# General Sir John Kotelawala Defence University

**Faculty of Management, Social Science and Humanities Department of Languages**

**BSc. in Applied Data Science Communication**

## Advanced SQL and Cloud Databases
## Assignment 01

**D/ADC/23/0012 - S.A.R Maleesha**

**D/ADC/23/0032 - W.A.K.N Wedagedara**

**D/ADC/23/0033 - K.G.K Chani**

**D/ADC/23/0039 - H.W Kaweesha**

# Task 01

CONTENT

# 01.Introduction of Dataset

Source: http://www.sports-reference.com

The dataset contains historical data on athletes who participated in the modern Olympic games from Athens (1896) to Rio (2016).In this Assignment, We used historical data from the modern summer Olympics, from the Sydney 2000 Games to the Rio 2016 Games.

The dataset consists of 271,116 rows and 15 columns. Each row represents a unique record of an athlete participating in a specific Olympic event. The columns include:

**Dataset Overview**

- **ID**: A unique identifier for each athlete.

- **Name**: The full name of the athlete.

- **Sex**: Gender of the athlete (M for male, F for female).

- **Age**: The age of the athlete at the time of the Olympics.

- **Height**: Height of the athlete (in centimeters).

- **Weight**: Weight of the athlete (in kilograms).

- **Team**: The country or team the athlete represented.

- **NOC**: National Olympic Committee code (three-letter country code).

- **Games**: The year and season of the Olympic Games (e.g., "2000 Summer").

- **Year**: The year of the Olympic Games.

- **Season**: Whether the event was in the Summer or Winter Olympics.

- **City**: The host city of the Olympics.

- **Sport**: The sport in which the athlete competed.

- **Event**: The specific event or discipline the athlete participated in.

- **Medal**: The medal won by the athlete (Gold, Silver, Bronze, or Unplaced).

## 2. Data Preprocessing and Modeling

Firstly, We imported the data set to Power BI and then transformed the dataset. We did the following steps to clean and process the data set

1. Replace NA values As "Unknown"(There were NA values in height, age, and weight columns)
2. Removing duplicates
3. There were NA values in the medal column also and we replaced those NA with "Unplaced"

After doing all these cleaning and preparation data set can be seen as follows,
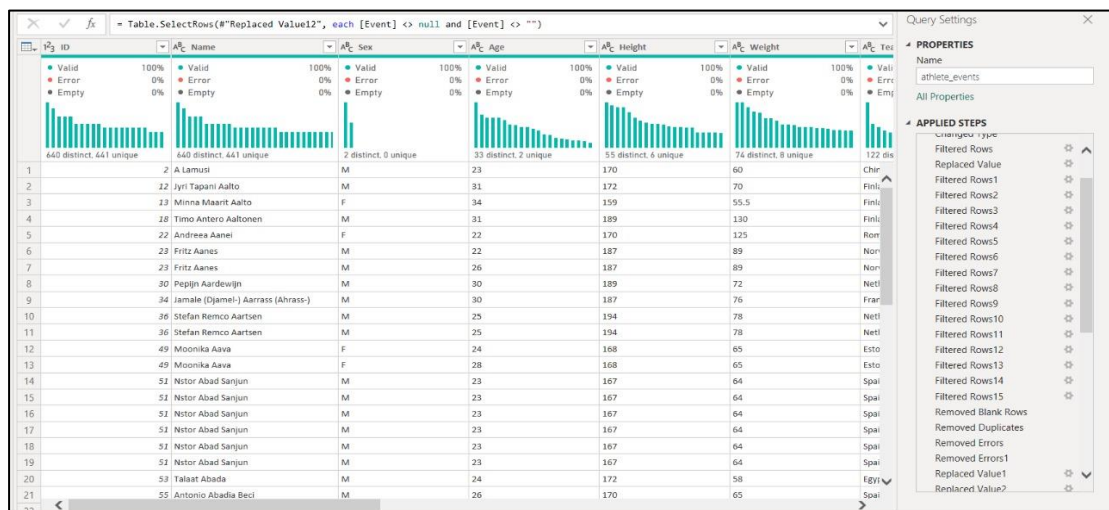


### 2.1 Data Modeling and Relationships:

After importing the dataset into Power BI, the next crucial step was to create relationships between the tables to enable meaningful analysis. The relationships were established as follows:

**1. Athlete Table:**

Contains detailed information about athletes, including their names, genders, dates of birth, and the specific events they participated in. Each athlete can participate in multiple events, but each event entry is linked to only one athlete. This relationship allows for a detailed analysis of how individual athletes performed across different events. It enables insights such as how many events an athlete participated in and how many medals they won in those events.

