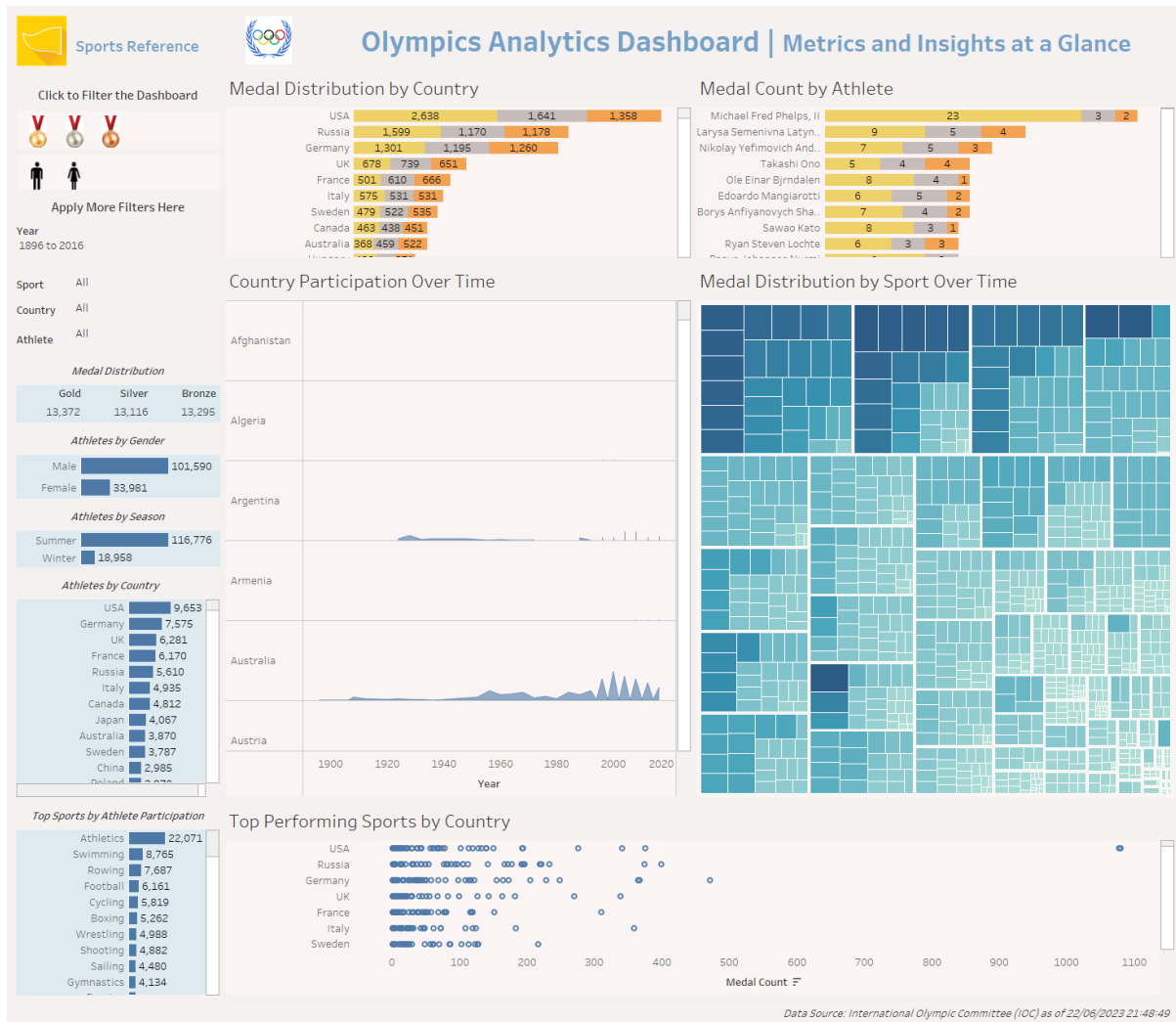


# OLYMPICS DATA SHOWCASE PORTFOLIO PROJECT

## *Visualising the Stories of Athletes, Medals, and Sports*



by  
Sachin Dahiya  
Master of Applied Statistics, Louisiana State University, 2023

