28061	M	Yorkshire Region	Lower Than A Level	40-50%	35-55	0	60	N	Pass	38.7																									
6	AAA	2014J	31600	F	South West Region	A Level or Equivalent	40-50%	0-35	0	60	N	Pass	27.75																									
7	AAA	2014J	46844	M	North Region	HE Qualification	30-40%	35-55	0	60	N	Pass	33.25																									
8	AAA	2014J	52765	F	Yorkshire Region	Lower Than A Level	0-10%	35-55	0	120	N	Pass	30.75																									
9	AAA	2014J	55104	F	West Midlands Region	A Level or Equivalent	60-70%	0-35	0	60	N	Pass	35.65																									
10	AAA	2014J	58071	M	East Anglian Region	HE Qualification	90-100%	55<=	0	60	N	Pass	37.45																									
11	AAA	2014J	58316	F	Wales	A Level or Equivalent	70-80%	0-35	0	120	N	Pass	36.45																									
12	AAA	2014J	62487	F	East Midlands Region	A Level or Equivalent	20-Oct	0-35	0	60	N	Pass	35.3																									
13	AAA	2014J	63165	M	South East Region	A Level or Equivalent	80-90%	35-55	0	60	N	Pass	26.05																									
14	AAA	2014J	65002	F	East Anglian Region	A Level or Equivalent	70-80%	0-35	1	60	N	Fail	10.1																									
15	AAA	2014J	70011	F	Scotland	HE Qualification	40-50%	0-35	0	60	N	Pass	40.65																									
16	AAA	2014J	75255	M	North Western Region	A Level or Equivalent	40-50%	0-35	0	120	N	Pass	41.5																									
17	AAA	2014J	79403	M	South Region	Lower Than A Level	80-90%	35-55	0	60	N	Withdrawn	14.6																									
18	AAA	2014J	85092	F	South West Region	Lower Than A Level	30-40%	0-35	0	60	N	Pass	23.2																									
19	AAA	2014J	85144	M	Scotland	A Level or Equivalent	50-60%	35-55	0	60	N	Pass	39.4																									
20	AAA	2014J	85302	M	London Region	Lower Than A Level	80-90%	35-55	0	60	N	Pass	35.5																									

Instructions: Complete **Column M** in the student_info worksheet. Use the header **student_module_gpa** for Column M. Using the GPAS pivot table, fill in the GPAs of each student.

Hints:

1. Use INDEX() + MATCH() (get the positions of the assessment and the student ID with MATCH() - one MATCH() for the column and another one for the row). Remember that there is one GPA for each student-module pair.
2. Make sure the formula performs an exact match (<https://support.microsoft.com/en-us/office/match-function-e8dffd45-c762-47d6-bf89-533f4a37673a>).

Note: Some students have a "Fail" as their final result even though they obtained a relatively high grade. This is because those students did not present the final exam and therefore were disqualified from receiving a degree. Recall that we may have missing values, and they can be ignored.

Answer.

Creating the dashboard

Exercise 8

Now we have all the data that we need! Load the `student_info` worksheet into Tableau and make a workbook (<https://www.linkedin.com/pulse/explain-difference-between-tableau-workbook-story-upendar-sheethala/>) (with several dashboards if needed) that allows the user to visualize the relationships between the *following metrics*:

- Average GPA (*The aggregate statistic must be "Average"*)
- Student count (*The aggregate statistic must be "Count distinct"*)

and the *following dimensions*:

- Age
- Gender
- Module
- Region
- Previous education

They should also allow for filtering the data by multiple dimensions; e.g. You should be able to answer "What is the average GPA of students who took the AAA module *and* were in Scotland?" with the dashboard.

Design factors to keep in mind:

- The dashboard pages **shouldn't be cluttered**. Keep it as clean as you can, and if you need to add several charts, distribute them in pages (dashboards).
- The Tableau charts should **employ the "Use as filter" functionality**. This way an arbitrary number of filters can be applied to the charts using all the dimensions without requiring the addition of new charts or filter controls. For instance, you can have a vertical bar graph with age range, and when you select one of the bars, the rest of the charts adjust to show you the students by region, gender, etc. for that age range.
- All the **charts must be of the appropriate type according to the data they represent**. In this case, going with all bar charts (or column charts) is a safe option. You do not have to create a choropleth map. It may be especially difficult to do so because the region data is not formatted for Tableau.

- Every chart must have a **descriptive title and chart axes must be clearly labeled**.

You will need to both *design and implement* the dashboards yourself, using the principles and tutorials you saw in previous cases. Paste the link to the Tableau Public dashboard in a new cell below. Also submit your completed Excel sheet in order to receive credit for your pivot tables.

Submission

Congratulations! You have completed this Extended Case! The last step is to submit your excel file. Please make sure that the following is complete.

1. Your Excel file has been rename to include your full name.
2. Your Tableau Link is inserted into Cell X308 in the ANSWER worksheet.

Answer.

Attribution

"Open University Learning Analytics dataset", Kuzilek J., Hlosta M., Zdrahal Z., 2017, Creative Commons Attribution 4.0, <https://www.nature.com/articles/sdata2017171>
(<https://www.nature.com/articles/sdata2017171>), Sci. Data 4:170171 doi: 10.1038/sdata.2017.171