



Checkpoint I: Project Proposal

Group: G04

Date: 2021/09/30

Domain

Domain description:

- Who doesn't like to watch the Olympic Games? Who doesn't like to support their country in the Olympics? Who doesn't like to have medalists from their country? Who doesn't feel honored when the participants from their country win? Who doesn't feel honored to be from the host country? Therefore, we decided it would be interesting to analyze the statistics and answer some of the most asked questions about the Olympics, its participants and how the context affects the outcome of the event.

Dataset

Dataset description:

- Our dataset consists on information about the Olympic Games (Summer and Winter) between 1896 and 2016, including the information about its participants.
- For our implementation of the visualization that will answer the concrete questions that we listed below we will be using two different datasets that together have all the information we need, as you can see in the examples we gave.
 - One dataset includes information about all athletes that have participated in the Olympics from the beginning until 2016.
This dataset includes attributes like "Name", "Sex", "Age", "Team", "Games", "Year", "Medal", etc.
 - The second dataset includes information about the host countries, such as the year, the date, the city, the type of games, etc...
- We will obtain this data through csv files that we found after an extensive search through different and reliable sources on the internet. These datasets were obtained free of charge. The URLs of the datasets we are going to use are the following ones:
 - https://www.kaggle.com/mysarahmadbhat/120-years-of-olympic-history/version/1?select=athlete_events.csv
 - https://www.kaggle.com/piterfm/olympic-games-hosts?select=olympic_hosts.csv

Example Questions

Questions that we are going to answer in our visualization:

- How does the number of medalists in the Olympic Games (Summer and Winter) vary according to the participants, per country?
- How does the gender of medalists in the Olympic Games (Summer and Winter) vary, per country?
- How being the host country of the Olympic Games affects the number of medalists in that country?

- How does the number of participants in the Olympic Games (Summer and Winter) evolve throughout the years?
- How does the number of women participants in the Olympic Games (Summer and Winter) evolve throughout the years?
- Which is the strongest sport of a country? (What is the sport that has more medalists per country)

Data Sample

Some examples of data that show that the above are possible and adequate.

Ex:

```
(from "athlete_events.csv")
ID;"Name";"Sex";"Age";"Height";"Weight";"Team";"NOC";"Games";"Year";"Season";"City"
;"Sport";"Event";"Medal"
```

```
1;"A Dijiang";"M";24;180;80;"China";"CHN";"1992
Summer";1992;"Summer";"Barcelona";"Basketball";"Basketball Men's Basketball";NA
```

```
2;"A Lamusi";"M";23;170;60;"China";"CHN";"2012
Summer";2012;"Summer";"London";"Judo";"Judo Men's Extra-Lightweight";NA
```

```
(from "host_per_event.csv")
Type;Country;Date;Athletes;Countries;City;Year

Summertime;Greece; 06 Apr - 15 Apr;241;14;Athens;1896

Summertime;France;14 May - 28 Oct;997;24;Paris;1900
```