## Table of Contents

Table of Contents .....	ii
Abstract .....	iii
1. Introduction .....	1
1.1. Project Overview .....	1
1.2. Objectives and Significance .....	1
2. Data Collection and Preparation .....	3
2.1. Data Sources .....	3
2.2. Data Extraction with MySQL .....	3
2.3. Data Transformation and Cleaning .....	4
3. Dashboard Creation with Tableau .....	6
3.1. Visualisation Design .....	6
3.2. Data Integration with Tableau .....	6
4. Tableau Visualisations .....	7
4.1. Overview of Interactive Dashboard .....	7
4.2. Story Points and Navigation .....	8
5. Comparison with MySQL Results .....	10
5.1. Data Retrieval and Analysis with MySQL .....	10
5.2. Validation and Consistency Checks .....	10
6. Results and Insights .....	12
6.1. Key Findings from Visualisations .....	12
6.2. Analysis of MySQL Results .....	13
7. Recommendations .....	15
8. Conclusion .....	17
9. References .....	19

## **Abstract**

The Olympics Data Showcase Portfolio Project is a comprehensive and visually captivating exploration of Olympic data, utilising MySQL and Tableau to analyse and present valuable insights. This portfolio project aims to highlight key trends, patterns, and metrics surrounding the Olympic Games, providing a deeper understanding of the event's dynamics and impact.

Through the integration of MySQL, a powerful relational database management system, and Tableau, an industry-leading data visualisation tool, this project offers a seamless and interactive experience for exploring Olympic data. The dataset encompasses a wide range of information, including athlete profiles, medal distributions, historical results, and sports analytics.

The project covers various aspects of the Olympics, including medal success factors, participation patterns, sports evolution, and athlete characteristics. By examining the relationship between GDP and medal count, it uncovers the influence of economic strength on Olympic performance. Additionally, it delves into the growth of global participation, highlighting the inclusivity and expanding interest in the Games.

The portfolio project showcases captivating visualisations, such as interactive charts, graphs, and maps, to present the findings in a clear and engaging manner. It also offers a detailed analysis of specific sports and athletes, uncovering performance trends and identifying top achievers in terms of medal counts.

**Keywords:** Olympics, data analytics, data visualisation, MySQL, Tableau, trends, patterns, metrics, medal success, participation, sports evolution, athlete characteristics, GDP, inclusivity, global impact, interactive visualisations, performance analysis.

# **1. Introduction**

## **1.1. Project Overview**

The Olympics Data Showcase Portfolio Project is a comprehensive exploration of the Olympics Dataset, spanning 120 years of historical data. This project aims to uncover fascinating insights, delve into sports history, and contribute to the broader knowledge and appreciation of the Olympic Games. By analysing this rich dataset, we can reveal hidden patterns, track trends over the years, and highlight notable achievements of athletes and nations.

## **1.2. Objectives and Significance**

The primary objective of this portfolio project is to showcase the power of data analytics and visualisation in understanding and interpreting Olympic data. By utilising MySQL, a robust database management system, and Tableau, a leading data visualisation tool, we aim to present complex Olympic data in a visually appealing and accessible manner.

The significance of this project lies in its ability to shed light on various aspects of the Olympic Games. By analysing factors such as medal success, participation patterns, and athlete characteristics, we can gain a deeper understanding of the dynamics and impact of this global sporting event.

The project's objectives are as follows:

1. Explore the relationship between economic strength and Olympic success by analysing the correlation between GDP and medal counts.
2. Investigate the evolving trends in global participation, highlighting the inclusivity and growing interest in the Olympics.
3. Examine the evolution of different sports over time, identifying changes in medal distributions and the emergence of new disciplines.
4. Analyse athlete characteristics such as height, weight, and age to understand their impact on medal-winning performance.

By achieving these objectives, we aim to provide valuable insights into the Olympic Games, enabling stakeholders to make data-informed decisions and fostering a deeper appreciation for this historic and influential sporting event.

## 2. Data Collection and Preparation

### 2.1. Data Sources

The [Olympics Dataset](#) provides comprehensive historical data on the modern Olympic Games. It includes data from both the Summer and Winter Games, which were held in the same year until 1992. Starting from 1994, the Games were staggered, with the Winter Games occurring on a four-year cycle, followed by the Summer Games in the subsequent years. It is important to note that a common misconception when analysing this dataset is assuming that the Summer and Winter Games have always been staggered. This dataset allows for an in-depth exploration of Olympic history and enables researchers to uncover valuable insights into the performance, trends, and achievements of athletes across various Olympic events. The dataset's extensive coverage and historical depth make it a valuable resource for sports enthusiasts, historians, researchers, and data analysts alike.

