Programming Course: Python Project, Part 2

Project Setup:

In the second part of the project, you should evaluate the data given in the data set *120 years of Olympic history: athletes and results* (given in the file `athletes_events.csv`) w.r.t. a given research question. You can also include data from additional data sets you can e.g. find on Kaggle (`www.kaggle.com/datasets`), Data World (`https://data.world/`) or other web sites.
After choosing a research question you should evaluate the data using *pandas* and visualize your results with the help of *Matplotlib*.
You will find some possible research questions below. You can either choose any of the given ideas or setup your own.
In the end, you will have to give a short presentation of around 10 - 15 minutes (including discussion).

Task description:

Download the CSV file `athletes_events.csv` from Ilias (or use the file you already downloaded for the first part of the project) and copy it to a new project directory.

Proceed as follows:

(1) For this project part, you should work in teams of two. So, find a team mate and write an email with the names of the group members to `rost@uni-mannheim.de`. If you would like to be assigned to a group, please write me an email.

(2) Specify the research question you would like to answer.

(3) Collect all data you need to answer the question.

(4) Write the Python code to filter the data and to create adequate visualizations for the results. It might be a good idea to create some intermediate data structures (like dictionaries, lists, etc.) with filtered data which can be used more easily to create your results.

(5) Cross check your results (e.g. with information from the internet).

(6) Prepare a presentation where you describe the steps you made (including the code you wrote) and the results you got. Don't forget to list all data sources you used.

(7) Upload all documents (slides, Python code) to Ilias.

The group presentations will take place in the Pipool, early in December. There will be an email with appointment suggestions soon.

Possible Research Questions:

- Which countries have been most successful (w.r.t. the number of (gold) medals won by these countries)?
  - Are there any single countries (or all countries of a continent) who were mainly successful in selected sports (or events)?
  - Is there any correlation with the gross domestic product of the respective countries?
  - Are there any trends (countries who became very successful over the years in selected sports)?
  - Are there any differences w.r.t. to success between summer and winter games?
- Explore the gender distributions in different sports and overall. How was the (overall) number and trend changing over the years? Evaluate this question w.r.t. to different sports and countries. Find out the proportion of male / female participants for different sports and countries.
- Find records:
  - How old was the youngest/oldest athlete in all olympic games? Break this down by sport.
  - For each country, find out which is the sport they're strongest.
  - What is the record of the number of participations in olympic games for a single person? Who was the recordsman? Recordswoman?

  - Is there any event that was offered in all olympic games (either summer or winter games)? Are there any events which were only offered for a restricted period (e.g. in the first decades)?
  - Find which sport has the tallest/shortest, min/max weighted athletes

General Hints:

- Take into account that there are missing values for some columns (e.g. Height, Weight)
- Each athlete has a unique id which is given in every row (representing a single participation of the respective athlete). However, the team (nation) of the athlete may differ over the years.
- The team names do not always directly correspond to the country names. You might have to manually setup a table (dictionary) with mappings.
- In order to group countries by regions you will have to use additional data sets, e.g. country data that include the respective continents.

Links:

- GDP: `https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?name_desc=false`
- General Country Data: `https://www.kaggle.com/datasets/manusmitajha/countrydatacsv`
- Countries by continent: `https://worldpopulationreview.com/country-rankings/list-of-countries-by-conti`