# Olympics Dataset

The '*dataset on the modern Olympic Games*' comprises all the Games from Athens 1986 to Rio 2016. The Olympics is more than just a quadrennial multi-sport world championship. It is a lens through which to understand global history, including shifting geopolitical power dynamics, women's empowerment, and the evolving values of society.

In this analysis, our goal is to shed light on major patterns in Olympic history. How many athletes, sports, and nations are there? Where do most athletes come from? Who wins medals? What are the characteristic of the athletes (e.g., gender and physical size)?

## Understanding data


Olympics data contains 271,116 rows and 17 columns with details of athletes, events and medals won.

Each row corresponds to an individual athlete competing in an individual Olympic event (athlete-events). The columns are: `ID - Unique number for each athlete

**Name** - Athlete's name

**Sex** - M or F

**Age** - Integer

**Height** - In centimetres

**Weight** - In kilograms

**Team** - Team name

**NOC** - National Olympic Committee 3-letter code

**Games** - Year and season

**Year** - Integer

**Season** - Summer or Winter

**City** - Host city

**Sport** - Sport

**Event** - Event

**Medal** - Gold, Silver, Bronze, or NA`

**Region** - Country that has won the title

**Notes** - Special Notes related to that event.

## Key Insights

In this exercise, we are going to explore the Modern Olympic dataset. The exercise mainly focuses on the analytics part, visualizing, trying to find out the most dominating countries for the past 120 yrs, etc.

Our main goal at the end of this exercise is to find out the answer to this question:

**Does hosting the Olympics improve performance on the medals table?**

Before proceeding further, let's see what EDA means.