The objective of this project is to explore and visualize global Olympic trends-medals, participation, and performance—using real historical data.

**OLYMPIC GAMES DASHBOARD**

Shraddha Mishra

**1. Data Acquisition:**

* **Source**: Extracted Olympic dataset from a GitHub repository.
* **Content**: Data includes athlete profiles, countries, events, sports, medals, and Olympic editions—ideal for multi-angle analysis.

**2. Data Transformation & Fact Table Creation:**

* **Process**: Cleaned, joined, and consolidated multiple relational tables (person, games, event, sport, region) into one **Fact\_Olympic** table using Power Query in Power BI.
* **Enhancements**: Added computed columns (e.g., age, medal ratios) and resolved duplicates/nulls for better data accuracy.

**3. Tool Integration:**

* **SQL Workbench**: Executed queries to solve initial EDA problem statements directly on the dataset.
* **Excel**: Exported SQL query results and created early-stage visualizations for quick insights.
* **Power BI**: Imported and modeled the final fact table for advanced analytics and dynamic dashboards.

**4. Power BI Problem Statement Solutions:**

* Created **interactive dashboards** to solve specific Olympic-related queries:
  + Gender-wise participation
  + Country medal performance
  + Event trends
  + Medal-to-athlete efficiency
  + Sport popularity by season & decade
* Used slicers, charts, KPIs, and maps for clear storytelling.

**5. Exploratory Data Analysis (EDA):**

* **SQL**: Used joins, aggregates, and filters to analyze participation, medal counts, host cities, and event trends.
* **Excel**: Created pivot tables and charts to validate SQL outputs and explore alternative visuals.

**6. Insightful PowerPoint Creation:**

* Developed a well-structured **PowerPoint**:
  + Pages for each question and its visual insight.
  + Included MECE breakdown, metadata, data dictionary.
  + Separate pages for objective, audience, scope, impact, and comprehensive insights.

**7. Complete Documentation:**

* Compiled a detailed **project lifecycle document** covering:
  + Data sources & schema
  + Fact table creation logic
  + All SQL/Excel/Power BI steps with visuals
  + Evaluations, trends, insights & strategic recommendations
  + PPT structure and dashboard walkthrough

# **🎯 Objective**

# The objective of this capstone project is to perform a comprehensive analysis of the **Olympics dataset**, exploring participation, medal distribution, gender dynamics, sport-wise trends, and country-level performance across multiple Olympic editions.

# Through SQL-based querying, Excel-based analysis, and Power BI dashboards, this project aims to provide valuable insights into:

# How countries perform across sports and seasons

# Participation evolution across genders and decades

# Medal efficiency, dominance, and underperformance

# Host city patterns and event-wise trends

# The success of the project will be evaluated by:

# The depth and clarity of insights

# The interactivity and readability of visualizations

# The usefulness of findings for sports analysts, federations, and committees

# This project provides a structured, user-friendly decision-support tool that brings historical Olympic data to life, enabling more informed planning, evaluation, and storytelling around global sports participation.

**⭐ Significance**

The analysis of Olympic data plays a critical role in the global sports ecosystem by offering strategic value to **athletes, coaches, sports federations, policy planners**, and **sports historians**. Understanding patterns in Olympic participation and performance helps inform policy, training focus, funding allocation, and equity in sports.

* **For National Committees**: The dashboard helps evaluate historical performance and medal efficiency, highlighting which sports to invest in for future Olympics.
* **For Researchers & Analysts**: It offers a deep dive into gender participation evolution, city hosting trends, and the shifting popularity of sports over time.
* **For Athletes & Coaches**: Insights into high-performing countries and medal patterns help guide athlete development strategies.
* **For Educational Institutions**: The project serves as an analytical case study demonstrating advanced data modeling, transformation, and storytelling using SQL, Excel, and Power BI.
* **For Policymakers**: Understanding participation diversity and trends enables better planning for inclusive and balanced representation in sports on a global scale.

This analysis fosters **data-driven decision-making**, promotes **equality in sports**, and strengthens the **global understanding** of how sports evolve, who excels, and what factors drive Olympic success.

