

In this project, you will explore a dataset containing detailed information about the modern Olympic Games, spanning from Athens 1896 to Rio 2016. The dataset, which is available on Moodle or Kaggle, includes data on athletes, events, countries, and medal counts over this time period.

Your task is to select a specific aspect of this dataset to visualize, such as:

- The evolution of participating nations and athletes over time
- Medal distributions by country, gender, or sport
- Trends in the performance of specific countries or regions
- The impact of political events (e.g., boycotts) on participation or results
- Any other topic of your interest related to the Olympic Games

The goal is to present your findings and insights in a visual format using charts, graphs, or infographics. Your final deliverable will be an A3 poster (or larger, depending on your preference) that clearly communicates your chosen aspect of the Olympic Games.

## 1 The Dataset

The dataset can be found here on Kaggle and consists of two csv files. Here, a brief description is given:

- The file `athlete_events.csv` contains individual athletes competing in an individual Olympic event.

Feature	Description
ID	Unique ID for each athlete
Name	Athlete's name
Sex	F/M
Age	Integer
Height	in cm
Weight	in kg
Team	Team name
NOC	3-letter code
Games	Year and season
Year	Integer
Season	Summer or Winter
City	Host City
Sport	Type of Sport
Event	Event
Medal	Gold, Silver Bronze, or NaN

- The file `noc_regions.csv` contains the country names corresponding to the National Olympic Committee (NOC) codes.

Feature	Description
NOC	3-letter code
Region	Country Name
Notes	notes, string

## 2 The Task

After exploring the dataset, formulate a key question/topic for your final project. This topic should focus on a specific aspect of the dataset, e.g.

- Alpine Skiing Men's Giant Slalom Medalists over the years

- Athletes with the most participation in the Olympic Summer Games (Overall/Per Sport etc.)
- Medalists from your home country
- All-time medal table for your favorite sport
- Niche Sports with least participants
- One-Hit-Wonders (Sports that were only part of few Olympic Games)
- ...

Use different visualization options to show in-depth analyses of your topic. Besides “basic” plots, think of including maps for geographic distributions, show time series, and be sensitive about choices of colors, marker styles, legend, and labeling (see Data Visualization 101). The more interesting, the better!

**Note:** Your poster needs to include at least three different kind of plots. A simple bar- or line-plot is not sufficient.

### 3 Getting to know the Dataset and the EDA

So to get started, you should first get to know the dataset. To give you a head start, there are some predefined questions you can answer, to get an initial understanding of the dataset.

To start off, use the Jupyter Notebook provided in the elearning course. The notebook contains some initial questions and code snippets to get you started with the dataset.

**GeoPandas:** For Windows users, with pip the installation does not work out-of-the-box. See this link for a guide on how to install GeoPandas on Windows accordingly.

**Done with those questions?** Start working on your own questions! What is of interest to you? Try to find a question that you can answer with the dataset and that is interesting to you.

### 4 The Final “Presentation”

The final presentation will take place as a poster session on January, 28<sup>th</sup>.

Be prepared to present your **printed** poster in a 5-minute pitch. Before the presentation, the poster has to be submitted as a PDF in the e-learning course.

#### 4.1 Poster Requirements

- choose a suitable topic and title and include it in your poster
- add an abstract to motivate your project (between 800–1000 character)
- add a paragraph with your key findings (between 1000–1500 character)
- include at least 3 different kinds of plots (created using Python) in your poster. Each plot should have a title, legend, labels, and should be self-explanatory
- include your name and the course title (“Applied AI with Python”) on the poster
- make sure the poster is visually appealing
- Optional: link or QR code to further information (e.g., GitHub repo with code)

**What does a poster need and how to design one?**

- this is a cool, but short, guideline on how to make a good scientific poster
- this short video explains how a good research poster is structured

**What to use to design your poster?**

- Powerpoint
- LaTeX
- see the links above for more options

**Poster Format:** The poster needs to be printed at a minimal size of A3.

## 4.2 Grading

The grading of the project will be based on several components and a combination (25% each) of the following factors:

1. visual presentation using the poster
2. design and usage (variety) of plots, formal requirements of plots
3. the topic of the project and the story told
4. the short verbal presentation of the poster

The grades of the project (and the overall grade of the course) will be communicated in a timely manner.