

# Capstone Project

## Sports Project Documentation

### 1. Project Overview

**Title:** Historical Data Analysis of the Olympic Games

**Domain:** Sports Analytics / Data Analysis

**Tools & Technologies:** MySQL, Microsoft Excel.

**Objective:** To extract insights from historical Olympic Games data by analyzing trends in games, sports, events, participants, regions, and medal distributions.

### 2. Problem Statement

The objective of this analysis is to gain valuable insights into the historical data of the Olympic Games by exploring various dimensions, including games, sports, events, participants, medals, and regional representation. The analysis spans multiple editions of the Olympics, identifying key patterns, trends, and notable observations. The final goal is to generate actionable insights and recommendations for improving the organization and diversity of future Olympic Games.

### 3. Dataset Description

The dataset comprises multiple normalized tables extracted from a relational MySQL database. It includes detailed records from various Olympic Games, capturing information about sports, events, participants, locations, and results.

#### *Table Summaries*

- **Sport Table** – List of all sports featured in Olympics.
- **Event Table** – Details of events under each sport, categorized by gender.

- **City Table** – Cities that hosted or participated in the Olympics.
- **Games Table** – Each Olympic edition with year and season type.
- **Games\_City Table** – Mapping of games and host cities.
- **NOC\_Region Table** – Country codes and region names.
- **Person Table** – Athlete profiles with demographic data.
- **Person\_Region Table** – Countries represented by athletes.
- **Games\_Competitor Table** – Participation of athletes in Olympic editions.
- **Medal Table** – Types of medals.
- **Competitor\_Event Table** – Largest table combining competitors, events, and medal outcomes.
- **Consolidated\_facts** - enormalized fact table combining all the essential columns from the above tables — optimized for charting and filtering in Excel or BI tools.

## 4. Methodology

Each of the following 18 EDA questions was approached with the following steps:

- **Step 1:** Frame SQL queries in MySQL Workbench.
- **Step 2:** Import query results into Excel.
- **Step 3:** Generate visualizations using pivot tables and charts.
- **Step 4:** Write insights and conclusions per question.
- **Step 5:** Create a final dashboard summarizing key visualizations.

## 5. EDA Questions and Analysis Summary

### *Sheet 1 – Frequency of Hosting Olympic Games*

EDA Question 1: Are there any trends or patterns in the frequency of hosting Olympic Games?

