



General Sir John Kotelawala Defence University
Faculty of Management, Social Sciences and
Humanities Department of Languages
BSc in Applied Data Science Communication

Advanced SQL and Cloud Databases

Year 2: Semester 4

18.08.2024

GROUP MEMBERS

D/ADC/23/0008	MAW Sammani
D/ADC/23/0022	HK Ahamat
D/ADC/23/0026	DMS Baddewithana
D/ADC/23/0036	MBSD Fernando

CONTENT

Task 01-Develop a comprehensive Power BI dashboard using historical data from the Sydney 2000 to Rio 2016 Summer Olympic Games

01.Introduction

02.DataSet

03.Explanation and preparation of dataset

3.1 Data Preprocessing

3.2 Data explanation

04.Relationship build with categories in Dashboard

05.Dashboard

06.Explore and analyze key aspects in Dashboard

07.Conclusion

01.Introduction

The modern Summer Olympic Games, held from Sydney 2000 to Rio 2016, highlight athletic excellence, global unity, and cultural exchange. This project aims to create a comprehensive Power BI dashboard to visualize and analyze key aspects of these five editions of the Olympics. The dashboard will offer an interactive platform for exploring data on participation trends, medal counts, and athlete performances. By using data visualization techniques, the project seeks to uncover patterns, highlight achievements, and identify areas for further exploration. This report outlines data preparation, dashboard development, and key findings, aiming to provide valuable insights for sports enthusiasts, researchers, and decision-makers. (wikipedia, 2024 last edit)

02.DataSet

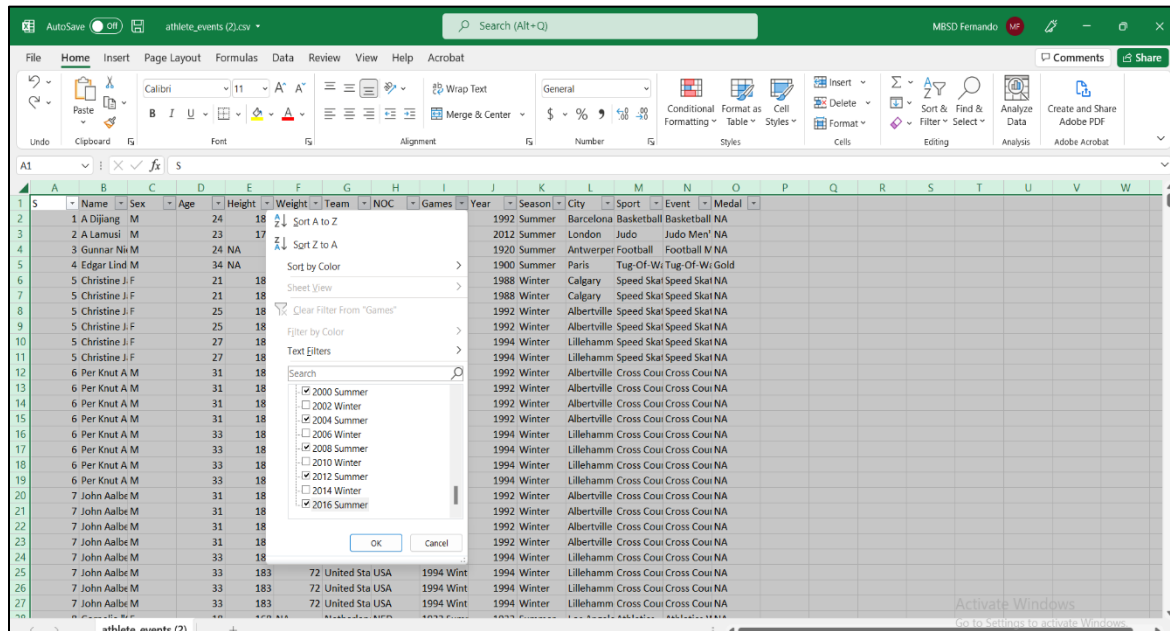
Source: [Olympic history: longitudinal data scraped from www.sports-reference.com \(figshare.com\)](https://www.sports-reference.com/olympic/history/longitudinal-data-scraped-from-figshare.com)

The historical Olympic dataset contains comprehensive data on the Summer Olympic Games from the Sydney 2000 Games to the Rio 2016 Games. This dataset is intended to provide a detailed overview of participation, performance, and achievements in the Summer Olympic Games over these five editions. It will be used to analyze trends, highlight key statistics, and develop visualizations that offer insights into various aspects of the Olympic Games.

03.Explanation and preparation of dataset

3.1 Data Preprocessing

First, filter the data to include only values from the Rio 2000 to the London 2016 Summer Olympic Games. Then, remove any duplicate values and null values.



3.2 Data explanation

The dataset contains 66,450 entries and 15 columns, representing data from the Summer Olympic Games between 2000 and 2016.

