

# **Zahida Khatoon**

**-Data Analytics-**

T---gyyy6t5trrdsw

Capstone Project- Sports Analytics



**This project aims to analyze the rich history of the Olympic Games using detailed historical data. It explores how the Games have evolved over time, focusing on the cities that hosted them, the changes in sports and events, and the introduction of new disciplines. It also examines the athletes who participated, their regions of origin, and how they were linked to each edition of the Games through the Games Competitor and Person Region data. A key aspect of the analysis is understanding medal distributions and athlete performance across different events using data from the Competitor Event table. Overall, the project provides insights into the transformation of the Olympics and highlights patterns in participation, hosting, and achievement over time.**

**This project is developed through the following key steps:-**

**1. Data Downloading and Transformation: -** The first step in the Sports Analytics project is to download the dataset from a portal or GitHub repository containing essential information on sports analytics, such as athlete details, event results, and country participation. If required, apply data transformation steps to enhance data quality, ensure consistency, and prepare it for accurate analysis.

**2. Connecting with Tools:-** Connect the dataset with tools like Power BI, Excel, and MySQL Workbench to easily work with the data and carry out analysis smoothly.

**3. Problem Statement with Power BI:-** In this step, key questions like top-performing countries or changes in Olympic participation are defined. The cleaned data is loaded into Power BI, tables are linked, and visuals like charts and maps are created to clearly show useful insights.

**4. Exploratory Data Analysis (EDA) Questions:-** In this step, EDA is done using Excel or MySQL Workbench based on the problem. The goal is to find useful patterns, relationships, and trends in the data to support better decision-making.

**5. Creating a Clear and Impactful PowerPoint:-** Create a complete PowerPoint presentation that covers the project’s objectives, methods, problem statements, solutions, and key visualizations. Present each problem separately with clear conclusions and insights.

**6. Detailed Documentation in Word File:-** A detailed report is created in a Word document covering all parts of the project, including data collection, cleaning, problem statements, tool integration, Power BI dashboards, EDA findings, and key visuals from the PowerPoint.

**7. Recording of the Complete Project by Explaining:-** The final step is to record a video of the full project, explaining data cleaning, SQL queries, Power BI dashboards, and presentation slides. It helps viewers understand your approach and shows your analysis and communication skills.

**Objective:-**

The goal of this Capstone Project is to use data analytics to uncover insights from historical Olympic Games data. The project analyzes information on athletes, sports, medals, and countries to explore trends, top performers, and participation changes over time.

It covers the full data analysis process—from downloading and cleaning the data, transforming it, and connecting it to tools like Excel, MySQL Workbench, and Power BI. EDA is done using Excel and SQL, and key insights are shown through clear Power BI dashboards.

Each problem is solved with visuals and conclusions, summarized in a presentation and a detailed report. A video walkthrough is also created to explain the entire project. This project helps build strong skills in data analysis, SQL, and visualization for real-world sports analytics.

**Significance:-**

This Sports Analytics Capstone Project is valuable for both learning and real-world use. It shows how data can reveal trends in athlete performance, medal counts, and country participation using Olympic data.

By using tools like Excel, MySQL, and Power BI, the project builds hands-on skills in data analysis and visualization. It helps learners think critically and solve problems using real data.

For the sports industry, the insights can help coaches and analysts improve training and plan strategies. Overall, this project prepares students for real data challenges in sports analytics.

**Table Explanation:-**

Sport Table

This table contains a comprehensive list of all sports featured in the Olympics, covering both the summer and winter editions. Each sport is further categorized into specific events, catering to a wide range of athletic disciplines.

Event Table

The Event table offers detailed information about the various events held within each sport. Events are categorized based on gender, with separate entries for Men's, Women's, and Mixed events.

City Table

This table provides an extensive list of cities from around the world that have hosted or participated in the Olympic Games over the years.

Games Table

The Games table records every instance of the Olympic Games since its inception in 1896. It includes essential details such as the year in which the Games were held, as well as distinguishing between Summer and Winter Olympics.

Games City Table

This joining table establishes the relationship between the Games and City tables, capturing instances where the Olympic Games were jointly hosted by multiple cities, as seen in the case of the 1956 Olympics held in both Melbourne and Stockholm.

NOC Region Table

This table consists of NOC codes, representing National Olympic Committees, and their corresponding countries. It acts as a reference to identify the countries participating in the Olympic Games.

Person Table

The Person table records information about individuals who have competed in the Olympics. It includes their names, gender, height, and weight, with height and weight data remaining consistent across different Olympic editions.

Person Region Table

This joining table establishes connections between individuals and the countries they represented while competing in the Olympics. It accommodates scenarios where a person may have competed for multiple countries in different Olympic Games.

Games Competitor Table

This table serves as a joining table, linking individuals to specific Olympic Games, showcasing the participants and their involvement in each event.

Medal Table

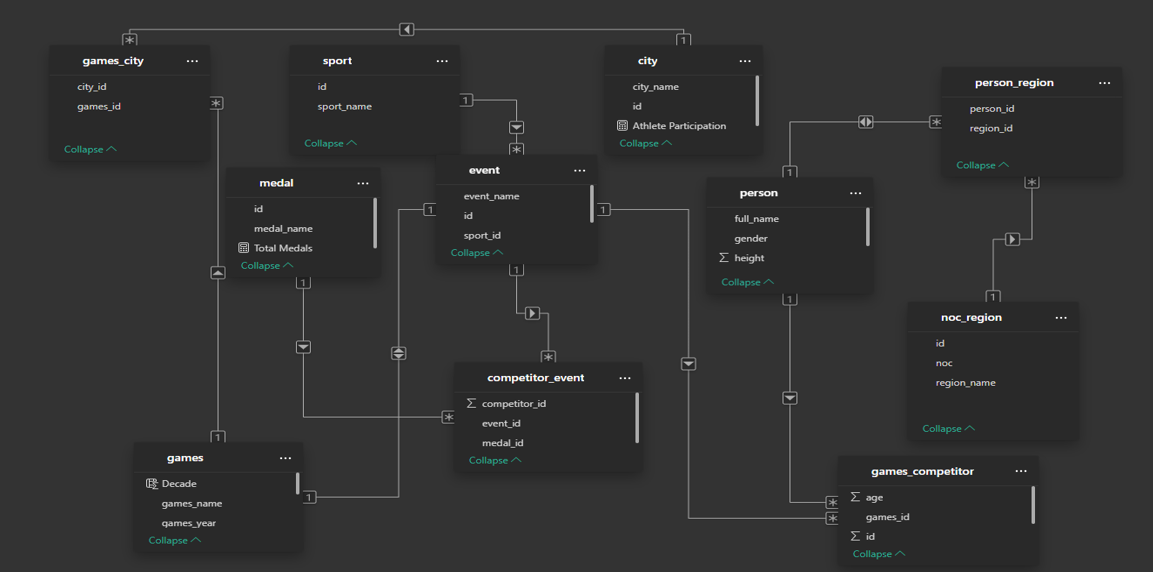
A small table listing the different types of medals awarded at the Olympics, including Gold, Silver, Bronze, and N/A (indicating no medal).

