# CPSC 1301 COMPUTER SCIENCE 1 FALL 2014

# Assignment 8 – Data Analytics

Posted on 11/30/2014

**Due date : Before midnight, Wednesday, December 8**
**Total Points : 50**

**You are encouraged to work on this assignment with a partner in your class.**

**READ ALL INSTRUCTIONS BELOW CAREFULLY**

**Assignment Overview**
"Data analytics (DA) is the science of examining raw data with the purpose of drawing conclusions about that information. Data analytics is used in many industries to allow companies and organization to make better business decisions and in the sciences to verify or disprove existing models or theories. Data analytics focuses on inference, the process of deriving a conclusion based solely on what is already known by the researcher." [http://searchdatamanagement.techtarget.com/definition/data-analytics]

The purpose of this assignment is to continue your study of computer programming and algorithms through the Python programming language. In this lab you will use several features you have learnt including – conditionals, loops, user-defined functions and lists and a new feature – reading from a file (Read sections 13.1 and 13.2 of chapter 13 in the textbook for information on how to do file input/output).

You will need three files in this assignment: `WorldCensus0-14.csv`, `WorldCensus15-64.csv`, and `WorldCensus64+.csv`. The files are provided in the Programing Assignment 8 Dropbox in CougarView. Each of these files has population information about the countries in the world. The number at the end of file name indicates what age group the date in that file belongs to. In each file, each line consists of a country name, male population for that country, and female population for that country separated by commas. It'd be a good idea to browse through these files before you start working on the assignment so you know what you are dealing with. To open the file, download it, right click and open it with Notepad++.

**Problem statement:**
Write a Python program that processes demographic data about the population of different countries of the world that are stored in three text files for three different age groups. It will accept user commands by providing a menu of options for the user to select from and produce outputs according to specifications detailed below.

**Program Specifications:**
Your program should do the following:

1. Define a function called `getLists` that takes a pointer to a file as a parameter and reads the data in the file into three lists and returns the three lists.

2. Define a function called `printMenu` that prints a menu for the users with the following options:
   A- Choose a country to display population information about that country.
   B- Choose a letter to display population information about all countries starting with that letter
   C- Male to female ratios among different age groups in ascending order
   D- List of countries with the highest male to female ratios among different age groups
   E- List of countries with the lowest male to female ratios among different age groups
   F- Numbers of countries with more males than females in each age group
   G- Quit

3. Define a main function where you
   a. Print a welcome message
   b. Call the getLists method three times to
      i. Read the information from the WorldCensus0-14.csv file into lists
      ii. Read the information from the WorldCensus15-64.csv file into lists
      iii. Read the information from the WorldCensus64+.csv file into lists

*Note that the country names are the same and appear in the same order in the three files. So you can use the same list for country names for all three calls.*

   c. Create three new lists called `totalMale`, `totalFemale`, `totalPopulation` that hold the total male population, total female population, and total population for each country. (all age groups put together)
   d. Display the menu
   e. Get user's choice
   f. Proceed according to user's choice. If the user choses an invalid option, prompt the user to enter their choice again.
   g. Print a good-bye message

4. Call the main function in your program.

***Note:*** *It is up to you what other functions to include in your program. You may want to have a function for each option or for some of them only. Before you start coding, make sure that you have thought your design thoroughly. Draw a flowchart and/or write a pseudocode so you will spend less time fixing errors when coding!*

**Sample outputs:**

**Option A:** For example, if the user enters "Turkey" for the country, this information should be displayed:

| Country | Age | Both Sexes | Male | Female | % Both | %Male | %Female | Male to Female |
|---|---|---|---|---|---|---|---|---|
| Turkey | Total | 81619392 | 41106823 | 40512569 | 100 | 100 | 100 | 1.014668386 |
| Turkey | 0-14 | 20839960 | 10660110 | 10179850 | 25.5 | 25.9 | 25.1 | 1.047177512 |
| Turkey | 15-64 | 55288903 | 27929494 | 27359409 | 67.7 | 67.9 | 67.5 | 1.02083689 |
| Turkey | 65+ | 5490529 | 2517219 | 2973310 | 6.7 | 6.1 | 7.3 | 0.846604962 |

