The Hashemite University
Prince Al-Hussein bin Abdullah II Faculty for IT
Department of Information Technology
Advance Programming for DS and AI (2010042351)

Assignment 1: Files and Exceptions
Due Date: 5-8-2023 11:59 PM
Max score: 100 points

Instructions:

1- Submit **one** python script file ({filename}.ipynb.
2- File naming: name your python script file with your [*your university ID]*. For example, if your team has two members the file's name must be their ID's separated by _.
3- send your code file to your instructor through the designated assignment.
4- Please keep in mind that **late submissions** will result in a **ZERO** score.
5- You must be able to discuss the details of your solution with your instructor.

The Dataset

Named Entity Recognition Data is used to classify words to different labels and is part pf de-identification of personal information that mitigates privacy risks to individuals.
The file contains of 40,000 records of phrase (sentence) that describe each record and their part-of-speeches (POS's) and tags.

File format:

The data format is txt (space separated) and each sentence is listed on multiple rows

Requirements:

Given a Named Entity Recognition dataset:

1. Use csv module to read the content of the two files "**file1.csv**" and "file2**.csv"**
   a. First file contains text data and,
   b. Second file contains its label (tags)

File1
Prime,Minister,Geir,Haarde,has,refused, to,resign,or,call,for,early,elections
Mr.,Karzai,has,led,Afghanistan,since,th e,Taleban,were,ousted,in,late,2001

File2
O,O,B-per,I-per,O,O,O,O,O,O,O,O,O,O
B-per,I-per,O,O,B-gpe,O,O,B-org,O,O,O,O,B-tim,

2. Combine the content of the two files to produce a new file named 'dataset-all.csv.' The structure of the new file is as shown in figure 1.
3. Use *re* module to replace numbers with comma-separator thousands to regular numbers. For example, the number 10000 should be replaced with "10,000" and write the result to a file named "**dataset-all.csv"**
   Hint: use `re.findall` and `re.replace` to accomplish this task
4. Finally, you must use exceptions to handle file reading and writing to avoid FileNotFound and ValueError exceptions

A	B	C
Sentence #	Word	Tag
Sentence: 1	Thousands	O
	of	O
	demonstra	O
	have	O
	marched	O
	through	O
	London	B-geo
	to	O
	protest	O
	the	O
	war	O
	in	O
	Iraq	B-geo
Sentence: 2	Families	O
	of	O
	soldiers	O
	killed	O
	in	O
	the	O
	conflict	O
	joined	O
	the	O
	protesters	O
Sentence: 3	They	O
	marched	O
	from	O

Figure 1: File structure