Midterm Examination

IST 5001 – Data Management with Python

Due: Friday, March 11, 2-16 by 11:59 PM

Possible Points: There are 4 programming problems each worth 62.5 points for a total of 250 possible points.

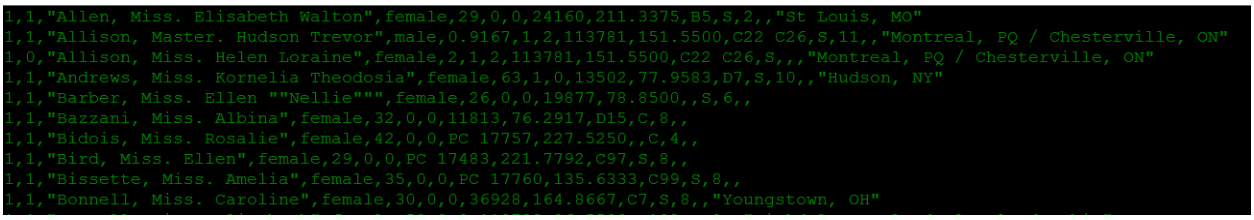
Submission: This is a bit complicated. Each problem asks you to take the "titanic3.csv" and transform it into other files using Python. Proceed as follows: put the output files and Python code for problem 1 into a folder named Problem 1. Create another folder named Problem 2 and put the material for problem two in it. And so forth until you have four folders. Copy these four folders into a zip folder named <your name>_midterm.zip. Upload the zipped folder into Blackboard. Note that I am asking you to have a python file for each problem instead of one file for the entire midterm. It took a lot of words to say this but I think you probably get it.

Input: All of these problems start with the same input file "titanic3.csv". That is, no processing done in Problem 1 carries to Problem 2, for example. Problem 2 begins a fresh with "titanic3.csv".

Problem 1: Classifying by Marital Status – The names in "titanic3.csv" all (I think) have titles such as: Mr. Mrs. Dr. Miss. Master. Etc. Also, the data includes the variable "sibsp" which reports whether the person has siblings (brother or sisters) or spouses (husbands or wives). Using these two fields, we will attempt to classify each person as married, single, or unknown. Use the following rules:

1. Titles of "Miss." and "Master." denote single people.
2. Titles of "Mrs.", "Mme.", "Sir", and "Lady" denote married people.
3. Titles of "Mr.", "Sir.", "Count.", are ambiguous. If you encounter one of these, look at the "sibsp" variable. If it is 1 or more, assume the person is married. If it is 0, then assume the person is of unknown marital status. Obviously, this is a weak rule, but this is just an exercise.
4. There are a scattering of other titles like military. Just assign all of those to unknown.

Once you have classified a person, his/her entire record from "titanic3." is to be written to an appropriate file. You will need a file for single, married, and unknown. Use csv files. Use the original record. You are just selecting particular records and writing them to a file. It will look something like this:



Problem 2: Also Known As Names – The names in "titanic3.csv" may offer an alternative name for people. These are like:

- (Lois Gornet) denotes a maiden name for Mrs. Michael Hilgers
- "Doc Hilgers" denotes an alternate name, traditional referred to "also known as" or aka names.
- ("Lolo Gornet") is an aka maiden name

You are to search the "titanic3.csv" for maiden and aka names and create three files: maiden names, aka names, and aka maiden names. Just write the maiden names or aka names to an output file. For example, the maiden names file might look like:

```
Bessie Waldo Daniels
Charlotte Lamson
Madeleine Talmadge Force
Helene DeLaudeniere Chaput
Sallie Monypeny
Helen Walton
Margaret Tobin
Caroline Lane Lamson
Emma Eliza Ward
Helen Churchill Hungerford
Charlotte Wardle Drake
Lucile Polk
```

