COSC 1306 – Computer Science and Programming Course

Programming Assignment # 3

Deadline: *04/16/2018 11:59 PM*

Problem: Reading Computer Science Career Fair file and modifying and summarizing the content.

The topics covered in this homework: Conditionals, Iterations, Strings, Lists, Reading/Writing Files.

Every semester, Computer Science Department organize a career fair. In the announcements, the company list (Employeers) are presented with other related information as a pdf file, as shown in Figure 1. The pdf file has 2 pages, in the first page the companies are listed based on alphabetical order whereas in the second page they are listed based on boot number.

The file "CF-Spring2018-Employers.csv" is created from this pdf file and our task is to read this csv file and manipulate the rows and columns to get the required data formats. The pdf file screenshot and CSV file may have some minor differences, please ignore that difference. And do not manually update the "CF-Spring2018-Employers.csv", imagine that you don't have a direct access to this csv file but you have only the access this file through the Python program.

1) Read the "CF-Spring2018-Employers.csv" file and save the data to lines data format. (You have several tasks to complete in the following but you can read the csv file only for once. Use all calculations and data manipulations from this lines/rows data format you have saved.)

2) By using these lines, find the related row that holds the information for the columns of the table. Print the columns as below:

```
0 Company
1 Booth
2 Full-Time
3 Full-Time Visa Sponsor
4 Part-Time
5 Internship
6 Freshman
7 Sophomore
8 Junior
9 Senior
10 Post-Bacs
11 MS
12 PhD
13 Alumni
```

UH Computer Science
Career Fair - Spring 2018

Employer List

Last Updated: 1/30/2018 2:03 PM

ALPHABETICAL ORDER

	Booth	Positions				Classifications							
Company	#	Full-Time	Full-Time Visa Sponsor	Part-Time	Internship	Freshman	Sophomore	Junior	Senior	Post-Bacs	MS	PhD	Alumni
AIG	10				Yes			Jr			MS		
Baylor College of Medicine	19	Yes	Yes										Recent
CGG	17	Yes	Yes								MS	PhD	Recent
Citi	27/28	Yes			Yes			Jr	Sr				
ExxonMobil	11	Yes			Yes	Fr	Soph	Jr	Sr	PB			
Flow-Cal, Inc.	16	Yes			Yes			Jr	Sr				All
Global Shop Solutions	18	Yes			Yes				Sr	PB			All
Harris County CTS	22	Yes			Yes			Jr	Sr	PB	MS	PhD	All
HCSS	29	Yes			Yes	Fr	Soph	Jr	Sr	PB	MS		Recent
Hitachi Consulting	13	Yes							Sr		MS		
HP Inc.	1	Yes			Yes			Jr			MS		Recent
INT, Inc.	20	Yes	Yes		Yes			Jr	Sr		MS	PhD	
JPMorgan Chase & Co	3	Yes			Yes			Jr	Sr				
Leidos	390	Yes			Yes	Fr	Soph	Jr	Sr	PB	MS		
McKesson	26	Yes							Sr				
MRE Consulting, Ltd.	2	Yes							Sr	PB	MS		All
NetIQ	7				Yes		Soph	Jr	Sr	PB			
PROS	21	Yes							Sr		MS	PhD	All
San Jacinto College	14				Yes		Soph	Jr	Sr	PB	MS		
SAS	4	Yes			Yes	Fr	Soph	Jr	Sr	PB	MS		Recent
Smartbridge	8	Yes							Sr	PB	MS		
Sogeti USA	15	Yes							Sr	PB	MS		
Southwest Research Institute	12	Yes			Yes			Jr	Sr	PB	MS	PhD	All
The Reynolds and Reynolds Company	23	Yes	Yes		Yes	Fr	Soph	Jr	Sr	PB			All
UH Enterprise Systems	9	Yes	Yes	Yes	Yes	Fr	Soph	Jr	Sr	PB	MS	PhD	All
U.S. Marine Corps	25	Yes			Yes	Fr	Soph	Jr	Sr	PB	MS		All
ValuD Consuting LLC	5	Yes							Sr	PB			All
Wipro	24	Yes							Sr	PB			

