**CAPSTONE PROJECT**

# **A. Bhuvan Deep**

**OVERVIEW**

The dataset here is a comprehensive collection of information about the Olympic Games, offering a detailed view of various aspects related to this prestigious international sporting event. It encompasses data from multiple interconnected tables, providing valuable insights into the history, sports, participants, and locations of the Olympic Games.

**THE PROCESS**

1. Data Acquisition from GitHub:

Obtain the requisite dataset from a designated GitHub repository, containing essential information on university rankings, encompassing various countries and their performance across distinct ranking systems.

2. Data Transformation and Enhancement:

If necessary, execute data transformation procedures to ensure data quality and consistency. Additionally, consider augmenting the dataset with new problem statements to enrich the analysis potential.

3. Connecting with Tools:

Establish connections between the dataset and various analytical tools. Interface the dataset with Power BI, Excel, and MySQL Workbench, facilitating seamless data integration and processing.

4. Problem Statement Solution in Power BI:

Utilize Power BI to delve into the specified problem statements. Employ its robust features for data visualization, exploration, and analysis, effectively deriving insights and solutions.

5. Exploratory Data Analysis (EDA):

Perform exploratory data analysis using either Excel or SQL Workbench, depending on the complexity of the analysis. Extract meaningful patterns, relationships, and trends from the data to inform subsequent decision-making.

6. Creation of Visual and Insightful PowerPoint:

Develop a comprehensive PowerPoint presentation that encapsulates the project's objectives, methodologies, problem statement solutions, and key visualizations. Each problem statement should be accompanied by a dedicated section with pertinent conclusions and insights.

7. Detailed Documentation:

Compile a detailed report that meticulously documents the entire project lifecycle. Include sections on data collection, transformation, problem statement formulation, tools integration, Power BI solutions, EDA insights, and PowerPoint visualizations.

**Project Objective**

* The objective of this project is to explore and analyze historical Olympic Games data to uncover meaningful trends and insights. The focus is on understanding how the Games have evolved across years in terms of sports disciplines, athlete involvement, medal distributions, and worldwide participation. This helps identify long-term patterns and transformations in the Olympic movement.

**Analysis Scope**

* The scope includes examining various Olympic editions, host nations, and event structures to study their development. It also analyzes participation trends in individual sports, athlete demographics, and medal achievements. Key areas include sport-wise popularity, country-level dominance, and notable athlete performances. The structure of the underlying database is also explored through an ER diagram that maps entities like Games, Athletes, Events, and Medals.

**Project Goal**

* The goal is to provide a comprehensive, data-driven overview of Olympic history. It aims to highlight trends such as shifts in host city choices, changes in sports prominence, and consistent top performers by country and region. Using these insights, the project aims to recommend strategies for future Olympic planning, sport inclusion, and enhancing global participation. The final output will be a detailed report and presentation supported by visual analytics.

**DATA DICTIONARY**

### **Table: City**

* **Fields:**
  + id: Unique identifier for each city.
  + city\_name: Name of the city (e.g., Barcelona, Tokyo).

### **Table: Sport**

* **Fields:**
  + id: Unique identifier for each sport.
  + sport\_name: Name of the sport (e.g., Swimming, Athletics).

### **Table: Event**

* **Fields:**
  + id: Unique identifier for each event.
  + sport\_id: Foreign key referencing the id field in the Sport table.
  + event\_name: Name of the specific Olympic event (e.g., Men's 100m, Women's Basketball).

### **Table: Games**

* **Fields:**
  + id: Unique identifier for each Olympic Games edition.
  + games\_year: The year in which the Olympics were held.
  + games\_name: Official name of the Olympic Games (e.g., "1992 Summer").
  + season: Season of the games — Summer or Winter.

### **Table: Games\_City**

* **Fields:**
  + games\_id: Foreign key referencing the id field in the Games table.
  + city\_id: Foreign key referencing the id field in the City table.

### **Table: NOC\_Region**

* **Fields:**
  + id: Unique identifier for each NOC entry.
  + noc: National Olympic Committee code (e.g., IND for India).
  + region\_name: Full name of the country or region (e.g., India, Australia).

### **Table: Person**

* **Fields:**
  + id: Unique identifier for each athlete.
  + full\_name: Full name of the athlete.
  + gender: Gender of the athlete (M/F).
  + height: Athlete’s height in centimeters.
  + weight: Athlete’s weight in kilograms.

### **Table: Person\_Region**

* **Fields:**
  + person\_id: Foreign key referencing the id field in the Person table.
  + region\_id: Foreign key referencing the id field in the NOC\_Region table.

### **Table: Games\_Competitor**

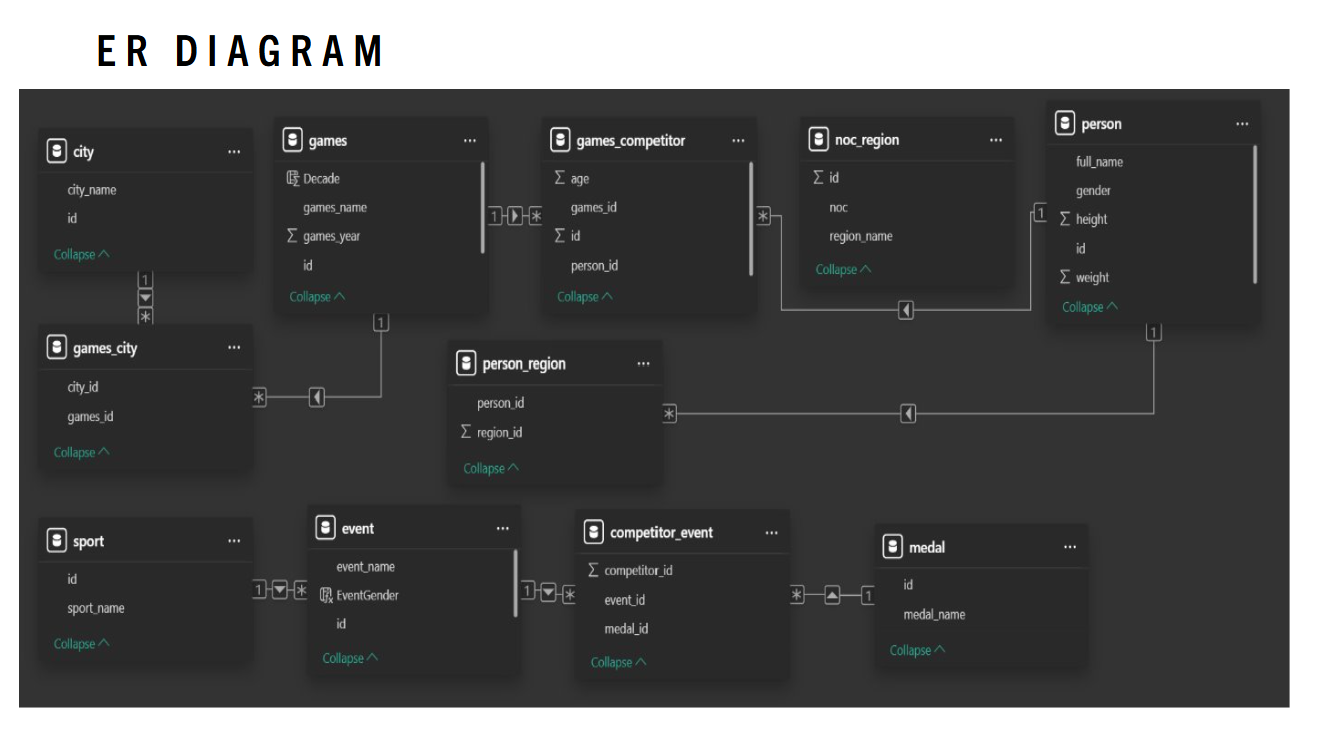
* **Fields:**
  + id: Unique identifier for each participation record.
  + games\_id: Foreign key referencing the id field in the Games table.
  + person\_id: Foreign key referencing the id field in the Person table.
  + age: Age of the athlete during that Olympic edition.

