

# SIT384 Cyber security analytics

## Credit Task 3.2C: Process data using Pandas

### Task description:

You are given a student result data file (result\_withoutTotal.csv). It has columns:

ID: student id

Ass1 ~ Ass4: assignment scores (out of 100); weight of ass1, ass2, ass3 and ass4 is 5%, 15%, 5%, and 15%, respectively.

Exam: examination score (out of 120); weight is 60%.

ID	Ass1	Ass2	Ass3	Ass4	Exam
1	89.1	50	85	88.9	65
2	95.1	82.5	90.5	94.5	52
3	74.3	54.4	63	63.9	31
4	89.8	81.3	82	90.4	37
5	91.3	98.8	92.5	95.9	79
...	...	...	...	...	...

(The above data is for demonstration purposes only. Please download the full version of result\_withoutTotal.csv.)

Total score can be calculated using formula:

$$\text{Total} = 5\% * (\text{ass1} + \text{ass3}) + 15\% * (\text{ass2} + \text{ass4}) + 60\% * \text{exam}$$

Read students' result data from file result\_withoutTotal.csv,

add:

- Total column:  $\text{Total} = 5\% * (\text{ass1} + \text{ass3}) + 15\% * (\text{ass2} + \text{ass4}) + 60\% * \text{exam}$ .
  - The max Total score is 100. Reduce the Total to 100 if the calculated Total > 100.
- Final column: Final = Total score rounded to the nearest integer.
  - To pass the unit, a student must achieve at least 50 of the Total and 40% of Exam which is 48 out of 120 (or  $\text{Total} \geq 50$  and  $\text{Exam} \geq 48$ ).
  - If a student failed the hurdle ( $\text{Exam} \geq 48$ ), the max Final is 44. No change to Final score if Final < 44.
- Grade column: N ( $\text{Final} \leq 49.45$ ), P ( $49.45 < \text{Final} \leq 59.45$ ), C ( $59.45 < \text{Final} \leq 69.45$ ), D ( $69.45 < \text{Final} \leq 79.45$ ) and HD ( $79.45 < \text{Final}$ ). Border values are as follows:

HD	
D	79.45
C	69.45
P	59.45
N	49.45

save:

- the result data file with the 3 new columns to a file called result\_updated.csv.
- the students' records with exam score < 48 to a file called failedhurdle.csv.

display:

- the result data file with the 3 new columns
- the students with exam score < 48
- the students with exam score > 100

(Hints: import pandas, use DataFrame, DataFrame.loc and display)

(Sample output as shown in the following figure is for demonstration purposes only.)

result\_updated:

	Ass1	Ass2	Ass3	Ass4	Exam	Total	Final	Grade
ID								
1	89.1	50.0	85.0	88.9	65	68.540	69	C
2	95.1	82.5	90.5	94.5	52	67.030	67	C
3	74.3	54.4	63.0	63.9	31	43.210	43	N
4	89.8	81.3	82.0	90.4	37	56.545	44	N
5	91.3	98.8	92.5	95.9	79	85.795	86	HD
..	...	...	...	...	...	...	...	...
87	71.8	95.0	70.0	98.6	49	65.530	66	C
88	96.0	97.6	65.0	100.0	76	83.290	83	HD
89	91.8	66.3	79.0	95.8	62	70.055	70	D
90	78.7	50.0	75.0	95.8	56	63.155	63	C
91	55.7	82.5	66.0	81.9	49	60.145	60	C

[91 rows x 8 columns]

failedhurdle.csv:

ID	Ass1	Ass2	Ass3	Ass4	Exam	Total	Final	Grade
3	74.3	54.4	63	63.9	31	43.21	43	N
4	89.8	81.3	82	90.4	37	56.545	44	N
15	66.3	53.7	53	81.9	30	44.305	44	N
24	57.7	76.3	71	87.7	35	52.035	44	N
25	84.7	65	73	88.9	34	51.37	44	N
26	84.7	53.8	75	78.1	36	49.37	44	N
33	64.2	50	18	0	0	11.61	12	N
42	81.5	43.8	0	0	0	10.645	11	N
44	71.9	61.3	76	94.5	38	53.565	44	N
47	50	71.3	56	93.8	34	50.465	44	N
54	76.1	50	50	50	33	41.105	41	N
60	73.9	53.2	74	95.9	34	50.16	44	N
78	52.9	53.2	50	50	36	42.225	42	N

Display output:

```
students with exam score < 48
```

	Ass1	Ass2	Ass3	Ass4	Exam	Total	Final	Grade
ID								
3	74.3	54.4	63.0	63.9	31	43.210	43	N
4	89.8	81.3	82.0	90.4	37	56.545	44	N
15	66.3	53.7	53.0	81.9	30	44.305	44	N
24	57.7	76.3	71.0	87.7	35	52.035	44	N
25	84.7	65.0	73.0	88.9	34	51.370	44	N
26	84.7	53.8	75.0	78.1	36	49.370	44	N
33	64.2	50.0	18.0	0.0	0	11.610	12	N
42	81.5	43.8	0.0	0.0	0	10.645	11	N
44	71.9	61.3	76.0	94.5	38	53.565	44	N
47	50.0	71.3	56.0	93.8	34	50.465	44	N
54	76.1	50.0	50.0	50.0	33	41.105	41	N
60	73.9	53.2	74.0	95.9	34	50.160	44	N
78	52.9	53.2	50.0	50.0	36	42.225	42	N

```
students with exam score > 100
```

	Ass1	Ass2	Ass3	Ass4	Exam	Total	Final	Grade
ID								
11	96.6	100.0	98.0	97.3	102	100.0	100	HD
84	93.6	100.0	96.0	100.0	106	100.0	100	HD

#### Submission:

Submit the following files to OnTrack:

1. Your program source code (e.g. task3\_2.py)
2. result\_updated.**txt** (save result\_updated.csv as **.txt**) with the 3 newly added columns Total, Final and Grade
3. failedhurdle.**txt** (save failedhurdle.csv as **.txt**) generated by your code
4. A screen shot of your program running (only “exam <48” and “exam >100” output required)

Check the following things before submitting:

1. Add proper comments to your code