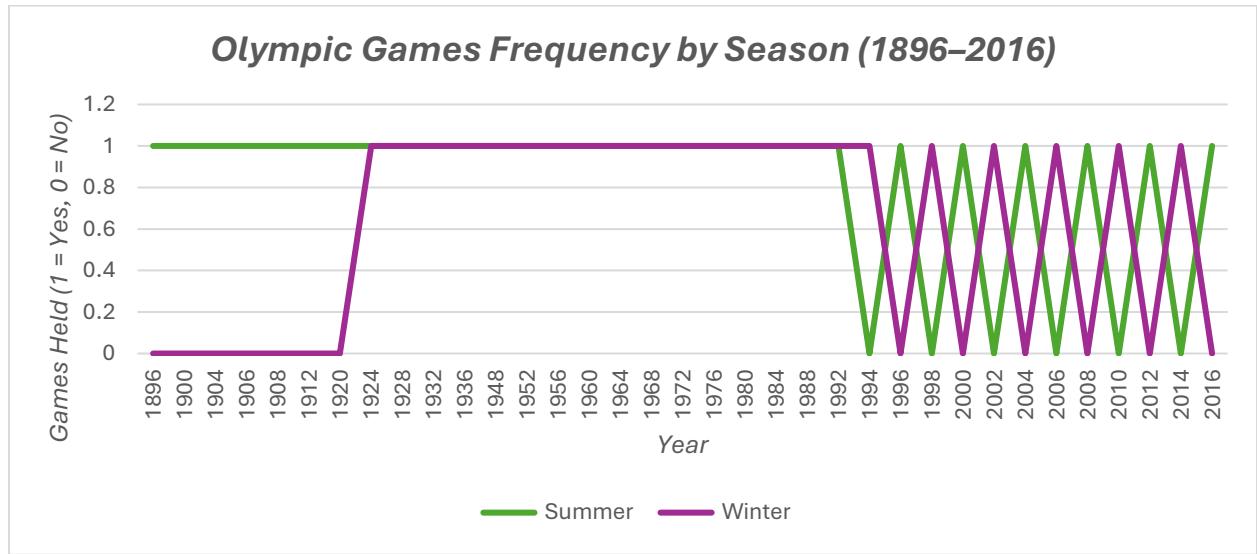
## Approach

To address this question, data was extracted from the consolidated\_fact table using the following SQL query:

```
SELECT games_year,  
       MAX(CASE WHEN games_season = 'Summer' THEN 1 ELSE 0 END) AS Summer,  
       MAX(CASE WHEN games_season = 'Winter' THEN 1 ELSE 0 END) AS Winter  
FROM consolidated_fact  
GROUP BY games_year  
ORDER BY games_year;
```

This query identifies whether Summer and/or Winter Olympic Games were held in each year between 1896 and 2016. The results were exported to Excel, and a line chart was generated to visually represent the occurrence of each Olympic season across time.

## Visualization



## Conclusion

The analysis of Olympic Games hosting patterns from **1896 to 2016** reveals the following key trends:

- **1896–1920:** Only **Summer Olympics** were held.
- **1924–1992:** Both **Summer and Winter Games** occurred every four years in the same year.
- **From 1994 onwards:** The **Winter Games** were shifted to occur **every two years**, alternating with the Summer Games.
- **Interruptions** were observed in **1916, 1940, and 1944** due to the World Wars.
- The post-1994 schedule reflects a **more consistent and predictable pattern**, allowing for better event focus, planning, and media attention.

### *Sheet 2 – Duration of Olympic Games*

EDA Question 2: How has the duration of Olympic Games changed over time?

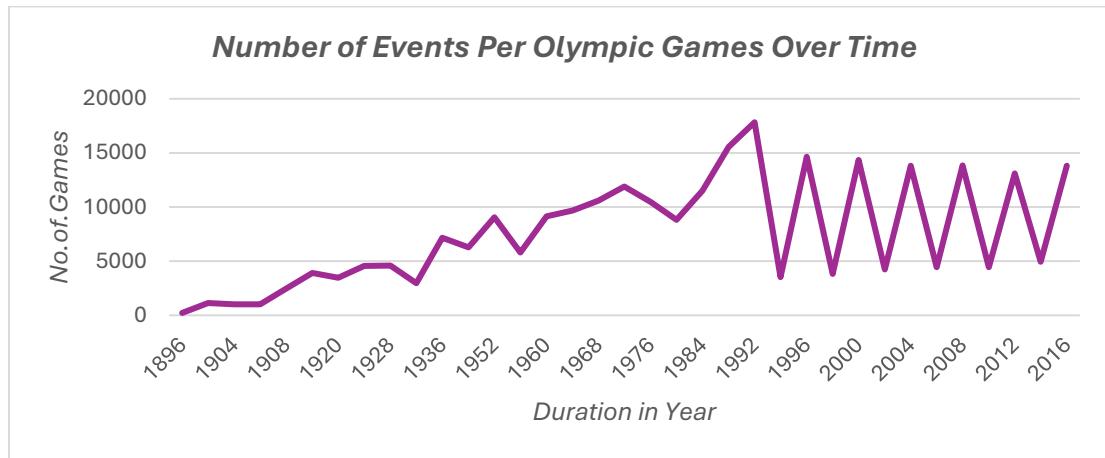
#### Approach

To explore how the scale and duration of the Olympic Games evolved, a query was executed on the consolidated\_fact table to count the number of events conducted in each edition of the Olympics:

```
SELECT games_year, COUNT(event_name) AS NumberOfEvents
FROM consolidated_fact
GROUP BY games_year
ORDER BY games_year ASC;
```

This query summarizes the total number of events held in each Olympic year. The result was exported to Excel, and a line chart was created to visualize how the number of events has changed over time.