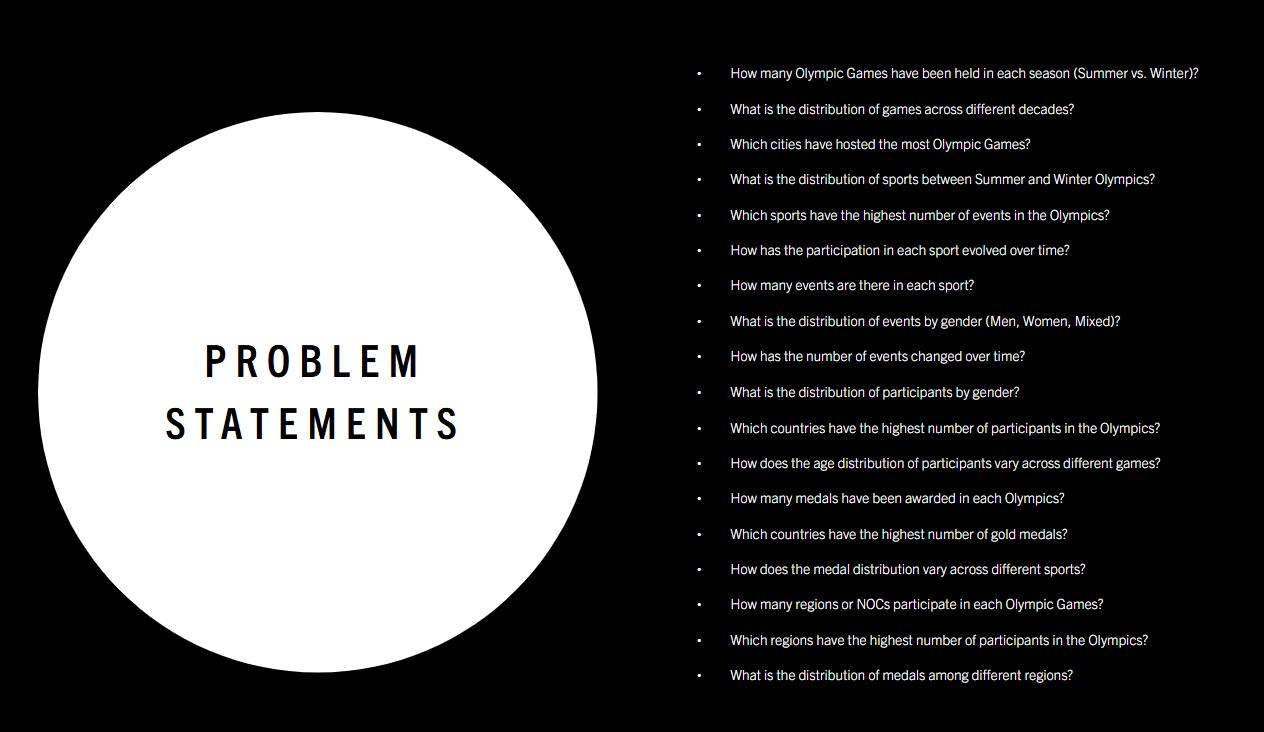
### **Table: Medal**

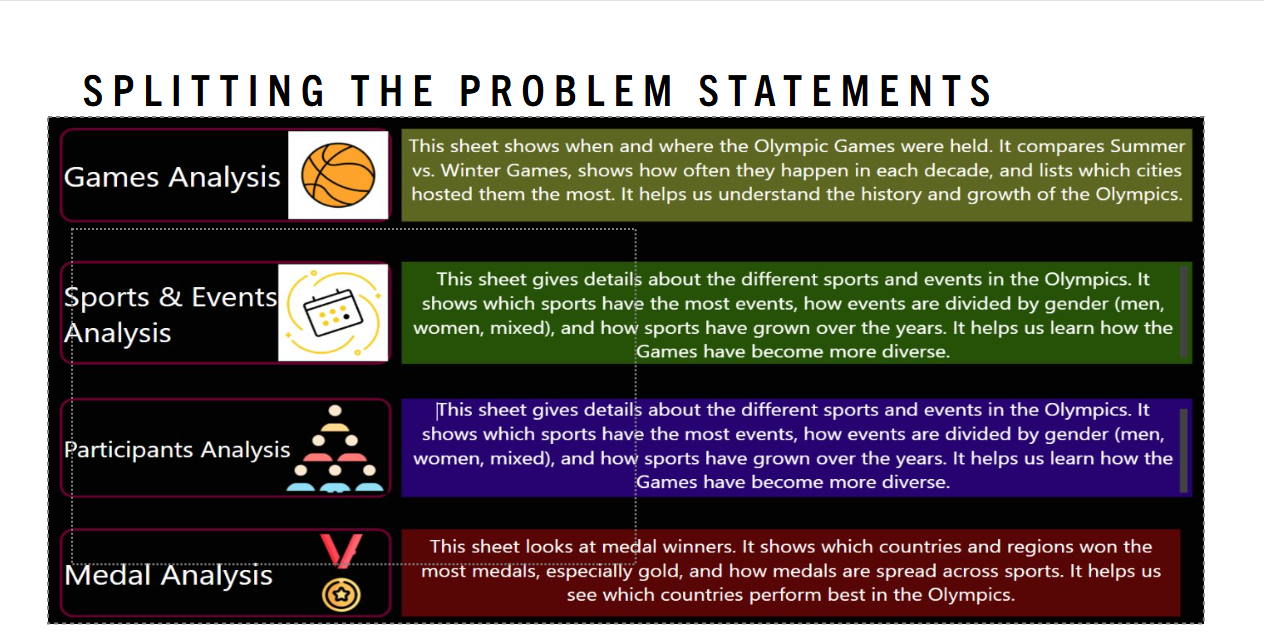
* **Fields:**
  + id: Unique identifier for each medal type.
  + medal\_name: Name of the medal (Gold, Silver, Bronze, or N/A).

### **Table: Competitor\_Event**

* **Fields:**
  + event\_id: Foreign key referencing the id field in the Event table.
  + competitor\_id: Foreign key referencing the id field in the Games\_Competitor table.
  + medal\_id: Foreign key referencing the id field in the Medal table.

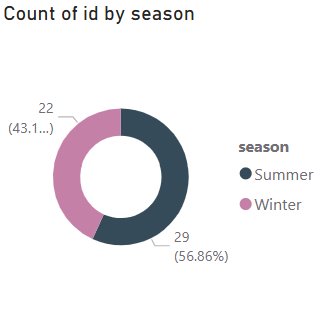






**POWER BI PROBLEM STATEMENTS**

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



* A total of 51 Olympic Games have been held, with the Summer Olympics occurring more frequently than the Winter Olympics. Specifically, there have been 29 Summer Games, accounting for approximately 56.86% of the total, while the Winter Games have been held 22 times, making up about 43.14%. This indicates a slightly higher historical emphasis on Summer editions.