BOOTH ORDER

	Booth	Positions				Classifications							
Company	#	Full-Time	Full-Time Visa Sponsor	Part-Time	Internship	Freshman	Sophomore	Junior	Senior	Post-Bacs	MS	PhD	Alumni
HP Inc.	1	Yes			Yes			Jr			MS		Recent
MRE Consulting, Ltd.	2	Yes							Sr	PB	MS		All
JPMorgan Chase & Co	3	Yes			Yes			Jr	Sr				
SAS	4	Yes			Yes	Fr	Soph	Jr	Sr	PB	MS		Recent
ValuD Consuting LLC	5	Yes							Sr	PB			All
NetIQ	7				Yes		Soph	Jr	Sr	PB			
Smartbridge	8	Yes							Sr	PB	MS		
UH Enterprise Systems	9	Yes	Yes	Yes	Yes	Fr	Soph	Jr	Sr	PB	MS	PhD	All
AIG	10				Yes			Jr			MS		
ExxonMobil	11	Yes			Yes	Fr	Soph	Jr	Sr	PB			
Southwest Research Institute	12	Yes			Yes			Jr	Sr	PB	MS	PhD	All
Hitachi Consulting	13	Yes							Sr		MS		
San Jacinto College	14				Yes		Soph	Jr	Sr	PB	MS		
Sogeti USA	15	Yes							Sr	PB	MS		
Flow-Cal, Inc.	16	Yes			Yes			Jr	Sr				All
CGG	17	Yes	Yes								MS	PhD	Recent
Global Shop Solutions	18	Yes			Yes				Sr	PB			All
Baylor College of Medicine	19	Yes	Yes										Recent
INT, Inc.	20	Yes	Yes		Yes			Jr	Sr		MS	PhD	
PROS	21	Yes							Sr		MS	PhD	All
Harris County CTS	22	Yes			Yes			Jr	Sr	PB	MS	PhD	All
The Reynolds and Reynolds Company	23	Yes	Yes		Yes	Fr	Soph	Jr	Sr	PB			All
Wipro	24	Yes							Sr	PB			
U.S. Marine Corps	25	Yes			Yes	Fr	Soph	Jr	Sr	PB	MS		All
McKesson	26	Yes							Sr				
Citi	27/28	Yes			Yes			Jr	Sr				
HCSS	29	Yes			Yes	Fr	Soph	Jr	Sr	PB	MS		Recent
Leidos	30	Yes			Yes	Fr	Soph	Jr	Sr	PB	MS		

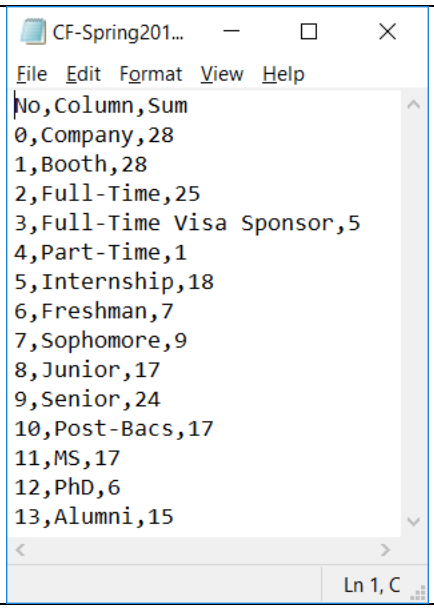
Figure 1: Pdf file screen shoot

3) By using these lines, filter the unique rows and print the following. In this section, your goal is to find and filter the list of the first page table data only.

```
0 AIG,10,,,,Yes,,,Jr,,,MS,,
1 Baylor College of Medicine,19,Yes,Yes,,,,,,,,,Recent
2 CGG,17,Yes,Yes,,,,,,,,MS,PhD,Recent
3 Citi,27/28,Yes,,,Yes,,,Jr,Sr,,,,
4 ExxonMobil,11,Yes,,,Yes,Fr,Soph,Jr,Sr,PB,,,
5 Flow-Cal Inc.,16,Yes,,,Yes,,,Jr,Sr,,,,All
6 Global Shop Solutions,18,Yes,,,Yes,,,,Sr,PB,,,All
7 Harris County CTS,22,Yes,,,Yes,,,Jr,Sr,PB,MS,PhD,All
8 HCSS,29,Yes,,,Yes,Fr,Soph,Jr,Sr,PB,MS,,Recent
9 Hitachi Consulting,13,Yes,,,,,,,Sr,,MS,,
10 HP Inc.,1,Yes,,,Yes,,,Jr,,,MS,,Recent
11 INT Inc.,20,Yes,Yes,,Yes,,,Jr,Sr,,MS,PhD,
12 JPMorgan Chase & Co,3,Yes,,,Yes,,,Jr,Sr,,,,
13 Leidos,390,Yes,,,Yes,Fr,Soph,Jr,Sr,PB,MS,,
14 McKesson,26,Yes,,,,,,,Sr,,,,
15 MRE Consulting Ltd.,2,Yes,,,,,,,Sr,PB,MS,,All
16 NetIQ,7,,,,Yes,,Soph,Jr,Sr,PB,,,
17 PROS,21,Yes,,,,,,,Sr,,MS,PhD,All
18 San Jacinto College  ,14,,,,Yes,,Soph,Jr,Sr,PB,MS,,
19 SAS,4,Yes,,,Yes,Fr,Soph,Jr,Sr,PB,MS,,Recent
20 Smartbridge,8,Yes,,,,,,,Sr,PB,MS,,
21 Sogeti USA,15,Yes,,,,,,,Sr,PB,MS,,
22 Southwest Research Institute,12,Yes,,,Yes,,,Jr,Sr,PB,MS,PhD,All
23 The Reynolds and Reynolds Company,23,Yes,Yes,,Yes,Fr,Soph,Jr,Sr,PB,,,All
24 UH Enterprise Systems,9,Yes,Yes,Yes,Yes,Fr,Soph,Jr,Sr,PB,MS,PhD,All
25 U.S. Marine Corps,25,Yes,,,Yes,Fr,Soph,Jr,Sr,PB,MS,,All
26 ValuD Consuting LLC,5,Yes,,,,,,,Sr,PB,,,All
27 Wipro,24,Yes,,,,,,,Sr,PB,,,
```

4) Calculate the following. Print the output and write this output to a new file as a comma separated format: "**CF-Spring2018-Employers_Summary.csv**".