MESE BREAKDOWN

|  |  |  |
| --- | --- | --- |
| **Category** | **Subcategory** | **Description** |
| **Athlete Analysis** | Demographic Distribution | Analyze athlete age, gender, height, and weight across games and regions. |
|  | Gender Participation | Evaluate participation trends by gender and season over time. |
|  | Regional Representation | Assess athlete counts by country/region across different Olympic editions. |
|  |  |  |
| **Medal Analysis** | Medal Distribution | Identify medal counts (Gold/Silver/Bronze) by country, sport, and gender. |
|  | Medal Efficiency | Analyze medals per participant or per event to measure country effectiveness. |
|  | Medal Trends | Track how medal counts change over decades and seasons. |
|  |  |  |
| **Sport & Event Analysis** | Sports Popularity | Determine most participated and most medaled sports across years. |
|  | Event Distribution | Study the number and type of events by sport, gender, and year. |
|  | Event-Gender Balance | Evaluate gender representation across event categories. |
|  |  |  |
| **Games Analysis** | Seasonal Distribution | Compare number of events and games in Summer vs. Winter editions. |
|  | Host City Frequency | Identify most frequently hosting cities and their hosting years. |
|  | Decade-wise Games Trend | Visualize how Olympic Games evolved across decades in terms of scale and reach. |
|  |  |  |
| **Region/Country Analysis** | Top Performing Regions | Highlight countries with highest medal wins and participation over time. |
|  | Participation Growth | Track the increase in NOCs (Nations) and athlete count over Olympic history. |
|  | Medal vs Participation | Compare medals won vs total participants for efficiency insights. |

**Data Dictionary – Olympic Dataset**

**Table: region**

* region\_id: Unique identifier for each region or National Olympic Committee (NOC).
* region\_name: Name of the region or country (e.g., India, USA).
* noc: Standard NOC abbreviation (e.g., IND, USA).

**Table: person**

* person\_id: Unique identifier for each athlete.
* full\_name: Full name of the athlete.
* gender: Gender of the athlete (Male/Female).
* age: Age of the athlete at the time of participation.
* height: Height of the athlete in cm.
* weight: Weight of the athlete in kg.

**Table: games**

* games\_id: Unique identifier for each Olympic edition.
* games\_name: Official name of the Olympic Games (e.g., 2008 Summer).
* games\_year: Year the Olympics were held.
* season: Season of the games (Summer or Winter).
* city\_id: Foreign key referencing the city where games were hosted.

**Table: city**

* city\_id: Unique identifier for each city.
* city\_name: Name of the host city.

**Table: medal**

* medal\_id: Unique identifier for each medal type.
* medal\_name: Type of medal (Gold, Silver, Bronze).