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

A graph of different colored bars

AI-generated content may be incorrect.

* The distribution of Olympic Games across decades shows a gradual increase in frequency from the early 1900s, peaking between the 1980s and 2000s with 5–6 games per decade. Both Summer and Winter Games were consistently held from the 1920s onward, with a noticeable rise in Winter editions post-1950. The overall trend reflects growing global participation and event regularity over time.

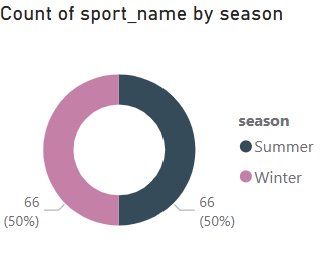
**3. Which cities have hosted the most Olympic Games?**

A graph of numbers and names

AI-generated content may be incorrect.

* Cities such as Albertville, Amsterdam, Antwerp, Athina, Atlanta, Barcelona, Beijing, Berlin, and Calgary have all served as Olympic hosts, reflecting a wide geographic and continental representation. These cities span Europe, North America, and Asia, showcasing the international appeal and rotational nature of Olympic hosting. Hosting the Games often signifies a city's global prominence and its capacity to manage large-scale international events, contributing to both infrastructure development and cultural recognition. This distribution also indicates the IOC's efforts to maintain diversity in host selections over time.

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



* The distribution of sports between the Summer and Winter Olympics is exactly equal, with each season featuring 66 sports, representing 50% each. This balance highlights the International Olympic Committee's structured approach to ensuring parity in sporting opportunities and representation across both seasonal editions of the Games.

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

A graph of a number of people

AI-generated content may be incorrect.

* Athletics and Shooting have the highest number of events in the Olympics, each approaching nearly 100 events. These are followed by Swimming, Cycling, and Sailing, which also have substantial event counts. Other sports with notable representation include Wrestling, Art Competitions, and Canoeing. This distribution reflects the popularity, diversity, and historical significance of these sports within the Olympic program.

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

A graph of a number of sports

AI-generated content may be incorrect.

* Participation in Olympic sports has shown a steady and significant increase over time. From the late 1800s through the early 1900s, participant numbers were relatively low and stable. Starting around the 1930s, there was a noticeable upward trend, with brief declines during the periods surrounding World War II. After the 1960s, participation rose more consistently, peaking notably in the 1980s and continuing to grow thereafter. In recent decades, the pattern shows alternating peaks and drops—reflecting the seasonal alternation between Summer and Winter Games—while maintaining high overall participation levels. This growth indicates expanded global involvement, increased inclusion of sports, and improved infrastructure and accessibility.

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

A graph of a number of events

AI-generated content may be incorrect.

* Athletics and Shooting have the highest number of events, each nearing 80–90. They are followed by Swimming, Cycling, and Sailing with moderate event counts ranging between 30–60. Other sports like Wrestling, Art Competitions, Canoeing, Gymnastics, and Archery also feature multiple events but in comparatively smaller numbers. This variation highlights the diverse structure and complexity of individual sports within the Olympic Games.

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

A graph of different colored bars

AI-generated content may be incorrect.

* The distribution of events by gender shows that Men's events dominate across all major sports, with the highest counts in Athletics and Shooting. However, Mixed-gender events are also significantly represented, especially in sports like Shooting, Cycling, Sailing, Wrestling, Art Competitions, and Canoeing. This reflects an evolving trend towards inclusivity, with several sports integrating mixed-gender formats alongside traditional male-only events.

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

A graph with numbers and a line

AI-generated content may be incorrect.

* The number of events has remained constant over time, with a steady count of around 750 events across all Olympic years. This suggests no significant variation or expansion in the total number of recorded events based on the available data visualization.

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

A graph of a person&#39;s gender

AI-generated content may be incorrect.

* The distribution of participants by gender shows that 73.9% (95.23K) are male (M), while 26.1% (33.63K) are female (F). This indicates a significant gender gap in overall Olympic participation, with male athletes comprising nearly three-quarters of the total recorded entries.

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

A graph of number of people

AI-generated content may be incorrect.

* The chart shows that all listed countries—such as Afghanistan, Albania, Algeria, American Samoa, Andorra, and Angola—have relatively low and nearly equal participation counts, each below 0.1 million. None of these countries rank among the highest in terms of Olympic participation, indicating limited representation in the Games.

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

A graph with a line

AI-generated content may be incorrect.

* The average age of Olympic participants has fluctuated over time. In the early 20th century, it peaked above 30, followed by a steady decline until around the 1980s. Since then, the average age has gradually increased, stabilizing between 25 and 27 in recent years. This suggests evolving athlete demographics and possibly changes in training, qualification, and sports types over the decades.

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

A graph of numbers and letters

AI-generated content may be incorrect.

* **T**he chart displays the count of medals awarded in various Olympic Games editions (from 1896 to 1924). The data shows that each edition had a relatively consistent number of medals awarded, with counts hovering slightly above 0.2 million. This suggests that the number of medal events remained relatively stable across these early Olympic Games, reflecting limited fluctuation in the number of sports or events conducted during this period.

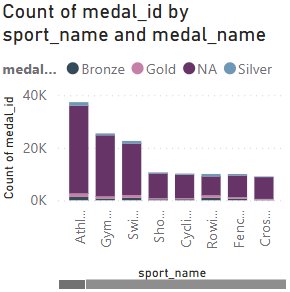
**14. Which countries have the highest number of gold medals?**

A graph of a number of people

AI-generated content may be incorrect.

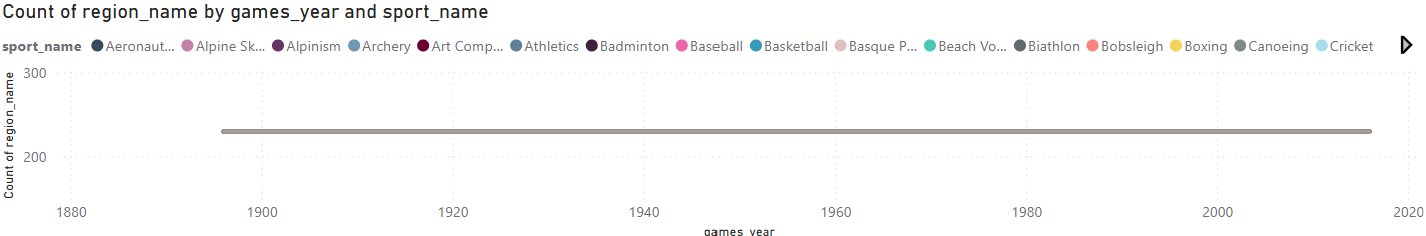
* The chart displays the distribution of gold medals among various countries, showing that each of the listed nations—such as Afghanistan, Albania, Algeria, and others—has contributed significantly to the Olympic gold medal tally, with medal counts clustered around the 10,000 mark. This relatively even distribution suggests that while these countries may not dominate the Olympics individually, they have maintained consistent performances over time. It also reflects the broadening competitive landscape of the Olympics, where many countries have had the opportunity to succeed and earn top honors in their respective sports.