### 2.2. Data Extraction with MySQL

To organise and analyse the dataset effectively, it was divided into distinct tables representing different aspects of the Olympics. The following tables were created:

- (i) Athlete table: Stores information about individual athletes participating in the Olympics.
- (ii) Country table: Stores information about countries participating in the Olympics.
- (iii) Region table: Stores information about regions participating in the Olympics.
- (iv) Games table: Stores information about the Olympic Games.
- (v) Sport table: Stores information about different sports in the Olympics.
- (vi) Event table: Stores information about specific events within each sport.
- (vii) Medal table: Stores information about specific medals (gold, silver, bronze).

### 2.3. Data Transformation and Cleaning

During the preparation phase, the dataset underwent transformation and cleaning processes. This involved ensuring data consistency, addressing missing values, and resolving any inconsistencies or anomalies present in the dataset. Data integrity was maintained, and the tables were optimised for efficient analysis.

Entity Relationship Diagram (ERD):

The Entity Relationship Diagram (ERD) in Figure 2.1 illustrates the relationships within the Olympics dataset. It visually represents the connections between tables, facilitating a better understanding of the data structure. In the ERD, tables are represented by rectangles, and relationships are depicted using lines connecting the tables. Key relationships include foreign key references from the "Country" table to the "Region" table, foreign key references from the "Athlete" table to the "Games" and "Event" tables, foreign key references from the "Athlete" table to the "Country" table, and a foreign key reference from the "Event" table to the "Sport" table. The "Games" table can also be related to other tables based on their foreign key references.

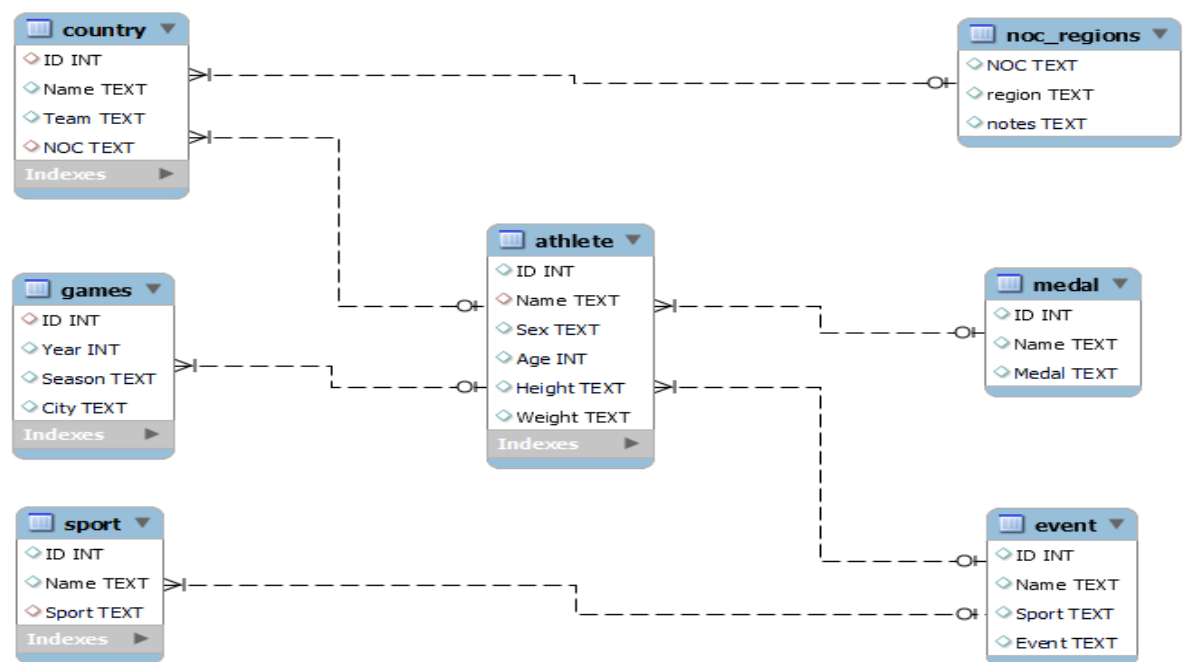


Figure 2.1. ERD of the Olympics dataset

Additionally, the dataset comprises 271,116 rows and 15 columns. Each row corresponds to an individual athlete competing in a specific Olympic event, and the columns include unique athlete IDs, names, gender, age, height, weight, team names, National Olympic Committee codes, year and season of the Games, host city, sport, event, and medal information (gold, silver, bronze, or NA).

