|  |  |
| --- | --- |
| CAPSTONE PROJECT  Olympic Sports Analysis | Abstract  The Olympics project showcase the pinnacle of human athletic achievement, featuring a diverse range of sports that test physical prowess, mental fortitude, and teamwork. Through fierce competition and global unity, they inspire individuals to push their limits and celebrate the boundless potential of the human spirit, embodying the universal pursuit of excellence.  Ajit Singh  Data Analyst |

# Overview

## Olympic sports analysis involves the systematic assessment of various facets of the Olympic Games. This encompasses evaluating the overall success of the Games, including factors like viewership and economic impact. It also delves into the performance of individual sports, examining metrics such as speed, accuracy, and endurance.

## Event analysis focuses on the structure and rules of specific competitions, while participant analysis assesses athlete performance and health. Medal analysis scrutinizes medal distribution and trends, and region analysis examines the representation and performance of different areas. This analysis provides valuable insights for organizers, athletes, and fans, shaping the future of the Olympic Games.

## Olympic sports analysis involves the examination and evaluation of various aspects related to the Olympic Games and the sports and events within them. Olympic sports analysis involves a comprehensive evaluation of various aspects of the Games, including the performance of athletes, the execution of events, and the broader impact on host cities and regions. This analysis can provide valuable insights for stakeholders, including organizers, athletes, coaches, and fans, and help shape future Olympic Games.

# Objective

## The objectives of Olympic sports analysis include evaluating the overall success of the Games, assessing their long-term impact, and examining the performance of individual sports and athletes. It aims to identify trends, rule changes, and technological advancements that may affect sports performance.

## The analysis also focuses on event structure, athlete health, and injury status. Additionally, it aims to understand medal distribution, performance trends among countries, and regional representation. Cultural, socioeconomic, and geopolitical factors are considered in this analysis. Overall, the objective is to provide valuable insights for stakeholders involved in the planning and execution of the Olympic Games.

## It aims to provide a comprehensive understanding of the various aspects of the Games, from the performance of athletes to the broader impact on host cities and regions. This analysis serves as a valuable resource for stakeholders involved in the planning, organization, and participation in the Olympic Games.

# Significance

## Olympic sports analysis holds significant importance in various aspects of the Games. It drives performance improvement for athletes through detailed assessments. Organizers use analysis for strategic planning, optimizing event formats, rules, and venues. Legacy planning benefits host cities by guiding infrastructure development and community engagement. Analysis informs policy and rule changes, ensuring fairness and safety. It enhances fan engagement by providing insights and statistics.

## Economically, it assesses the impact of hosting the Games. Additionally, it inspires and motivates athletes and fosters national pride. Trends and patterns inform sports development strategies. Health monitoring ensures athlete well-being, and considerations of culture and society promote inclusivity.

## Ultimately, analysis enables data-driven decision-making for a successful Olympic Games. Olympic sports analysis plays a vital role in improving performance, planning and executing successful Games, assessing long-term impacts, and making data-driven decisions that benefit athletes, organizers, and the broader community.

# Data Dictionary

## Table: CITY

## id (integer): Unique identifier for the city.

## city\_name (string): Name of the city.

## Table: COMPETITOR\_EVENT

## event\_id (integer): Unique identifier for the event.

## competitor\_id (integer): Unique identifier for the competitor.

## medal\_id (integer): Unique identifier for the medal won in the event.

## Table: EVENT

## id (integer): Unique identifier for the event.

## sport\_id (integer): Unique identifier for the sport associated with the event.

## event\_name (string): Name of the event.

## Table: GAMES

## id (integer): Unique identifier for the games.

## games\_year (integer): Year in which the games took place.

## games\_name (string): Name of the games.

## season (string): Season of the games (e.g., "Summer" or "Winter").

## Table: GAMES\_CITY

## games\_id (integer): Unique identifier for the games.

## city\_id (integer): Unique identifier for the city hosting the games.

## Table: GAMES\_COMPETITOR

## id (integer): Unique identifier for the record.

## games\_id (integer): Unique identifier for the games.

## person\_id (integer): Unique identifier for the person/competitor.

## age (integer): Age of the competitor.

## Table: MEDAL

## id (integer): Unique identifier for the medal.

## medal\_name (string): Name of the medal.

## Table: NOC\_REGION

## id (integer): Unique identifier for the record.

## noc (string): National Olympic Committee code.

## region\_name (string): Name of the region associated with the NOC.

## Table: PERSON

## id (integer): Unique identifier for the person/competitor.

## full\_name (string): Full name of the person.

## gender (string): Gender of the person (e.g., "Male" or "Female").

## height (float): Height of the person.

## weight (float): Weight of the person.

## Table: PERSON\_REGION

## person\_id (integer): Unique identifier for the person/competitor.

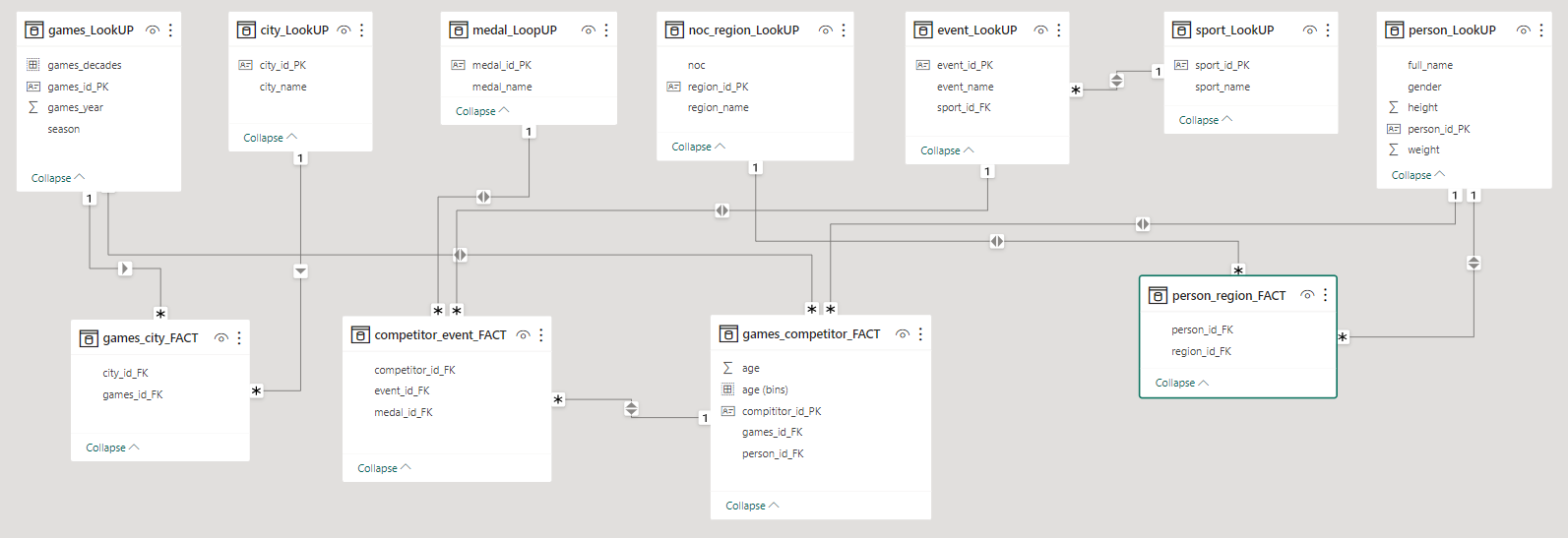
## region\_id (integer): Unique identifier for the region associated with the person.

## Table: SPORT

## id (integer): Unique identifier for the sport.

## sport\_name (string): Name of the sport.

# ER-Diagram

****

# Process

## 1. Data Acquisition from GitHub: Obtain the requisite dataset from a designated GitHub repository containing essential information on Olympic Games.

## 2. Data Transformation and Enhancement: Performed data transformation and prepared data to ensure consistency, data integrity and data quality.

## 3. Establishment of connection with Tool: Establish connections between the dataset and various analytical tools. Interface the dataset with Power BI, Excel and MySQL, Workbench data integration and processing

## 4. Problem Statement Solution in Power BI: Utilize Power BI to delve into the specified problem statement. its robust features for data visualization, exploration and analysis effectively insights and solutions

## 5. Exploratory Data Analysis (EDA): Perform exploratory data analysis using either Excel or SQL, Workbench, depending on the complexity of analysis. Extract meaningful patterns, relationship and from the data to inform subsequent decision making.

## 6. Creation of Visual and Insightful PowerPoint: PowerPoint presentation that encapsulates the projects’ objective, methodologies, problem statement solutions, and key visualization.