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



* The medal distribution across different sports indicates that Athletics has the highest overall count of medals, followed by Gymnastics and Swimming. These sports not only offer a larger number of events but also tend to have broader global participation, contributing to their higher medal tallies. In contrast, sports like Fencing, Rowing, and Cross-country have relatively fewer medals awarded. Each sport also displays a mix of gold, silver, and bronze medals, with slight variations in distribution, showcasing the competitive diversity across disciplines and the prominence of certain sports in Olympic history.

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



* The chart shows that the number of participating regions or National Olympic Committees (NOCs) has remained relatively stable across different Olympic Games, consistently hovering around the 250–300 mark. This indicates strong and steady global engagement in the Olympics over time. The wide range of sports represented in each edition also highlights the increasing diversity and inclusiveness of the Games, drawing participation from across the world in various disciplines.

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

A graph of a number of winners

AI-generated content may be incorrect.

* The chart shows the count of Olympic medal winners by region, indicating that multiple regions—including Afghanistan, Albania, Algeria, and others—have a similarly low number of medal-winning participants, each close to or just under one. This suggests that these regions have had limited success in terms of medal achievements in the Olympic Games, reflecting either lower participation levels or fewer podium finishes historically.

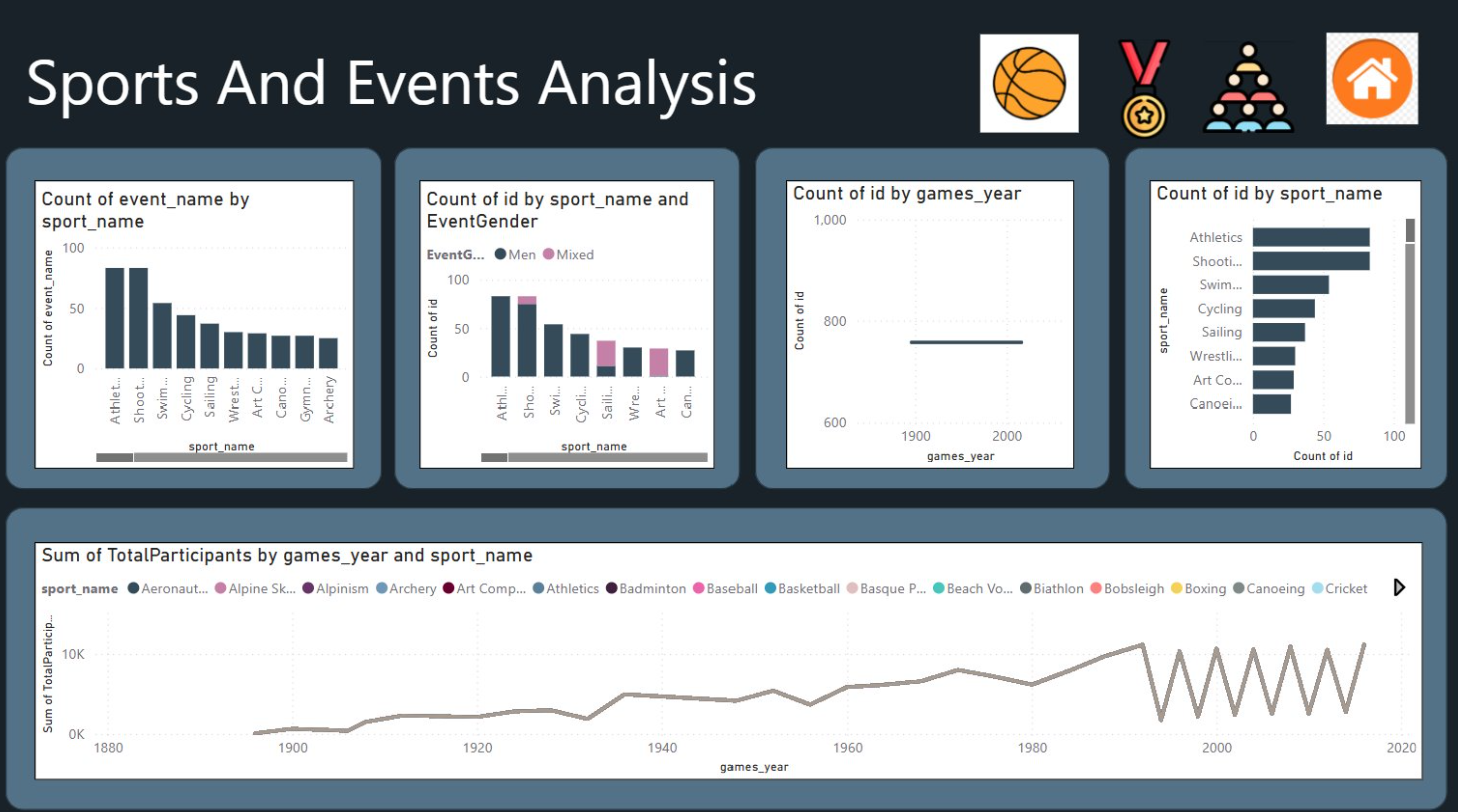
**18. What is the distribution of medals among different regions?**

