Python Project

February 24, 2021

1 Goals:

- practice basic python commands
- practice using functions
- modularity and abstraction and separation of concerns
- go over the process from receiving data sets to generating results

2 Tasks

2.1 Data Management (exploring, cleaning, creating data sets)

1. (function) exploring data set : show the columns tags of gapminder data set + data types of columns
2. (function) data integrity check : find out of there is an empty numerical cell, if yes, replace it with the mean value of the column
3. (function) create a new .csv file for each country and save in folder called countries. In each file the columns are years and rows are gdp, population, life expectancy.
4. (function) create a new .csv file ties countries with continents.
5. (function) use (3,4) to create .csv files for continents gpd, population, life expectancy

2.2 Analysis Tasks (mathematical, classification)

1. mathematical: (function) calculate the growth rate of population, gdp, life expectancy for a country.
2. (function) save the growth rates back into the corresponding .csv file of the country. it adds a new column with growth rates for each year.

3. use the function in (1,2 iterative) with list of countries extracted from .csv of countries to populate all .csv files of countries.

4. (function) classify a country into developed/developing based on final year of life expectancy

5. (function) saves a country status in .csv files continents

6. use (4,5 iterative) to add a new column in continents.csv called status, and set it developed/developing.

2.3 Results in jpeg or pdf format (visualization and result saving)

1. (function) visualization : plot the population, gpd, life for one country

2. (function) visualization : plot population, gdp, life for multiple countries - use (1, iterative)

3. (function) visualization : plot growth rate of gdp,pop,life expectancy for one country

4. (function) visualization : plot growth rates for multiple countries

5. (function) saving plots : takes a plot and a name and save it to a folder called results.