Competitor Event Table

This table represents the largest one, providing crucial details about the combination of competitors, the events they participated in, and the medals they received, if any.

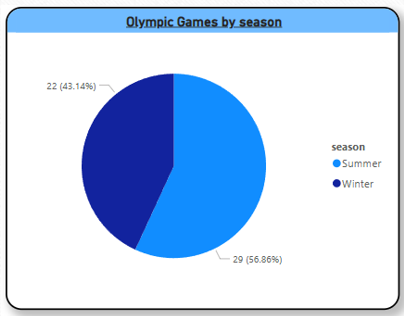
**ER-Diagram:-**

This ERD represents the structure of Olympic Games data. The games table connects to host city details and participating competitors. Each sport includes multiple events, which link to athletes via the competitor event table, capturing their participation and any medals won. Athlete details like name, gender, and height are stored in the person table, which links to regional data through person\_region and noc\_region. This model allows analysis of events, athletes, medals, and regional participation in the Olympics.

****

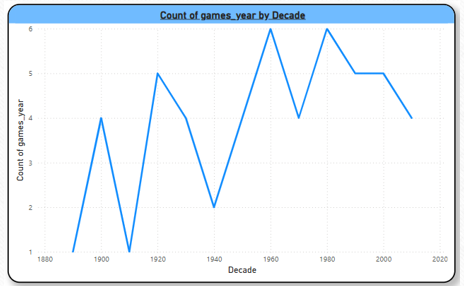
1. How many Olympic Games have been held in each season (summer vs. winter)?

**Power BI Problem Statements:-**



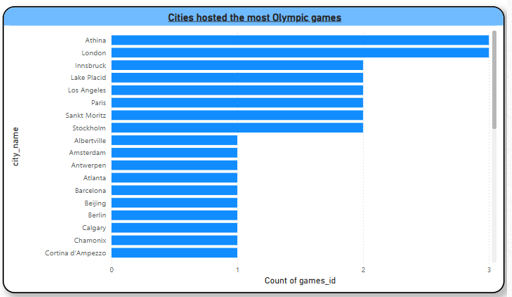
The analysis shows that the **Summer Olympics** have been held more often than the **Winter Olympics.** This is because they **started earlier**, include a **wider variety of sports**, and attract **greater global participation**. In contrast, the Winter Olympics involve more region-specific sports, limiting the number of participating countries. These factors make the Summer Games more frequent and globally popular.

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



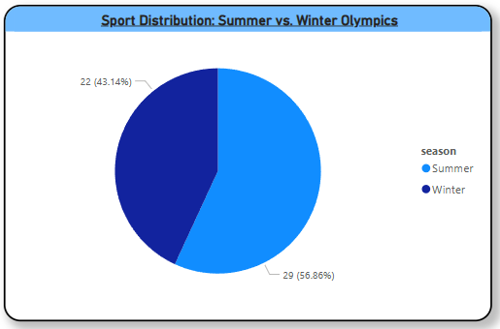
The chart shows that the number of Olympic Games increased over the decades, reflecting the event’s growth. However, there were some declines during certain periods due to global events like wars. While the overall trend was upward, there has been a slight drop in recent years, showing both expansion and occasional disruption in Olympic history.

3. Which cities have hosted the most Olympic Games?



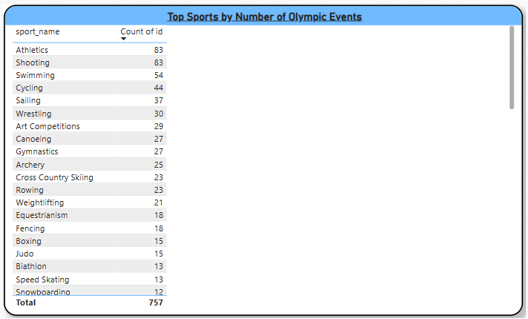
The chart highlights which cities have hosted the Olympic Games the most times, showing patterns in host city selection. Some cities like London, Paris, or Los Angeles have hosted multiple times, reflecting their strong infrastructure and international status. However, the distribution also shows a conscious global effort by the Olympic Committee to rotate hosting duties among different continents and cultures. This not only spreads economic and cultural benefits but also promotes global unity, diversity, and inclusion by involving countries from all parts of the world in hosting such a prestigious event.

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



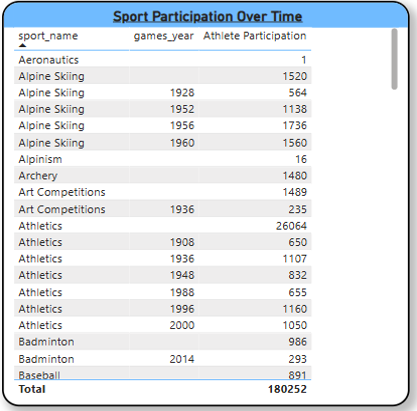
The Summer Olympics have significantly more sports than the Winter Olympics because summer provides better conditions for a wide range of outdoor and team events. Sports like athletics, swimming, football, and basketball thrive in warmer weather and are practiced globally. In contrast, Winter Olympics require snow and ice, limiting participation to fewer countries. This makes the Summer Games more diverse and inclusive, attracting more athletes and nations. The broader variety of events reflects global interest and encourages wider representation, making the Summer Olympics a much larger and more varied sporting event overall.

