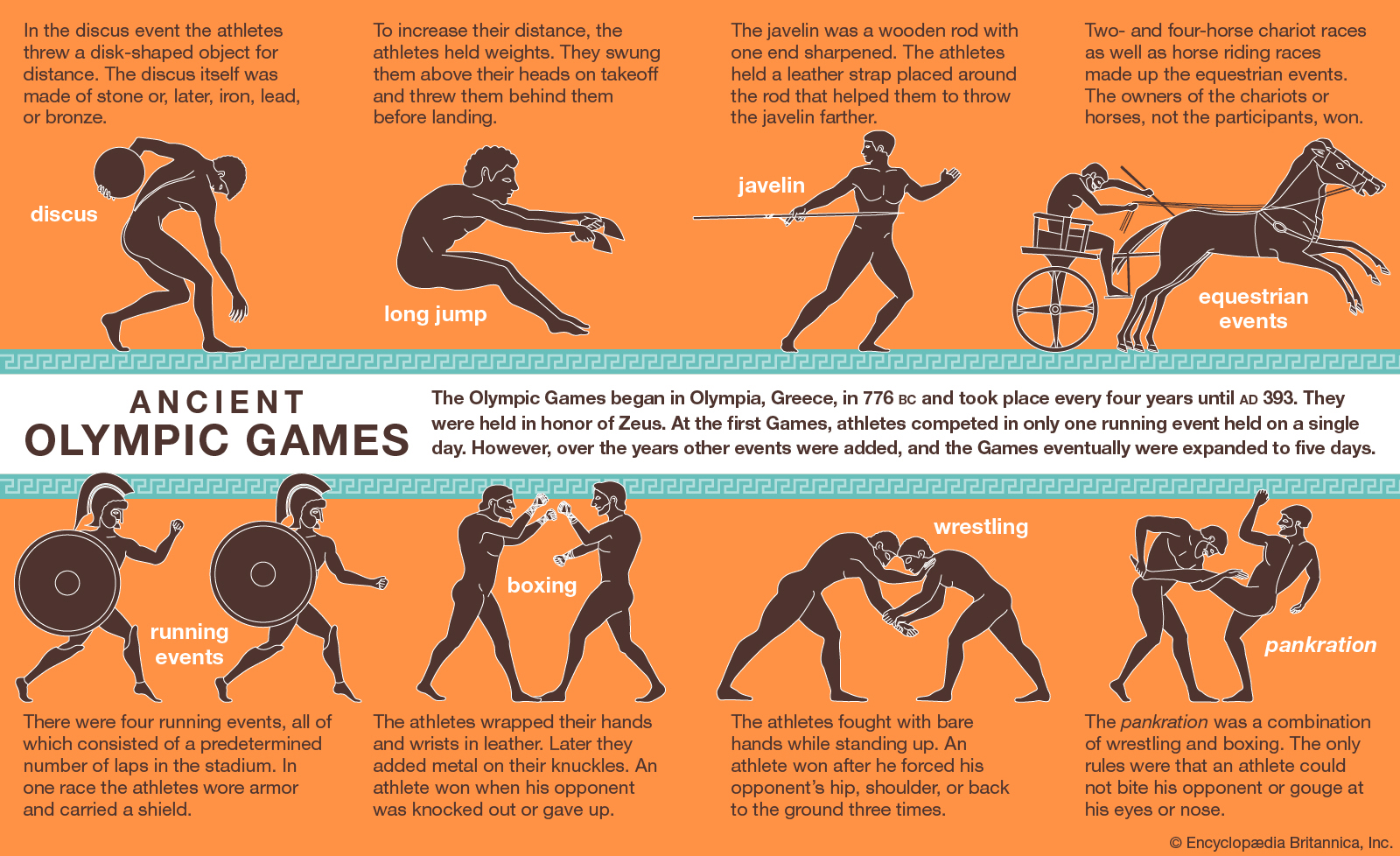
# 

Olympic Participants Analysis 1896 - 2016



# **BLACK PANTHERS: GROUP MEMBERS**

Simon Mmari

Christine Kandeo

Ted Kimani

Stacy Keago

Peter Kirimi

This is the project of Black Panthers Group

In this project, we are required to analyse data with **CRISP-DM** process. The CRISP-DM process is below.