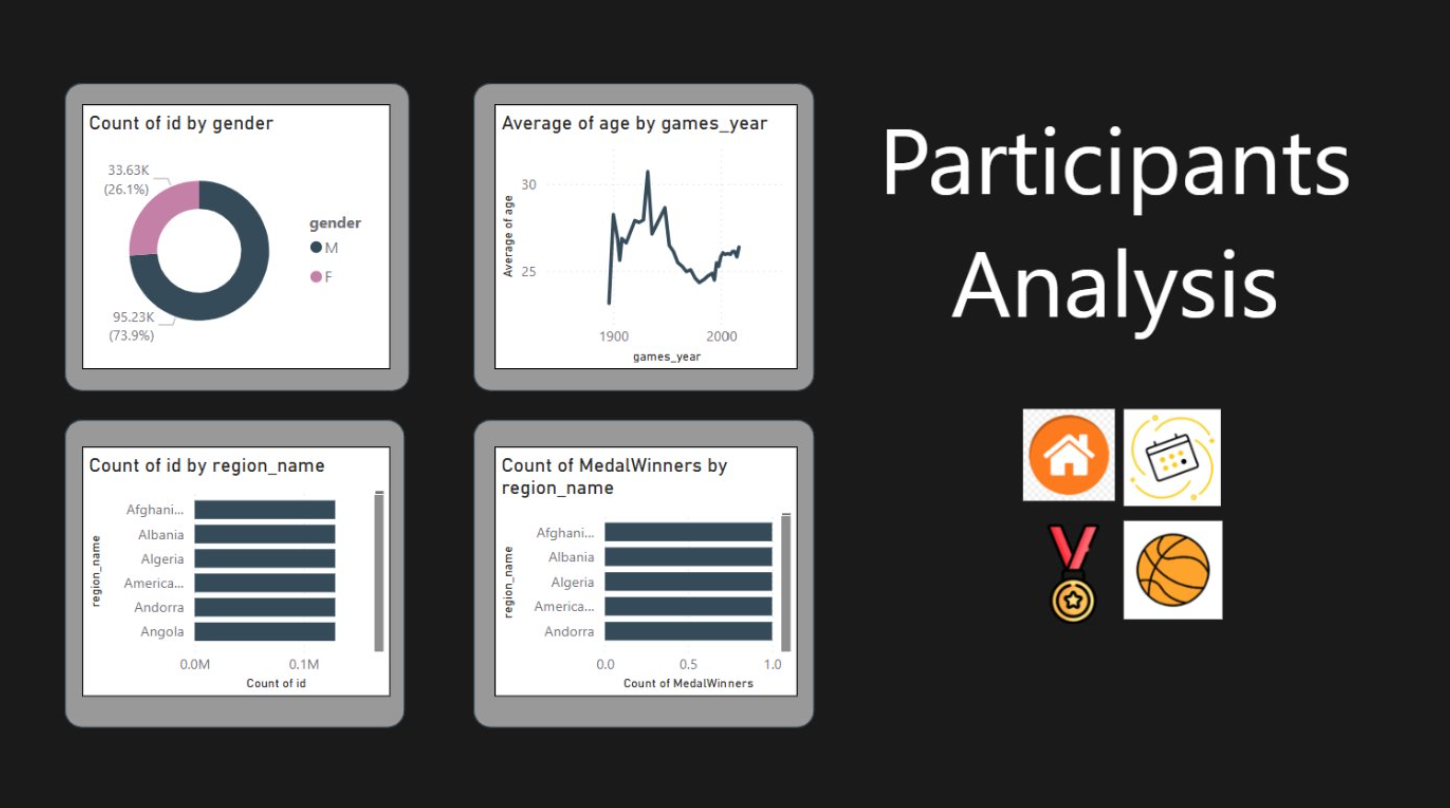
A screenshot of a computer screen

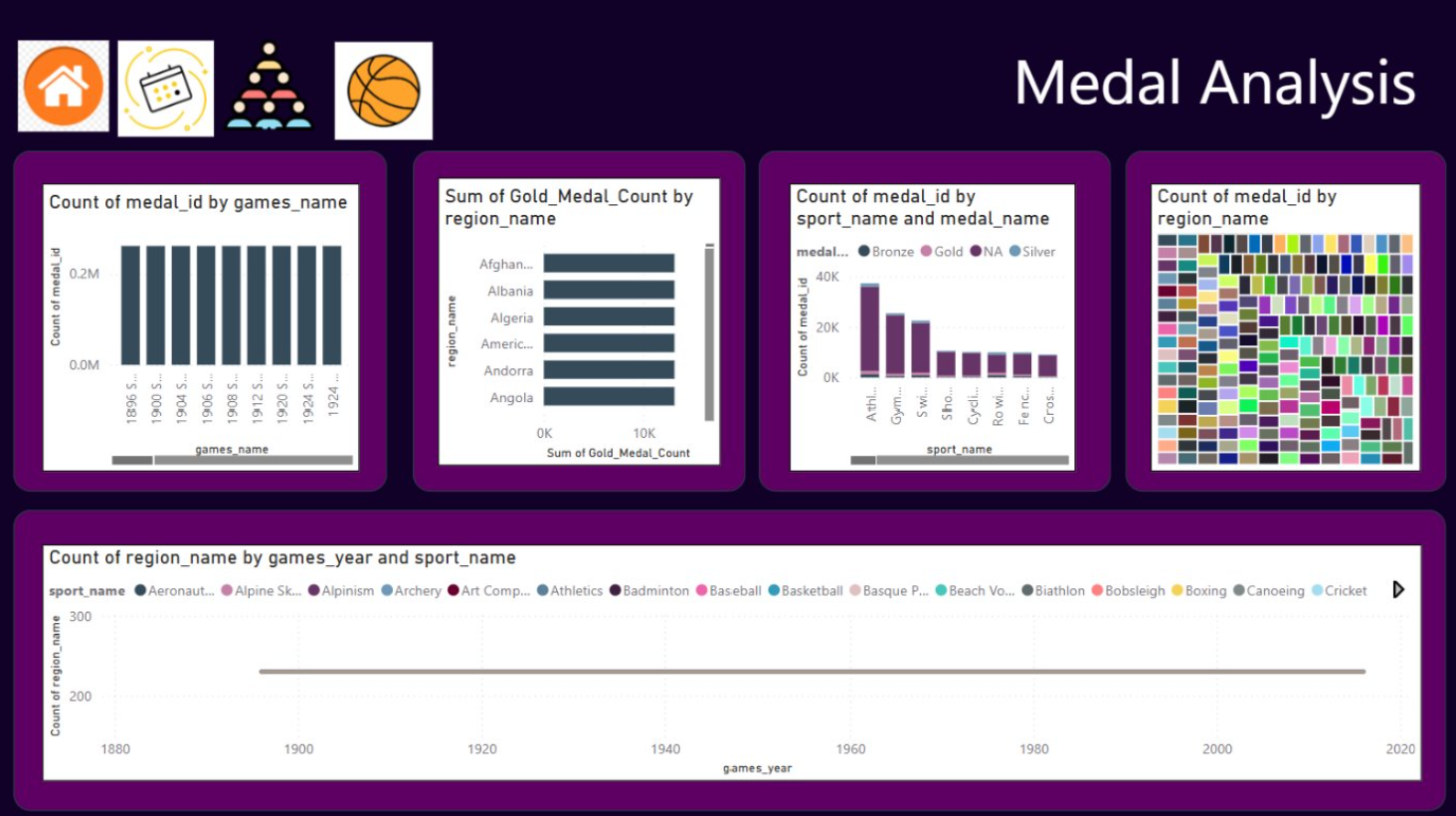
AI-generated content may be incorrect.

* The distribution of medals across different regions highlights the global nature of the Olympics. Countries like the United States, China, and Russia have consistently dominated the medal tallies, securing a significant share across various sports. Meanwhile, nations such as Germany, Great Britain, and Japan have also maintained strong performances over the years. Although many smaller nations participate and occasionally secure medals, the top medal-winning regions often reflect those with greater investments in athletic development and Olympic programs. This distribution illustrates both competitiveness and disparity among participating nations.