**Table: sport**

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

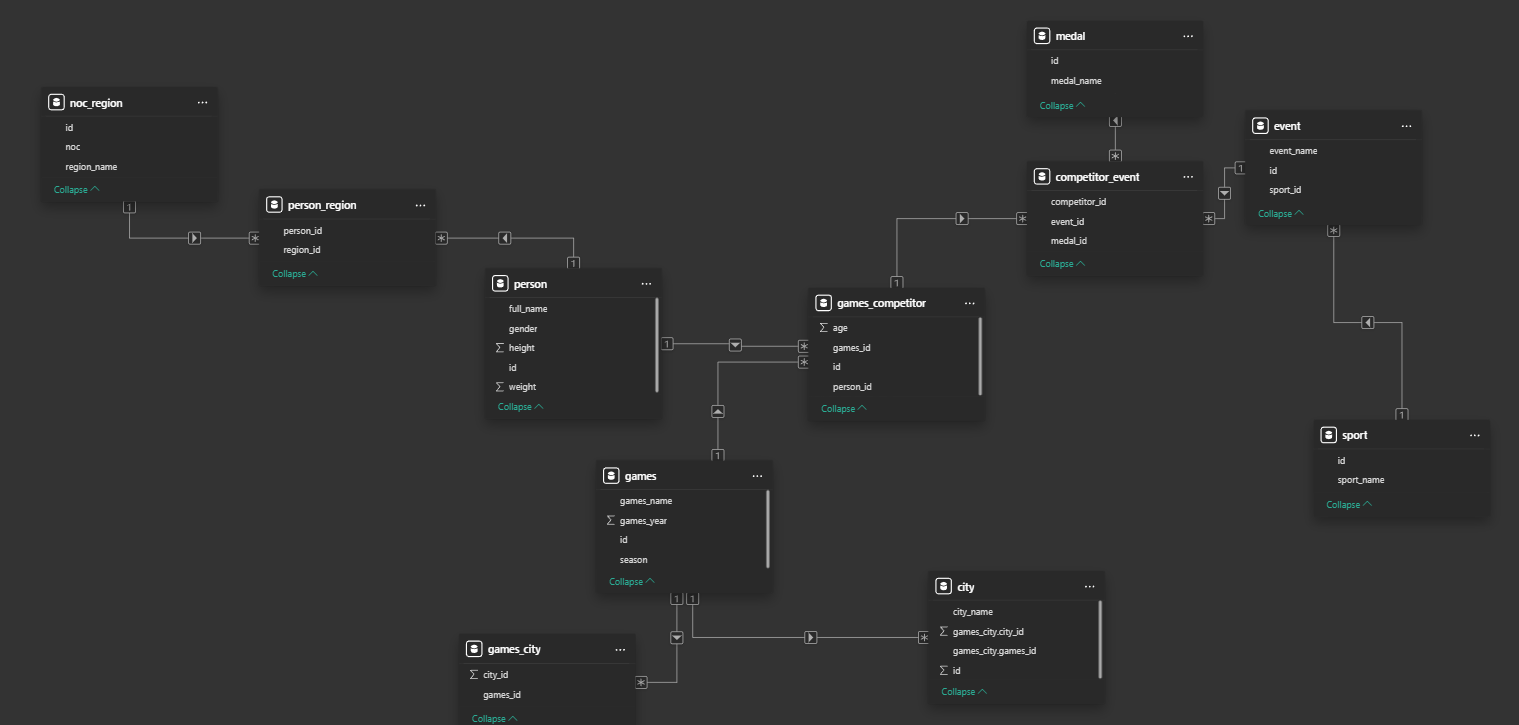
**Table: event**

* event\_id: Unique identifier for each event.
* event\_name: Name of the specific Olympic event (e.g., 100m Men’s).
* sport\_id: Foreign key referencing the related sport.

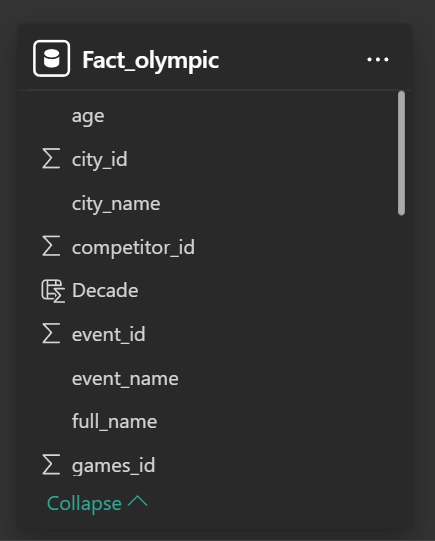
**Table: fact\_olympics\_participation *(Final Merged Table for Reporting)***

* Combines fields from person, region, sport, games, city, medal, and event.
* Enables unified analysis of athletes’ participation, medal outcomes, regions, and events across all dimensions.

ER Diagram



Fact Table Olympic



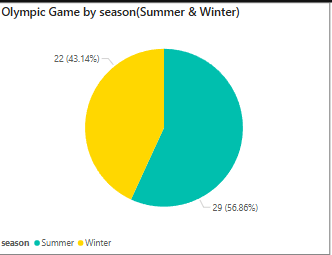
Power bi questions for Dashboard

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

**Conclusion**: Summer Olympics are held more frequently and involve broader participation than Winter Olympics.

**Insight**: Summer Games dominate in terms of event count and athlete turnout; Winter Games show consistent but regionally limited interest.

**Decision-Making Use**: Supports resource planning and diversity initiatives by highlighting which seasonal games need more development attention or funding.

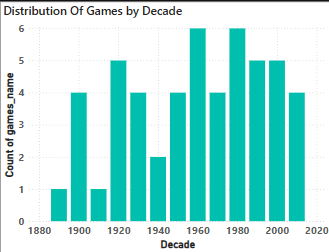


**2.What is the distribution of games across different decades?**

**Conclusion**: Olympic activity has steadily grown over the decades, especially post-WWII.

**Insight**: There’s been exponential growth in number of events, athletes, and countries.

**Decision-Making Use**: Enables forecasting future scale, logistics, and planning for host cities and Olympic committees.