## 7. Detailed Documentation: Compile a detailed report that meticulously document the entire project lifecyle. Include sections on data collection,

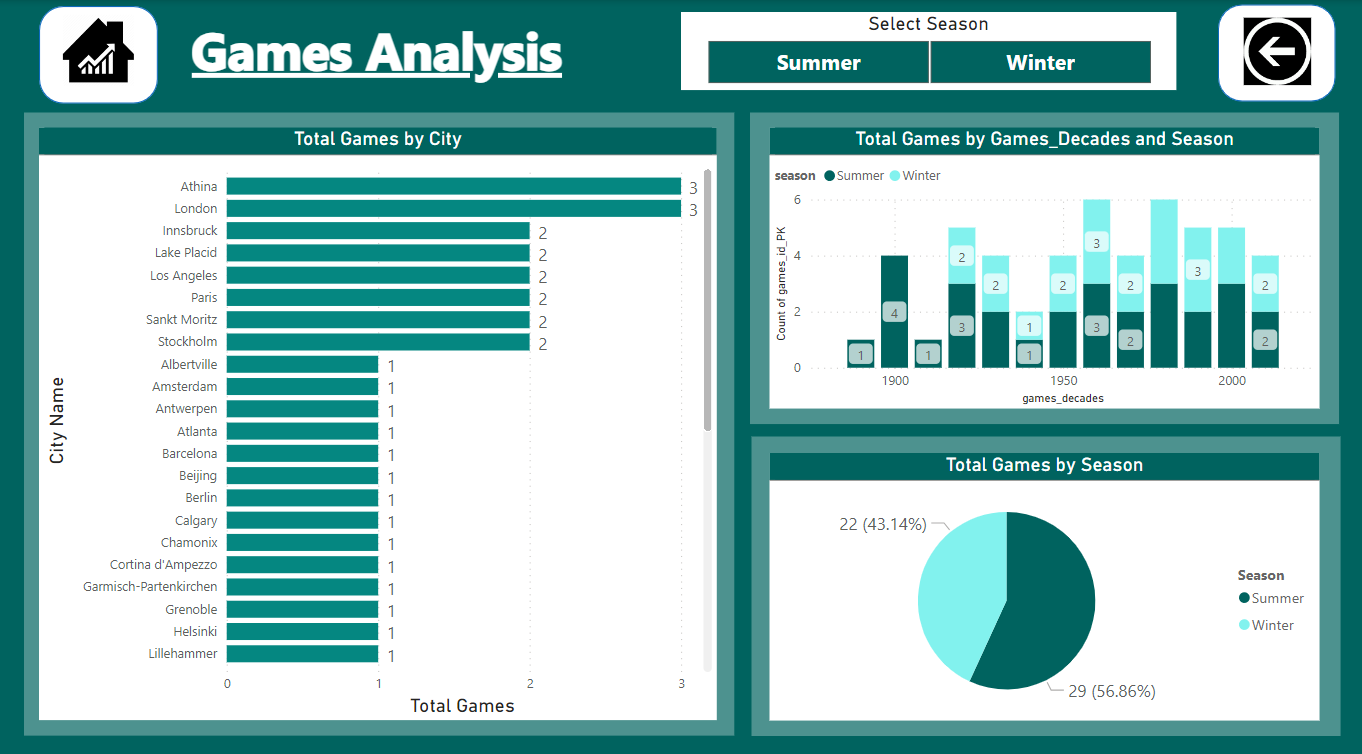
## transformation, problem statement formulation, tools integration, Power BI solution.

## EDA insights and PowerPoint visualization

Power BI – Problem Statement



# Games Analysis



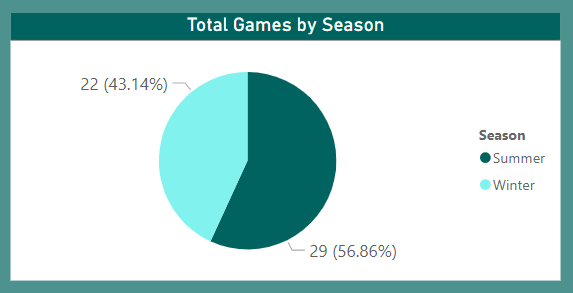
## 1. How many Olympic Games have been held in each season (Summer vs. Winter)?

## We used a Pie chart visualization to present this information. This chart displays the total Olympic Games for each season. This allows us to easily see how many games have been held in Summer compared to Winter over the years.

## The purpose of using a Pie chart in this context is to visually show the relative distribution of Olympic Games between Summer and Winter over a period of time. Each slice of the Pie chart would represent the percentage or proportion of games held in each season. This visual representation makes it easy for the audience to quickly understand and compare the number of Summer and Winter Olympic Games.

## Summer – 29(56.58%)

## Winter – 22(43.14%)



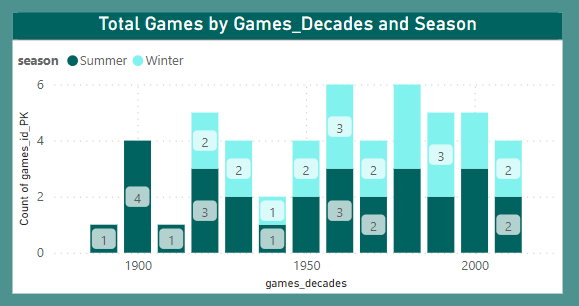
## 2. What is the distribution of games across different decades?

## Distribution Chart: This is a type of visual tool used to display data in a graphical form. In this context, it's being used to show information about the number of video games released over different time periods.

## Visualizes: This means that the chart helps to represent or illustrate the data in a way that makes it easier to understand at a glance. Instead of looking at raw numbers, you can see trends and patterns in the data.

## Across Different Decades: The data is organized based on periods of ten years, starting from a specific point in time. For example, one decade might span from 1990 to 1999, the next from 2000 to 2009, and so on.

## Each Row Represents a Decade: In the chart, there will be rows, and each row corresponds to a specific decade. For instance, one row might represent the 1980s, another row the 1990s, and so forth.

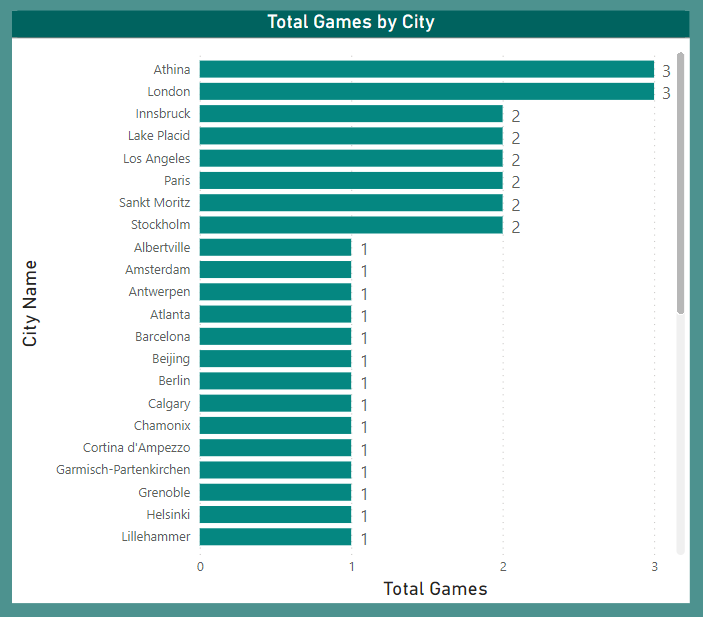


## 3. Which cities have hosted the most Olympic Games?

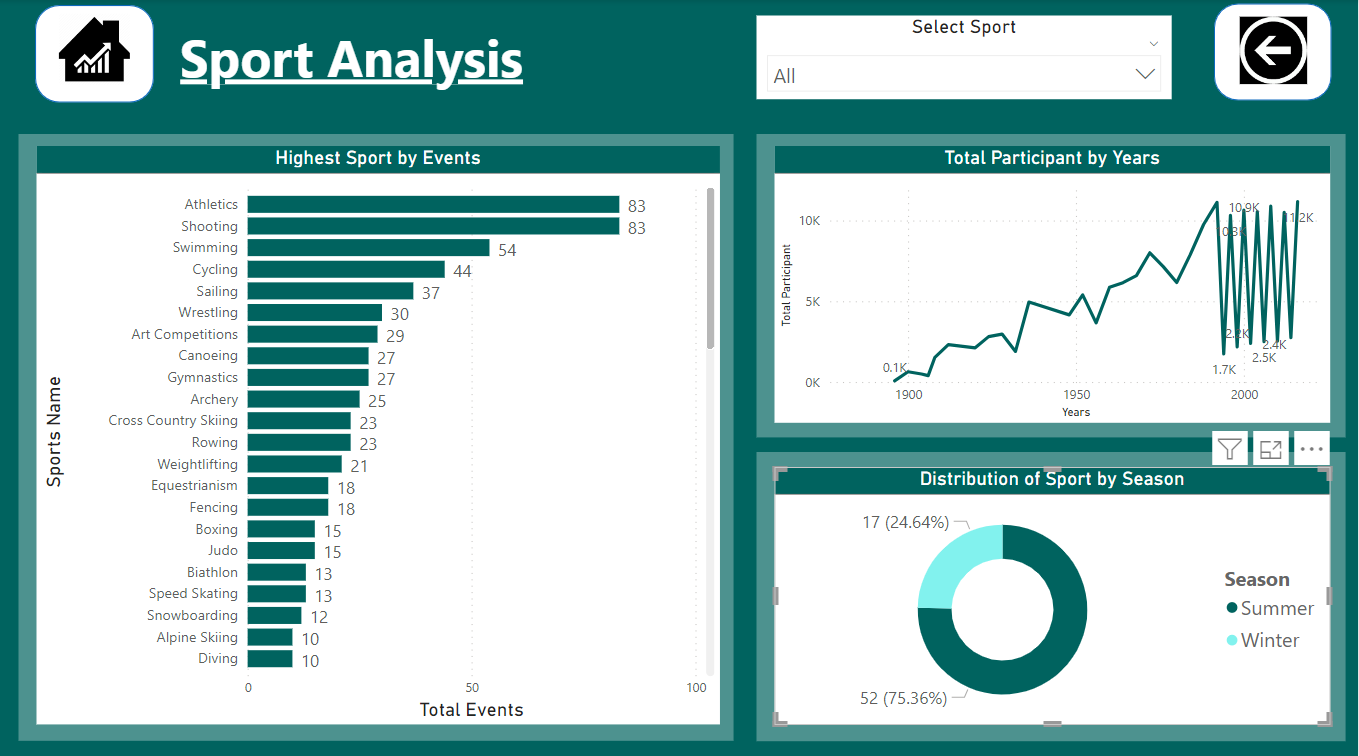
## The Bar Chart displays the list of cities that have hosted the Olympic Games, along with the number of times they have served as hosts.