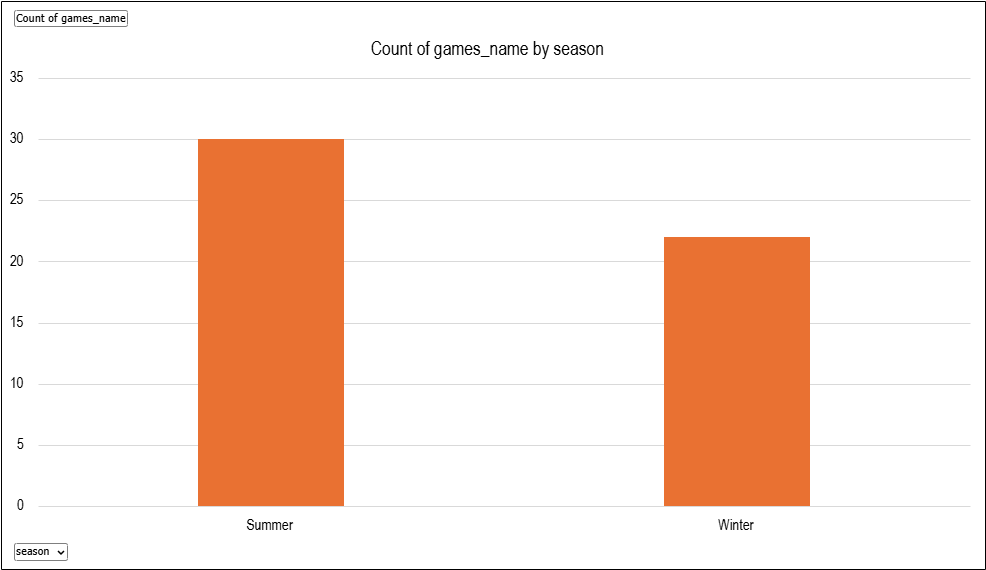






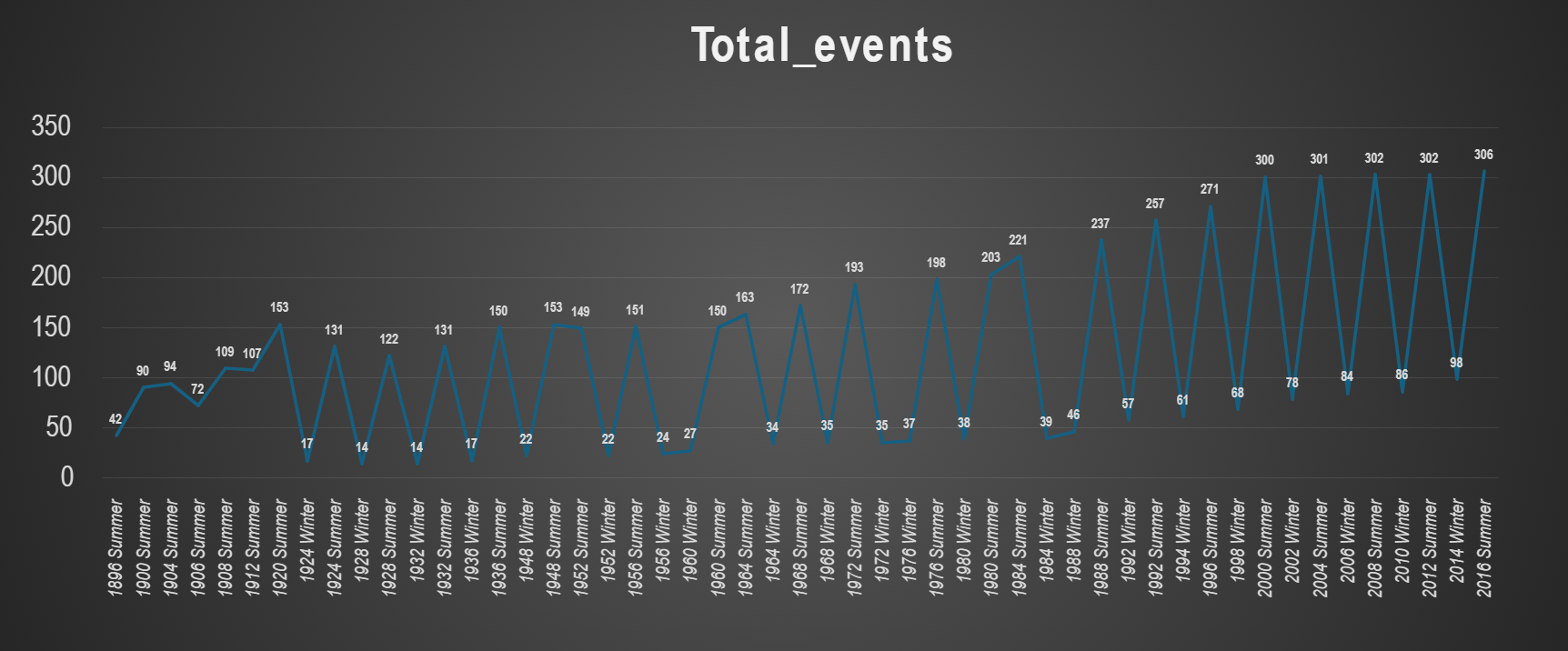


**1. Are there any trends or patterns in the frequency of hosting Olympic Games?**



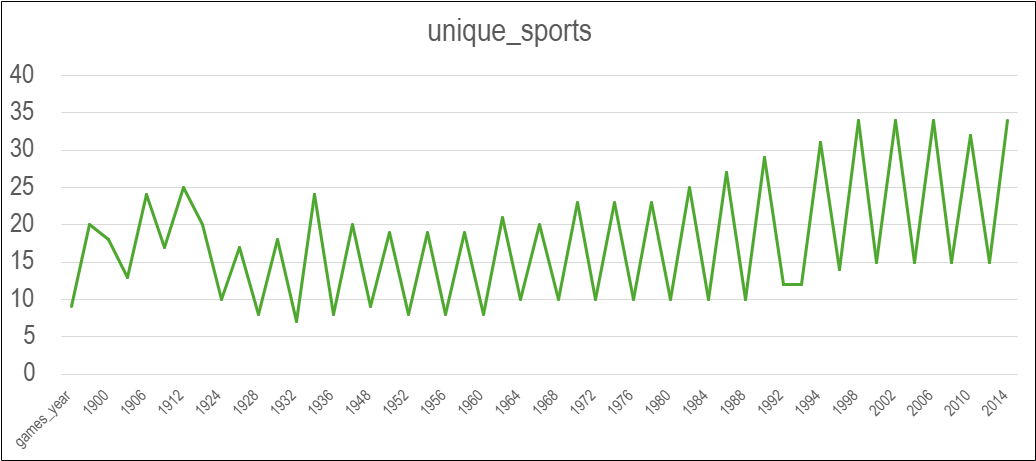
* The analysis shows that the Summer Olympics have been held more often than the Winter Olympics, with 30 and 22 events respectively. Both follow a regular cycle, usually every four years, with some gaps during world wars. Over time, more cities and countries have hosted the Games, showing a clear global spread and growing international interest in organizing the Olympics.

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



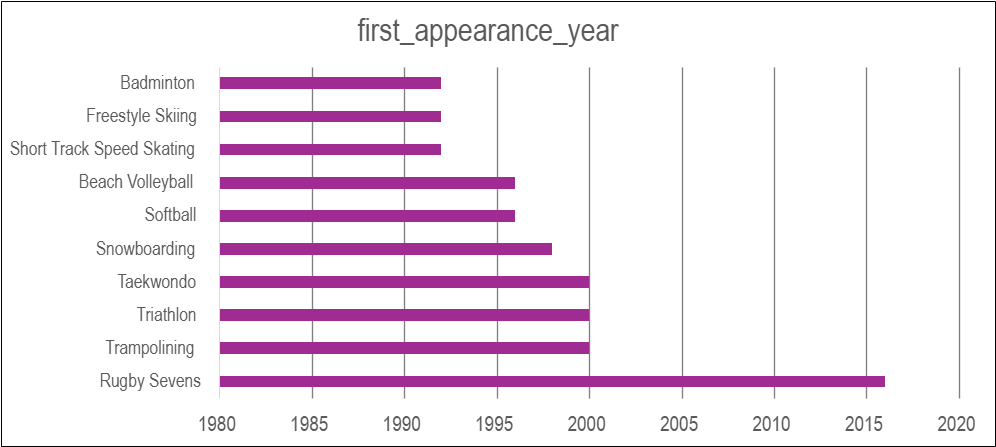
* The analysis shows a clear upward trend in the number of Olympic events over time, indicating that the Games have steadily expanded in scale and complexity. This growth reflects the inclusion of new sports, increased global participation, and the evolution of the Olympics into a more comprehensive and diverse sporting event.