1. **ID:** A unique identifier for each athlete.
2. **Name:** The name of the athlete.
3. **Sex:** The gender of the athlete (M for male, F for female).
4. **Age:** The age of the athlete at the time of the event.
5. **Height:** The height of the athlete in centimeters.
6. **Weight:** The weight of the athlete in kilograms.
7. **Team:** The country or team the athlete represented.
8. **NOC:** The National Olympic Committee code for the athlete's country.

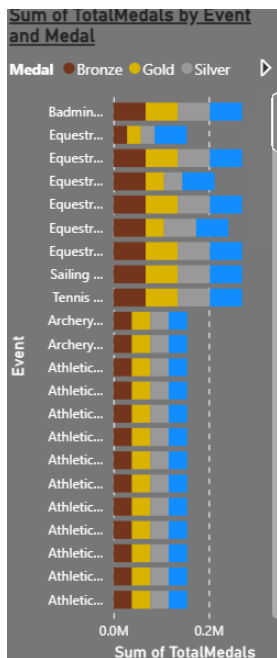
- 9.**Games:** The specific Olympic Games the athlete participated in (year and season).
- 10.**Year:** The year of the Olympic Games.
- 11.**Season:** The season of the Olympic Games (Summer or Winter).
- 12.**City:** The host city of the Olympic Games.
- 13.**Sport:** The sport in which the athlete competed.
- 14.**Event:** The specific event within the sport.
- 15.**Medal:** Indicates the medal won by the athlete (Gold, Silver, Bronze, or 0 if no medal was won).

04.Relationship build with categories in Dashboard

Tables in the Data Model

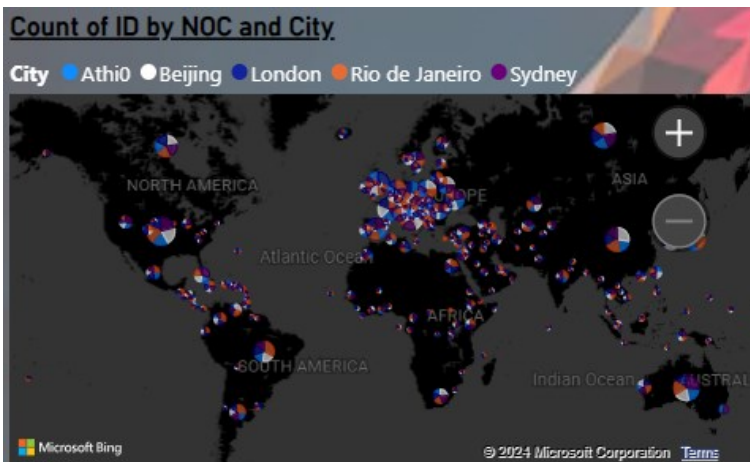
1. **Gender Performance:** This table summarizes performance data by gender, showing how many medals were won by male or female athletes from each team.(sex, Team, Total Medals)
2. **Athlete Participation By Year:** This table appears to track the number of athletes participating in each year of the Olympic Games. (Total Athletes, year)
3. **Age Group Distribution:** This table groups athletes into different age categories and could be used to analyze trends in age distribution over time. (Adult, Senior, Youth)

06.Explore and analyze key aspects in Dashboard



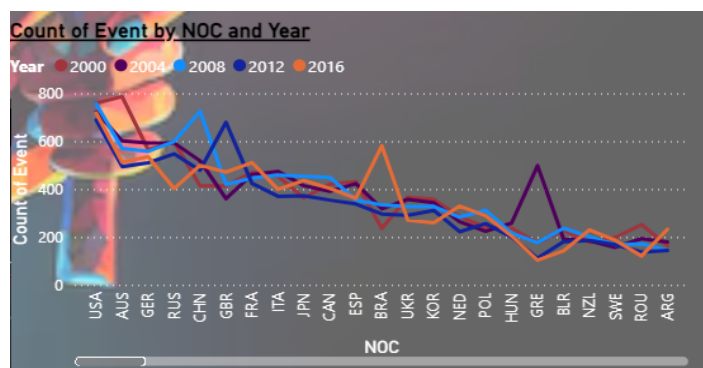
Sum of Total Medals by the event

- The chart shows the total medals (Bronze, Silver, Gold) won across various Olympic events. "Athletics" dominates with the highest medal counts. The bar segments visually represent the distribution of each medal type within the events, highlighting differences in success across sports.



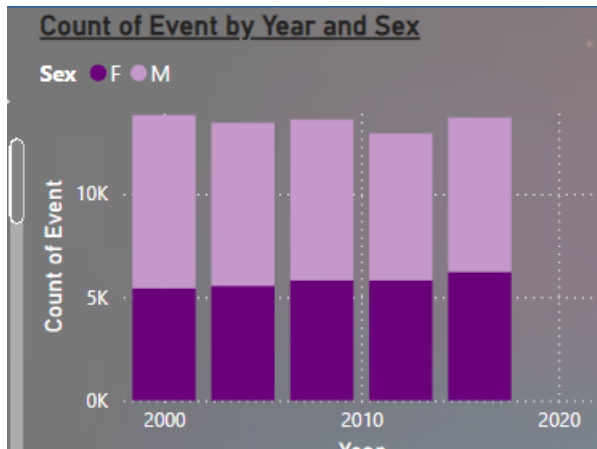
Count of ID by NOC and City

- The map visualizes the count of IDs (athletes or teams) by National Olympic Committees and host cities (Athens, Beijing, London, Rio de Janeiro, Sydney).
- This chart shows geographic distribution and concentration of participating athletes.
- Africa** has shown a relatively consistent performance, with a slight increase in medal counts over the years.