## The name of the city: This is the name of the specific location where the Olympic Games have been held.

## The number of times the city has served as a host: This indicates how many times the city has been chosen to host the Olympic Games.



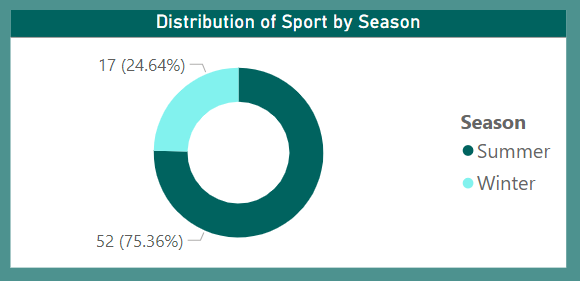
# Sport Analysis

****

## 4. What is the distribution of sports between Summer and Winter Olympics?

## This Power BI report visualizes the distribution of sports between the Summer and Winter Olympics. It provides a clear overview of how many sports are featured in each type of Olympic event.

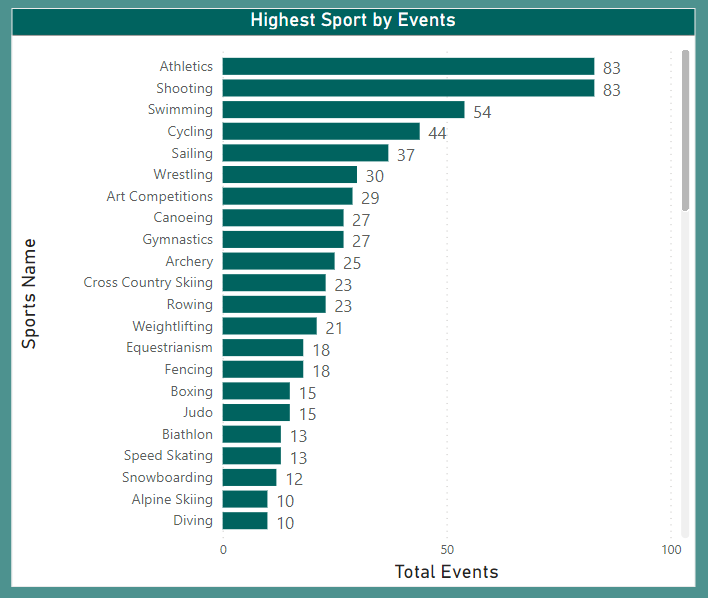
## This Power BI report visualizes the distribution of sports between the Summer and Winter Olympics. It provides a clear overview of how many sports are featured in each type of Olympic event.



## 5. Which sports have the highest number of events in the Olympics?

## The visual displays a breakdown of sports with the highest number of events in the Olympics. This information provides valuable insights into the diversity and popularity of sports in the Olympic Games.

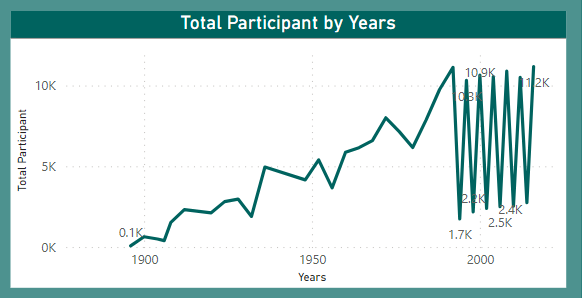
## the visual is giving you a clear picture of how many events each sport contributes to the Olympics. This information can be quite valuable because it sheds light on the range of sports represented and which ones are considered most significant in the context of the Olympic Games.



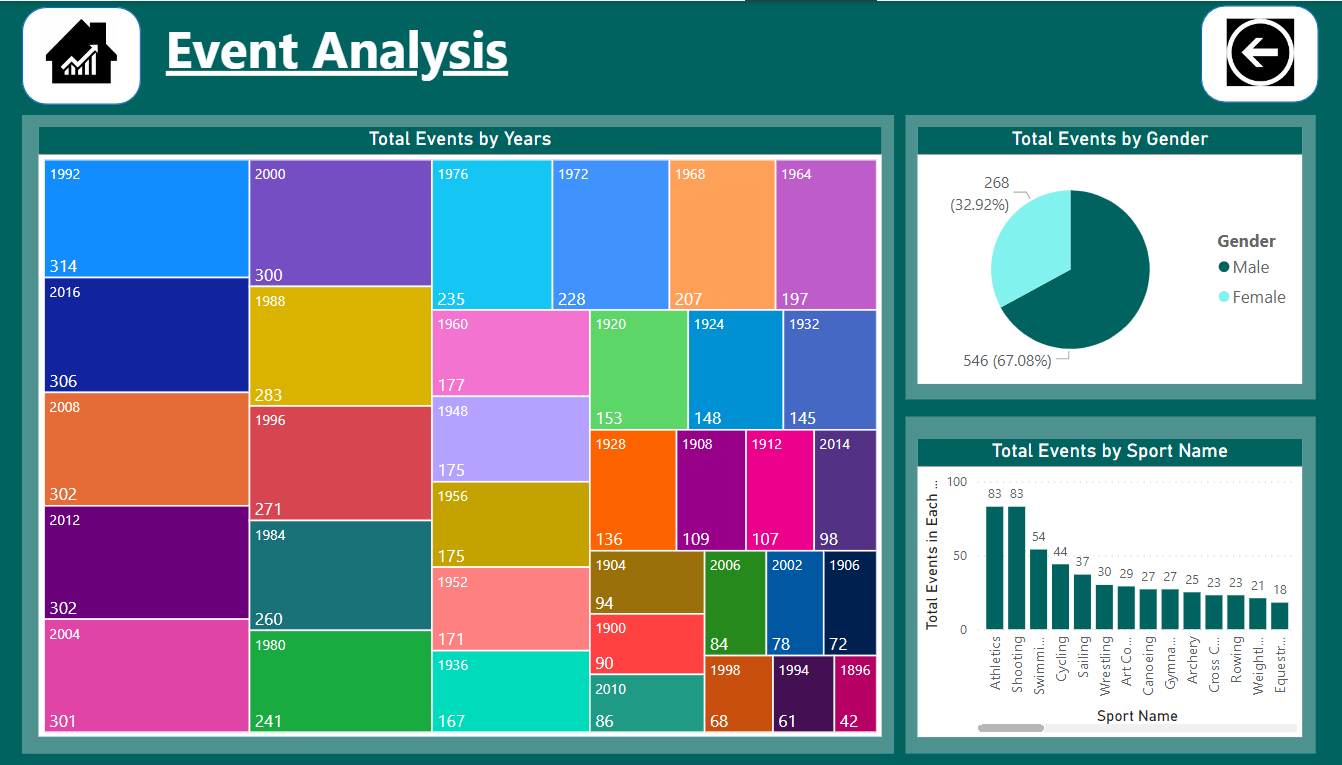
## 6. How has the participation in each sport evolved over time?

## Line chart display the evolution of sports participation over time. This process involves ultimately sharing insights to understand how different sports have changed in popularity throughout the years.

## Display sports participation trends allows for a clear and concise way to analyze and communicate how different sports have changed in popularity over a specified period. This information can be valuable for sports organizations, policy makers, and anyone interested in understanding the dynamics of sports engagement over time.



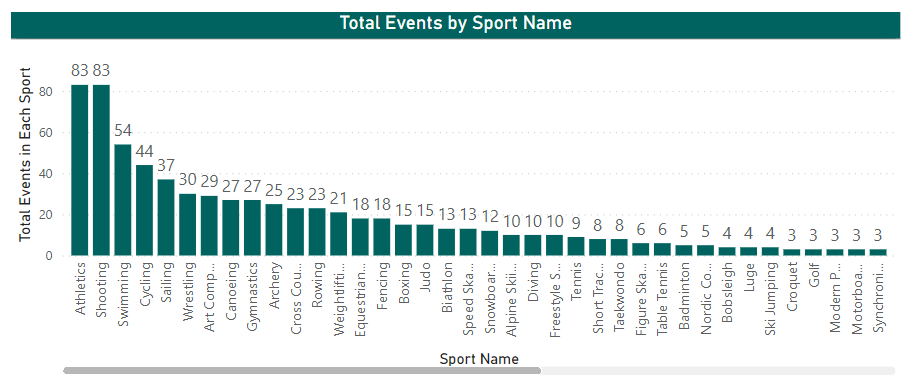
# Event Analysis



## 7. How many events are there in each sport?

## Bar chart display total events in each sport. This can be a valuable tool for decision-making, reporting, and gaining insights into the sports landscape.