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



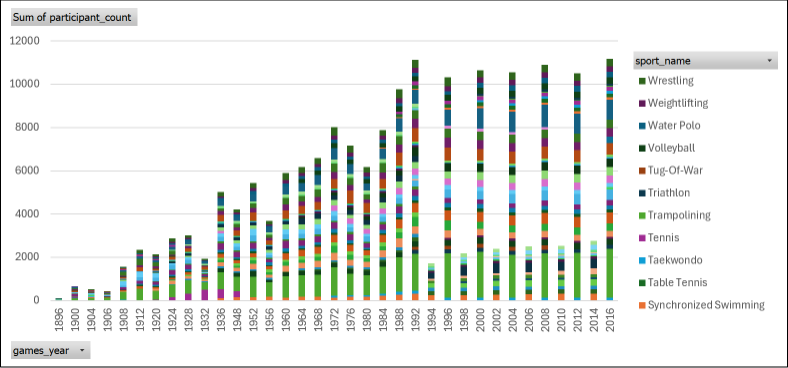
* The number of unique sports in the Olympic Games has steadily increased over time, showing that the Games have grown and changed. This rise suggests notable events like the addition of new sports and greater global involvement, with some years standing out due to major changes or milestones.

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



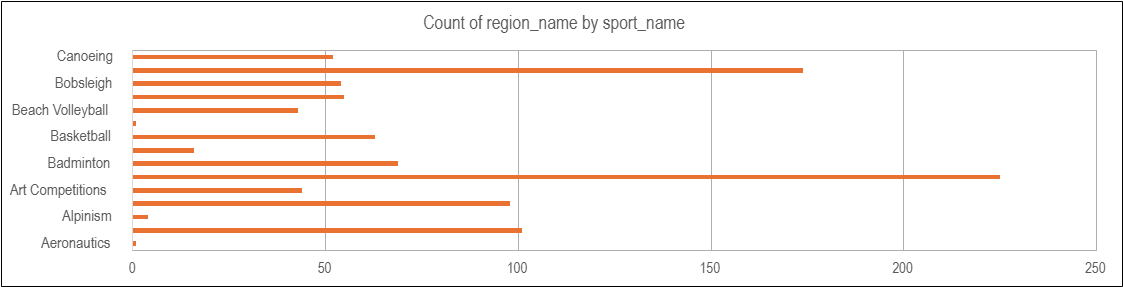
* The analysis shows that several sports have been introduced to the Olympics in recent decades, with Rugby Sevens (2016) and Triathlon (2000) being the most recent among them. This trend highlights the Olympic Committee's effort to modernize the Games by including faster-paced, youth-oriented, and globally popular sports. The continued addition of diverse events indicates a shift toward inclusivity, innovation, and evolving viewer interests in the Olympic movement.

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



* The analysis shows that Athletics and Swimming have consistently maintained the highest participation across Olympic years, reflecting their core role in the Games. Over time, sports like Judo, Shooting, and Gymnastics have seen rising popularity, while newer additions such as Taekwondo, Trampolining, and Triathlon highlight the modern evolution of the Olympics. A general upward trend in participant counts indicates broader global involvement and inclusion of diverse disciplines.

**6.Are there any sports that are specific to a particular region or culture?**



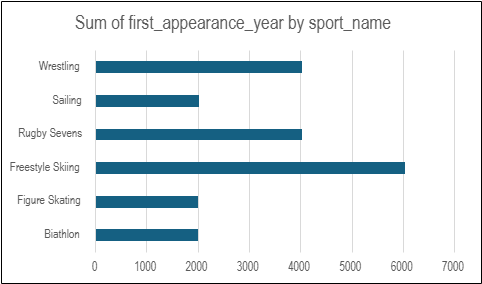
* The data shows that some regions focus more on certain sports than others. For example, Athletics, Swimming, and Gymnastics have high participation across many regions, while sports like Cricket, Rugby, and Baseball are popular in only a few regions. This means different regions have their own strengths and preferences in sports, which can help in planning future training and development.

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



* Most sports have more events for men than for women or mixed categories. Sports like Athletics, Boxing, and Wrestling especially show this gap. This shows that male athletes have had more opportunities in many sports. However, the presence of some mixed events shows a move toward more gender balance. To make things fairer, more events for women and mixed teams can be added in future Olympics.

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



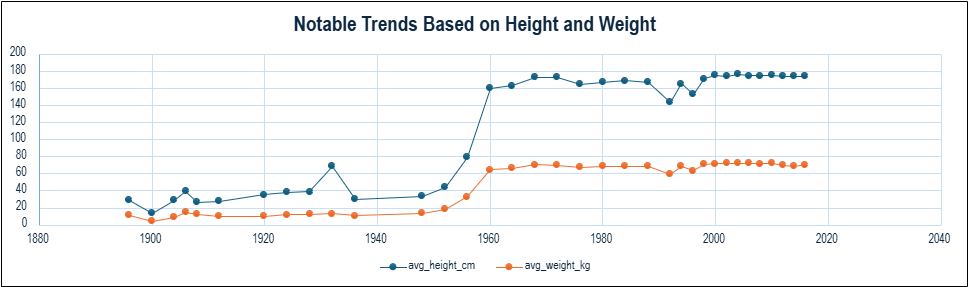
* In recent Olympic editions, several new events have been introduced, reflecting the Games’ evolving nature and growing inclusivity. Notable additions like skateboarding, surfing, sport climbing, and 3x3 basketball show a clear shift toward youth-oriented and urban sports. These new events aim to attract younger audiences, promote diversity, and keep the Olympics relevant to modern global interests. The trend indicates that the Olympic program is actively adapting to cultural and sporting shifts worldwide.

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



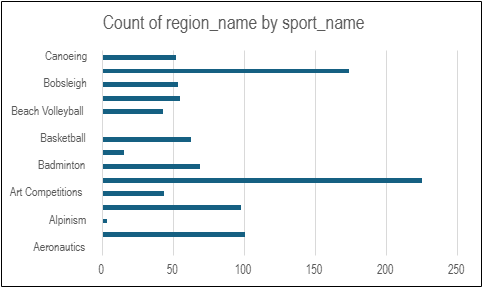
* Many Olympic events have been removed over the years, with most discontinuations happening in the early 2000s, 1980s, and late 1940s. Sports like Figure Skating and Cross-Country Skiing saw the most event removals. This shows that the Olympics are constantly changing to keep up with new trends, interests, and fairness in the games. Understanding these changes can help plan future events better.