**Option B:** For example, if the user enters "U" for the letter, this information should be displayed:

| Country | Age | Both Sexes | Male | Female | % Both | %Male | %Female | Male to Female |
|---|---|---|---|---|---|---|---|---|
| Uganda | Total | 35918915 | 17841632 | 18077283 | 100 | 100 | 100 | 0.986964247 |
| Uganda | 0-14 | 17480254 | 8714354 | 8765900 | 48.7 | 48.8 | 48.5 | 0.994119714 |
| Uganda | 15-64 | 17695815 | 8799507 | 8896308 | 49.3 | 49.3 | 49.2 | 0.989118969 |
| Uganda | 65+ | 742846 | 327771 | 415075 | 2.1 | 1.8 | 2.3 | 0.789666928 |
| Ukraine | Total | 44291413 | 20333333 | 23958080 | 100 | 100 | 100 | 0.848704612 |
| Ukraine | 0-14 | 6204822 | 3191247 | 3013575 | 14 | 15.7 | 12.6 | 1.058957219 |
| Ukraine | 15-64 | 31041037 | 14831434 | 16209603 | 70.1 | 72.9 | 67.7 | 0.914978239 |
| Ukraine | 65+ | 7045554 | 2310652 | 4734902 | 15.9 | 11.4 | 19.8 | 0.488004187 |
| UnitedArabEmirates | Total | 5628805 | 3861930 | 1766875 | 100 | 100 | 100 | 2.185740361 |
| UnitedArabEmirates | 0-14 | 1167751 | 597476 | 570275 | 20.7 | 15.5 | 32.3 | 1.04769804 |
| UnitedArabEmirates | 15-64 | 4405595 | 3229383 | 1176212 | 78.3 | 83.6 | 66.6 | 2.745579028 |
| UnitedArabEmirates | 65+ | 55459 | 35071 | 20388 | 1 | 0.9 | 1.2 | 1.720178536 |
| UnitedKingdom | Total | 63742977 | 31688615 | 32054362 | 100 | 100 | 100 | 0.98858979 |
| UnitedKingdom | 0-14 | 11041339 | 5660891 | 5380448 | 17.3 | 17.9 | 16.8 | 1.052122611 |
| UnitedKingdom | 15-64 | 41529500 | 21037700 | 20491800 | 65.2 | 66.4 | 63.9 | 1.026639924 |
| UnitedKingdom | 65+ | 11172138 | 4990024 | 6182114 | 17.5 | 15.7 | 19.3 | 0.807171139 |
| UnitedStates | Total | 318892103 | 157082861 | 161809242 | 100 | 100 | 100 | 0.970790414 |
| UnitedStates | 0-14 | 61801455 | 31580349 | 30221106 | 19.4 | 20.1 | 18.7 | 1.044976613 |
| UnitedStates | 15-64 | 210911644 | 105197868 | 105713776 | 66.1 | 67 | 65.3 | 0.995119766 |
| UnitedStates | 65+ | 46179004 | 20304644 | 25874360 | 14.5 | 12.9 | 16 | 0.784739951 |
| Uruguay | Total | 3332972 | 1609739 | 1723233 | 100 | 100 | 100 | 0.934138912 |
| Uruguay | 0-14 | 701427 | 356851 | 344576 | 21 | 22.2 | 20 | 1.035623491 |
| Uruguay | 15-64 | 2167037 | 1067756 | 1099281 | 65 | 66.3 | 63.8 | 0.971322164 |
| Uruguay | 65+ | 464508 | 185132 | 279376 | 13.9 | 11.5 | 16.2 | 0.662662505 |
| Uzbekistan | Total | 28929716 | 14396548 | 14533168 | 100 | 100 | 100 | 0.990599434 |
| Uzbekistan | 0-14 | 7208572 | 3693838 | 3514734 | 24.9 | 25.7 | 24.2 | 1.050958053 |
| Uzbekistan | 15-64 | 20340654 | 10113829 | 10226825 | 70.3 | 70.3 | 70.4 | 0.988951019 |
| Uzbekistan | 65+ | 1380490 | 588881 | 791609 | 4.8 | 4.1 | 5.4 | 0.743903872 |

**Option C:** There should be four columns with the male to female ratios for the countries listed in ascending order: One for each age group and one for total population. The results should be printed 15 rows at a time. (The country names do not need to be included)

**Option D:** Four country names should be displayed one for each age group and one for total population. The ratios should be also displayed along with the country names. It should be clear which country goes with which age group.

**Option E:** Similar to Option D but the lowest ratios and the corresponding country names are displayed.

**Option F:** Four numbers for each age group and one for total population should be displayed. It should be clear which number goes with which age group.

**EXTRA CREDIT (20pts) -** if you provide a GUI using TKinter in your program

**What to submit (Deliverables):**
Take **five** screenshots of your program's outputs:

1. When it first starts - the menu and the prompt to the user to choose an option.
2. When the user chooses Option B and enters the letter U.
3. When the user chooses Option D
4. When the user chooses Option E
5. When the user chooses Option F

Copy these screenshots into a Word or OpenOffice file. In this file, provide answers to these questions on your thoughts on completing this assignment :

1) What was the most challenging part ?
2) What was the most enjoyable part ?
3) What topics did you master when completing this assignment ?

Save your file as <last name>_<initial>_Assignment8.py>, or <last name>_<last name>_Assignment8.py> (if working as a pair).

Submit the file in the **Assignment 8 Dropbox** in CougarView. Both you and your partner must submit the same files.

**Grading Rubric:**

| | |
|---|---|
| Screenshots | 5 points |
| Answers to questions | 5 points |
| Style and documentation | 5 points |
| Program | 35 points s |

REMINDER ABOUT ACADEMIC INTEGRITY

Any work turned in for individual credit must be entirely the work of the student submitting the work (unless it is a sanctioned pair or group effort). All work must be your own. You may share ideas but submitting identical assignments (for example) will be considered cheating. You may discuss the material in the course and help one another with debugging, however, I expect any work you hand in for a grade to be your own. A simple way to avoid inadvertent plagiarism is to talk about the assignments, but don't read each other's work or write solutions together. Having access to another person's work on the system or giving access to your work to another person is not allowed. It is your responsibility to keep your work confidential.

No cheating in any form will be tolerated. Please be aware that anyone caught cheating or plagiarizing in this class will receive a "0" for the assignment/exam and may receive a "0" for the course.