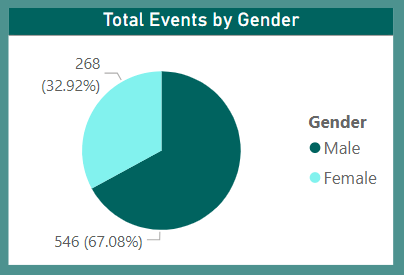
## This means that for each sport, you count the total number of events that have taken place. For example, in athletics, you might have events like sprinting, long jump, and javelin throw. The total number of these events would be displayed on the bar chart.



## 8. What is the distribution of events by gender (Men, Women, Mixed)?

## pie chart displaying the distribution of events by gender, with each slice representing a different gender and the size of each slice corresponding to the proportion of events for that gender.

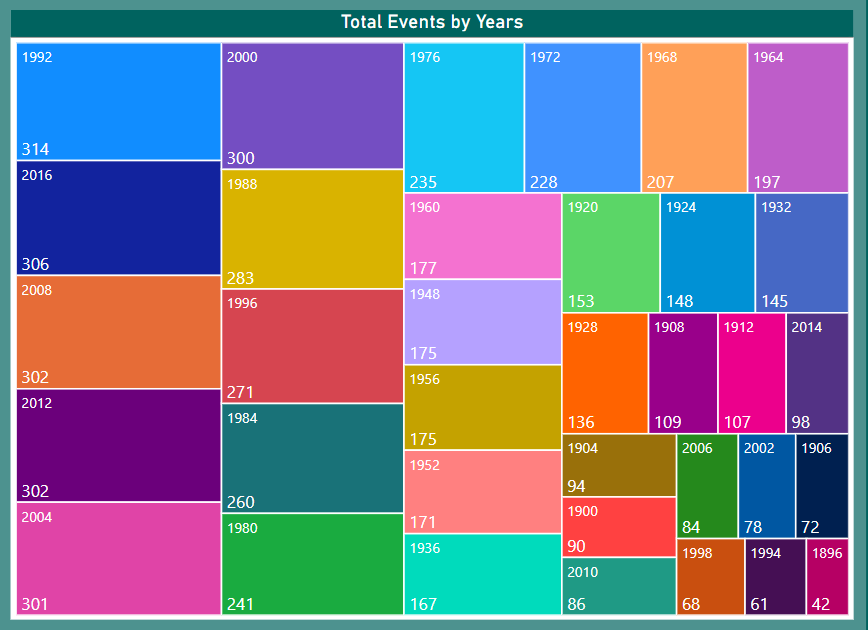
## the pie chart is being used to display information about events categorized by gender. Each slice of the pie chart represents a different gender category. For example, there might be slices for categories like "male," "female," "non-binary," and so on. The size of each slice corresponds to the proportion of events that are associated with that particular gender category.



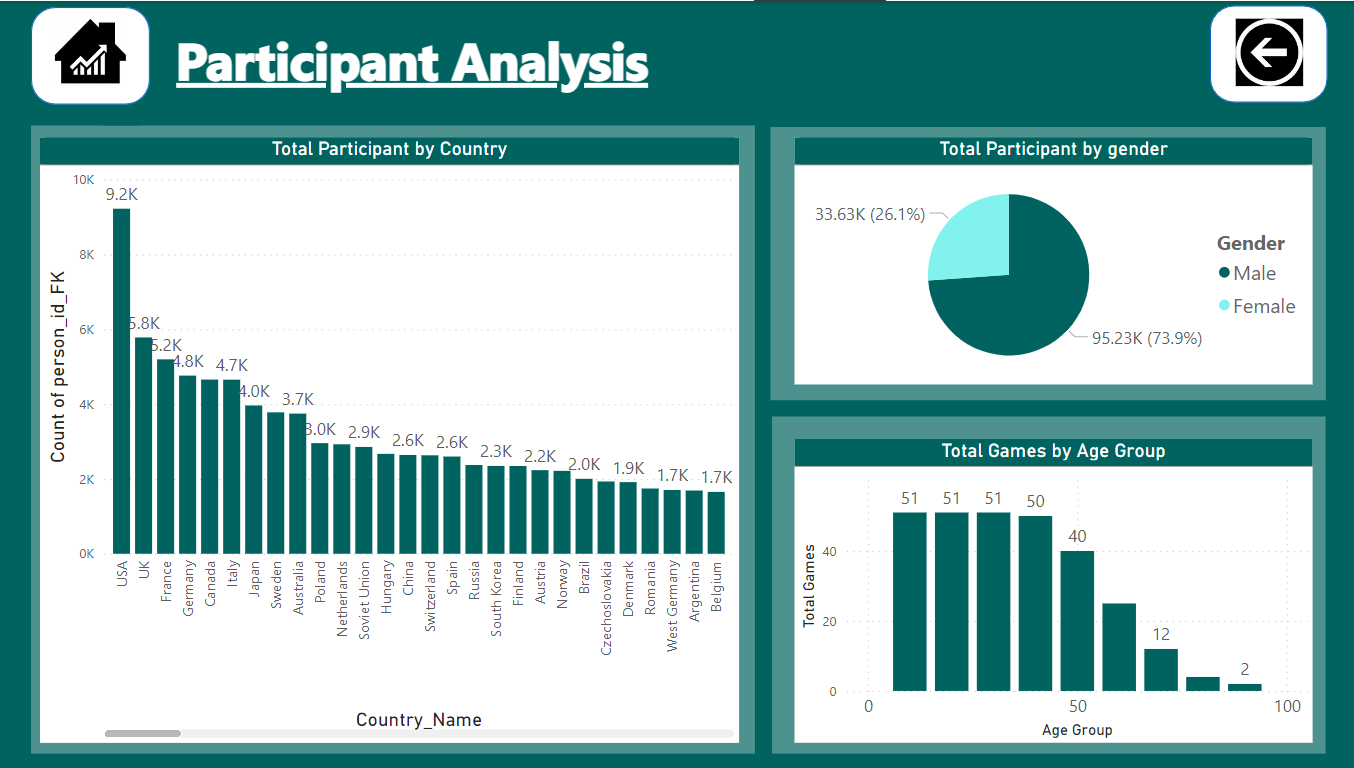
## 9. How has the number of events changed over time?

## This visual representation of how the number of events has changed over time.

## you can quickly grasp how the distribution of events across different sports has evolved over the years, allowing for easy comparison and identification of trends. This visualization tool is particularly useful for researchers, analysts, and enthusiasts who want to gain insights into the historical development of Olympic sports.

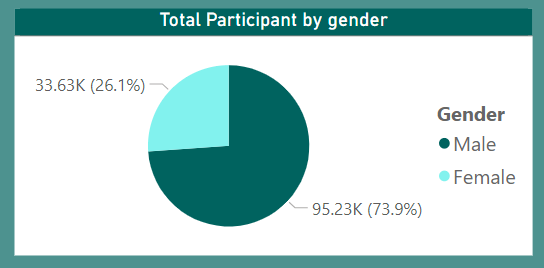


# Participant Analysis



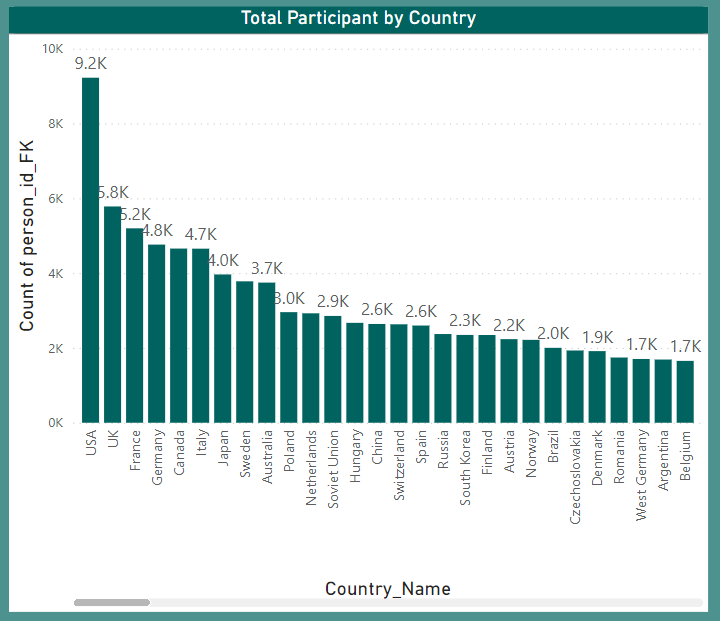
## 10. What is the distribution of participants by gender?

## Pie chart helping to visualize the distribution of participants by gender. A pie chart is a circular graph that is divided into slices to represent data. Each slice corresponds to a category, and the size of each slice is proportional to the quantity it represents. In this case, the pie chart is being used to visualize the distribution of participants by gender.