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



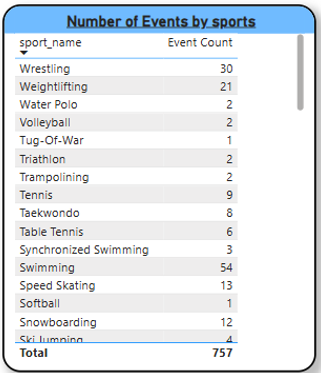
The analysis reveals that certain sports, particularly athletics and shooting, dominate the Olympic program in terms of the number of events. This is because these sports encompass a wide range of disciplines—such as sprints, long-distance runs, relays, jumps, and throws in athletics, and different firearm types and distances in shooting. Each discipline is treated as a separate medal event, increasing their overall count. As a result, these sports not only contribute significantly to the total medal tally but also attract a large number of athletes, making them central to the Olympic Games in terms of scale and impact.

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



The table clearly indicates that Athletics consistently attracts the highest number of athletes across Olympic Games, highlighting its central role and popularity in the event. Athletics includes a wide variety of events such as running, jumping, and throwing which increases opportunities for athlete participation and contributes to its dominance. In comparison, sports like Alpine Skiing show spikes in participation during certain Winter Olympics, reflecting seasonal interest and regional strengths. Meanwhile, some sports have very low athlete numbers, suggesting limited global reach or fewer event categories. Overall, the data underscores the prominence of Athletics and the varying levels of athlete engagement across different sports.

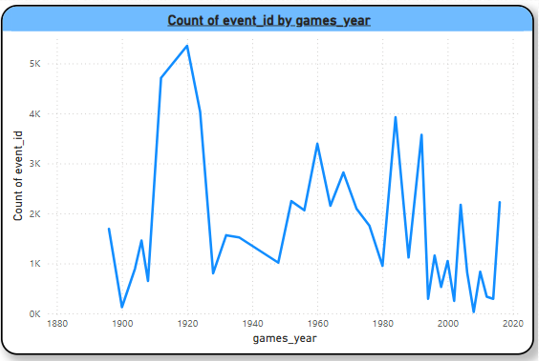
7. How many events are there in each sport?



The table illustrates that sports such as Swimming and Wrestling have a notably high number of events in the Olympics. This is largely due to the many weight categories, distances, and styles (like freestyle, backstroke, or Greco-Roman wrestling), which allow for a broad range of competitions within each sport. Their structure not only increases medal opportunities but also reflects their popularity and long-standing presence in the Games. In contrast, sports with fewer events, such as Equestrian or Modern Pentathlon, have a narrower competitive scope, often due to limited categories or specialized participation. This variation shows how different sports contribute uniquely to the diversity and scale of the Olympic program.

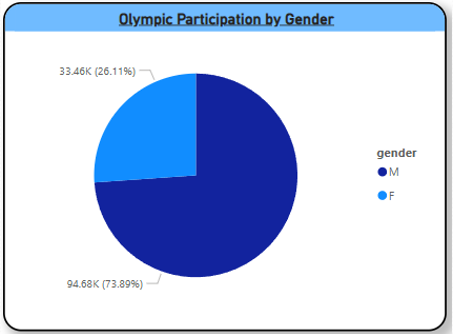
8. What is the distribution of events by gender (Men, Women, and Mixed)?

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



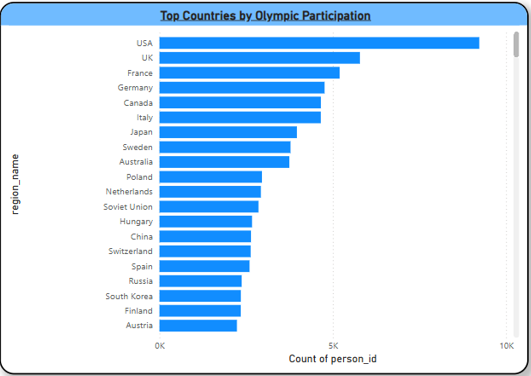
The chart shows that the number of Olympic events has varied over time due to the addition of new sports, removal of less popular ones, and global events like wars or pandemics. Peaks reflect efforts to expand and modernize the Games, while drops indicate periods of restriction or transition. This highlights the Olympics’ ability to adapt to changing global trends.

10. What is the distribution of participants by gender?



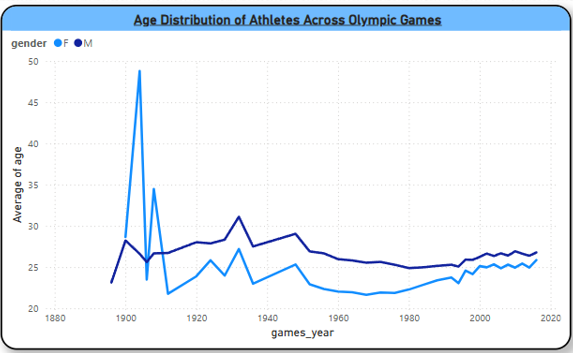
The analysis reveals that historically, male athletes have significantly outnumbered female athletes in the Olympic Games. However, over the years, there has been a steady increase in female participation. This positive trend reflects ongoing efforts by the International Olympic Committee and global sports organizations to promote gender equality. The introduction of more women's events and mixed-gender competitions has helped reduce the gender gap, signaling progress toward a more inclusive and balanced Olympic movement.

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



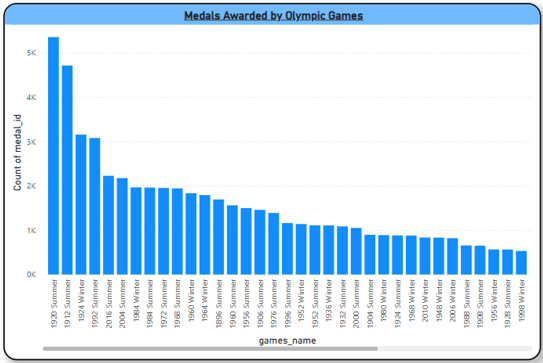
Countries with strong sports infrastructure, government support, and a culture of athletic development such as the United States, China, and Russia consistently send a large number of athletes to the Olympics. Their high participation is not only due to larger populations but also because of long-term investments in training facilities, coaching, and talent scouting. These nations often have well-organized national sports federations that identify and nurture athletes from a young age. Their consistent involvement highlights how resources, population size, and commitment to sports play a key role in Olympic representation.

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



The chart shows that Olympic athletes' average age has stayed mostly between 23 and 28 years over time, reflecting the typical peak performance period. Male athletes usually have a slightly higher average age than females, possibly due to differences in sport type or career length. This suggests a stable and consistent age trend across Olympic Games.