## Visualization



## Conclusion

The analysis of Olympic Games data from **1896 to 2016** reveals a dramatic expansion in the **scale and complexity** of the Games:

- In **1896**, the Olympics featured **only 219 events**.
- By **2016**, the number of events increased to **13,806**, demonstrating substantial growth.
- A significant surge occurred post-1950s, coinciding with **global participation growth**, enhanced international coordination, and diversification of sports.
- Noticeable **dips** in 1994, 1998, 2002, and alternate years reflect **Winter Olympic editions**, which traditionally feature fewer events than their Summer counterparts.
- The overall trend points to the Olympics becoming **longer, larger, and more sophisticated**, reinforcing their prominence on the global stage.

## *Sheet 3 – Notable Events by Edition*

EDA Question 3: Are there any notable events or occurrences associated with specific Olympic Games?

## Approach

To identify notable Olympic events based on participant count, a query was run on the consolidated\_fact table. This query grouped data by Olympic edition (game\_name), sport\_name, and event\_name, then counted the number of participants for each combination:

```
SELECT game_name, sport_name, event_name, COUNT(games_competitor_id) AS Participants
```

```
FROM consolidated_fact
```

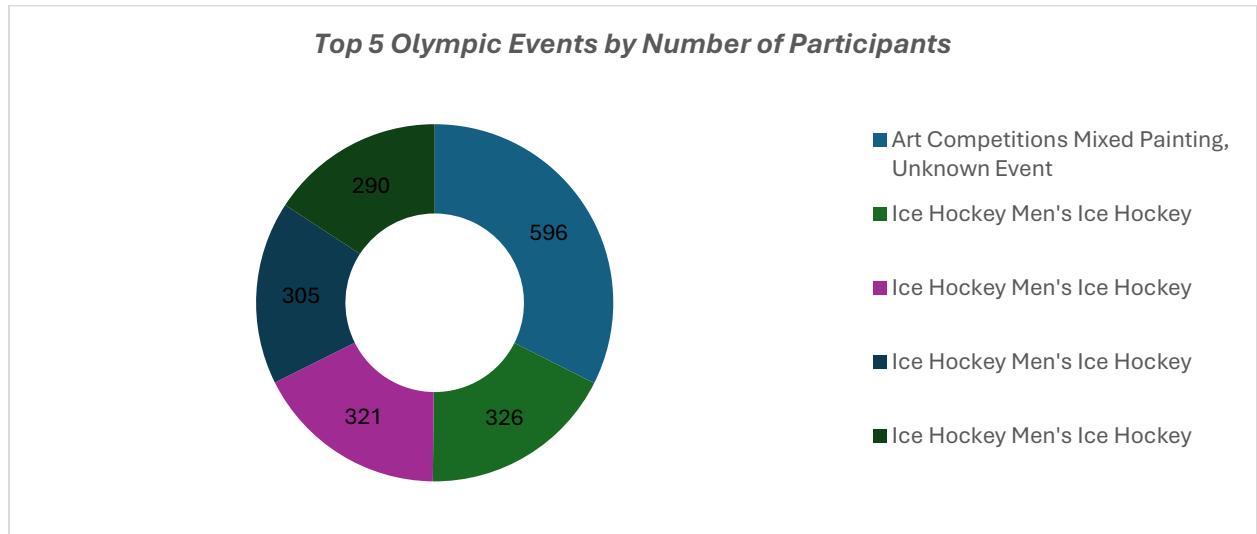
```
GROUP BY game_name, sport_name, event_name
```

```
ORDER BY Participants DESC
```

```
LIMIT 5;
```

The output was exported to Excel, where a donut chart was used to showcase the **top 5 Olympic events with the highest number of participants**.