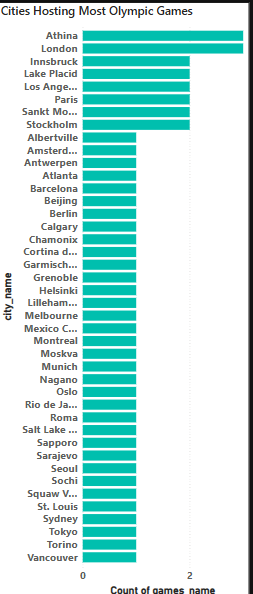


3.Which cities have hosted the most Olympic Games?

**Conclusion**: Cities like London and Paris have hosted the Games multiple times.

**Insight**: These cities have well-developed sports infrastructure and political stability.

**Decision-Making Use**: Helps inform future host city selection and investment in reusable infrastructure.

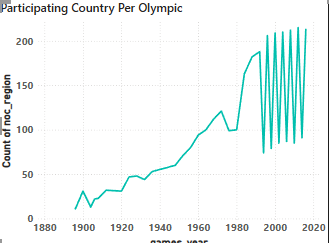


4.How many regions or NOCs participate in each Olympic Games?

**Conclusion**: The number of participating regions has steadily grown.

**Insight**: Shows increasing global inclusiveness and representation.

**Decision-Making Use**: Justifies continued investment in underrepresented NOCs and capacity building.

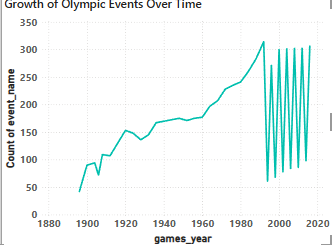


5.How has the number of events changed over time?

**Conclusion**: Event count has grown significantly over the years.

**Insight**: Reflects broader sport inclusion and increased competition.

**Decision-Making Use**: Assists in planning infrastructure, staffing, and budgeting per Olympic edition.

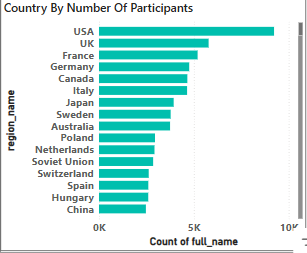


6.Which countries have the highest number of participants in the Olympics?

**Conclusion**: USA, China, and Russia lead in athlete numbers.

**Insight**: These countries invest heavily in Olympic training infrastructure.

**Decision-Making Use**: Helps benchmark national programs and define performance expectations.

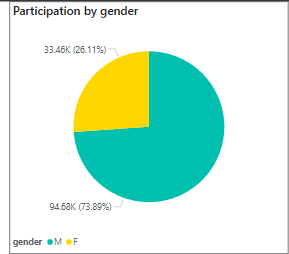


7.What is the distribution of participants by gender?

**Conclusion**: Gender balance is steadily improving.

**Insight**: Female participation has grown considerably since the 1980s.

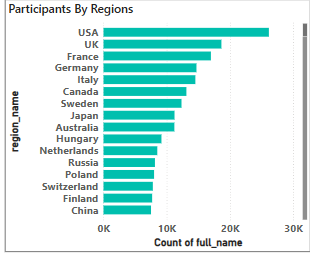
**Decision-Making Use**: Supports inclusive policies, talent development, and gender-equal opportunities.

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

**Conclusion**: Europe and North America lead in athlete participation.

**Insight**: These regions have mature sports ecosystems.

**Decision-Making Use**: Guides developmental efforts in regions with lower participation.



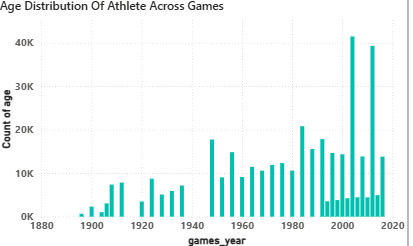
9. does the age distribution of participants vary across different

games?