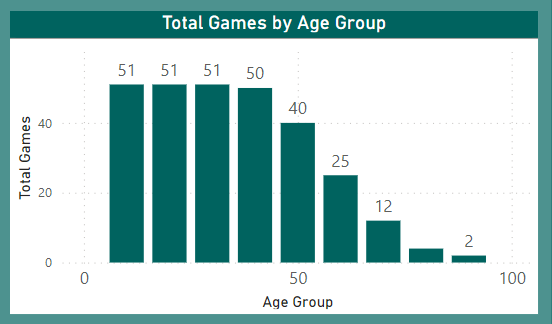
## 11. Which countries have the highest number of participants in the Olympics?

## This visual representation of countries with the highest number of participants helps us understand which countries have consistently sent the largest contingents to the Olympic Games. By visually representing this data, we can see at a glance which countries have consistently sent the largest groups of people to participate in the Olympic Games over a period of time. This information can provide insights into the level of commitment and investment that different countries have towards Olympic competition. It can also give us an idea of which countries have a strong tradition or interest in sports and the Olympic movement.

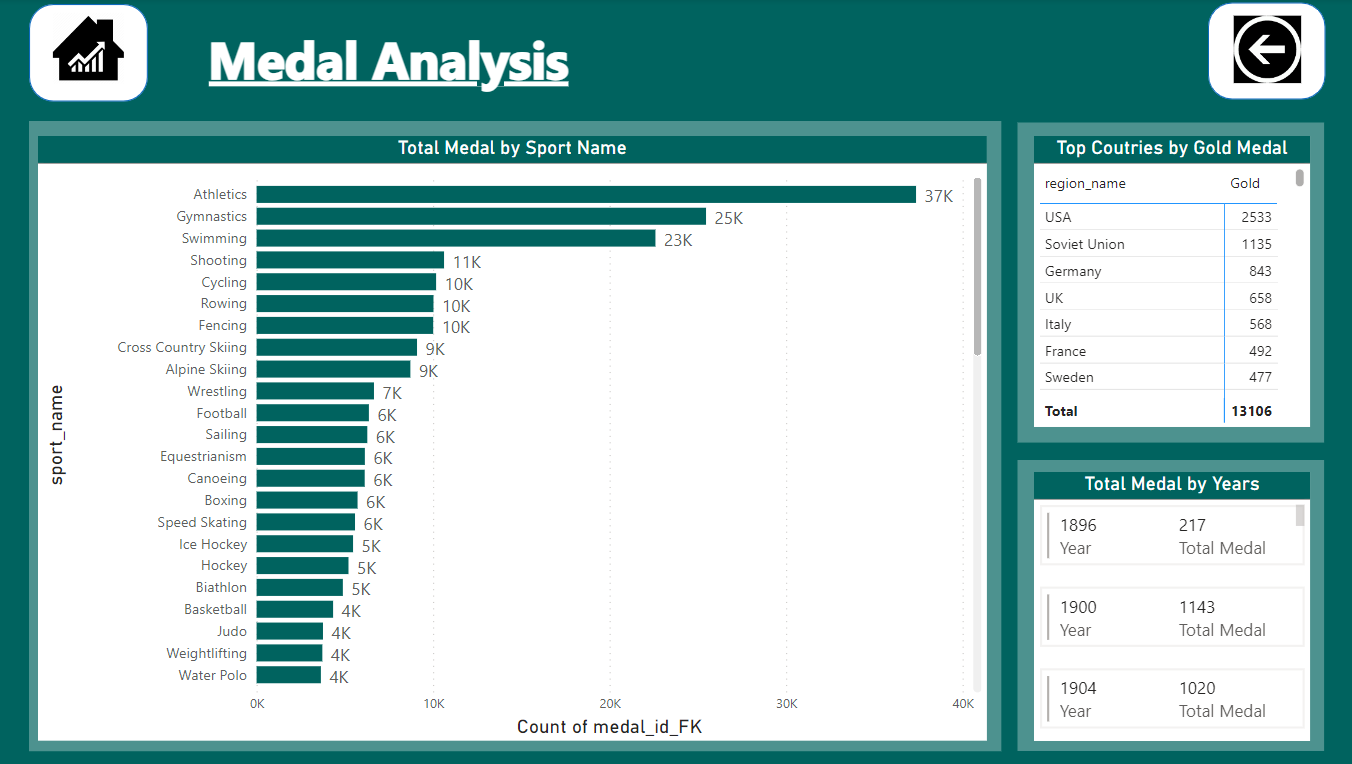


## 12. How does the age distribution of participants vary across different games?

## This analysis uses explore the age distribution of participants in various games. By visualizing the data, we aim to gain insights into the demographics of players across different gaming experiences. This chart examines the age distribution of participants in different games, with the aim of extracting meaningful insights about the demographics of players in various gaming experiences.

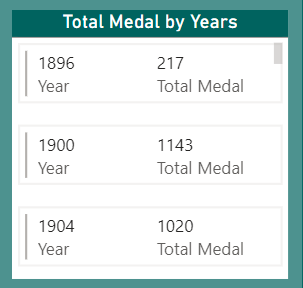


# Medal Analysis



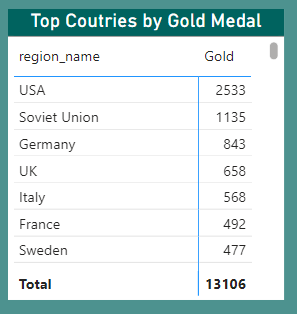
## 13. How many medals have been awarded in each Olympics?

## This Power BI report provides a visual representation of the number of medals awarded in each Olympic year. It offers insights into the historical trends of medal distribution across different Olympic games. This chart serves as a powerful tool for exploring and understanding the historical data related to Olympic medals, allowing users to gain insights into the trends and patterns of medal distribution over the years.



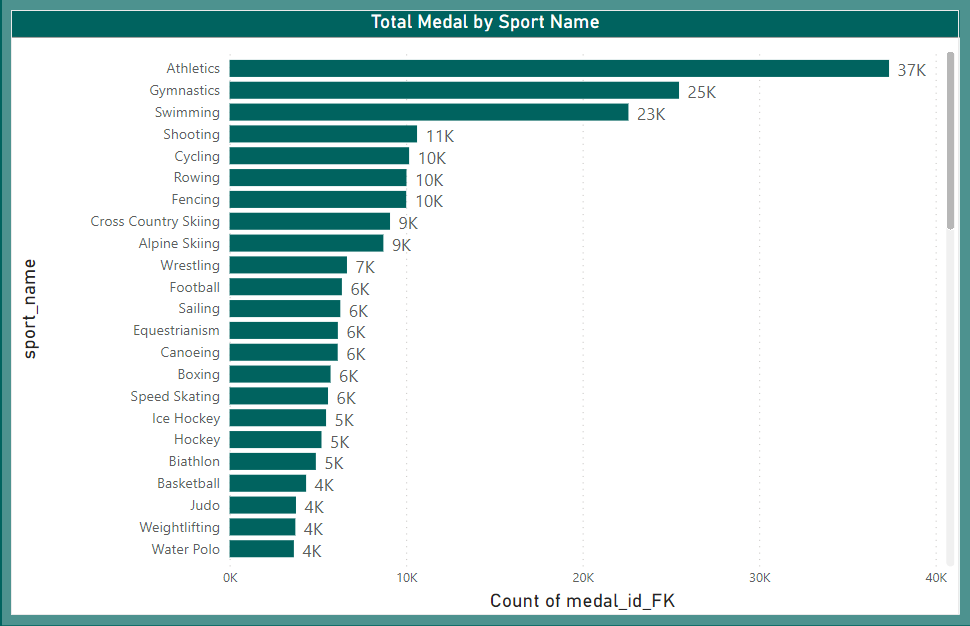
## 14. Which countries have the highest number of gold medals?

## In this, we have visualized the distribution of gold medals across different countries. The data provides us with a comprehensive overview of which countries have achieved the highest number of gold medals. You would be able to quickly identify which countries have achieved the highest number of gold medals in the given context, providing a clear and visual overview of the medal distribution.