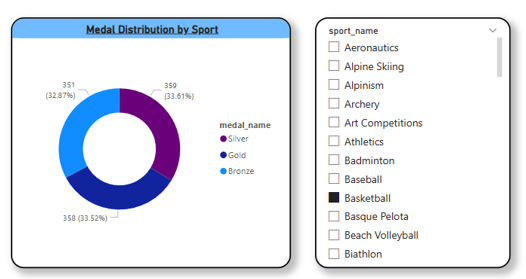
13. How many medals have been awarded in each Olympics?



The trend in medal counts over time shows a general increase, largely driven by the expansion of Olympic events and greater athlete participation. As new sports and disciplines have been added to the Games, more medal opportunities have emerged, contributing to the overall rise. However, occasional dips or fluctuations are visible in certain years, often caused by global events like wars or boycotts, or changes in the Olympic program that affect the number of competitions. This pattern reflects both the Olympics’ continued growth and the impact of historical and organizational factors on medal distribution.

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

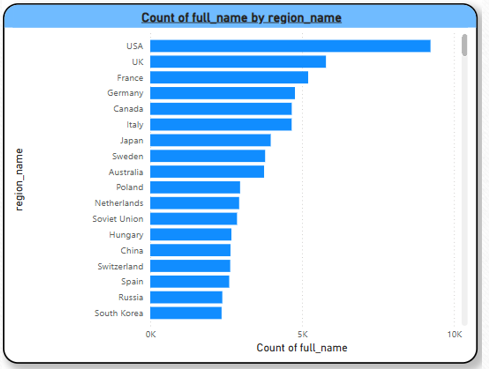
15. How does the medal distribution vary across different sports?



Sports such as Athletics and Swimming dominate the total medal count in the Olympics because they consist of many individual events and categories—like different distances, techniques, and formats. This wide range allows more athletes to compete and win medals, significantly boosting the overall count. On the other hand, sports with fewer events, such as Modern Pentathlon or Trampoline, have limited medal opportunities due to their narrow structure. This highlights that the number of medals awarded in each sport is directly tied to how many events it offers, making event variety a major factor in a sport’s medal contribution.

16. What is the distribution of medals among different regions?

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



The chart reveals that countries like the USA, UK, and France have consistently had the highest Olympic participation, indicating their long-standing commitment to sports and strong infrastructure for athlete development. These Western nations have historically invested in training, facilities, and international competition, leading to larger athlete delegations. In comparison, countries like China, Russia, and South Korea also show strong participation but with relatively lower numbers. This pattern highlights how Western nations have played a dominant role in shaping Olympic history through consistent and widespread athlete representation.

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

**EDA Problem Statements:-**

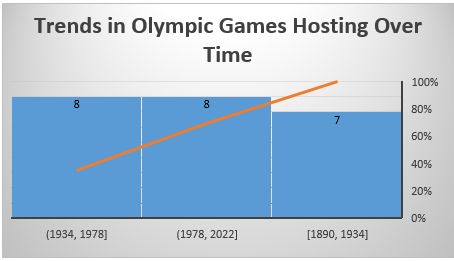
1. Are there any trends or patterns in the frequency of hosting Olympic Games?

This chart provides a comprehensive view of the number of Olympic Games hosted in each decade, clearly separated by season summer and winter. By using the **games\_count** column to represent the total number of Olympic events held per decade, the chart offers valuable insights into how the scale, frequency, and structure of the Games have evolved over time.

It highlights several key historical and cultural developments. For example, the emergence of the Winter Olympics in the 1920s marked a significant expansion of the Olympic program, diversifying the types of sports represented and attracting a new segment of athletes and audiences. The chart also captures the profound impact of global events—most notably World War II, during which both summer and Winter Games were canceled in the 1940s. These gaps in the timeline serve as visual markers of how external geopolitical factors have shaped Olympic history.

Additionally, the chart reflects the steady regularization of the Games in later decades. From the 1980s onward, the Olympics have been held more consistently every four years, with fewer interruptions. This trend signals stronger international cooperation, improved global infrastructure, and a growing commitment to maintaining the Olympic cycle regardless of global challenges. The increasing consistency and number of hosted Games also point to the rising global interest in sports, cultural exchange, and international unity through athletic competition.

Overall, the chart not only quantifies Olympic history but also tells a broader story of global change, resilience, and the enduring importance of the Olympic tradition.

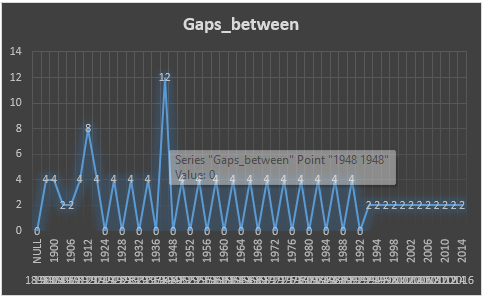


2. How has the duration of Olympic Games changed over time?

This provides a detailed look at the regularity—or irregularity—of game releases over time by focusing on the **gaps\_between** values. Each value indicates the number of years between consecutive releases, serving as a direct measure of how consistently new games were launched. A **gaps\_between** value greater than 1 reveals that a game release was skipped or delayed beyond the expected interval, breaking the usual release pattern.

These irregularities can arise from a wide range of factors. For instance, production delays due to technical challenges, changes in development teams, or extended testing periods can push back release dates. Strategic decisions—such as shifting to align with new console launches, avoiding market competition, or building anticipation—may also lead to longer gaps. Additionally, external influences like economic downturns, pandemics, or company restructurings can disrupt development timelines.

By surfacing and analyzing these gaps, the query provides valuable insights into the dynamics behind the game development and publishing process. It helps detect periods of stability versus turbulence, identify long-term trends, and understand how both internal and external forces have influenced the cadence of releases. This information is especially useful for industry analysts, developers, and fans interested in the historical context and decision-making behind each title’s launch.



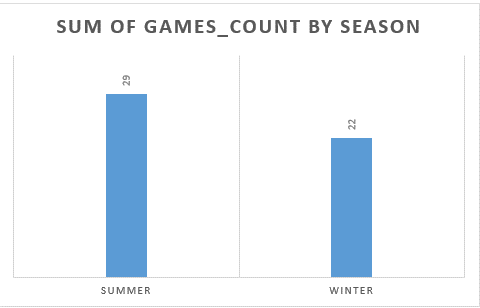
3. Are there any notable events or occurrences associated with specific Olympic Games?