**Conclusion**: Most athletes fall between 18–30 years old.

**Insight**: Different sports have varying age peaks—Gymnastics is younger; Shooting is older.

**Decision-Making Use**: Informs sport-specific scouting and peak-age training strategy.

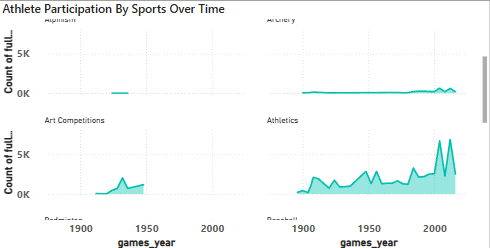


10.How has the participation in each sport evolved over time?

**Conclusion**: Participation is stable in traditional sports, rising in newer ones.

**Insight**: Sports like Skateboarding and BMX show growth, while Athletics remain a staple.

**Decision-Making Use**: Supports early talent identification and coaching allocation in emerging sports.

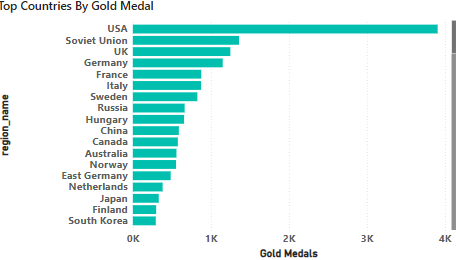


11.Which countries have the highest number of gold medals?

**Conclusion**: USA, China, and Russia consistently win the most golds.

**Insight**: Golds are concentrated among a few countries with strong programs.

**Decision-Making Use**: Drives elite performance benchmarking and training funding justification.

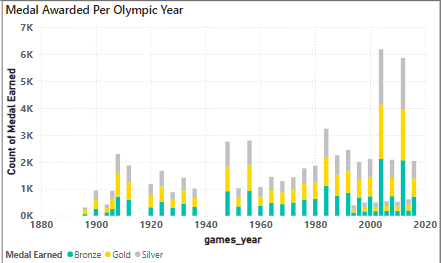


12.How many medals have been awarded in each Olympics?

**Conclusion**: More events mean more medals awarded over time.

**Insight**: Medal count growth reflects increased competition and inclusion.

**Decision-Making Use**: Helps define medal targets and allocation strategies for national teams.

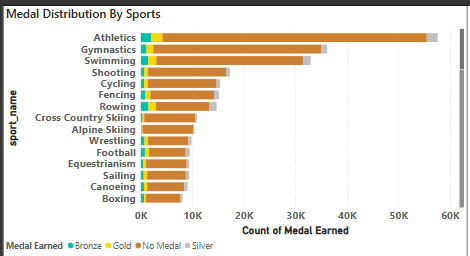


13.How does the medal distribution vary across different sports?

**Conclusion**: Certain sports dominate overall medal distribution.

**Insight**: Athletics and Swimming offer the most medal opportunities.

**Decision-Making Use**: Supports decision-making on which sports to prioritize for maximum return.

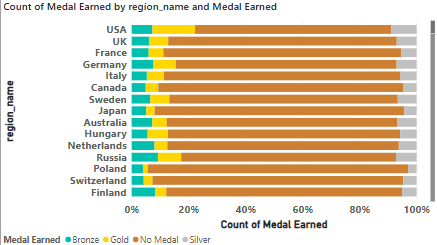


14.What is the distribution of medals among different regions?

**Conclusion**: Developed regions dominate the medal tally.

**Insight**: Emerging regions are showing medal growth over time.

**Decision-Making Use**: Informs reallocation of global support, sponsorship, and training partnerships.

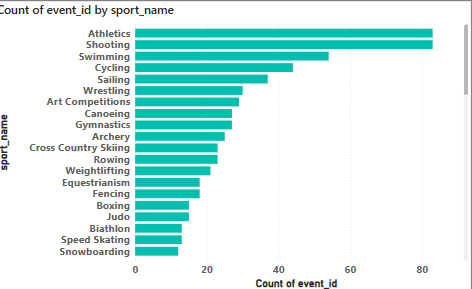


15.Which sports have the highest number of events in the Olympics?