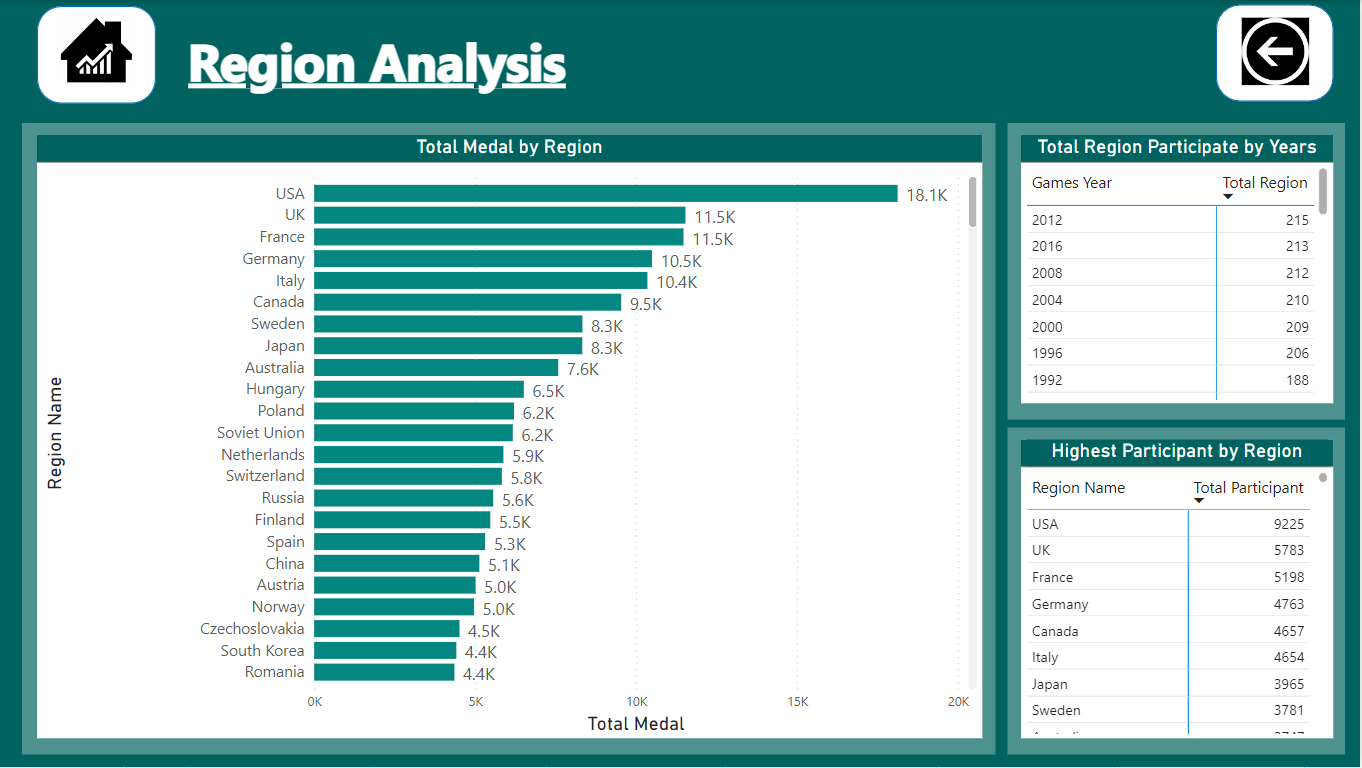


## 15. How does the medal distribution vary across different sports?

## Through this analysis, we gained valuable insights into the distribution of medals across different sports. Into the distribution of medals across different sports" specifies what the analysis focused on. It means that the study was centered around understanding how medals (presumably in a competitive event like the Olympics or a similar sporting event) were divided or awarded among various sports. This could involve looking at which sports received more or fewer medals, which countries excelled in certain sports, or any other patterns related to how medals were allocated.



# Region Analysis



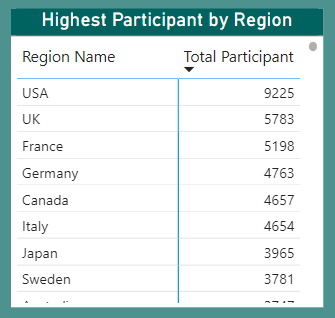
## 16. How many regions or NOCs participate in each Olympic Games?

## This report provides an insightful view of the number of NOCs participating in each Olympic Games over a specified period. The report is designed to help visualize the global participation trends in the Olympic Games. The report offers valuable and informative perspectives on the number of NOCs involved in various Olympic Games. It likely provides data and analysis that can give readers a better understanding of how different countries have been participating in the Olympics.



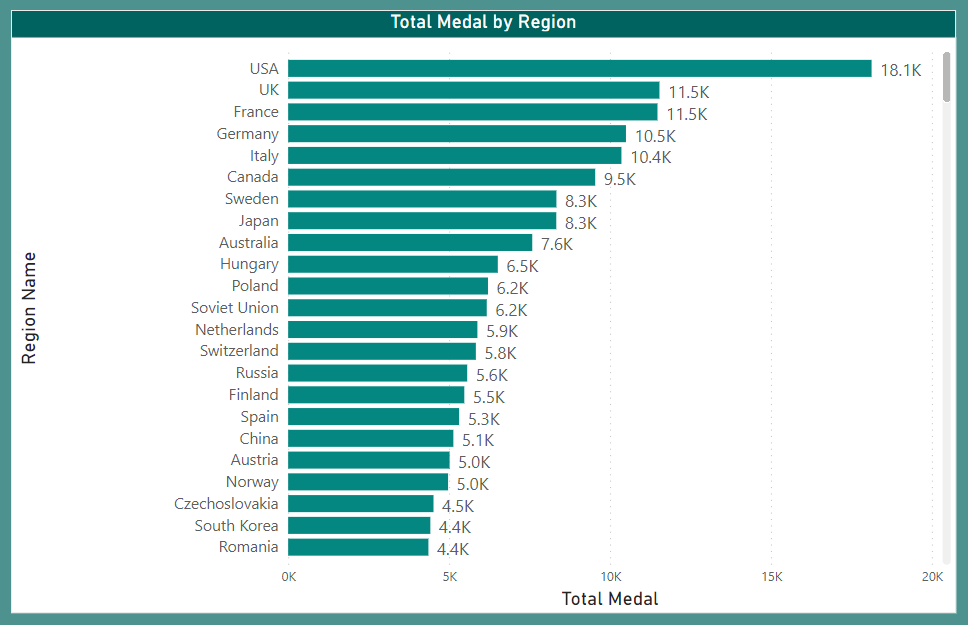
## 17. Which regions have the highest number of participants in the Olympics?

## This report aims to analyze and visualize participant data from the Olympics to identify the regions with the highest number of participants. the report's purpose is to examine data related to participants in the Olympics, process it to find regions with the most participants, and then present this information in a visual format. This could be useful for various purposes, such as understanding the distribution of athletic talent across different regions, planning for future Olympic events, or even for marketing and sponsorship considerations.



## 18. What is the distribution of medals among different regions?

## In this, we are using to analyze and visualize the distribution of medals among different regions, such as region, to gain a better understanding of which regions have received the most medals in a particular competition. we're using a method to examine and display information about how medals are distributed among different regions. This will help us see which regions have performed the best in a particular competition. For instance, if we're talking about the Olympics, it would show us which countries have won the most medals.



EDA – Problem Statement

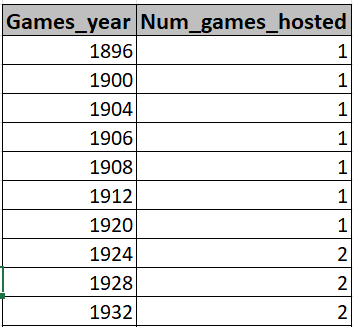


## 1.Are there any trends or patterns in the frequency of hosting Olympic Games?

## Each row represents a year in which the Olympic Games were hosted, along with the number of games hosted in that year. The result will be sorted by the year in ascending order.

## This means that in the dataset or table, each row corresponds to a specific year in which the Olympic Games took place. Alongside each year, there is information about how many Olympic Games were hosted in that particular year.

## This indicates that after you perform some sort of operation or analysis on the dataset or table, the final output will be arranged in a particular way. Specifically, it will be sorted based on the year, from the earliest year to the latest year. This is known as ascending order.



## 2.How has the duration of Olympic Games changed over time?

## It retrieves information about the Olympic Games, specifically the name and year of each game, as well as the following year's game and the duration between consecutive games.

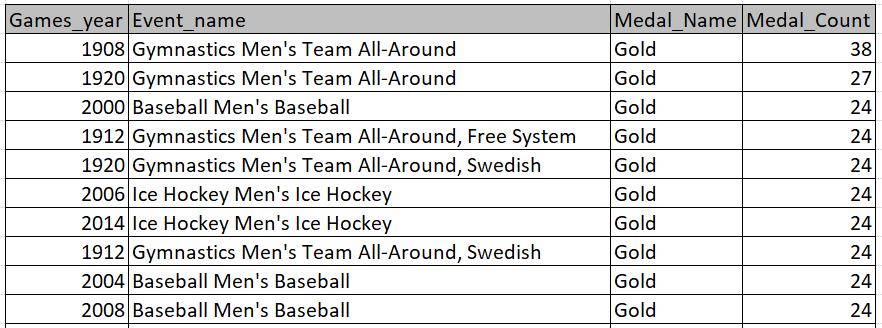
## Retrieves information about the Olympic Games: This means that the process or system is designed to collect data or details about various instances of the Olympic Games. This information could include a variety of things such as dates, locations, participants, events, and more.

## Specifically, the name and year of each game: This indicates that one of the pieces of information it gathers is the title (or name) of each individual Olympic Games, along with the year in which each of them took place. For example, "2016 Rio Olympics" or "2020 Tokyo Olympics".



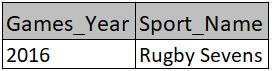
## 3.Are there any notable events or occurrences associated with specific Olympic Games?

## It specifically focuses on events where a gold medal was awarded. It retrieves information about notable events or occurrences associated with specific Olympic Games, focusing on events where gold medals were awarded. The result will include the year of the Games, the event name, the name of the medal (which will always be gold in this case), and the count of gold medals awarded for each combination of Games year and event name.



