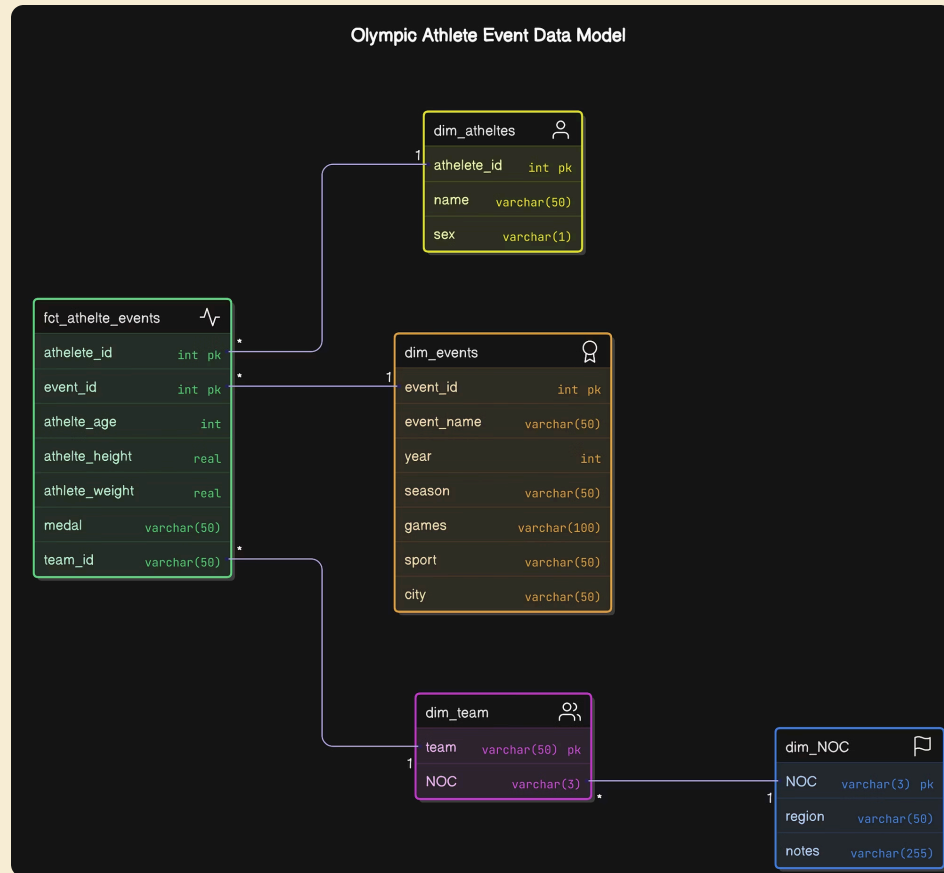# Olympic Athlete Events Data Project

This project analyzes a historical dataset covering 120 years of Olympic history, from Athens 1896 to Rio 2016. The data was scraped from sports-reference.com and aims to sharpen skills in data cleaning, manipulation, modeling, visualization, and documentation. The project is part of the IEEE data engineering roadmap at Mansoura University.

AW **by Abdulrahman Wael**

Database Schema Overview

Olympic Athlete Event Data Model

dim_atheltes
athelete_id	int pk
name	varchar(50)
sex	varchar(1)

fct_athelte_events
athelete_id	int pk
event_id	int pk
athelte_age	int
athelte_height	real
athlete_weight	real
medal	varchar(50)
team_id	varchar(50)

dim_events
event_id	int pk
event_name	varchar(50)
year	int
season	varchar(50)
games	varchar(100)
sport	varchar(50)
city	varchar(50)

dim_team
| team | varchar(50) pk |
| NOC | varchar(3) |

dim_NOC
NOC	varchar(3) pk
region	varchar(50)
notes	varchar(255)

Design Notes

1. "dim" denotes a dimension table, and "fct" denotes a fact table.

2. The attribute (NOC) has been assigned to the team entity itself.

3. The columns for Age, Height, and Weight were excluded from the Athlete dimension table as these attributes vary with each record; thus, it is more convenient to handle them as measurements.

4. Measurements in the fact table include athlete_age, athlete_height, athlete_weight, and medal; other columns function as foreign keys referencing the dimension tables.

