TASK

In this assignment, you will perform some basic data analysis on a dataset obtained from Gapminder project (<https://www.gapminder.org/tools/>). Gapminder collects authentic facts and statistics of all countries worldwide and then plots the data in easy to understand visualization tools.

You have been provided a dataset file Emissions.csv which contains CO2 emissions data extracted from a Gapminder dataset. Download Emissions.txt file from the unit Interact site. The file contains comma-separated data of annual CO2 emissions (per capita) from 195 countries for a period of 1997 to 2010. CO2 emissions are measured in metric tones. It is a plain text file as shown in screenshot below. First line contains data headers, and then each line contains data for one country. To clearly understand data structure, you can also open the csv file in a spreadsheet software.

Your program will read this data file and perform the following jobs:

(1) Read all the data from file and save it into a Python dictionary. Each key in the dictionary should be a country name as read from the file, and value of that key will be a Python list containing emission data for that specific country. Once all the file is read, dictionary will contain 195 keys Each key will correspond to a Python list containing 14 numbers (emission data from 1996 to 2010). You should use this dictionary for the next three jobs.

(2) Calculate worldwide statistics (min, max, average) for a user-selected year. See example in the sample-run below.

(3) Extract data for up to three user-selected country and save it to a new file Emissions\_subset.csv. New file should have exactly same format as the source file, i.e. first line of headers and then up to 3 lines for selected countries. See the sample-run below for an example.

(4) Plot the emissions data from a user-selected country. You should use Python plotting library matplotlib for drawing the plots. The links below contain examples on how to draw simple plots using this library.

**Important**: Other than matplotlib, you CAN NOT use any other Python library (pandas, numpy etc.) for this assignment. Only use Python built in functions.

Your program should be able handle invalid inputs and errors such as

* File Emissions.csv does not exist or can’t be read
* Output file can’t be saved
* Incorrect Year provided by user
* Incorrect country name provided by user

A sample run of the program is given below to clearly demonstrate all the requirements. Take a note of two things (1) Emission statistics are displayed in 6 decimal places. (2) User-input country names should be case insensitive.

A Simple Data Analysis Program  
  
\*\* Job 1 \*\*  
All data from Emissions.csv has been read into a dictionary.  
  
  
\*\* Job 2 \*\*  
Select a year to find statistics (1997 to 2010): **2019**  
  
Sorry that is not a valid year.  
  
Select a year to find statistics (1997 to 2010): **2003**  
  
In 2003, countries with minimum and maximum CO2 emission levels were: [Afghanistan] and [Qatar] respectively.  
Average CO2 emissions in 2003 were 5.265891  
  
  
\*\* Job 3 \*\*  
Write up to three comma-separated countries for which you want to extract data: **jupiter, mars, saturn**  
ERR: Sorry “jupiter” is not a valid country  
  
Write up to three comma-separated countries for which you want to extract data: **germany, france, brazil, neptune**  
ERR: Sorry, at most 3 countries can be entered.  
  
Write up to three comma-separated countries for which you want to extract data: **germany, sri lanka**  
  
Data successfully extracted for countries Germany, Sri Lanka saved into file Emissions\_subset.csv  
  
  
\*\* Job 4 \*\*  
Select the country to visualize: **denmaRk**  
  
Plot for Denmark opens in a new window.  
  
Thanks for using this program.