

OLYMPIC DATA ANALYSIS

Priyanka Awatramani

D17A/05

Sneha Lalwani

D17A/39

Sartha Tambe

D17A/66

Description of Problem Statement:

The 'modern Olympics' comprises all the Games from Athens 1896 to Rio 2016. For years, Olympics have been held around the world, and participants from around the world compete to succeed in different sports. We wanted to work with this data because we want to look for answers to the questions about which countries are superior in which branches, which countries are leading women or men in which sports and bringing medals to their country. This dataset provides an opportunity to ask questions about how the Olympics have evolved over time, including questions about the participation and performance of women, different nations, and different sports and events.

Motivation:

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.

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 characteristics of the athletes (e.g: gender and physical size)?

Dataset:

This is a historical dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016. The athlete events dataset contains information about Olympic Athletes and events that they have competed in, including biological data (Age, Sex, Height, Weight etc.) and event data (Year, Season, City,

Sport etc.). The file athlete_events.csv contains 271116 rows and 15 columns. Each row corresponds to an individual athlete competing in an individual Olympic event.

```
"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
"3","Gunnar Nielsen Aaby","M",24,NA,NA,"Denmark","DEN","1920 Summer",1920,"Summer","Antwerpen","Football","Football Men's Football",NA
"4","Edgar Lindenau Aaby","M",34,NA,NA,"Denmark/Sweden","DEN","1900 Summer",1900,"Summer","Paris","Tug-Of-War","Tug-Of-War Men's Tug-Of-War","Gold"
"5","Christine Jacoba Aaftink","F",21,185,82,"Netherlands","NED","1988 Winter",1988,"Winter","Calgary","Speed Skating","Speed Skating Women's 500 metres",NA
"5","Christine Jacoba Aaftink","F",21,185,82,"Netherlands","NED","1988 Winter",1988,"Winter","Calgary","Speed Skating","Speed Skating Women's 1,000 metres",NA
"5","Christine Jacoba Aaftink","F",25,185,82,"Netherlands","NED","1992 Winter",1992,"Winter","Albertville","Speed Skating","Speed Skating Women's 500 metres",NA
"5","Christine Jacoba Aaftink","F",25,185,82,"Netherlands","NED","1992 Winter",1992,"Winter","Albertville","Speed Skating","Speed Skating Women's 1,000 metres",NA
"5","Christine Jacoba Aaftink","F",27,185,82,"Netherlands","NED","1994 Winter",1994,"Winter","Lillehammer","Speed Skating","Speed Skating Women's 500 metres",NA
"5","Christine Jacoba Aaftink","F",27,185,82,"Netherlands","NED","1994 Winter",1994,"Winter","Lillehammer","Speed Skating","Speed Skating Women's 1,000 metres",NA
"6","Per Knut Aaland","M",31,188,75,"United States","USA","1992 Winter",1992,"Winter","Albertville","Cross Country Skiing","Cross Country Skiing Men's 10 kilometres",NA
"6","Per Knut Aaland","M",31,188,75,"United States","USA","1992 Winter",1992,"Winter","Albertville","Cross Country Skiing","Cross Country Skiing Men's 50 kilometres",NA
"6","Per Knut Aaland","M",31,188,75,"United States","USA","1992 Winter",1992,"Winter","Albertville","Cross Country Skiing","Cross Country Skiing Men's 10/15 kilometres Pursuit",NA
"6","Per Knut Aaland","M",31,188,75,"United States","USA","1992 Winter",1992,"Winter","Albertville","Cross Country Skiing","Cross Country Skiing Men's 4 x 10 kilometres Relay",NA
"6","Per Knut Aaland","M",33,188,75,"United States","USA","1994 Winter",1994,"Winter","Lillehammer","Cross Country Skiing","Cross Country Skiing Men's 10 kilometres",NA
"6","Per Knut Aaland","M",33,188,75,"United States","USA","1994 Winter",1994,"Winter","Lillehammer","Cross Country Skiing","Cross Country Skiing Men's 30 kilometres",NA
"6","Per Knut Aaland","M",33,188,75,"United States","USA","1994 Winter",1994,"Winter","Lillehammer","Cross Country Skiing","Cross Country Skiing Men's 10/15 kilometres Pursuit",NA
"6","Per Knut Aaland","M",33,188,75,"United States","USA","1994 Winter",1994,"Winter","Lillehammer","Cross Country Skiing","Cross Country Skiing Men's 4 x 10 kilometres Relay",NA
"7","John Aalberg","M",31,183,72,"United States","USA","1992 Winter",1992,"Winter","Albertville","Cross Country Skiing","Cross Country Skiing Men's 10 kilometres",NA
"7","John Aalberg","M",31,183,72,"United States","USA","1992 Winter",1992,"Winter","Albertville","Cross Country Skiing","Cross Country Skiing Men's 50 kilometres",NA
"7","John Aalberg","M",31,183,72,"United States","USA","1992 Winter",1992,"Winter","Albertville","Cross Country Skiing","Cross Country Skiing Men's 10/15 kilometres Pursuit",NA
"7","John Aalberg","M",31,183,72,"United States","USA","1992 Winter",1992,"Winter","Albertville","Cross Country Skiing","Cross Country Skiing Men's 4 x 10 kilometres Relay",NA
```

Dataset Link:

https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results#athlete_events.csv

Analysis:

Analysis performed in Hive:

- Performance of different countries in a specific event or sport
- Performance of a specific country over the years
- Comparison of how well countries played in a particular year
- Measuring the participation in Winter / Summer Olympics

Analysis performed in Impala:

- The 10 countries with the highest number of Olympic medals in 120 years
- First gold of various sports in the history of Chinese Olympics
- Participation of female athletes in previous Olympic Games
- Comparison of age of athletes distributed over the years