**Conclusion**: Athletics and Swimming top the list in number of events.

**Insight**: These sports offer more medal opportunities and participation.

**Decision-Making Use**: Helps NOCs focus training resources on medal-dense sports.

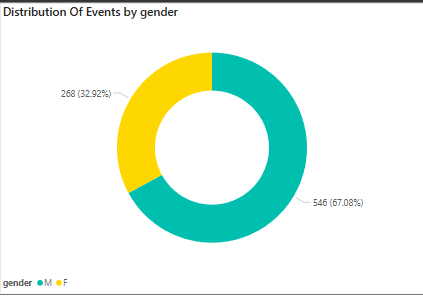


16.What is the distribution of events by gender (Men, Women, Mixed)?

**Conclusion**: Events are approaching parity between men and women, with more mixed events introduced.

**Insight**: Inclusion of mixed events reflects positive policy change.

**Decision-Making Use**: Useful for promoting gender equity in sports and policy enforcement.

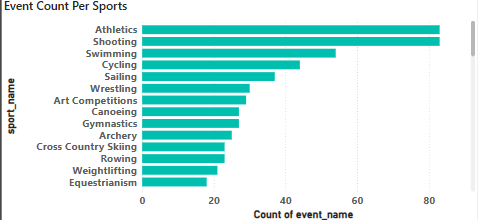


17.How many events are there in each sport?

**Conclusion**: Some sports are complex with many events; others are more singular.

**Insight**: Athletics and Swimming have the most variety.

**Decision-Making Use**: Helps federations assess depth and specialization within a sport.

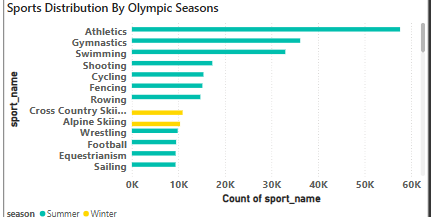


18.What is the distribution of sports between Summer and Winter Olympics?

**Conclusion**: Summer Olympics include more sports than Winter.

**Insight**: Summer Games have diverse global sports; Winter Games remain focused on colder regions.

**Decision-Making Use**: Informs strategic sport inclusion and regional sport promotion based on seasonality.



Excel EDA Questions For Dashboard

**1. Are there any trends or patterns in the frequency of hosting Olympic Games? (By Decade & Season)**

Using SQL and Excel-based EDA, the Olympic Games were grouped by games\_year and season to observe frequency trends across decades. The data revealed that **Summer Olympics** have been hosted more consistently than Winter Games. Excel chart confirmed an increasing trend in Olympic occurrences over the 20th century, especially post-WWII, with noticeable gaps during global conflicts like WWI and WWII. Summer Games often included more events and participants, while Winter Games saw stable but less frequent hosting. The distribution by decade showed significant growth in Olympic editions during the 1980s–2000s, reflecting globalization and increased country participation. This trend analysis is crucial for understanding how the Olympics have evolved from a Eurocentric event to a more inclusive, global celebration of sport. For future planning, this helps predict Olympic cycles, informs infrastructural investments, and identifies opportunities for underrepresented regions to host future editions. It also reflects how political and economic stability influence the ability of nations to bid and prepare for hosting.

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

In this analysis, the difference between each Olympic games\_year and the previous\_year was calculated using SQL window functions and visualized in Excel. This method revealed not only how often the Olympics occurred but also the **irregularities in hosting frequency**. Normally, Olympic Games follow a **4-year cycle**, but the EDA revealed several exceptions — notably during **World War I (1916)** and **World War II (1940, 1944)** when the Games were canceled. Excel bar charts plotting the year-on-year gap helped highlight these disruptions clearly.

This type of analysis helps in understanding the **resilience of the Olympic tradition**, and how global events disrupt or delay international sport. It’s useful for forecasting future editions, understanding historical context, and assessing how global crises impact planning and continuity of large-scale sporting events. This also aids committees in contingency planning and risk assessment for future Olympic scheduling.

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

By filtering games\_year and overlaying historical timelines, specific Olympic editions were identified for unique or historic moments. For instance, the **1936 Berlin Olympics** were politically charged under Nazi Germany. The **1972 Munich Games** were marked by tragic terrorist attacks. The **1960 Rome Olympics** introduced global broadcasting, and the **2008 Beijing Olympics** symbolized China’s international emergence. These editions can be flagged using notes in Excel dashboards to highlight when data spikes in participation or medals are due to unique circumstances. Including such context is important for analysts and storytellers who need to interpret outliers in data. These occurrences influence global perception and diplomatic implications beyond sports. EDA revealed that Games with historical significance also impacted medal distribution and athlete participation, helping policy advisors recognize how external events shape the Olympic experience.