- Key Columns: Athlete ID (Primary Key), Name, Gender, Age,Height,Weight.

```
1  Athlete = DISTINCT(SELECTCOLUMNS(
2      'athlete_events_01',
3      "ID", [ID],
4      "Name", [Name],
5      "Sex", [Sex],
6      "Age", [Age],
7      "Height", [Height],
8      "Weight", [Weight]
9  ))
```

2. **Event Table:**

   Includes data on the events held during the Olympics, such as the type of event, year, location, and the athletes who participated. This relationship is crucial because it allows for complex analysis, such as determining how different countries performed in specific events across multiple Olympic Games

   o Key Columns: Event ID (Primary Key), Year, Sport, Medal, Athlete ID.

   o
   ```
   1  Event = DISTINCT(SELECTCOLUMNS(
   2      'athlete_events_01',
   3      "Sport", [Sport],
   4      "Event", [Event],
   5      "Games", [Games],
   6      "Year", [Year],
   7      "Season", [Season],
   8      "City", [City]
   9  ))
   10
   ```

3. **Team Table:**

   Captures data about the teams or countries participating in the Olympics, including country codes (NOCs) and team names. Each team or country can have multiple athletes, but each athlete is linked to only one team. This structure is crucial for analyzing country-specific performance, including total medals won and individual athlete achievements within the same country.

   o Key Columns: Team ID (Primary Key), NOC, Team Name,

   ```
   1  Team = DISTINCT(SELECTCOLUMNS(
   2      'athlete_events_01',
   3      "Team", [Team],
   4      "NOC", [NOC]
   5  ))
   6
   ```

**DAX Measures**:

- **Total Medals**: A DAX measure was created to calculate the total number of medals won by each athlete or team.

- **Medal Breakdown**: Separate DAX measures were developed to count gold, silver, and bronze medals, facilitating detailed analysis by medal type.

   ```
   1  Total Medals = COUNTROWS('athlete_events_01')
   ```
   ```
   1  Total_Gold_Medals = CALCULATE(COUNTROWS('athlete_events_01'), 'athlete_events_01'[Medal] = "Gold")
   ```
   ```
   1  Total_Silver_Medals = CALCULATE(COUNTROWS('athlete_events_01'), 'athlete_events_01'[Medal] = "Silver")
   ```

# 3. Dashboard Overview

The Power BI dashboard was divided into four main pages to provide a structured and comprehensive view of the Olympic data:

1. **Medal Count Overview**
2. **Participation Statistics**
3. **Trends and Patterns**
4. **Top 10 Categories**

## 3.1. Medal Count Overview

**Pie Chart - Medal Count by Year**: Shows the distribution of medals across the five Olympic Games, using tooltips for detailed insights.
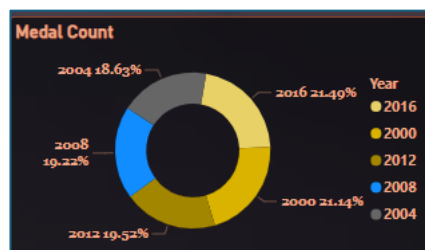


**Bar Chart - Medal Count by Athlete Name**: Highlights the top athletes by the total number of medals won, with tooltips providing a medal breakdown.
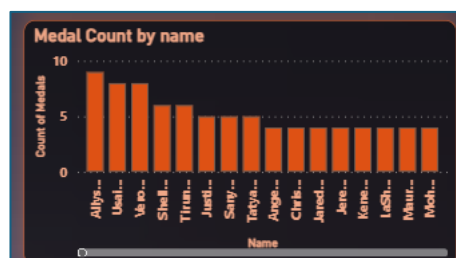


**Stacked Bar Chart - Medal Count by Team**: Displays the distribution of gold, silver, and bronze medals by country, with detailed tooltips.
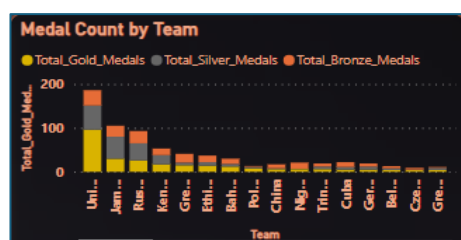
**Line Chart - Total Medals by NOC**: Tracks the performance of countries in terms of total medals won over time, with tooltips for year-by-year analysis.
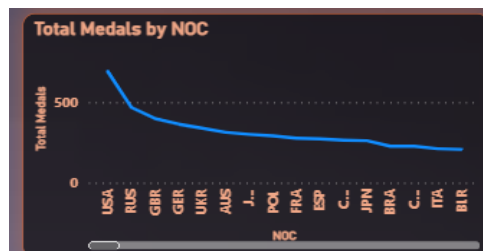