This analyzes Olympic Games data by organizing it into meaningful historical segments, specifically by **decade** and **season** (summer or winter). The main goal is to examine how the frequency and distribution of the Games have evolved over time. To determine the decade for each event, the query uses a simple mathematical transformation: it divides the year of each Olympic event by 10, rounds it down to remove the remainder, and then multiplies it by 10. This effectively groups years like 1996 and 1992 under the 1990s, ensuring consistency in the time-based analysis.

Once the decade is calculated, the data is grouped by both decade and season. The query then counts how many Olympic Games were held within each group using an aggregate function, typically **COUNT ().** This count reflects the total number of events (summer or winter) conducted in each ten-year period.

By ordering the results first by decade and then by season, the output provides a structured, chronological timeline that offers a clear visual of Olympic hosting trends. This makes it easy to identify historical shifts—such as the introduction and growing presence of the Winter Olympics, periods of interruption like during the World Wars, and the increasing regularity of the Games in recent decades.

Overall, this approach helps contextualize Olympic history, allowing for the observation of long-term patterns, disruptions, and the growing global reach of the Games. It’s especially useful for understanding how external factors and organizational changes have influenced the Olympic calendar across both summer and winter editions.



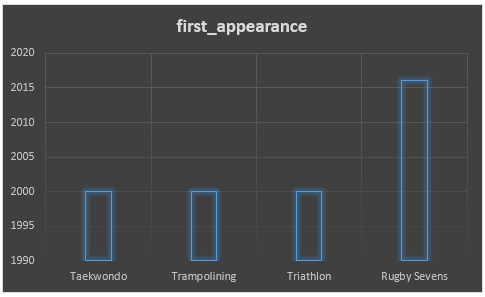
4. Are there any emerging sports that have been recently added to the Olympics?

This is crafted to identify sports that were introduced into the Olympic Games in the year 2000 or later, offering valuable insights into the modernization and expansion of the Olympic program in the 21st century. To achieve this, the query joins several key tables—typically those that store information about Olympic events, the sports they belong to, and the specific years the Games were held. These joins allow the query to accurately link each sport to its corresponding Olympic appearances.

The core of the logic lies in grouping the data by sport name and using the **MIN ()** function on the Olympic year field. This ensures that for each sport, only its **earliest appearance** is captured, effectively identifying its Olympic debut. By doing so, the query filters out repeated instances of the same sport across multiple Games and focuses solely on when each sport first entered the Olympic stage.

The **HAVING** clause is then applied to narrow the results to only those sports whose first appearance occurred in the year 2000 or later. This time-based filtering is crucial for focusing on the most recent changes to the Olympic lineup—reflecting global shifts in sports popularity, evolving definitions of athletic competition, and efforts by the International Olympic Committee (IOC) to appeal to newer, younger audiences.

Finally, the results are sorted in ascending order by year, providing a clear timeline of how and when new sports have been integrated into the Olympics. This not only showcases the increasing diversity of the Games but also highlights broader cultural and societal changes influencing what is considered an Olympic sport today. The query, therefore, serves as a useful tool for tracking the evolution of the Olympic program and understanding the dynamic nature of international sporting events.



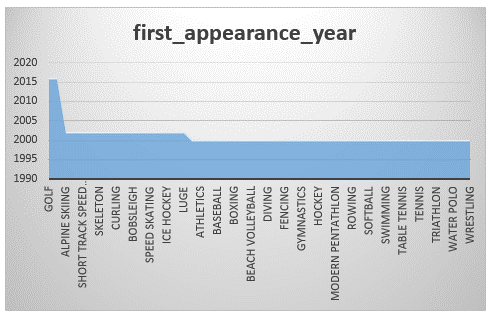
5. How has the popularity of certain sports changed over the years?

This is focused on identifying the **first Olympic appearance of each sport between the years 2000 and 2025**, providing a snapshot of how the Olympic program has expanded and evolved in the modern era. The query begins by creating a **Common Table Expression (CTE)** called first appearances, within this CTE, several key tables are joined such as competitor\_event, games\_competitor, games, event, and sport to accurately link each Olympic competition entry to the sport it belongs to and the year it took place.

A **WHERE** clause filters the dataset to include only records from Olympic Games held between **2000 and 2025**, narrowing the focus to a recent and relevant time frame. The data is then **grouped by sport ID**, and the **MIN ()** function is applied to determine the **earliest year** each sport appeared during this period. This step ensures that, even if a sport appeared multiple times, only its first appearance within the defined range is considered.

In the **main query**, the results from the CTE are joined with the sport table to retrieve the actual **names of the sports** corresponding to each sport ID. The final result lists each sport along with its first Olympic year, **sorted in descending order** so the most recently introduced sports appear at the top.

This query is particularly useful for analyzing **how the Olympic Games have adapted in recent decades**, showing which new sports have been added and when. It reflects changes in global interest, the rise of emerging sports, and the International Olympic Committee's efforts to keep the Games relevant and inclusive in the 21st century.



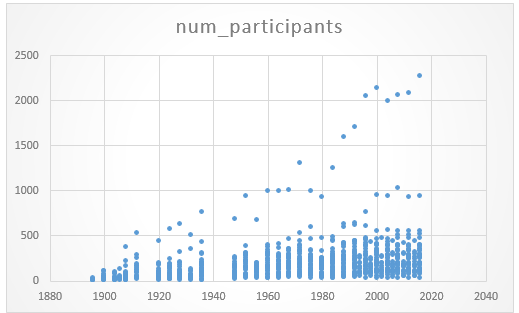
6. Are there any sports that are specific to a particular region or culture?

This is designed to analyze **athlete participation in Olympic sports over time** by calculating the number of unique competitors involved in each sport for every Olympic year. The process begins with a **subquery** that joins several key tables: competitor\_event (linking athletes to events), games\_competitor (linking athletes to specific Olympic Games), and event (associating each competition with a particular sport). These joins ensure that each athlete is accurately connected to both the sport they competed in and the specific edition of the Games they participated in.

The subquery then **groups the data by sport ID and games ID**, allowing for a breakdown of participation on a per-sport, per-year basis. It uses the COUNT (DISTINCT person\_id) function to calculate the number of **unique athletes** who competed in each sport during each Olympic year. This step ensures that even if an athlete competed in multiple events within the same sport, they are counted only once per sport per Games, providing a clean measure of individual participation.

