

# Python Programming

## Files Homework

**Mostafa S. Ibrahim**

*Teaching, Training and Coaching since more than a decade!*

*Artificial Intelligence & Computer Vision Researcher*

*PhD from Simon Fraser University - Canada*

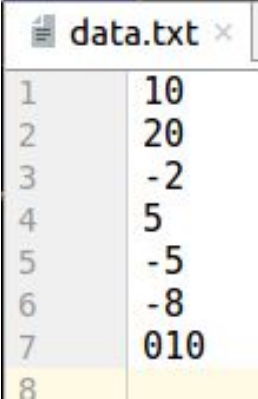
*Bachelor / Msc from Cairo University - Egypt*

*Ex-(Software Engineer / ICPC World Finalist)*



# Problem #1: Max \* Sum of a file

- Given a file of integers, each on a separate line (data.txt)
- We need to read and postprocess them by converting all to positive
- After that: print their sum \* their max
- Purpose: Mix reading the file with **list comprehension**
- Output for this file
  - 1200



| data.txt |     |
|----------|-----|
| 1        | 10  |
| 2        | 20  |
| 3        | -2  |
| 4        | 5   |
| 5        | -5  |
| 6        | -8  |
| 7        | 010 |
| 8        |     |

# Problem #2: Compute Differences V1

- **Background**
- As a udemy instructor, you would like to keep information about your competitors relevant courses
- When you access a course page, there are 2 numbers: total reviews / total users: We can use python to read URL and extract this information
  - Sometimes Udemy server rejects the request (stop crawlers)

## Learn Pro Advanced Python Programming

Become a Professional Python Programmer by learning Advance Level Programming & using them to make real Applications

4.3 ★★★★★ (107 ratings) 625 students

# Problem #2: Compute Differences V1

- You have the **URLs** of several courses. You developed a script that does:
  - Each 30 minutes, for each url it queries the udemy page and extract the 2 numbers
  - Then it **appends** a single line in the file as following
    - `date#r1 s1#r2 s2#r3 s3`
    - (r1, s1) is the rating & total students for the first course and so on
    - Each entry is **tab separated** (viewed above with #)



# Problem #2: Compute Differences V1

- Your task: we need to generate a new file with the following
  - 1) A single line per day: the latest valid day in the file
  - 2) Append extra value per course info: the total **number of increased students** from a day to another
- Assumptions
  - If a row is invalid, it will have the word NA as part of it
  - There will be no missing days in the middle, though it doesn't matter!
  - The data otherwise will be correct
- Find in the section resources: Input and expected output files
  - Generate output file named: courses-output-current datetime
  - E.g. courses-output-**2021-01-24 11:05:58**.txt

# Problem #2: Compute Differences V1

- Below is an output example

```
2020-11-18 19:15:39 327900 1134545 → 62239 253423 → 43974 201344 → 30250 105988 → 20828 77170
2020-11-19 16:27:01 328074 1135359 +814 → 62309 253615 +192 → 43988 201410 +66 → 30263 106042 +54 → 20836 77188 +18
2020-11-20 19:01:01 328265 1135862 +503 → 62367 254028 +413 → 44025 201593 +183 → 30280 106127 +85 → 20847 77283 +95
2020-11-21 7:48:56 → 328274 1137324 +1462 → 62371 254243 +215 → 44026 201691 +98 → 30281 106164 +37 → 20849 77323 +40
2020-11-22 17:59:56 → 328286 1138787 +1463 → 62373 254729 +486 → 44027 201926 +235 → 30283 106262 +98 → 20848 77401 +78
2020-11-23 19:58:34 → 328297 1140661 +1874 → 62381 255195 +466 → 44027 202123 +197 → 30284 106378 +116 → 20849 77471 +70
2020-11-24 13:42:04 → 328306 1141681 +1020 → 62381 255536 +341 → 44027 202246 +123 → 30284 106451 +73 → 20849 77557 +86
```

*“Acquire knowledge and impart it to the people.”*

*“Seek knowledge from the Cradle to the Grave.”*