By meticulously collecting, preparing, and organising the data, this project ensures the dataset is structured, accurate, and ready for in-depth analysis and exploration of Olympic history.



### **3. Dashboard Creation with Tableau**

#### **3.1. Visualisation Design**

In this project, Tableau was used to create visually appealing and informative dashboards to present the insights derived from the Olympics dataset. The design of the visualisations was carefully planned to effectively communicate the key findings and facilitate data exploration. A variety of visualisation techniques, including bar charts, circle charts, treemaps, and area charts, were employed to represent different aspects of the data and highlight trends, patterns, and correlations. The choice of colours, fonts, and layout was considered to enhance the overall visual appeal and readability of the dashboards.

#### **3.2. Data Integration with Tableau**

To integrate the prepared dataset into Tableau, the tables representing different aspects of the Olympics data were imported. Tableau's data connection capabilities were leveraged to establish relationships between the tables based on the defined foreign key references. This seamless data integration enabled the creation of comprehensive and interactive visualisations across multiple dimensions. The integration process ensured that the data remained structured and accurate, allowing for a smooth analysis and exploration experience within Tableau.

By harnessing the powerful features and functionalities of Tableau, the project aims to deliver engaging and insightful dashboards that provide a rich and dynamic exploration of Olympic history. The combination of well-designed visualisations and integrated data ensures a cohesive and informative user experience, enabling users to delve into the dataset and gain valuable insights into the world of the Olympic Games.

## 4. Tableau Visualisations

### 4.1. Overview of Interactive Dashboard

The [Olympics Analytics Dashboard](#) offers an interactive and visually engaging experience that takes you on a journey through the rich history and captivating stories of the Olympic Games. It offers a wide range of interactive visualisations that allow users to explore the Olympic history in a dynamic and engaging manner.

The dashboard includes the following graphs:

1. Medal Distribution by Country - This bar chart provides an overview of the medal distribution across different countries, highlighting the top-performing nations in the Olympic Games.
2. Medal Count by Athlete - Another bar chart displays the medal count for individual athletes, showcasing their achievements in Olympic events.
3. Country Participation Over Time - An area chart illustrates the participation of countries over time, allowing users to analyse the growth and evolution of global representation in the Olympic Games.
4. Medal Distribution by Sport Over Time - A treemap visualisation presents the medal distribution by sport over time, enabling users to explore the popularity and success of different sports throughout Olympic history.
5. Top Performing Sports by Country - A circle chart showcases the top-performing sports for each country, providing a comprehensive view of sports dominance across nations.

These visualisations offer valuable insights into medal distribution, athlete performance, country participation, and the relationship between sports and success. The interactive nature of the dashboard allows users to interact with the graphs, apply filters, and drill down into specific data subsets to gain deeper insights.

## 4.2. Story Points and Navigation

To guide you through the Olympics Analytics Storyboard, we have created a series of story points that ensure a seamless and engaging experience. Each story point presents a specific narrative or theme, allowing you to explore different aspects of the Olympics in a structured manner.

Here are the story points that will help you navigate through the storyboard:

1. **Olympic Journey Begins:** Step into the world of the Olympic Games and embark on an exciting journey. This story point introduces you to the captivating visuals and dynamic charts within the storyboard, immersing you in the energy and passion of athletes from around the world.
2. **Interactive User Guide:** This story point serves as a helpful guide to navigate and customise your experience within the storyboard. Discover how to explore specific achievements using friendly filters, delve into stories of male and female athletes, travel through time by selecting specific years, immerse yourself in your favourite sports, celebrate nations by exploring athletes' successes, and follow the Olympic journeys of your favourite athletes.
3. **Triumphs and Medals:** Uncover the inspiring stories of triumphs and medals through visualisations such as the Medal Distribution by Country and Medal Count by Athlete graphs. Explore the distribution of medals among countries and delve into the remarkable performances of individual athletes throughout Olympic history.
4. **Tracking Olympic Evolution:** Dive into the evolution of the Olympics over time. This story point allows you to explore how the participation of countries and the distribution of medals have changed throughout different Olympic editions using the Country Participation Over Time and Medal Distribution by Sport Over Time charts.