In the **main query**, the results from the subquery are joined with the sport and games tables to retrieve more descriptive information specifically, the **sport names** and the **corresponding years** of the Games. This enriches the output, making it easier to interpret the data in human-readable form.

Finally, the output is **sorted alphabetically by sport name** and then **chronologically by year**, creating a structured and easy-to-navigate dataset. This organization allows for a clear view of how athlete participation has changed over time across different sports. It can highlight trends such as the growth or decline in popularity of specific sports, gender or policy-based inclusion patterns, and the overall expansion of athlete involvement in the Olympic movement.



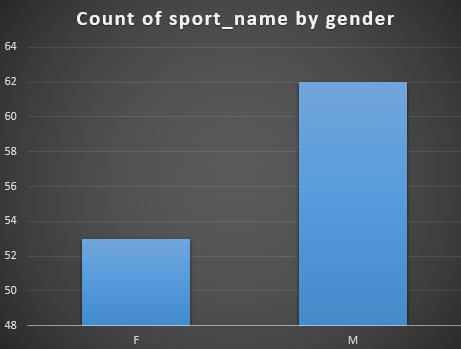
7. Are there any sports that have a higher number of events for one gender compared to others?

This is designed to analyze **gender representation across Olympic sports** by determining how many unique events exist for male and female athletes within each sport. To achieve this, the query performs **multiple joins** across several key tables: it links athlete data to their respective events, the sports those events belong to, and personal details such as **gender.** These joins ensure that each record accurately reflects the gender of the competitor, the sport they participated in, and the specific event.

Once the necessary data is assembled, the query **groups the results by both sport name and gender**, enabling a gender-based breakdown within each sport category. It then uses the COUNT (DISTINCT event\_id) function to calculate the number of **unique events** associated with each gender in each sport. This approach ensures that events are not double-counted even if multiple athletes of the same gender participated in them.

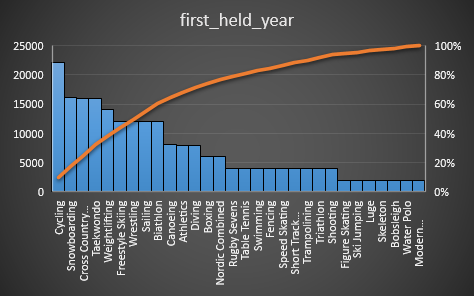
The results are **sorted alphabetically by sport name and then by gender**, making it easy to compare how opportunities are distributed across sports and between genders. For example, some sports may show a nearly equal number of events for both male and female athletes, indicating a balanced structure. Others may reveal a clear disparity, with more events offered to one gender—often a legacy of historical gender norms in sport.

This analysis offers valuable insights into the state of gender equality in the Olympic Games. It helps identify where progress has been made and where further efforts may be needed to ensure equal representation and opportunity. These insights are particularly useful for policymakers, sports organizations, and advocates working toward a more inclusive and equitable Olympic program.

****

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

This is designed to identify Olympic events that were introduced from the year 2000 onwards, offering insight into how the Olympic Games have evolved in the modern era. It begins by determining the earliest year each event appeared in the Olympics, typically by selecting the minimum year associated with each event in the games-event mapping table. This step reveals the debut year of every event. The query then filters this data to include only those events whose first appearance occurred in or after the year 2000, focusing on more recent additions to the Olympic program. To make the results more informative, it joins this filtered data with the event and sport tables, allowing it to retrieve the names of the events along with the sports they belong to. Finally, the results are ordered in descending order based on the debut year, ensuring that the most recently introduced events appear at the top. This approach not only highlights the newest Olympic events but also helps showcase emerging sports and trends that reflect the changing interests and values of the global sporting community.

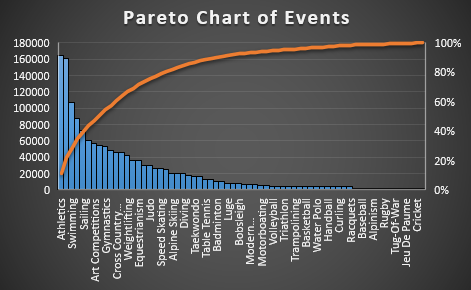


9. Are there any events that have been discontinued or removed from the Olympics?

This identifies the most recent year each Olympic event was held prior to 2025, helping to reveal which events have been discontinued or haven't occurred in recent Games. It begins with a subquery that calculates the latest year each event took place using the **MAX ()** function on the games year, grouped by event ID. This gives the most recent occurrence of every event up to (but not including) 2025.

The main query then joins this result with the event and sport tables to retrieve detailed names for each event and the sport it belongs to, making the output more informative and readable. Finally, the results are sorted in descending order by the most recent year, placing the most recently held events at the top and the least recent ones at the bottom.

This structure allows users to easily identify which events are still active and which may have been phased out or temporarily removed from the Olympic program. It can be particularly useful for analyzing shifts in event popularity, sport evolution, or changes in Olympic scheduling over time.

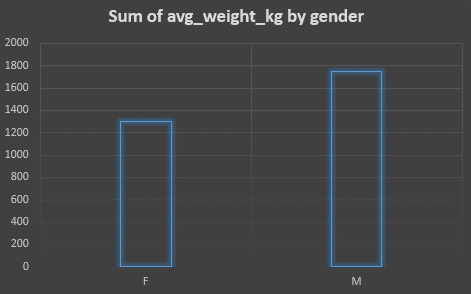


10. Are there any notable trends in the height and weight of participants over time?

This provides an overview of the average physical characteristics of Olympic athletes—specifically height and weight—grouped by gender and year of the Games. It begins by joining the person table with game participation data, allowing access to each athlete’s physical attributes along with their gender and the year they competed. To ensure accuracy, the query filters out any records where height or weight data is missing.

Once the data is cleaned, it is grouped by games year and gender, allowing for gender-specific analysis over time. The query then calculates the average height and weight for each group using the **AVG ()** function, rounding the results to two decimal places for readability. Additionally, it counts the number of unique athletes included in each group, giving context to the averages and indicating the sample size.