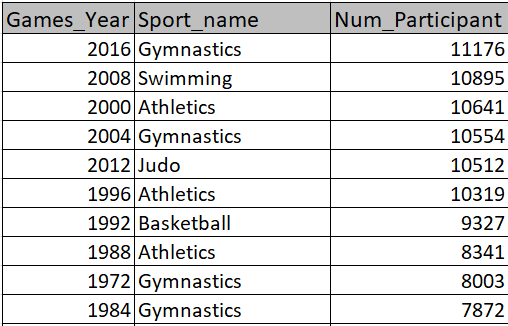
## 4.Are there any emerging sports that have been recently added to the Olympics?

## It does this by identifying sports that have not been included in the Olympics before the most recent games year. It retrieves distinct combinations of games year and sport name, but only includes sports that have not been part of the Olympics before the most recent games year. This allows you to identify emerging sports that have been recently added to the Olympics.



## 5.How has the popularity of certain sports changed over the years?

## It retrieves information about the number of participants in each sport for each year. It provides a report that shows the number of participants in each sport for each year, ordered by the popularity of the sports (in terms of participant count) in descending order. This information can be useful for understanding how the popularity of different sports has changed over time.

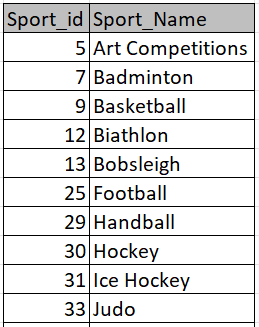


## 6. Are there any sports that are specific to a particular region or culture? **Note:** In light of the absence of pertinent data in the dataset, I will now proceed to examine a fresh perspective –

## : -Are there any sports of the sports events that are only held in one city?

## This display is unique sports and their corresponding IDs, but only for sports that have a single associated city. It providing a list of sports along with their unique identification numbers, but it's only showing sports that are linked with a single city. This could be useful in scenarios where you want to focus on sports that are strongly tied to a particular location.

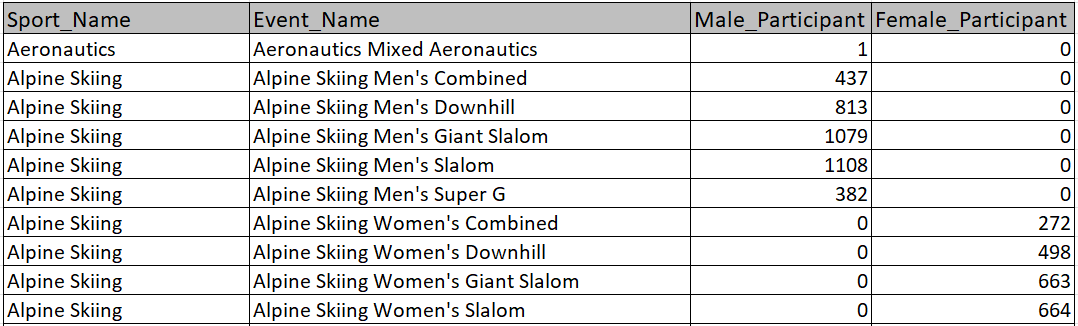
## This means that the display is showing different sports (like basketball, soccer, etc.) and for each sport, it is also showing a specific identification number that uniquely identifies that sport. But only those sports are shown which are connected to one and only one city.



## 7.Are there any sports that have a higher number of events for one gender compared to others?

## It identifies sports and events where there is an imbalance in the number of male and female participants. It calculates these counts and presents the results with the sport name, event name, number of male participants, and number of female participants.

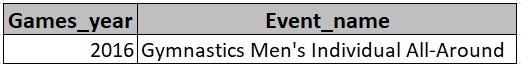
## This statement is describing a process or system that is designed to analyze sports and events to determine if there is a disparity or unevenness in the number of male and female participants. It then goes on to explain that once this imbalance is identified, the system performs some calculations to quantify the difference, and finally, it provides a report that includes specific details about the sport and event, along with the corresponding counts of male and female participants.



## 8. Are there any new events that have been introduced in recent editions of the Olympics?

## It identifies events that have not existed in previous editions of the Olympics. Then, it further narrows down the results to only include events from the most recent edition of the Olympic. This list of distinct combinations of the year and event name for events that have been introduced in recent editions of the Olympics.

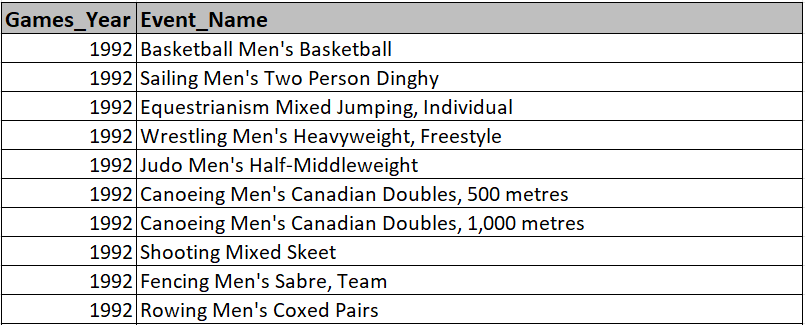
## the process is all about finding and listing the new events introduced in the most recent Olympic Games, using a combination of the year and the event name to uniquely identify each of them.



## 9. Are there any events that have been discontinued or removed from the Olympics?

## The result of the logic described in the SQL query will be a list of unique combinations of games\_year and event\_name for events that have not been discontinued or removed from the Olympics. These combinations represent specific Olympic Games years along with the names of the events that took place in those years.

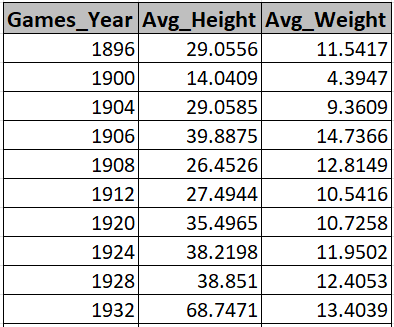
## For example, if there were Olympic Games held in 2020 and 2024, and various events took place in each of those years



## 10. Are there any notable trends in the height and weight of participants over time?

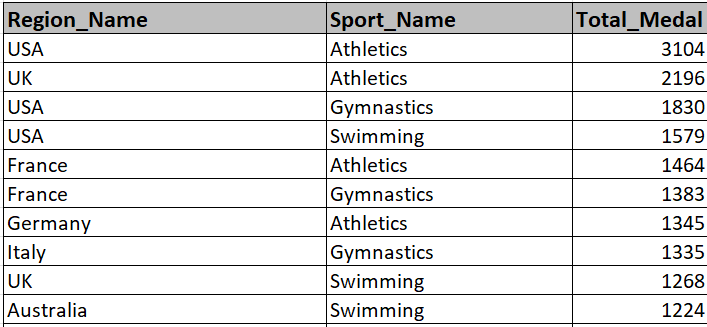
## Each row will represent a unique year, and it will show the average height and weight of participants for that year. This information can be used to identify any notable trends in the height and weight of participants over time.

## this approach allows for a systematic and structured way to capture and analyze information about the average height and weight of participants in a study, with the aim of discerning any meaningful patterns or trends over time.



## 11. Are there any dominant countries or regions in specific sports or events?

## "Which regions or countries are dominant": This part of the query is asking about places in the world that are particularly strong or successful in certain sports. For example, some countries might be exceptionally good at swimming, while others might excel in track and field events. It emphasizes that the question is not asking about overall performance across all sports, but rather, it's focusing on individual sports or categories of sports. For instance, someone might be interested in knowing which countries tend to perform best in sports like soccer, basketball, or swimming. In this context, the number of medals won in a particular sport is used as a measure of success. For example, in the Olympic Games, countries are awarded gold, silver, and bronze medals for first, second, and third place respectively in each event.

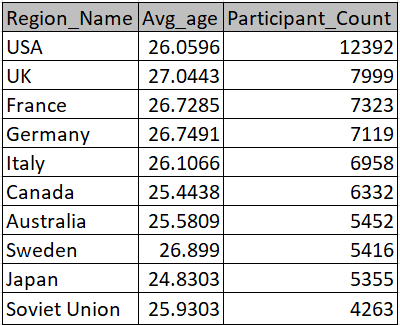


## 12.What factors contribute to the success or performance of participants from different countries?

**Note:** In light of the absence of pertinent data in the dataset, I will now proceed to examine a fresh perspective.

## : - What is the average age of participants and the count of participants in each region?

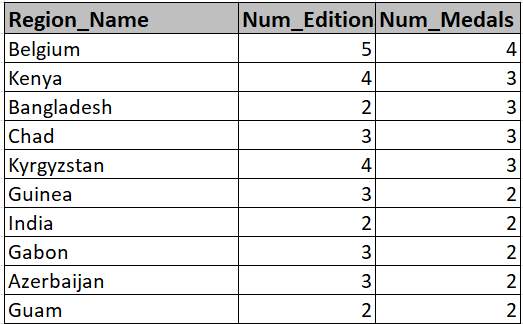
## It displays the average age of participants and total participant, it show you the average age of all the participants and also tell you how many participants there are in total.



## 13. Are there any countries that consistently perform well in multiple Olympic editions?

## It includes countries that have participated in at least two Olympic editions and displaying the region name, the number of editions they participated in, and the number of medals they have won, sorted by the number of medals in descending order. This helps identify countries that have consistently performed well in multiple Olympics.