5. Duplicate entries within each dimension table have been eliminated.

Analytical Questions Using SQL

Top Gold Medal Country (Winter Olympics)

Which country won the most gold medals in Winter Games?

Athlete Height Distribution

What is the distribution of athlete heights across different sports?

Top Medalists

Who are the top 10 athletes with the most total medals?

Host Cities

Which cities have hosted both Summer and Winter Olympics?

Participation Trends

How has athlete participation changed over the years?

Analytical Questions Using Python (Pandas)

Silver Medals in Summer Olympics

Which country won the most silver medals in Summer Games?

Gender Distribution Over Time

How has the gender distribution of athletes evolved across Olympic Games?

Sport Participation Changes

Which sports have seen the largest increases or decreases in participation over time?

Age Data Analysis

What are the age statistics for athletes across all sports?

Data Completeness Over Time

How complete is the age, weight, and height data historically, and are there biases or gaps?

Key Findings from the Analysis

Top Medal Countries and Athletes

Canada leads Winter Olympics gold medals with 289. Top athletes include Heikki Ilmari Savolainen (39 medals) and Joseph Stoffel (38).

Host Cities & Participation

42 cities have hosted both Summer and Winter Games. The peak participation year was 1992 with 16,413 athletes.
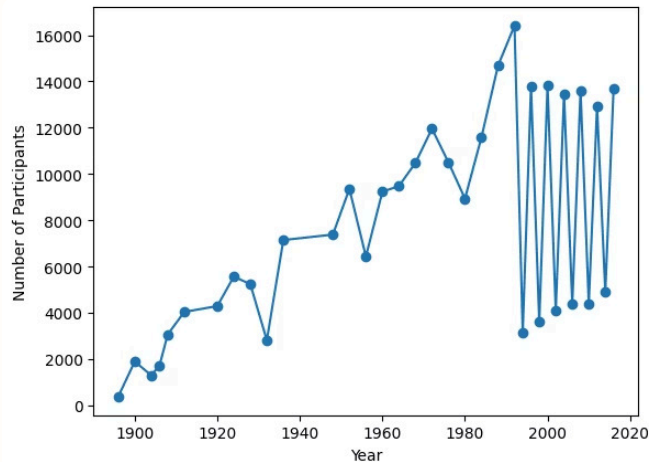
Trends and Data Quality

USA has most Summer silver medals (1333). Male athletes dominate but female participation is rising. Gymnastics shows fluctuating participation. Data completeness improved notably from the 1950s to 1970s.
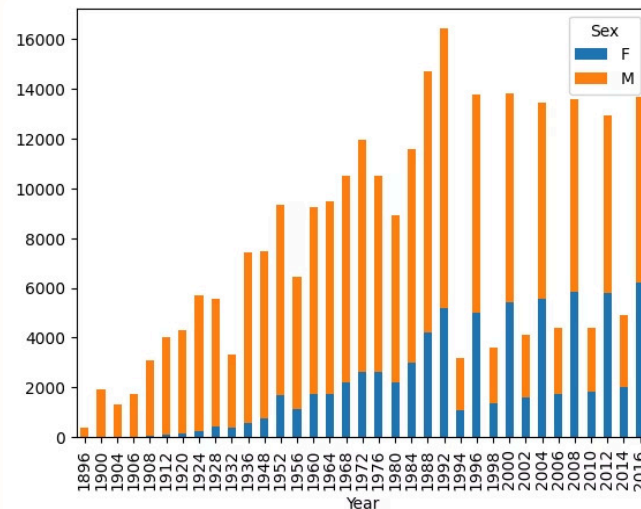
Participation Trends and Data Completeness Over Time

This set of graphs illustrates key aspects of Olympic athlete data over the years.

- The overall participation trend shows the total number of athletes competing across different years.

- The trend breakdown by sex highlights the evolving gender distribution among athletes during the Olympic Games.

- Data completeness over time reveals how the availability of critical athlete information such as age, height, and weight has improved historically.

Lessons Learned and Future Work

Skills sharpened

1. data cleaning

2. data manipulation

3. data modeling and database design

4. data visualization

5. project documentation and presentation

Future Enhancements

Plans include adding interactive dashboards with Tableau or Plotly, expanding time-series and career progression analyses, getting more data using web scrapping techniques from the source website, and building machine learning models.

Special thanks to my leader: Mohamed Wael.