Problem 3: Grouping by Life Boats – The people who survived may be associated with a life boat number. Group survivors by the boat in which they were rescued. Write to a file as shown:

```
Lady. (Lucille Christiana Sutherland) (Mrs Morgan) Duff Gordon, 01
Sir. Cosmo Edmund (Mr Morgan) Duff Gordon, 01
Miss. Laura Mabel Francatelli, 01
Mr. Abraham L Salomon, 01
Mr. Charles Emil Henry Stengel, 01
Miss. Elisabeth Walton Allen, 02
Mrs. Robert Clifford (Malvina Helen Lamson) Cornell, 02
Mrs. Walter Donald (Mahala Dutton) Douglas, 02
Miss. Emilie Kreuchen, 02
Miss. Bertha LeRoy, 02
```

Note that the boats are grouped together. Here are a few hints:

1. You need to put a zero in front of the single digit boat numbers otherwise they will sort like 11, 12, 2, 3, 4. You want 02, 03, 04, 11, 12.
2. There are boats with alphabetic names. Sort those to the end.

```
Master. Assad Alexander Thomas, 16
Mr. Elmer Zebley Taylor, 5 7
Mrs. Elmer Zebley (Juliet Cummins Wright) Taylor, 5 7
Mr. Peter Denis  Daly, 5 9
Miss. Constance Willard, 8 10
Mr. Thomson Beattie, A
Mr. George Alexander Lucien Rheims, A
Mr. Richard Norris II Williams, A
Mrs. Stanton (Rosa Hunt) Abbott, A
```

3. Some people have multiple boats. I have no idea what that means. Use the first number and ignore the others.
4. Write the names as shown.
5. Comma separate the names and the boat.
6. Write the results to a csv file.
7. In order to do this problem, I used the "defaultdict" class from the "collections" module. Read about this at: https://docs.python.org/3.4/library/collections.html#defaultdict-objects

Problem 4: Classifying by Age – The ages are integer with the exception of children under 1 who are given a decimal number (0.925). In some cases, all we want are age categories. In this case we want:

Age in Years (Integer except for less than 1)	Classification
0 to 12	"Child"
12 to 19	"Teenager"
20 to 55	"Adult"
56 and greater	"Senior"

Read a record from "titanic3.csv", determine the passengers age, then replace this number with the classification. Write this record to another csv file. It will look like:

1	1	Allen, Miss. Elisabeth Walton	female	Adult	0	0	24160	211.3375	B5	S	2		St Louis, MO
1	1	Allison, Master. Hudson Trevor	male	Child	1	2	113781	151.55	C22 C26	S	11		Montreal, PQ / Chesterville, ON
1	0	Allison, Miss. Helen Loraine	female	Child	1	2	113781	151.55	C22 C26	S			Montreal, PQ / Chesterville, ON
1	0	Allison, Mr. Hudson Joshua Creighton	male	Adult	1	2	113781	151.55	C22 C26	S		135	Montreal, PQ / Chesterville, ON
1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	Adult	1	2	113781	151.55	C22 C26	S			Montreal, PQ / Chesterville, ON
1	1	Anderson, Mr. Harry	male	Adult	0	0	19952	26.55	E12	S	3		New York, NY
1	1	Andrews, Miss. Kornelia Theodosia	female	Senior	1	0	13502	77.9583	D7	S	10		Hudson, NY
1	0	Andrews, Mr. Thomas Jr	male	Adult	0	0	112050	0	A36	S			Belfast, NI
1	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	Adult	2	0	11769	51.4792	C101	S	D		Bayside, Queens, NY
1	0	Artagaveytia, Mr. Ramon	male	Senior	0	0	PC 17609	49.5042		C		22	Montevideo, Uruguay
1	0	Astor, Col. John Jacob	male	Adult	1	0	PC 17757	227.525	C62 C64	C		124	New York, NY
1	1	Astor, Mrs. John Jacob (Madeleine Talmadge Force)	female	Teenager	1	0	PC 17757	227.525	C62 C64	C	4		New York, NY
1	1	Aubart, Mme. Leontine Pauline	female	Adult	0	0	PC 17477	69.3	B35	C	9		Paris, France
1	1	Barber, Miss. Ellen "Nellie"	female	Adult	0	0	19877	78.85		S	6		
1	1	Barkworth, Mr. Algernon Henry Wilson	male	Senior	0	0	27042	30	A23	S	B		Hessle, Yorks
1	0	Baumann, Mr. John D	male	Unknown	0	0	PC 17318	25.925		S			New York, NY