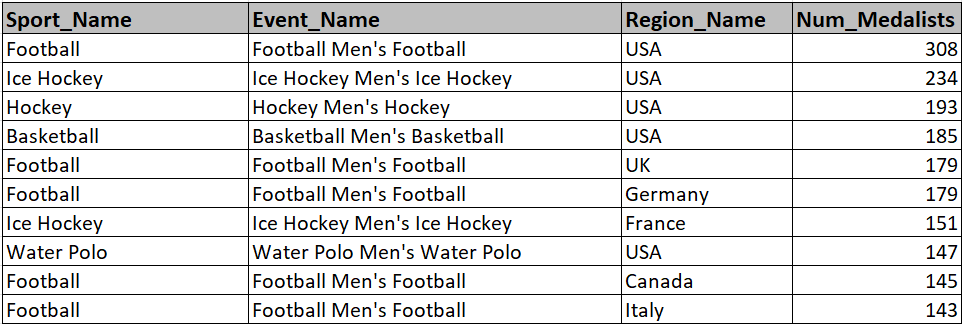
## countries (represented by regions in the noc\_region table) that have consistently performed well in multiple Olympic editions. It calculates the number of editions a country has participated in (num\_editions) and the total number of medals won by that country (num\_medals). The query then filters the results to only include countries that have participated in at least two editions and orders the result by the number of medals in descending order.



## 14. Are there any sports or events that have a higher number of medalists from a specific region?

## The results will only include rows where there is more than one medalist from the same region in a specific sport and event combination. The table will be sorted in descending order based on the number of medalists. This means that the combinations with the highest number of medalists from the same region will appear at the top of the result set.

## Once you've identified the rows that meet the criteria mentioned in the first part, the data will be arranged in a specific order. This order is determined by the number of athletes from the same region who have won medals. The rows with the highest number of medalists will come first, and it will go in descending order, meaning the number of medalists will decrease as you go down the table.

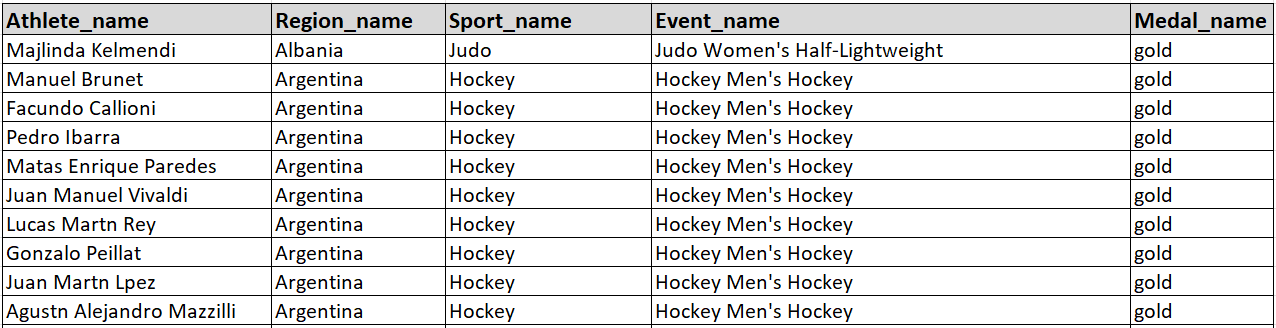


## 15. What are some notable instances of unexpected or surprising medal wins?

## Athletes who won gold medals between 2014 and 2016: This part of the statement is specifying a time frame. It's looking for athletes who achieved the highest level of success in their respective sports by winning a gold medal in a competition that took place between the years 2014 and 2016.

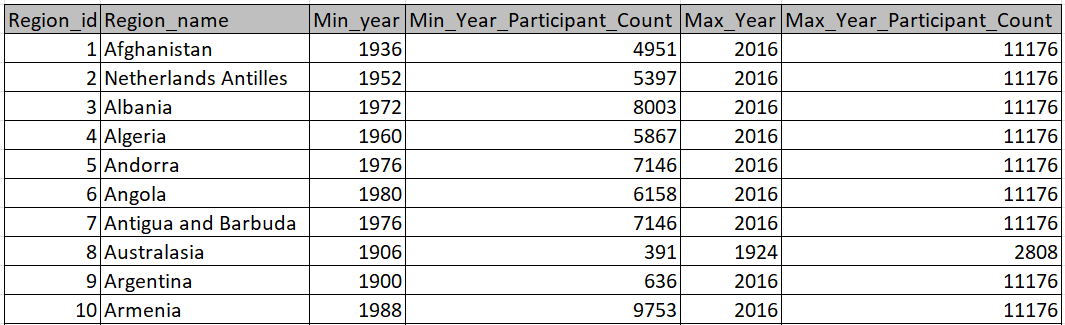
## had a previous unexpected or surprising bronze medal win in the specified time frame: This part of the statement is adding a condition about the athlete's prior performance. It's looking for athletes who, before winning the gold medal in the specified time frame (2014-2016), had previously won a bronze medal in a way that was unexpected or surprising.

## For example, let's say there is an athlete named John. In 2012 (which is not within the specified time frame of 2014-2016), John won a bronze medal. This alone doesn't meet the criteria, as it wasn't within the specified time frame. However, if John then went on to win a gold medal between 2014 and 2016, he would meet the criteria described in the statement.



## 16. Are there any regions that have experienced significant growth or decline in Olympic participation?

## It describes a system that can gather specific details about different regions in the context of the Olympic Games. This includes unique IDs, names, the earliest and latest years of Olympic participation, and the total count of individual participants from each region. This information is valuable for tasks like statistical analysis, historical research, and event planning related to the Olympics.



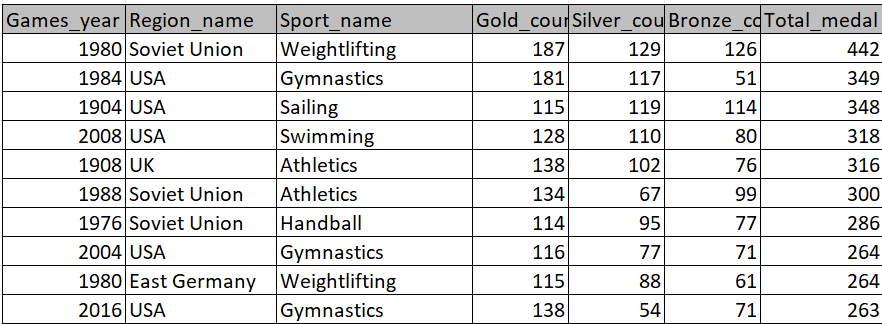
## 17.How do cultural or geographical factors influence the performance of regions in specific sports?

**Note:** In light of the absence of pertinent data in the dataset, I will now proceed to examine a fresh perspective.

## What is the distribution of gold, silver, and bronze medals won by regions in each year for different sports, along with the total number of medals won?

## It including the year of the games, the region, the sport, and various counts related to the medals awarded (Gold, Silver, Bronze) for each combination of year, region, and sport.

## To determine, this could include things like the total number of gold, silver, and bronze medals awarded, the total number of medals won by each country, or any other relevant information about the medal counts. Taken together, the statement is describing a dataset that contains detailed information about Olympic Games. For each combination of year, region, and sport, the dataset includes information about the medals awarded (Gold, Silver, Bronze) and potentially other related counts or statistics.

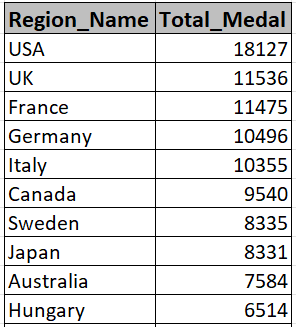


## 18.Are there any regions that have had a notable impact on the overall medal tally?

## To determine which regions or countries have made a notable impact, you count the total number of medals won by competitors from each region. This means adding up the gold, silver, and bronze medals earned by athletes representing a specific region or country.

## By calculating the total medals won by each region or country, you can assess their contribution to the overall medal tally. Regions or countries with a higher medal count are likely to have a more significant impact on the final results.

## After counting the total medals for each region, you can analyze the data to identify which regions or countries have had a notable impact. This could be done by setting a threshold or criteria, such as a minimum number of medals won, to classify a region as having a significant influence on the medal tally.



# Conclusion

## "In conclusion, the Olympic Games and sport showcased an incredible display of athleticism, determination, and sportsmanship. Athletes from around the world pushed the boundaries of human performance, breaking records and inspiring millions. Not only did we witness outstanding individual achievements, but also remarkable team efforts that highlighted the power of unity in sports. The competition was fierce, yet it was the spirit of camaraderie and mutual respect that truly defined the essence of the Olympic Games.

## The host city and organizers deserve commendation for their impeccable execution, providing a platform for athletes to shine and creating memorable moments for spectators worldwide. The Games served as a testament to the unifying power of sports, transcending borders, cultures, and backgrounds.

## As we reflect on these extraordinary Games, it is important to celebrate not only the medalists, but all the athletes who dedicated themselves to reaching this pinnacle of competition. Their journeys are a testament to the dedication, sacrifice, and passion required to excel at the highest level.

## Ultimately, it will be remembered for its exceptional moments, leaving a lasting legacy of inspiration for generations to come. It reinforced the belief that in the pursuit of excellence, barriers can be overcome, and dreams can be realized. These Games were a true celebration of human achievement and the enduring spirit of the Olympic movement.