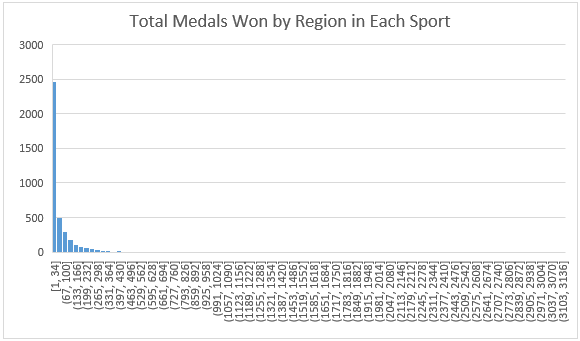
The final result is ordered by year and gender, creating a chronological view of trends in athlete physiques. This can reveal how the physical profiles of male and female Olympians have changed over time possibly in response to evolving training methods, sport-specific demands, or changes in athlete selection and qualification standards.



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

This provides a comprehensive summary of how many medals each country has won across all Olympic sports. It begins by creating a Common Table Expression (CTE) called medal counts, which calculates the total number of medals won by each country in each sport. This is done by joining tables that link events, competitors, and country or region data, while filtering to include only rows where a medal was awarded (i.e., medal\_id IS NOT NULL). The data is grouped by both sport and country IDs to get the medal counts.

In the main part of the query, these grouped totals are joined with the sport and noc\_region tables to retrieve the actual sport and country names instead of IDs. The final result displays a list of sports, the countries that have won medals in those sports, and their corresponding medal counts. The output is sorted first by sport name (alphabetically) and then by the number of medals in descending order within each sport. This allows for easy comparison of country performance across different Olympic disciplines.

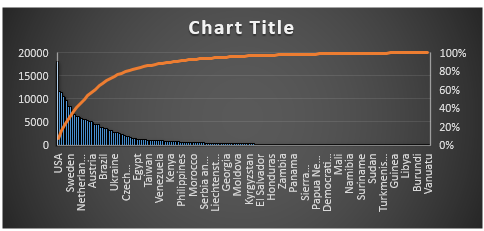


12. What factors contribute to the success or performance of participants from different countries?

This provides a detailed analysis of Olympic medal-winning athletes by country, combining performance metrics with demographic insights. It begins by joining several related tables—typically including athlete, event, region, and medal data—to create a comprehensive dataset that links each athlete to their respective country, event, and medal outcome. A key filter is applied to include only those athletes who have actually won a medal, ensuring the analysis focuses solely on medalists.

For each country, the query calculates several important metrics. These include the **total number of medals** won, as well as the **number of unique athletes** who earned those medals—offering a sense of how widespread the success is across a nation’s participants. It also computes **average age, height, and weight** of the medalists, which provides insight into the physical and age-related characteristics of top-performing athletes from each country. Additionally, the query examines **gender distribution** among medal winners, giving a clearer picture of how balanced each country’s medal success is between male and female athletes.

The results are grouped by country (or region) and then **sorted in descending order by total medal count**, so that the most successful countries appear at the top. This output not only highlights national medal performance but also reveals demographic trends and physical profiles of medal-winning athletes across different nations, making it useful for analysts, coaches, and researchers interested in Olympic competitiveness and athlete development.

****

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

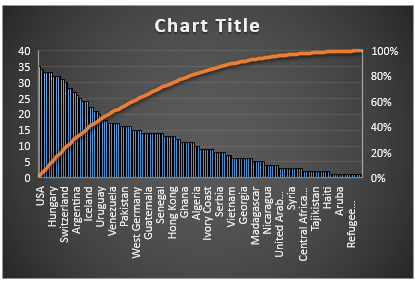
This is designed to identify countries that have consistently performed well at the Olympic Games by focusing on how many **separate editions** of the Games they have won **at least 10 medals**. It provides insight into not just one-time successes, but sustained high performance across different Olympic years.

The process begins with an **inner query** that joins key tables such as competitor event, noc\_region, and games to link individual athletes with their respective countries and the specific edition of the Olympics in which they competed. It includes a filter condition to consider only **medal-winning entries** (e.g., where medal\_id IS NOT NULL), ensuring that the analysis is limited to athletes who contributed to a country's medal count.

This inner query then **groups the data by country and Olympic year**, counting the total number of medals each country won in each edition of the Games. After this aggregation, it applies a **HAVING clause** to retain only those (country, year) pairs where the country won **10 or more medals** in that edition.

The **outer query** then takes this filtered set of high-performing editions and groups the data again, this time **only by country**. It counts how many times each country appears in the filtered results effectively, **how many Olympic Games each country has won at least 10 medals.** The final result is sorted in descending order, with countries that have had the **most editions with 10+ medals** appearing at the top.

This query is especially useful for identifying **consistent Olympic powerhouse’s** nations that don’t just succeed once, but reliably achieve high medal counts across multiple Games. It's a valuable tool for sports analysts, historians, and policy makers interested in trends of long-term athletic excellence.

****

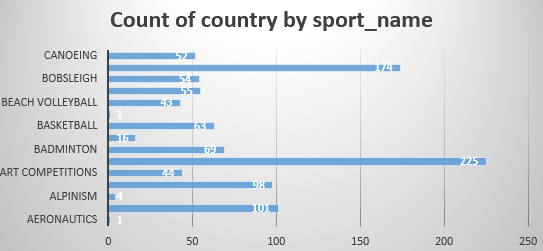
14. Are there any sports or events that have a higher number of medalists from a specific region?

This provides a detailed breakdown of how many medals each country has won in each Olympic sport, offering a clear view of national strengths across different disciplines. It begins with a **subquery** that calculates the total number of medals earned by each country in each sport. This is done by joining relevant tables—typically including competitor event, event, and person region to link athletes to their sports and countries. The subquery filters the data to include only **medal-winning performances** (i.e., where medal\_id IS NOT NULL) and then **groups the results by sport ID and region ID** (representing countries), counting the total medals per group.

Once the medal counts are calculated, the **main query** joins this aggregated data with the sport table and the noc\_region table to retrieve the **names of sports and countries**, replacing the numerical IDs with human-readable values. This makes the output much more accessible and meaningful for users.

The final result displays a list of sports, each followed by the countries that have won medals in that sport, along with their corresponding **medal totals**. The data is **sorted alphabetically by sport name**, and within each sport, countries are listed in **descending order of medal counts**, so the top-performing nations for each sport appear first.

This query is useful for understanding the distribution of Olympic success across sports and nations. It can highlight **specialized national strengths**, such as a country dominating in swimming, gymnastics, or athletics, and is valuable for comparative analysis, strategic sports development, and historical insights into Olympic competition.