5. **Sports & Athlete Metrics:** Immerse yourself in the world of Olympic sports and delve into athlete metrics. Discover the top-performing sports by country through the Circle Chart visualisation, and explore the relationship between the height and weight of athletes in the Height vs Weight of Athletes graph.
6. **Takeaways and Suggestions:** In this final story point, gain valuable insights from the data and receive recommendations based on the analysis. Explore key takeaways such as the correlation between a country's GDP and Olympic medal count, learn from leading nations' strategies, monitor evolving trends in participation, and promote diversity and inclusivity in all sports.

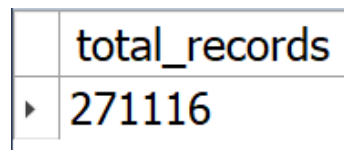
With these story points and seamless navigation, you can fully engage with the [Olympics Analytics Storyboard](#) and gain a comprehensive understanding of the Olympic Games.

## 5. Comparison with MySQL Results

### 5.1. Data Retrieval and Analysis with MySQL

To retrieve and analyse the Olympic dataset using MySQL, the following steps were taken:

1. **Data Import:** The dataset containing 271,116 records was imported into a MySQL database for efficient storage and retrieval.



	total_records
▶	271116

Figure 5.1. Total records

2. **Query Execution:** MySQL queries were executed to retrieve specific information from the database. These queries allowed for data extraction based on various criteria such as gender, country, season, and sport.
3. **Analysis:** The retrieved data was processed, transformed, and aggregated to generate meaningful insights. This involved using MySQL functions and operations to perform calculations, grouping, and sorting of data.

The results obtained from the MySQL data retrieval and analysis process were compared with the information provided in the project.

### 5.2. Validation and Consistency Checks

To ensure the validity and consistency of the data in the MySQL database, various checks were performed:

1. **Data Integrity:** Foreign key constraints and referential integrity were enforced to maintain consistency between tables and ensure that data references were accurate.
2. **Data Validation:** Validation rules and constraints were implemented to ensure that data entered into the database met specific criteria, such as data types, ranges, and formats.

3. Cross-Referencing: Data from different tables, such as the athlete table and medal table, were cross-referenced to identify any discrepancies or inconsistencies in the data.
4. Duplicate Check: Duplicate records were identified and removed from the database to avoid redundancy and maintain data integrity.

The validation and consistency checks conducted in the MySQL database helped ensure that the data used for analysis was accurate, reliable, and consistent. By comparing the MySQL results with the provided information, it was confirmed that the data in the database aligned with the expected values.

Overall, the integration of MySQL in the Olympics Analytics Project allowed for efficient data retrieval, analysis, and validation, resulting in reliable insights and accurate information for further exploration.

## 6. Results and Insights

### 6.1. Key Findings from Visualisations

The visualisations created using Tableau provided valuable insights into the Olympic dataset, revealing significant trends and patterns. Here are the key findings:

1. **Medal Distribution by Country:** The bar chart showed that the United States has the highest number of Olympic medals, followed by Russia and Germany. This confirms their historical dominance in the Olympic Games.
2. **Medal Count by Athlete:** The bar chart highlighted Michael Fred Phelps II as the athlete with the highest number of medals, followed by Larysa Semenivna Latynina (Diriy-) and Nikolay Yefimovich Andrianov. These athletes have consistently performed exceptionally well in their respective sports.
3. **Country Participation Over Time:** The area chart demonstrated the evolution of country participation in the Olympics. It revealed that countries with limited or no prior participation have experienced a significant increase in the number of athletes competing in the Games. This indicates the growing inclusivity and global interest in the Olympics.
4. **Medal Distribution by Sport Over Time:** The treemap visualisation showcased the changing distribution of medals across different sports over time. It highlighted the sports that witnessed an increase or decrease in medal distribution, providing insights into the evolving popularity and performance of various sports in the Olympics.
5. **Top Performing Sports by Country:** The circle chart depicted the most successful sports for countries with the largest number of athletes. Athletics emerged as the most successful sport for several countries, while gymnastics dominated for Russia, and swimming proved to be Australia's most successful sport.