This line chart shows that most medals are won by athletes aged 20 to 30, peaking in the mid-20s, with a decline in medals beyond age 30.
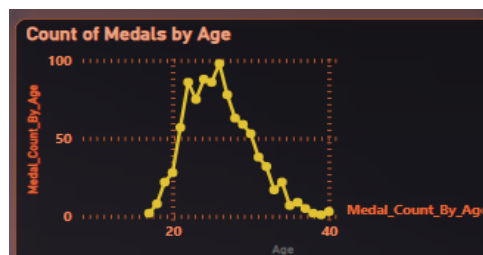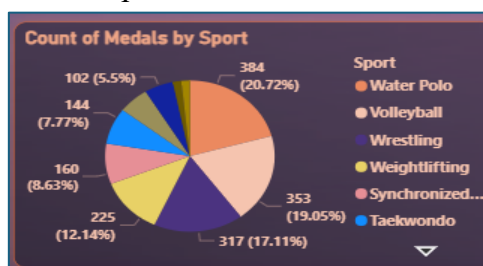


### 3.2 Participation Statistics

**Map - Participants by Country**: Provides a geographical representation of athlete participation, enhanced with tooltips.



**Pie Chart - Count of Medals by Sport**: Shows the distribution of medals across different sports, with detailed tooltips for each sport.

**Line Chart - Medal Count by Age**: Analyzes how the age of athletes relates to their performance, using tooltips to reveal trends in medal-winning ages.
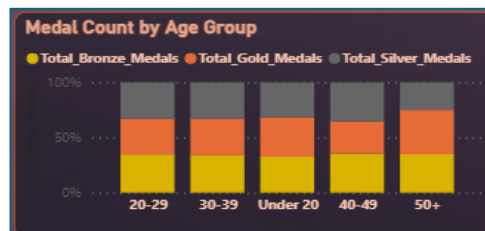


**This pie chart** shows athlete **participation across five Summer Olympic Games**, with each Games having around 11,000 athletes. The 2016 Games had the highest participation at 20.78%.
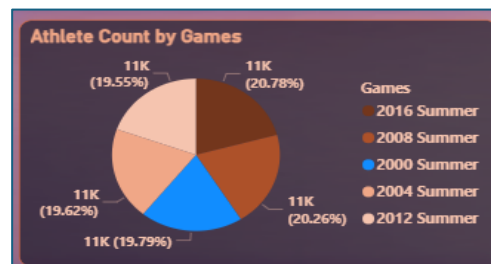


**The line and column chart**, This chart shows a stable count of participating countries (NOCs) over the years, with a fluctuating athlete count, peaking in 2016.
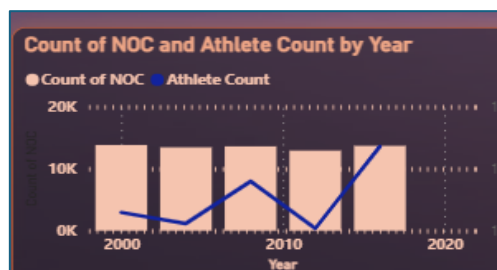


### 3.3. Trends and Patterns

**Participation vs. Medals Over Time:** The combined bar and line chart illustrates how athlete participation correlates with medal counts over different Olympic Games. Tooltips reveal specific athlete and medal counts for each year, enabling a detailed examination of trends.
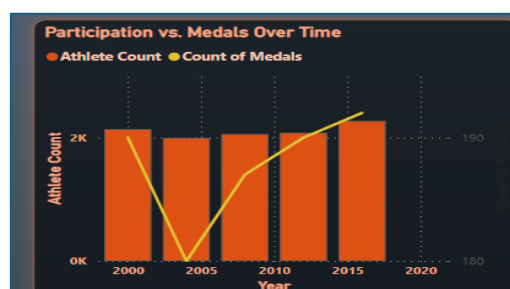
**Medal Wins by Country and Sport:** This horizontal bar chart compares the medal counts of various countries across sports, with color-coded bars representing different Olympic years. Tooltips provide additional context by displaying the exact number of medals won per sport and country for each year.
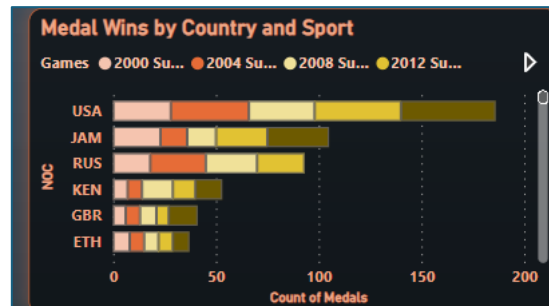