The links to Resource Finding for this project are:

1.Github[(Link)](https://github.com/Jerome-debug/Black_Panthers_Module1_Project)

2.Google Drive[(Link)](https://drive.google.com/drive/u/0/folders/1klJyVUQQgjqBOvYUmV8DL2XsfcoSheKC)

3.Jira Board([Link](https://kandeo.atlassian.net/jira/software/projects/BPOD/boards/5))

4.Tableu([Link](https://public.tableau.com/views/PresentationVisualsforParticipants/TotalNumberoffemaleparticpantspercountry?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link))

#### CRISP-DM (Cross-Industry Standard Process for Data Mining)

* Business Understanding
* Data Understanding
* Data Preparation
* Modeling
* Evaluation
* Deployment

## **1.0 Business Understanding**

Equal gender representation among athletes and leaders has been a challenge for the longest time.

We venture into this research to show women's representation in the Worldwide Olympics over the years.Starting from when the first modern Olympics was held in 1896 in Athens, Greece, where women were deliberately barred from taking part.

**Introduction**

From sexualisation and policies on the testosterone levels and women having to fight to bring their breastfed children to the pandemic-restricted Games, analysts say that discrimination remains rife.

The idea of gender equality actually conceals the fact there is so much more to be done.Yes these inferior groups are included in the pitch and arena but their experiences are very concerning. Much of this has to do with a history of women’s exclusion in sport.

**When the first modern Olympics was held in 1896 in Athens, Greece, women were deliberately barred from taking part.**

At the time, the founder of the IOC, Baron Pierre de Coubertin, argued that an Olympics with women would be “impractical, uninteresting, *unaesthetic* and indecent”.

The Games, he said, were created for “the solemn and periodic exaltation of male athleticism” with “female applause as a reward”.Straight away, you can see the kind of exclusion that women were up against as part of the Olympic movement.If they were allowed in at all, it was to applaud the male athletes who were actually taking part.”

## **Overview**

***Stereotypes***

Women have fought against gender inequality over the years. Under pressure from athletes such as **French rower Alice Milliat** who even launched a separate Olympics for women, the IOC began including more and more female events.

Still, for years, women were “confined to more aesthetic events” or “even play and dance routines” such as swimming, figure skating and fencing. The idea around this was that it was more suitable for female biology and less threatening to dominant images of femininity around the time.

“Women weren’t expected to run too far because they might sweat and we don’t want them sweating. They might not throw things as far because we don’t want them to damage their internal organs.” Over time, the IOC did cede ground to women athletes – albeit reluctantly.

It was only in 2012 that the global sporting body allowed women to compete in all sports on the Olympic programme and it was only in 2014 that it committed to gender parity at the Games. This year, women make up 48.8 percent of the 11,000 Olympians, up from 45.6 percent in 2016.

## Objectives

In this research we focus on Gender inequality as our problem statement in the Olympic games over the years.

**Brief Overview**

Sexist ideals also plague the IOC’s regulations on testosterone set by World Athletics, the rules say female athletes who are intersex or have differences in sex development need to artificially reduce their testosterone levels to below 5 nanomoles per litre if they want to take part in the middle-distance running events of 400m to 1,500m**.**

This has resulted in the disqualification of several women. Two Namibian runners – Christine Mboma and Beatrice Masilingi – were pulled from the 400m race, despite being considered medal contenders.

South Africa’s Caster Semenya, Burundi’s Francine Niyonsaba and Kenya’s Margaret Wambui were also barred from competing in the 800m race – the three had swept the event in the 2016 Summer Games in Rio de Janeiro, with Semenya winning the gold, and Niyonsaba and Wambu taking the silver and bronze, respectively.

Professors from the University of Notre Dame in the United States, say the testosterone rules are “problematic”, “sexist” and “based on questionable science”. They imply testosterone, a primarily male hormone, is “the all-powerful secret sauce of athletic performance”, she says. “And that’s just not true. There are so many other variables – including genetics, hormones, training and nutrition, and even just how the athlete feels on the morning of an event.”

Research conducted [found](https://www.chi.ac.uk/news/women-still-overlooked-top-sport-jobs-new-study-finds) that women make up just 22 percent of executive boards in international sporting organisations and only seven percent of president or chair roles.

Meanwhile, only 10 percent of accredited coaches at the Summer and Winter Olympics over the past decade have been women. In terms of equal gender representation for athletes at the Olympics, there’s been huge progress made . But we still have a long way to go.

## Business Success criteria

We are going to analyze the female participation in different sports in the Olympics.

Greater gender equality (measured using the Gender Inequality Index) is associated with higher participation and medal counts in the Olympic Games from 1996 through the years.

## Assessing the Situation

1. **Resource Inventory**

This is a historical dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016. I scraped this data from [www.sports-reference.com](http://www.sports-reference.com/) in May 2018.

