Data Proc 2024: Assignment 04 (Part 1)

- Due: 8 Nov 2024 (before class: 1315 PM)
- Complete the assignment on the server using your user account.
  - You must create the scripts in the correct location in your home directory.

Assignment 04: Goals

- 1) Simple python processing
  - read command-line arguments via sys.argv
  - open and read a short file
  - write a loop to process the file contents
  - process some file contents (parse to numbers and add)
  - write results to a different file
- 2) Simple bash scripting
  - Copy a (py) script file and input files to a directory
  - Execute the script and output results.
  - Move the results to a specially named file (with date/time)

Example scripts in $DP24

- See the examples in:
  - $DP24/lectures/04/examples

```
riveale@dataproc2023vm:/usr/share/dataproc/2024/lectures/04/examples$ ls
00_sysargv.py                09_process_strings.py
01_python_formatting.py      exfile1.txt
02_python_opentextfile.py    exfile2.txt
03_python_textbylines.py     exfile2_wspace.txt
04_python_withas_open.py     run_examples.sh
05_python_otherreadlines.py  testout.bin
06_python_binaryread.py      testout.scratch
07_python_writingfile.py     testout.scratch.bin
08_python_writebinfile.py
```

I strongly recommend you look at these and make sure you understand them.

Assignment 04

- Create a directory in your home:
  - $HOME/dataproc2024/assignments/04/
- You will create/complete the following scripts:
  - $HOME/dataproc2024/assignments/04/problem0.py
  - $HOME/dataproc2024/assignments/04/problem1.py
  - $HOME/dataproc2024/assignments/04/problem2.py
  - $HOME/dataproc2024/assignments/04/problem3.py
  - $HOME/dataproc2024/assignments/04/problem4.py
- Extra credit:
  - $HOME/dataproc2024/assignments/04/problem5.py
  - $HOME/dataproc2024/assignments/04/problem6.py

Assignment 04: script templates

- Skeletons for the scripts are stored in:
  - $DP24/assignments/04/

```
riveale@dataproc2023vm:/usr/share/dataproc/2024/assignments/04$ ls
exfile1.txt  exfile2.txt  problem0.py  problem1.py  problem2.py
```

I recommend copying these into your home directory and modifying them.

exfile1.txt and exfile2.txt are example input files.

Assignment 04: Task 1

$HOME/dataproc2024/assignments/04/problem0.py

Your script problem0.py should print to standard output exactly:

```
10 10 10 30
Hello, my name is Bobert.
Then, he said to me "goodbye"; I took my leave.
Pi is divisible by 7 (3.1415...)
```

Note: there are no hidden spaces/newlines.

Assignment 04: Task 2

$HOME/dataproc2024/assignments/04/problem1.py

(see/copy: $DP24/assignments/04/problem1.py)

- Write a script that:
  - Takes 1 command line argument (filename to read in).
  - Reads the file indicated by the command-line argument.
  - Example file: exfile1.txt (see the format)
  - Format:
    - no header row(s)
    - separated by spaces
    - line format: FirstName LastName AgeYrs Number
- Output: Prints to standard output the average age of all entries in the file.

Assignr

$HOME/dat

(see/copy: $

- Write a scri
  - Takes 1 co
  - Reads the
  - Example fi
  - Format:
    - no header row(s)
    - separated by spaces
    - line format: FirstName LastName AgeYrs Number

```
John James 21 2902820191
Jimmy John 22 0002820192
Shota Ootani 30 0000000000
Brett Hall 19 B029699933
Scooby Doo 7 D398300022
Kanako Dempo-Evans 31 2000000292
Biff McSkiff 53 F0009388299
Ariana Grande 29 2901000000
Michael Jackson 70 50000002000
Glover Glover 23 2588823388
```

- Output: Prints to standard output the average age of all entries in the file.

Assignment 04: Task 3

$HOME/dataproc2024/assignments/04/problem2.py

(see/copy: $DP24/assignments/04/problem2.py)

- Write a script that:
  - Opens/reads file specified in command line argument.
  - Example file: exfile2.txt (see the format)
  - Format:
    - 1 header row!
    - separated by commas (,)
    - line format: Name,Surname,AgeYrs,StudentID,University
- Output: Count of number of people with university and number of people without university, and the average age of each of those groups. Print a header row and then 2 rows, using whitespace (' ') separation. E.g.:

  InUniv Count AvgAgeYrs

  Yes COUNT AVGAGE

  No COUNT AVGAGE

```
Name,Surname,AgeYrs,StudentID,University
Bob,Saget,21,29392002020,Kyoto University
Johnny,Depp,60,,
Nicholas,Cage,65,,
Miki,Yawata,21,B03989020202,Kyoto University
Don,Draper,35,,
Arnold,Schwarzanegger,75,,
Harvey,Specter,40,,
Kris,Donalds,24,77777229,Tokyo University
Kagami,Matthews,18,20019020001920,Harvard University
```

- separated by commas (,)
  - line format: Name,Surname,AgeYrs,StudentID,University

- Output: Count of number of people with university and number of people without university, and the average age of each of those groups. Print a header row and then 2 rows, using whitespace (' ') separation. E.g.:

  InUniv Count AvgAgeYrs

  Yes COUNT AVGAGE

  No COUNT AVGAGE