Count of Event by NOC and Year

- The line chart at the bottom right shows the count of events by National Olympic Committees (NOCs) and year.
- It provides a trend analysis across different countries over the years, showcasing

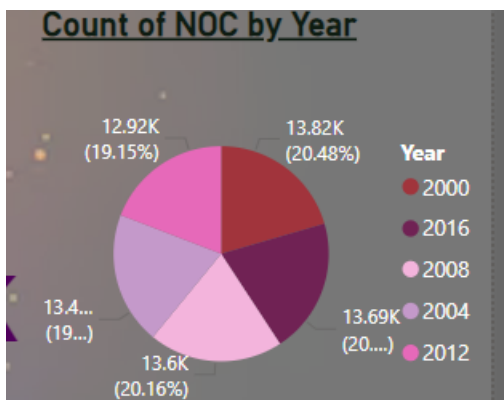


Count of event by year and sex

- The chart displays the distribution of Olympic events by gender over the years 2000, 2010, and 2020.
- It shows a consistent number of events across these years, with a gradual increase in female participation.



The infographic summarizes key Olympic metrics, highlighting 210 active National Olympic Committees (NOC), 337 sports events, 579 total medals awarded, and 67.47K athlete participations across five Summer Olympic Games. These figures underscore the global diversity, wide array of athletic disciplines, and the significant scale of the Olympic Games.

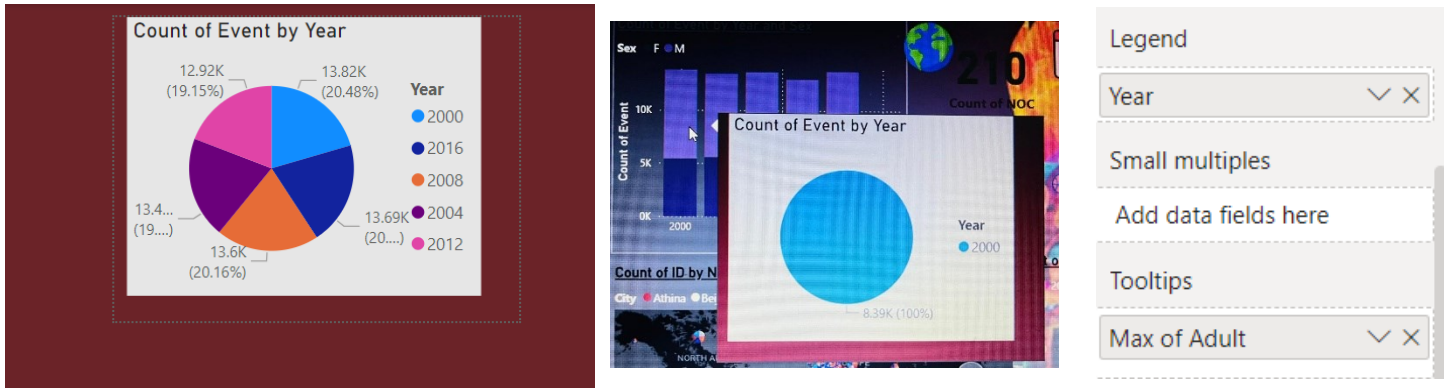


Count of NOC by Year

- The pie chart shows a gradual increase in National Olympic Committees (NOCs) participating in the Summer Olympics from 2000 to 2016, peaking in Rio 2016, reflecting the growing global reach of the Games.

Crate tooltips:

In our dashboard, tooltips are provided in two ways: a dedicated tooltips page for detailed explanations and direct tooltips within visualizations for quick, on-the-spot insights.



07.Conclusion

The Power BI dashboard developed for analyzing the Summer Olympic Games from Sydney 2000 to Rio 2016 effectively highlights key trends and patterns across five Olympic events. By leveraging interactive visualizations, users can explore participation trends, medal counts, and athlete performances with ease, using filters by year, country, sport, and gender. The dashboard successfully brings attention to notable trends, such as the increased participation of female athletes and the rise of certain countries in medal standings.

Through data cleaning, preprocessing, and strategic categorization, the dashboard ensures accurate representation and analysis of the Olympic data. The use of clear labels, tooltips, and a well-chosen color scheme enhances the user experience, making the dashboard both informative and engaging. Overall, this project demonstrates the power of data visualization in uncovering insights and tells a compelling story of the Olympic Games' evolution over the years.

References

wikipedia. (2024 last edit). Retrieved from https://en.wikipedia.org/wiki/2000_Summer_Olympics

Data set link:

[Olympic history: longitudinal data scraped from www.sports-reference.com \(figshare.com\)](#)