EXPECTED PROGRAM OUTPUT:	**EXPECTED CSV FORMAT:**

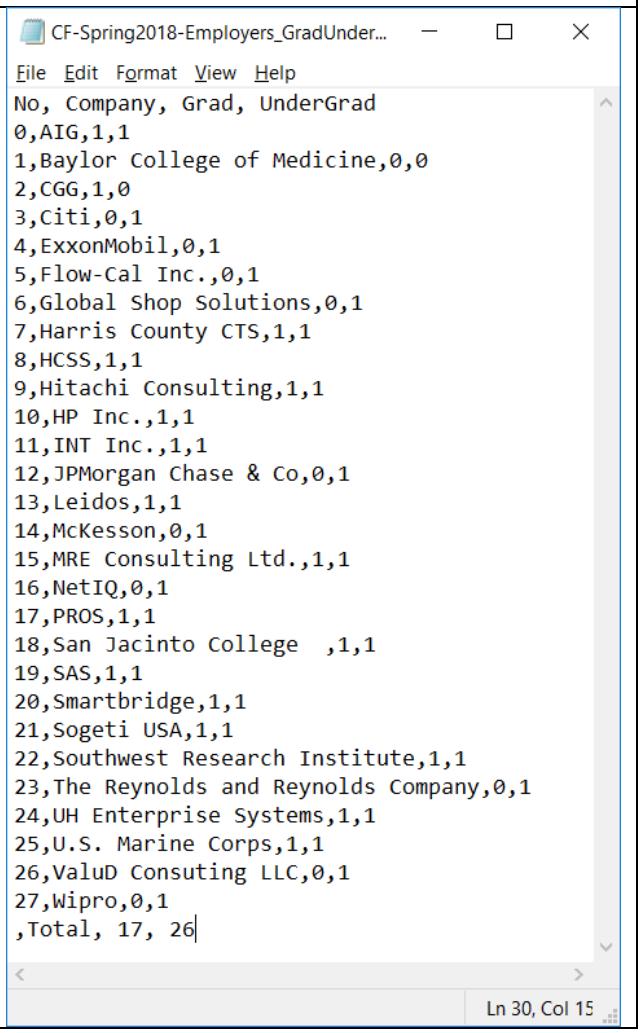
`No Column Sum`
`0 Company 28`
`1 Booth 28`
`2 Full-Time 25`
`3 Full-Time Visa Sponsor 5`
`4 Part-Time 1`
`5 Internship 18`
`6 Freshman 7`
`7 Sophomore 9`
`8 Junior 17`
`9 Senior 24`
`10 Post-Bacs 17`
`11 MS 17`
`12 PhD 6`
`13 Alumni 15`
	

4) Calculate the following. You need to merge related columns (`Freshman, Sophomore, Junior, Senior, Post-Bacs, MS,PhD)` to create new columns, grad and undergrad. Print the output and write this output to a new file as a comma separated format: "**CF-Spring2018-Employers_GradUnderGrad.csv" as below.**

EXPECTED PROGRAM OUTPUT:

```
No Company Grad UnderGrad
0 AIG 1 1
1 Baylor College of Medicine 0 0
2 CGG 1 0
3 Citi 0 1
4 ExxonMobil 0 1
5 Flow-Cal Inc. 0 1
6 Global Shop Solutions 0 1
7 Harris County CTS 1 1
8 HCSS 1 1
9 Hitachi Consulting 1 1
10 HP Inc. 1 1
11 INT Inc. 1 1
12 JPMorgan Chase & Co 0 1
13 Leidos 1 1
14 McKesson 0 1
15 MRE Consulting Ltd. 1 1
16 NetIQ 0 1
17 PROS 1 1
18 San Jacinto College   1 1
19 SAS 1 1
20 Smartbridge 1 1
21 Sogeti USA 1 1
22 Southwest Research Institute 1 1
23 The Reynolds and Reynolds Company 0 1
24 UH Enterprise Systems 1 1
25 U.S. Marine Corps 1 1
26 ValuD Consuting LLC 0 1
27 Wipro 0 1
 Total 17 26
```

EXPECTED CSV FORMAT:



Requirements

Submit to Blackboard a single python file named "**PeopleSoftId_PA3.py**". *Include in the header the Programming Assignment number, your name, and PSID*. Use meaningful variable names and constants as needed. Make sure that your program print the expected output and write the CSV files in the same format. Compare your CSV files with the expected output CSV files given to you.

CF-Spring2018-Employers_Summary.csv == CF-Spring2018-Employers_Summary_EXPECTED_OUTPUT.csv

CF-Spring2018-Employers_GradUnderGrad.csv == CF-Spring2018-Employers_GradUnderGrad_EXPECTED_OUTPUT.csv

There may be extra rows and noisy data, you may also need to handle other possible issues so please start early.