

Graded assignment

Information

Instructions

- Answer the questions in a single Jupyter notebook.
- You are allowed to team up with 1 other student. If you do so, you **MUST** state the names of both team members at the top of your Jupyter notebook.
- You are allowed to use any package available on the [Python Package Index](#).
- Several questions require data cleaning and enrichment efforts. In some cases there may be several different processing options, resulting in different sets of results. Thus: if you are in such an ambiguous situation, then briefly explain how you choose to process the data. If your processing steps are reasonable, then your answer will be counted as correct.
- Submit a clean notebook that contains **ONLY (!)** the code or explanations that are relevant for the questions asked. I recommend to first solve the tasks and then revise your code to make it clean and [pythonic](#).
- Your results must be fully reproducible.

Data

You are provided with 3 datasets related to Olympic Games:

- `athletes.txt` contains biographic information of athletes (e.g. the name and sex)
- `games.tsv` contains basic information about of the Olympic Games (e.g. the location and opening dates)
- `results.csv` contains information about individual results of athletes in the Olympic Games since 2012: Which athlete participated in which Olympic Games, sports and disciplines, achieving which position?

If you want to get a better idea of what these information mean and where they come from, see: [athletes](#), [games](#), [results](#). Only minor renaming has been performed on the raw data from these sources.

Exercises

Exercise 1

- What is the name of the most successful athlete since 2012, as measured by the number of gold medals?
- How many different athletes have participated in the Olympic Games since 2012?
- What are the 10 countries with the lowest ratio of female participants? Represent this information using a bar chart.

Exercise 2

- What are the three most common firstnames of female athletes from the country 'Republic of Korea'?

- Who is the tallest female athlete? Provide the name, height, country and sport.
- What are the 10 sports with the highest average weight of participants? Provide the sport and the weight.
- What are the 3 sports with the largest number of participants that are severely underweight? Provide the sport and the number of severely underweight participants. Severe underweight is defined by a body-mass-index (BMI) smaller than 16, where BMI is calculate as $weight/height^2$ in the unit kg/m^2 .

Exercise 3

For the following two questions you need to calculate the age of participantants. We define age as the number of completed years at the beginning of the Olympic Games at which the athlete participates. To calculate the age you need to (1) leverage the information about the date of birth and you (2) need to consider the date when the Olympic Games were opened.

- Who is the oldest athlete? Provide the name and the age.
- What are the 10 sports with the highest average age of participants? Provide the sport and the average age.

Exercise 4

Calculate the medal table for the Summer Olympic Games 2016 in Rio de Janeiro. See the [official medal table](#) as a reference. Your medal table should have the same structure (same column names and same sorting) as the official medal table. Display the top 10 countries.

Hint: In team events such as Basketball, all team members of the winning team receive a gold medal, but for the medal table it should only count as one gold medal. The same logic also applies to silver and bronze medals. You can recognize team events either by the fact that multiple participantants have won the same medal in one event event (combination of sport and discipline), or via the column `team` (which is then non-missing).

Exercise 5

Not surprisingly, large countries tend to win more medals than small countries. Your task is now to calculate a different version of the 2016 medal table that shows the number of gold, silver and bronze medals per 1 million inhabitants. Display the top 10 countries of this new medal table.

To solve this exercise, retrieve the countries population in 2016 using the Worldbank API. You may consult the [API documentation](#) and the code examples from the class.