Note that the Winter and Summer Games were held in the same year up until 1992. After that, they staggered them such that Winter Games occur on a four year cycle starting with 1994, then Summer in 1996, then Winter in 1998, and so on.

A common mistake people make when analyzing this data is to assume that the Summer and Winter Games have always been staggered.

### **Content**

The file athlete\_events.csv contains 271116 rows and 15 columns. 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 centimeters
* 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

### Acknowledgements : The Olympic data on [www.sports-reference.com](http://www.sports-reference.com/) is the result of an incredible amount of research by a group of Olympic history enthusiasts and self-proclaimed 'statistorians'. Check out their [blog](http://olympstats.com/) for more information

**ASSUMPTIONS AND CONSTRAINTS**

There were no constraints in the data since it was derived from the official Olympics page, and we assumed the data was accurately recorded.

**DATA MINING GOALS**

Our main goal is to portray Gender inequality in the Olympic Games over the years.

This project details the concepts and steps of data mining and exploratory data analysis done to produce insights and recommendations. This is done by performing univariate and bivariate analysis, and by creating visualizations to better portray our findings. The focus will be on the gender dispersion per country, sport, season and other variables that may be of interest.

A generic process is introduced, which outlines data mining goals and techniques, supported by example scenarios.

Various applications of manufacturing environments are shown in which data mining has been applied successfully, and potential areas in which the outlined mechanisms are capable of being applied.

#### **DATASETS USED**

* + [athlete\_events.csv](https://github.com/cstorm125/information_value/blob/master/data/120-years-of-olympic-history-athletes-and-results/athlete_events.csv)
  + Noc\_regions.csv

# 2.0 DATA DESCRIPTION / UNDERSTANDING

We have two datasets available for this project showing Olympic games over the years.

1. **athlete\_events.csv** - This is a historical dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016
2. **noc\_regions.csv** -This dataset shows the Regions and their NOC(A **National Olympic Committee** (NOC) is a national constituent of the worldwide Olympic movement.

### 

### **Verifying Data Quality**

The Data Provided was correct and up to date

## **3.0 DATA PREPARATION**

These are the steps followed in preparing the data

#### **Loading Data**

* Loaded the datasets from the CSV and then created a python notebook using google colab
* We joined the two datasets for easy analysis.

#### **Cleaning Data**

* Replace Null values with the median value for efficiency.
* Dropping the rows with missing values in Regions
* Dropped irrelevant data columns to our research

**Modelling**

The modeling techniques employed are driven by the company’s data mining goal,thus using python to answer our data mining goal. Use historical information about previous services offered to generate a model that is highly understandable. At its simplest, this involves clustering the olympics participation and grouping it by Gender.

The terms and columns associated with this research include:

* ID - Unique number for each athlete
* Name - Athlete's name
* Sex - M or F
* Age - Integer
* Height - In centimeters
* 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

**Evaluation**

* The study produced what are hoped to be better recommendations and an improved insight on strongholds and weak points.
* The evidence that there is gender inequality in the olympic games.However at the same time,we have witnessed great improvements to bring gender equality as per the required percentage.
* Now we are fairly confident of both the accuracy and relevancy of the project results and so is continuing to the deployment phase.

**Deployment**

In terms of equal gender representation for athletes at the Olympics, there’s been huge progress made.But we still have a long ways to go

**Data Insights**

* The more participants a country has going to the Olympics(on average) the more medals they can expect to win at the games
* The country with the tallest competitor on record was china with the shortest coming from Bangladesh
* An above average height and weight for participants has a positive correlation with winning more medals with Gold winners having the highest values for all medal winners
* Basketball, volleyball, water polo have the highest average heights of participants with Gymnastic, Trampolining and diving having the lowest average heights
* The oldest participants were mostly playing Roque(bougie croquet) at ~53 years, while the youngest were in rhythmic gymnastics at ~18 years
* The heaviest participants were in Tug-of-War, Bobsleigh & Basketball, with the featherweights in Rhythmic Gymnastics, Synchronized swimming & Gymnastics

**Conclusions**

* The number of athletes, events, and nations has grown dramatically since 1896, but growth leveled off around 2000 for the Summer Games.
* Geographic representation in the Games has grown since 1896, although Africa, Southeast Asia, the Middle East, and South America are still very under-represented.
* Female participation increased dramatically, and this trend started during the Cold War.
* The size of Olympians have become more extreme over time. In most sports this means taller and heavier, but in a few sports such as gymnastics, athletes have become smaller.
* More sporting activities have been added to the Olympics over the years.