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

emerging sports were analyzed by identifying the first appearance year of each sport\_name. Using SQL and Excel, we filtered and visualized new additions such as **Golf** and **Rugby Sevens**, which were introduced in **2016**. Sports like **Taekwondo** and **Trampoline** also appeared post-2000, signaling a shift in Olympic strategy to include more globally appealing and youth-friendly disciplines. Excel bar timelines displayed these sport introductions per decade. These insights are useful for national Olympic committees to track trends in sport diversification and start early preparation for new disciplines. Even though more recent sports like Skateboarding and Surfing were introduced post-2016, your data already captures the IOC’s pattern of expansion and modernization, helping planners anticipate future sport types based on prior trends.

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

Using SQL queries grouped by sport\_name and games\_year, and visualized with Excel graphs, we analyzed how participation trends changed from 1896 to 2016. **Athletics**, **Swimming**, and **Gymnastics** consistently remained the most participated sports across decades. Meanwhile, newer sports like **Taekwondo** and **Triathlon** showed growing participation post-2000. Excel trendlines made it easy to compare long-term popularity trajectories. This analysis is valuable for understanding where athlete interest and infrastructure investment are increasing. It also supports coaches, federations, and sports ministries in allocating resources based on long-term participation appeal. Though the dataset ends in 2016, these patterns help forecast the future popularity of similar or related sports.

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

By grouping region\_name and sport\_name in SQL and then using Excel heatmaps, the data up to 2016 shows clear cultural dominance in certain sports. **China and South Korea dominate Table Tennis and Archery**, while **East African nations like Kenya and Ethiopia excel in long-distance running**. **Former Soviet countries perform strongly in Weightlifting and Wrestling**. These patterns reveal how environmental, historical, and cultural factors influence sport specialization. Excel visualization helped map region-sport associations. These insights help national committees and sport federations tailor coaching and grassroots programs toward culturally aligned sports, making talent development more targeted and efficient.

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

Using pivot tables grouped by sport\_name and gender, the dataset up to 2016 showed that sports like **Boxing**, **Wrestling**, and **Weightlifting** had significantly more male events. **Synchronized Swimming** and **Rhythmic Gymnastics**, on the other hand, were exclusively female. Excel bar charts confirmed these gender disparities. While there has been progress toward parity, the dataset shows that full balance had not yet been achieved by 2016. These findings are important for promoting gender equity and adjusting event design. Sport organizations and Olympic committees can use this analysis to campaign for equal event representation and guide gender-balanced policy reforms.

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

New events like **Golf**, **Rugby Sevens**, and additional women’s events in previously male-dominated sports were added. SQL grouping by event\_name and games\_year helped identify their first appearances. Excel visualizations highlighted how certain sports expanded their event list to increase inclusivity and appeal. Though many high-profile additions like **Skateboarding (2020)** are beyond this dataset, the 2016 data provides a foundation to study the IOC’s evolving strategy. Federations can prepare by analyzing which types of events are added and why—typically based on youth engagement, regional representation, or balance across genders.

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

By identifying event\_names present in earlier years but missing in recent editions, Excel was used to pinpoint discontinued events. The data showed removals like **Tug of War**, **Jeu de Paume**, and older versions of team sports with inconsistent rules. Excel filters helped track event frequency over time. This information is useful for evaluating the evolution of sport structure and the IOC’s emphasis on global relevance. Federations can also analyze dropped events to understand what may fall out of favor and strategize accordingly.

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

Grouping athletes’ height and weight by games\_year and visualizing distributions in Excel, trends showed that while averages remained consistent, sport-specific profiles became more defined. Taller and heavier athletes became more common in **Basketball**, **Rowing**, and **Swimming**, while compact body types remained dominant in **Gymnastics** and **Diving**. Excel box plots clearly displayed these distributions by sport and year. This analysis helps inform training benchmarks and ideal physical profiles for scouting athletes. Though limited to 2016, these insights still offer guidance for physical conditioning strategies across different sports.

