

UCS538 – Data Science Fundamentals

Assignment-04: Merging of Results

- **Dead Line:** 13 Sep 2020 | 23:59:59
- **Submission Link:** bit.ly/ucs538-assignment04
- **Submission Guidelines:** You need to submit -
 - One python (.py) file for Program 1 [**Required**]
- Sample testing files are available in “**Input files for Assignment04**” folder
- **Note:** Multiple submissions are allowed, but **latest submission** will be considered for the evaluation.

Program 1: Develop a command line program to merge all the results available in multiple files.

	Input File(s) Format						Output (Merged) File Format			
Case I (Multiple Input Files)	A.csv		B.csv		C.csv					
	Roll No	Marks	Roll	Mark	No	Mark	Roll No	A	B	C
	100	3	100	3	100	3	100	1	3	3
	101	4	101	4	101	4	101	3	3	3
	102	5	102		102		102	5	0	0
	101	3	101	3	101	3	105	0	1	0
	100	1	105	1	106	7	106	0	0	7
Case II (Single Input File)	A.csv									
	Roll No	Marks					Roll No	A		
	100	3					100	1		
	101	4					101	3		
	102	5								
	101	3								
	100	1								

Description:

- Column names may be different in input files, but first column is id/roll-num and second column is marks.
- Input file(s) may contain duplicate roll numbers, you need to consider the **last entry only**.
e.g. In **Case I**, the file A.csv contain two values for roll no 100. You need to considered only last entry for roll no 100 i.e. 1.
- Put “0” (Zero) for missing values.
 - e.g. In **Case I** the file A.csv contain no entry for roll no 105 and 106. You need to put “0” (Zero) in output file in Column “A”.
 - e.g. In **Case I** the file C.csv contain no entry for roll no 102. You need to put “0” (Zero) in output file in Column “C”.
- In **Case II**, If only one input file is present, then remove duplicate entries and considered only last entry.
- Marks must be of numeric type (Except for missing value). For the non-numeric value of marks, make the entry of **File Name**, **Roll No** and **Marks** in the log file.
- **Hint:** Find the union of all roll numbers, available in all input files.

Run the program through command line: **python merge.py <File1> <File2> n**

Example:

- python merge.py file1.csv
- python merge.py file1.csv myFile.csv
- python merge.py file1.csv myFile.csv marksFile.csv
- python merge.py file1.csv myFile.csv marksFile.csv abc.csv

Input/Output:

- Input File(s) Name → file1.csv | file1.csv file2.csv | file1.csv file2.csv file3.csv
- Output File Name →
 - Result file: Merged result file
 - “result-” + str(time.time()) + “.csv”
 - e.g. → “**result-20200909.csv**”
 - Log file: Non numeric entry for marks with file name, roll no and marks.
 - “log-” + str(time.time()) + “.csv”
 - e.g. → “**log-20200909.csv**”
 - A.csv, 102012, ab
 - B.csv, 102, x
 - B.csv, 119, ‘y’

Check for:

- Correct number of parameters
- Show appropriate message for wrong inputs.
- Handling of “File not Found” exception
- Input file(s) contain only two columns.
- Output file name will be “result-” + str(time.time()) + “.csv”
- Log file name will be “log-” + str(time.time()) + “.csv”