**Medal Distribution by Age Group:** The pie chart shows how medals are distributed among different age groups, with the majority won by athletes aged 20-29. Tooltips provide detailed medal counts for each age group, offering insights into which age groups are most successful.
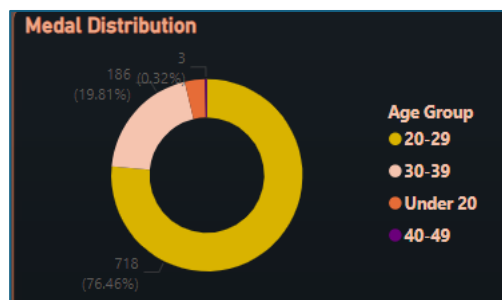


**Medal Trends by Gender Over Time:** This horizontal bar chart highlights the gender distribution of Olympic medals across different years. Tooltips offer a breakdown of medal counts by gender for each year, allowing users to explore how male and female athletes have performed over time.
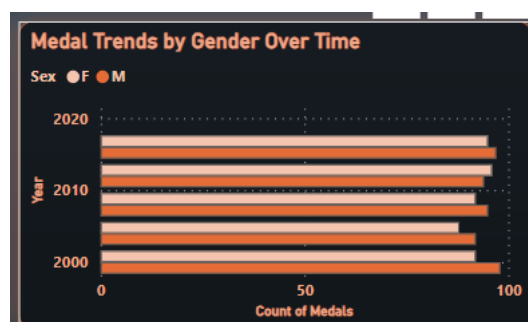


### 3.4 Top 10 category

**Top 10 Countries by Total Medals:** This bar chart ranks the top 10 countries by total medals, with the USA leading.
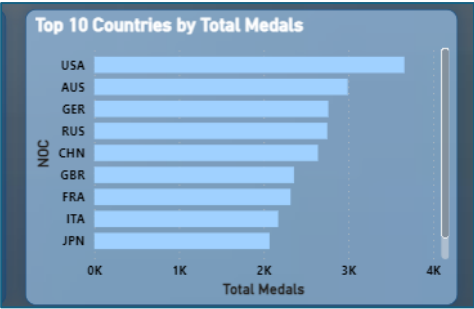
**Top 10 Most Participated Events:** This chart shows the top 10 events with the highest athlete participation, led by Football (Men's).



**Top Medal Winners:** The table lists top athletes by total medals, with Michael Phelps leading with 30 medals.

| Name | Total Medals | Total_Bronze_Medals | Total_ |
|------|------|------|------|
| Michael Fred Phelps, II | 30 | 2 | |
| Fabian Hambchen | 26 | 1 | |
| Yang Wei | 26 | | |
| Alberto Busnari | 24 | | |
| Daniele Matias Hyplito | 24 | | |
| Marian Drgulescu | 24 | 2 | |
| Dimitri Karbanenko | 23 | | |
| Enrico Pozzo | 22 | | |
| **Total** | **263** | **6** | |

# 4. Conclusion

Recommendations for Further Analysis

This report provides a comprehensive analysis of the modern Summer Olympic Games, focusing on the five Games from Sydney 2000 to Rio 2016. Through various visualizations, we have explored key aspects such as the distribution of medals, participation trends, and the performance of top athletes and countries.

**The Medal Count Overview** highlighted the countries and athletes with the most significant achievements, with the USA leading in total medals and Michael Phelps standing out as the

top individual medalist. **The Participation Statistics** section shed light on the events with the highest athlete participation, revealing Football (Men's) as the most popular event.

The analysis also identified patterns and trends over time, such as the consistency in top-performing countries and the dominance of certain sports by specific nations. The visualizations provided insights into the age distribution of medal winners and the gender-based differences in performance.

**Recommendations for further analysis** include a deeper dive into gender disparities, age group performance, and the impact of hosting the Games, which could reveal additional insights into the factors driving Olympic success.

This report not only showcases the achievements of athletes and countries but also opens the door to more detailed analyses that can provide a richer understanding of the dynamics at play in the Olympic Games. The findings can inform future strategies for countries aiming to enhance their Olympic performance and contribute to ongoing discussions about the role of sports in society.