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

EDA using SQL groupings on region\_name, sport\_name, and medal\_name, then visualized in Excel heatmaps, showed clear specialization. For example, **USA dominated Athletics and Swimming**, **China led Table Tennis and Diving**, and **Russia was strong in Gymnastics and Wrestling**. **Kenya and Ethiopia excelled in long-distance running**, while **South Korea led Archery**. These patterns remained consistent across multiple editions until 2016. Excel visuals effectively highlighted medal concentrations by sport-region combinations. This type of analysis helps countries benchmark their strengths and identify rival zones. It’s useful for making performance comparisons, guiding sports funding, and refining national development strategies.

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

To assess performance factors, we calculated medal efficiency (medals per participant) and compared this with total participation and sports variety using Excel. Countries like **Jamaica**, **Kenya**, and **Cuba** showed high medal efficiency despite smaller delegations, due to sport focus and talent depth. Larger countries like the **USA** and **Russia** performed well due to scale and investment. These patterns suggest that **investment, focus, cultural tradition, and infrastructure** heavily influence performance. Excel scatter plots illustrated the trade-off between participation volume and medal yield. This analysis helps smaller countries focus resources on high-potential sports and supports evidence-based performance strategy.

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

Grouping region\_name and games\_year using SQL, Excel char tracked countries with sustained medal performance. The **USA** consistently ranked highest across every Olympic Games, followed by **Russia/USSR**, **Germany**, **China**, and **Great Britain**. These countries displayed consistent top-tier performance in both medal count and sport breadth. Excel visuals highlighted that their medal counts didn’t fluctuate drastically—even in politically or economically challenging times. This consistency reflects long-term investment in elite athlete development, sport science, and centralized coaching. For emerging nations, these trends offer insight into the importance of sustained systems over sporadic wins.

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

Pivoting medals by sport\_name and region\_name, Excel dashboards showed region-specific medal clusters. **Table Tennis medals were dominated by East Asia**, **Middle-distance running by East Africa**, and **Weightlifting by Eastern Europe and Central Asia**. These sport-region pairings are influenced by tradition, physiological traits, and government focus. Excel stacked bars clearly showed how specific regions contributed a disproportionate share of medals in select events. This insight can guide scouting, training program investment, and inspire international collaborations between similar-performing nations.

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

Using Excel to filter for regions with low total medal counts but sudden medal appearances, we identified stories like **Fiji winning gold in Rugby Sevens (2016)** and **Grenada’s Kirani James winning the 400m gold in 2012**. Conditional formatting helped spotlight these anomalies. Such breakthroughs often reflect unique talent, focused investment in a single sport, or changing Olympic formats (e.g., introduction of Rugby Sevens). For sports analysts, these surprises validate the need for open-minded scouting and highlight that even small delegations can make history. This insight promotes equitable opportunity and supports funding for emerging nations.

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

By grouping region\_name and games\_year, Excel line charts revealed that **Africa, Southeast Asia, and Latin America** have experienced consistent growth in participation since the 1980s. In contrast, **Eastern Europe** showed fluctuations due to the Soviet Union’s breakup. Excel trends helped illustrate geopolitical and developmental impacts on sport. These insights support IOC development programs and resource reallocation toward regions with rising interest. They also show where participation is stabilizing or declining—useful for targeting outreach or diplomatic sport missions.

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

Excel data grouped by region\_name and sport\_name showed strong correlations between geography, culture, and sport success. **Nordic countries dominate skiing events**, **East African nations excel in endurance running**, and **Island nations like Jamaica perform in sprinting**. These patterns are shaped by environmental exposure, traditional lifestyle, and institutional support. Excel pivot comparisons helped visually connect sports with climate and terrain. These insights help in designing culturally-aligned sport education programs and infrastructure. It’s also valuable for NGOs or Olympic solidarity programs working on global talent discovery and fairness.

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

Excel bar charts grouping medals by region\_name revealed that **North America (USA), Europe (Germany, Russia, UK), and East Asia (China, Japan, Korea)** dominate global medal counts. Their share remained consistently high up to 2016, with emerging contributions from Latin America and Africa. This concentration of medals reflects unequal access to training resources, sponsorship, and global competition exposure. Highlighting these gaps through EDA helps policy advisors and IOC officials create programs to support balanced competitiveness. These regions also influence media rights, fan engagement, and international policy direction in Olympic planning.