

# Dataset

Find a dataset (or multiple) that you will explore. Assess the quality of the data it contains and how much preprocessing / data-cleaning it will require before tackling visualization. We recommend using a standard dataset as this course is not about scraping nor data processing.

<https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results>

# Problematic

Frame the general topic of your visualization and the main axis that you want to develop. • What am I trying to show with my visualization? • Think of an overview for the project, your motivation, and the target audience.

This project aims to analyze trends in Olympic performances over time, highlighting aspects such as the evolution of athlete physiques, the dominance of certain countries in specific sports, and the historical progression of medal distributions.

The Olympic Games have evolved significantly since their inception. This project seeks to visualize how countries, sports, and athletes have changed over time. Motivated by the growing interest in data-driven sports analytics, this project provides a historical perspective on the Olympics using interactive visualizations.

Target Audience: Sports enthusiasts interested in Olympic history. Data analysts and researchers exploring long-term trends in sports. General audiences who are curious about how sports and athletes have evolved. Journalists looking for historical insights on Olympic performances.

# Exploratory Data Analysis

Pre-processing of the data set you chose • Show some basic statistics and get insights about the data

Pre-processing of the dataset: ...

Basic Statistics & Insights:

- Total number of athletes: ~270,000.
- Average height and weight have increased over time in many sports.
- The USA, USSR, and China lead in overall medal counts.
- Some sports have become more inclusive (e.g., more female athletes over time).
- Certain countries dominate specific sports (e.g., Kenya in long-distance running, USA in swimming and track & field).

# Related Work

What others have already done with the data? • Why is your approach original? • What source of inspiration do you take? Visualizations that you found on other websites or magazines (might be unrelated to your data). • ~~In case you are using a dataset that you have already explored in another context (ML or ADA course, semester project...), you are required to share the report of that work to outline the differences with the submission for this class.~~

## Related Work Existing Visualizations & Studies:

- Several analyses on Kaggle explore medal distributions, athlete demographics, evolution of the Olympics overtime (Variation of age along time, variation of weight along time, ...). While they provide good insights on the dataset with static basic graphs, they lack the dynamic in their visualization.
- The official Olympic website provides basic medal tables but lacks cool visual analytics.
- News agencies like The New York Times and The Guardian have produced Olympic-related infographics, often focusing on recent Games rather than long-term trends.

## Originality of This Approach:

- Most existing visualizations focus on static tables, whereas this project will include interactive elements.
- Combining athlete physique trends with country performance offers a unique perspective.
- Analyzing how the Olympics have changed from 1896 to 2016 provides deeper historical insights compared to event-specific reports.

## Inspiration Sources:

- The "PyeongChang 2018 Medal Map": <https://coolmaps.esri.com/Olympics/2018/>
- Previous Kaggle projects on Olympic data
- City leaders survey: <https://cityleaderssurvey.urbis.com.au/>
- Behalf studio: <https://onbehalfof.studio/>