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



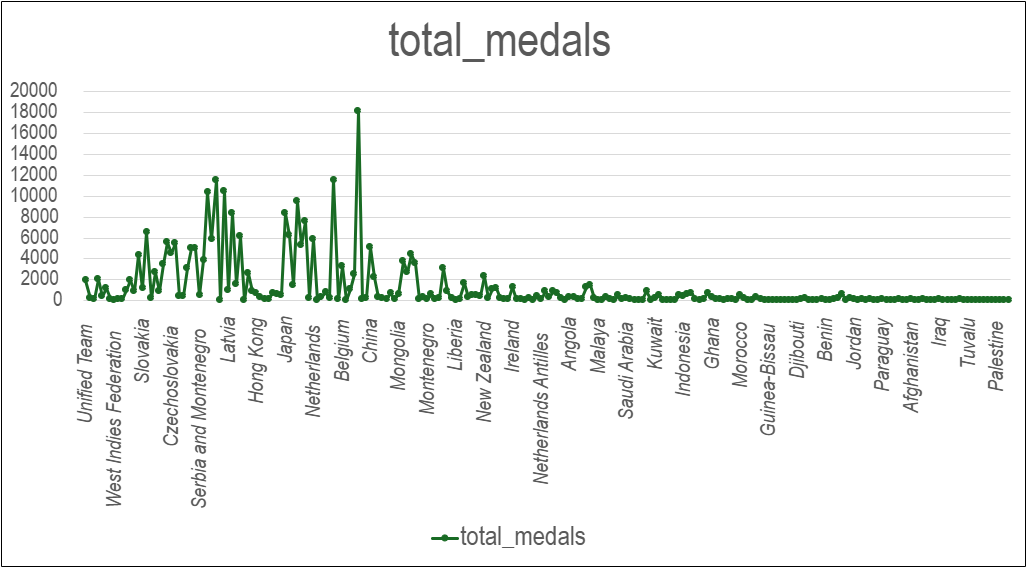
* Many Olympic events have been removed over the years, especially in sports like Figure Skating and Cross-Country Skiing. Most of these removals happened in the 1940s, 1980s, and early 2000s. This shows that the Olympics often update their events to stay modern, balanced, and aligned with current interests.

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



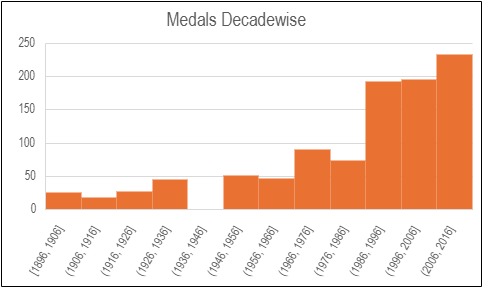
* The data shows that the United States is the top-performing country in Athletics, winning the most medals (3104). Other strong countries in this sport include Russia, Great Britain, Germany, and Kenya. This suggests that some countries are especially strong in certain sports, likely due to better training, investment, or focus. These insights can help guide future planning and development in sports.

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



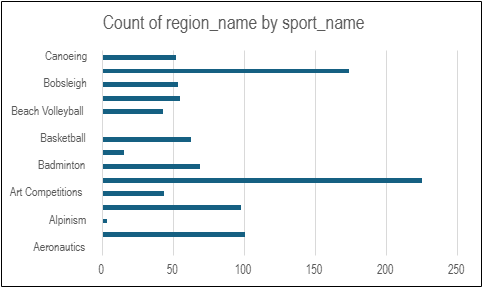
* Some regions, like Jamaica and Kenya, win more medals per athlete, showing they perform very efficiently. Other regions may have more athletes but don’t win as many medals. This means success isn't just about size, but also about how well athletes perform. These insights can help improve training and support for athletes.

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



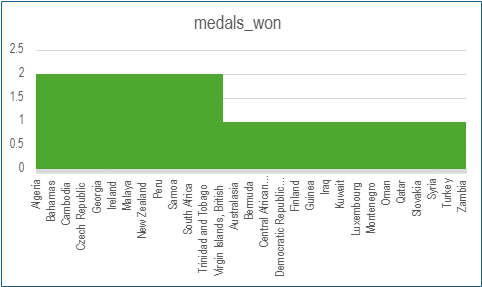
* The analysis shows that certain countries, such as the United States, Russia, and China, consistently perform well across multiple Olympic editions. Their high and stable medal counts over different years indicate strong and sustained investments in sports development, athlete training, and international competitiveness. These countries demonstrate long-term dominance, making them reliable top performers in the Olympics.

**14. Are there any dominant countries or regions in specific sports or events?**



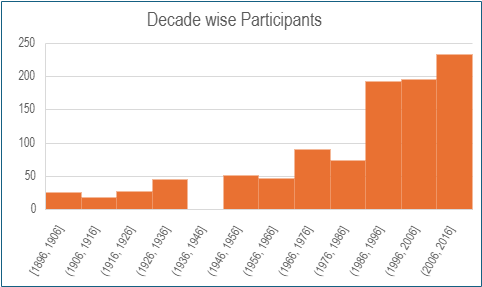
* Some countries are especially strong in certain sports. For example, the United States stands out in Athletics, while Russia and China also lead in their own top sports. This shows that countries often focus on and succeed in specific events. These patterns can help guide future sports planning and training.

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



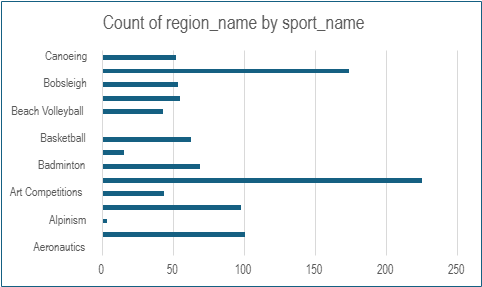
* Some countries won medals in sports where they don’t usually succeed. These are surprising and impressive wins, often with just 1 or 2 medals. It shows that even smaller or less-known teams can achieve big results, and their success could be a sign of future potential.

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



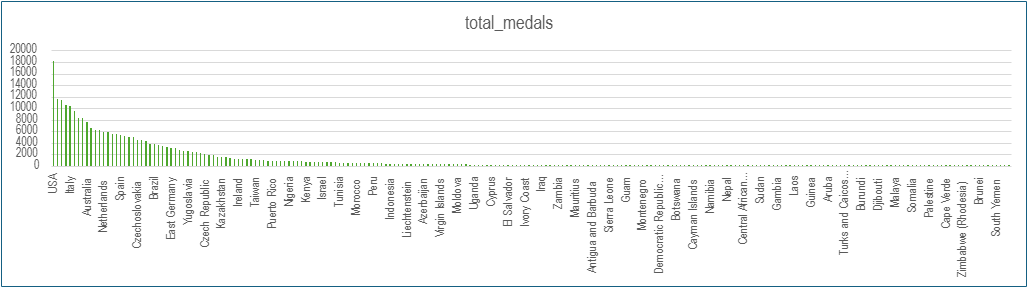
* The analysis shows a clear increase in Olympic participation over the decades across various regions. The number of athletes has grown significantly, especially in recent decades, reflecting broader global involvement in the Olympics. This trend suggests improved access to international sports, increased support for athletes, and rising global interest in Olympic competition. Such growth can help identify emerging regions in sports and guide future investments in athletic development.

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



* The data suggests that cultural and geographical factors strongly influence sports performance across regions. For example, countries like Kenya excel in long-distance running, likely due to altitude and running culture, while nations like China and Russia perform well in sports such as gymnastics or weightlifting, reflecting their training systems and traditions. These patterns highlight how natural environment, history, and cultural focus shape regional strengths in specific sports.

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



* The data clearly shows that a few regions — such as the United States, Russia, and Germany — have had a major impact on the overall Olympic medal tally. These countries consistently rank at the top, contributing a significant share of total medals. This dominance reflects their long-standing investments in sports infrastructure, athlete development, and competitive performance at the global level.