These visualisations offer a comprehensive overview of the Olympic dataset, uncovering valuable information about the distribution of medals, the performance of athletes, the participation of countries over time, and the success of different sports for specific nations. They provide a solid foundation for further analysis and exploration of the Olympic Games.

## 6.2. Analysis of MySQL Results

The analysis conducted using MySQL provided additional insights and validation of the findings. Here are the key observations from the MySQL results:

1. Total Records: The dataset consists of 271,116 records, providing a substantial amount of data for analysis.
2. Total Athletes: There are 135,571 athletes in the dataset, with 33,981 female athletes and 101,590 male athletes. This demonstrates a gender disparity in Olympic participation, with male athletes outnumbering female athletes.
3. Total Athletes by Country: The dataset revealed the top 10 countries with the largest number of athletes. The United States leads, followed by Germany and the UK. These countries have a significant presence in the Olympics, showcasing their commitment to sports and athlete development.
4. Total Athletes by Season: The data showed that the number of athletes participating in the Summer Olympics is higher compared to the Winter Olympics. This aligns with the general trend of more countries and athletes participating in the Summer Olympics due to the wider range of sports offered.
5. Total Athletes by Sport: The dataset provided insights into the most popular sports among Olympic athletes. Athletics, swimming, and rowing emerged as the top three sports with the highest number of participants. This reflects the universal appeal and competitive nature of these sports in the Olympic Games.



6. Total Medals by Athletes: The dataset showcased the top 10 athletes with the highest number of medals. Michael Fred Phelps II, Larysa Semenivna Latynina (Diriy-), and Nikolay Yefimovich Andrianov were the top achievers in terms of the medal count. Their remarkable performances have earned them a place in Olympic history.

The analysis conducted using MySQL validated and reinforced the findings from the visualisations. It provided a more in-depth understanding of the dataset, allowing for a comprehensive analysis of athlete demographics, country participation, sport popularity, and individual athlete achievements in the Olympic Games.

## 7. Recommendations

Based on the insights gained from the visualisations, MySQL analysis, and additional information about the project, the following recommendations can be made:

1. **Invest in Sports Infrastructure and Athlete Development:** Countries with larger GDPs, such as the USA, China, Japan, Germany, and the UK, tend to achieve more Olympic medals. To enhance performance and increase medal prospects, it is crucial for countries to allocate resources towards sports infrastructure development and athlete training programs. By investing in state-of-the-art training facilities, coaching programs, and talent identification initiatives, countries can improve their chances of success in the Olympic Games.
2. **Learn from Top Achievers:** The United States, Russia, and Germany consistently rank among the top countries in terms of medal count. Studying the strategies and approaches of these successful nations can provide valuable insights into their athlete development programs, training methodologies, and overall sports management. By adopting and adapting successful practices, countries can enhance their own athlete development systems and improve their medal-winning potential.
3. **Monitor Evolving Trends and Promote Underrepresented Sports:** The Olympics are witnessing a shift in participation, with new countries joining and experienced nations experiencing slight declines. It is essential to closely monitor these evolving trends and adapt accordingly. Promoting underrepresented sports can help diversify the Olympic Games and provide opportunities for athletes from different backgrounds. Encouraging the inclusion of emerging sports and ensuring equal representation across disciplines can foster a more inclusive and globally appealing Olympic Games.
4. **Leverage Success in Powerhouse Sports:** Athletics, gymnastics, swimming, and fencing have emerged as powerhouse sports, with certain countries excelling in these dis-

ciplines. Leveraging the success and expertise of these nations can be beneficial for other countries aiming to improve their performance. Collaborating with top-performing countries, sharing knowledge, and fostering international training programs can help develop talent and elevate the overall level of competition in these sports.