****

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

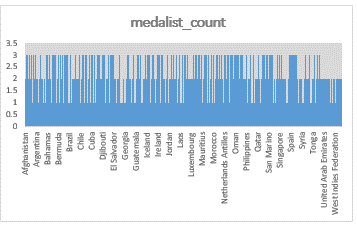
This is designed to identify **sports in which a country had a relatively small number of medal-winning athletes specifically, three or fewer.** The goal is to spotlight instances where a country achieved medal success in a sport, but with a very limited number of individual athletes, which may suggest either standout individual performances or strong reliance on a few key competitors.

The query works by joining several core tables such as those representing **athletes, events, sports, and regions (countries)** to connect each athlete to the sport they competed in and the country they represented. A **filter condition** is applied to select only those entries where a **medal was awarded** (i.e., medal\_id IS NOT NULL), ensuring the focus remains strictly on medal-winning performances.

Once this dataset of medalists is constructed, the query **groups the data by both country and sport**, then uses a **COUNT (DISTINCT athlete\_id)** to calculate how many unique athletes from each country won medals in each sport. It then applies a **HAVING clause** to keep only those (country, sport) pairs where the number of medal-winning athletes is **three or fewer.**

The final output includes the **country name, sport name**, and the **number of distinct medal-winning athletes**, and is **sorted** first by the athlete count (ascending), then by country, and finally by sport name. This ordering makes it easy to see the most minimal medal-winning efforts at the top.

This query is particularly useful for identifying **concentrated or standout performances**, such as when a single athlete or a small group carries a nation’s success in a particular sport. It can also provide insights into talent distribution, training focus, or reliance on elite individuals within national Olympic programs.



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

This is designed to **count how many unique participants represented each country in each edition of the Olympic Games**, and it does so in a structured and performance-conscious way by breaking the task into clear stages.

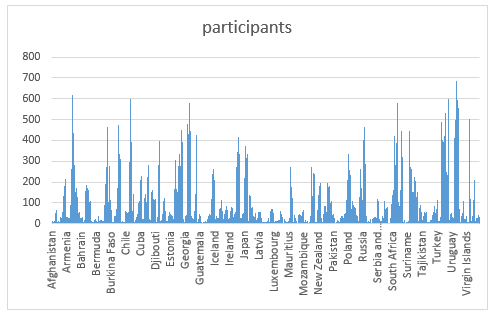
The query begins by identifying **unique combinations of each participant, the specific Olympic Games they participated in, and the region (country) they represented**. This is important because an athlete may have multiple records due to participating in different events within the same Games. By isolating only distinct combinations of athlete ID, games ID, and region ID, the query avoids **double-counting participants** who competed in multiple events during the same Olympic edition.

Next, this distinct set of athlete-game-country combinations is **joined with the games table** to retrieve the **year** of the Games, and with the noc\_region table to get the **country name.** These joins enrich the data with contextual information that makes the final output more informative and human-readable.

After this, the query **groups the data by year and country**, and performs a **count of distinct participants** for each group. This final aggregation tells us how many individual athletes each country sent to each edition of the Games.

This approach is not only logically sound but also **computationally efficient**, as it narrows down the dataset before aggregation. Rather than trying to count distinct participants across multiple joined tables in a single complex step, the query **prepares a clean and reduplicated intermediate result**, making the final grouping and counting simpler and faster.

Overall, this query provides valuable insights into **Olympic participation levels over time**, helping analysts understand trends in athlete delegation sizes by country and edition. It can be used to track growth, investment in sports, or participation disparities between nations.



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

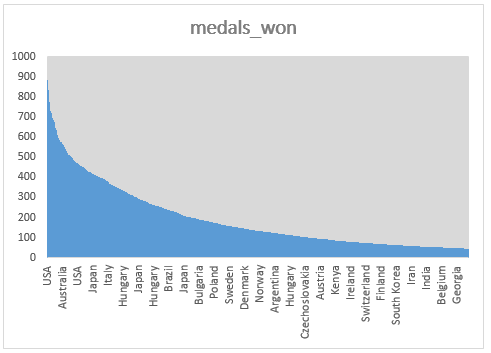
This is designed to analyze **how many medals each country (region) won in each Olympic season,** offering a seasonal view of national medal performance. It follows a logical, step-by-step approach to ensure accuracy and clarity in the results.

The query begins by identifying **the total number of medals earned by competitors from each region (country) in each edition of the Games**. This is typically done by joining tables such as competitor event, person region, and possibly event, filtering only for **medal-winning entries** those where medal\_id IS NOT NULL. The data is then **grouped by region ID and games ID** to calculate the medal count per country for each Olympic edition.

Once these medal counts are determined, the query joins this intermediate result with the games table to retrieve details about each Olympic edition, specifically the **season** (either 'summer' or 'winter'). It also **joins with the** noc\_region table to obtain the **country name** corresponding to each region ID, making the output more readable and meaningful.

In the final step, the query **orders the results by season** and then by **medal count in descending order**. This ensures that for each Olympic season, the countries with the **highest medal totals appear first**, allowing easy comparison of national performance within each season.

This query is especially useful for spotting **seasonal trends in Olympic success** for instance, which countries dominate in Summer Games versus Winter Games and for tracking a country’s **historical performance across different Olympic editions**. It serves as a valuable tool for sports analysts, historians, and organizations interested in the evolution of international competitiveness at the Olympics.



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

This is designed to calculate **how many unique participants from each country competed in the Olympic Games each year**, providing a year-by-year view of athlete representation by nation.

The query starts by **joining data from tables** that link each competitor to the Olympic Games they participated in and to the region (or country) they represented. This typically involves tables like competitor event, games, and noc\_region, ensuring that each athlete is correctly associated with both the edition of the Games **(year)** and their **national Olympic committee (NOC)** and country.

To avoid counting the same athlete multiple times within a single year, the query uses a COUNT (DISTINCT athlete\_id) or equivalent to ensure that only **unique participants** are counted for each country in each Games edition. This prevents inflation of participant numbers due to athletes competing in multiple events within the same Olympics.

Once the data is prepared, it is **grouped by country and year**, meaning that the result will show, for each country and for each Olympic year, the total number of distinct athletes who took part. The query then **orders the results by country and year**, making it easy to follow a chronological progression of athlete participation for each nation.

This kind of analysis is especially useful for tracking **trends in Olympic participation**, such as growth in team sizes, the emergence of new countries in the Games, or the impact of global events on national representation. It offers a clear, organized way to assess how actively each country has engaged in the Olympics over time.