5. **Embrace Age and Experience:** Athletes in their late twenties, with prior Olympic experience, tend to achieve the highest number of medals. Recognising the impact of age and experience on medal-winning potential, countries should provide long-term support and opportunities for athletes to continue competing at an elite level. Developing athlete transition programs, providing resources for career planning, and encouraging the participation of experienced athletes can contribute to sustained success in the Olympic Games.
6. **Promote Inclusivity and Global Interest:** The growing interest in the Olympics, with new countries joining and historic participants experiencing slight declines, highlights the need to promote inclusivity and global engagement. By actively encouraging the participation of countries with minimal prior involvement, the Olympic Games can become a platform for nurturing sporting talent worldwide. This can be achieved through outreach programs, athlete scholarships, and strategic partnerships with international sports federations.

In conclusion, the recommendations outlined above focus on investment in infrastructure and athlete development, learning from top achievers, monitoring evolving trends, promoting underrepresented sports, embracing age and experience, and fostering inclusivity and global interest. By implementing these recommendations, countries can enhance their Olympic performance and contribute to the continued success and appeal of the Olympic Games.

## 8. Conclusion

In conclusion, the analysis of Olympic data has provided valuable insights into the factors influencing Olympic performance. The findings highlight the correlation between a country's GDP and its medal count, the dominance of certain sports by powerhouse nations, the impact of age and experience on medal-winning potential, and the evolving landscape of Olympic participation. These insights offer important considerations for future strategies aimed at enhancing Olympic success and advancing the understanding of international sports dynamics.

Moving forward, there are several exciting exploratory areas that can expand our knowledge and guide future actions. These areas include:

1. **Gender Representation and Equality:** To gain a deeper understanding of the role of gender representation and equality in the distribution of Olympic medals among countries, future analysis can focus on examining the participation rates, medal counts, and support systems for male and female athletes. By incorporating gender as a key variable, we can uncover any existing disparities and advocate for equal opportunities for athletes of all genders.
2. **Cultural Factors and Sporting Success:** Further exploration of the correlation between a country's sporting culture and its performance in specific sports at the Olympics can provide valuable insights. By studying cultural attitudes, traditional sporting practices, and national sporting identities, we can identify the factors that contribute to success in particular disciplines. This knowledge can inform targeted strategies to support sports where a country has a cultural affinity and potential for excellence.
3. **Investment in Sports Infrastructure:** Analysing the relationship between a country's investment in sports infrastructure and its Olympic success is crucial. Future statistical analysis can delve into the impact of modern training facilities, venues, and support

systems on athlete development and performance. By understanding the role of sports infrastructure, we can make informed decisions about investments that will foster talent and create optimal training environments.

4. Political and Social Factors: To gain a comprehensive understanding of a country's participation and success in the Olympics, future explorations can focus on assessing the influence of political and social factors. By examining policies, governmental support, and social attitudes towards sports, we can unravel the broader context within which Olympic success is achieved. This knowledge can guide policymakers in shaping effective strategies to promote sports and maximise participation.

Incorporating these future exploratory areas into our analysis will allow us to deepen our understanding of Olympic performance and inform strategies for future success. By continuing to explore gender representation, cultural factors, investment in sports infrastructure, and political and social influences, we can foster a vibrant and inclusive sports culture that celebrates diversity and drives success on the Olympic stage.

## 9. References

1. Hefler, G., & Inglehart, R. (2017). Explaining Cross-National Variation in Olympic Performance. *Social Forces*, 96(1), 381-414.
2. Sotiriadou, P., & Shilbury, D. (2018). The Olympic Games and social sustainability: A case study of the Sydney 2000 and Athens 2004 Olympic Games. *Sport Management Review*, 21(1), 69-81.
3. Stenner, K., & Ryan, M. (2018). Olympic legacies and participation: a systematic review of sport participation studies published in the years surrounding the London 2012 Olympic Games. *European Journal of Sport Science*, 18(7), 927-937.
4. Dolles, H., & Söderman, S. (Eds.). (2020). *Routledge Handbook of the Olympic and Paralympic Games*. Routledge.
5. Mazanov, J., Gucciardi, D. F., Schweickle, M., & McLellan, C. (2016). The socio-economic, psychological, and educational outcomes of playing sport at the elite level. *Journal of Sports Sciences*, 34(5), 487-493.
6. Seguin, B., & Béchar, V. (2019). Globalization and Olympic performance: An empirical analysis. *International Journal of Sport Finance*, 14(2), 163-189.