### Visualization



### Conclusion

The analysis highlights several significant insights and anomalies in Olympic history:

- The **Art Competition** with **596 participants**, the highest count in the dataset. This event is historically notable, as art competitions were once a part of the Olympics but have since been discontinued.
- **Ice Hockey** dominates the list with participation counts between **290 and 326**, underscoring its status as a major team sport in the Winter Olympics.
- **Football (Men's)**, although not in the top five here, also shows high and consistent participation in broader data, further supporting its popularity.
- **Basketball**, in contrast, appears less frequently in the top rankings by participant count.

Overall, the findings emphasize that **team sports like Ice Hockey and Football consistently draw large numbers of participants**, reflecting their enduring appeal and logistical scale in the Olympic Games.

#### *Sheet 4 – Emerging Sports*

EDA Question 4: Are there any emerging sports that have been recently added to the Olympics?

#### Approach

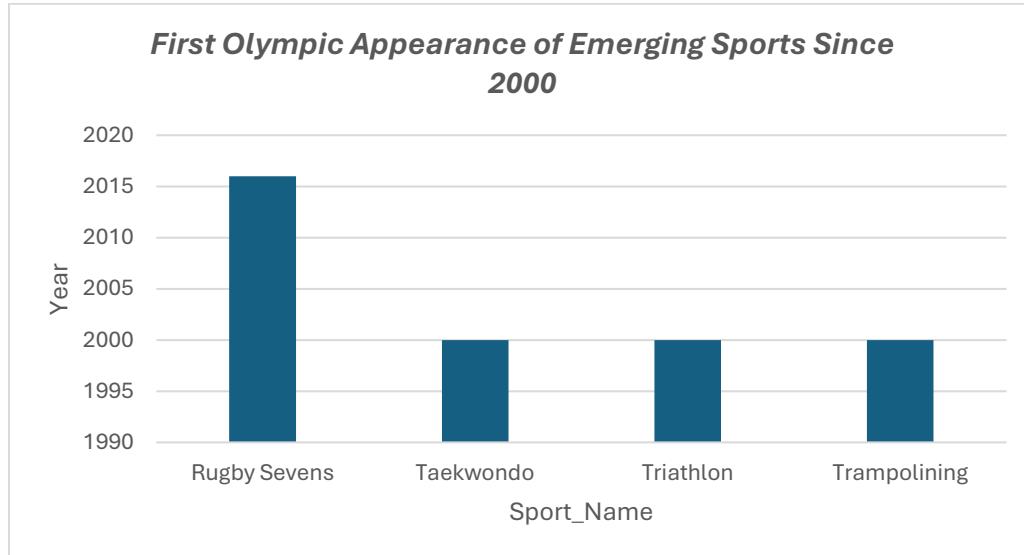
To identify newly added sports in the Olympic Games since the year 2000, the following SQL query was executed on the consolidated\_fact table. The goal was to retrieve the earliest year each sport appeared, then filter only those whose first appearance was **from 2000 onwards**:

```
SELECT sport_name, MIN(games_year) AS first_appearance
FROM consolidated_fact
GROUP BY sport_name
HAVING first_appearance >= 2000
```

**ORDER BY first\_appearance DESC;**

The resulting data was exported to Excel and visualized using a bar chart, showing the timeline of these emerging sports' first Olympic appearances.

### **Visualization**



### **Conclusion**

This analysis highlights the inclusion of **emerging sports** in the Olympic Games post-2000, emphasizing the evolving nature of the event lineup:

- **Rugby Sevens** made its **Olympic debut in 2016**, the most recent among the group.
- **Taekwondo, Triathlon, and Trampolining** were all introduced during the **2000 Sydney Olympics**, marking a significant point of diversification.
- These additions reflect the International Olympic Committee's strategic efforts to **modernize and globalize** the Games by embracing sports with increasing popularity and youth appeal.
- The inclusion of these sports demonstrates the Olympics' **adaptability to global sporting trends** and its intent to engage **new audiences**.

## *Sheet 5 – Popularity Trends of Sports*

EDA Question 5: How has the popularity of certain sports changed over the years?

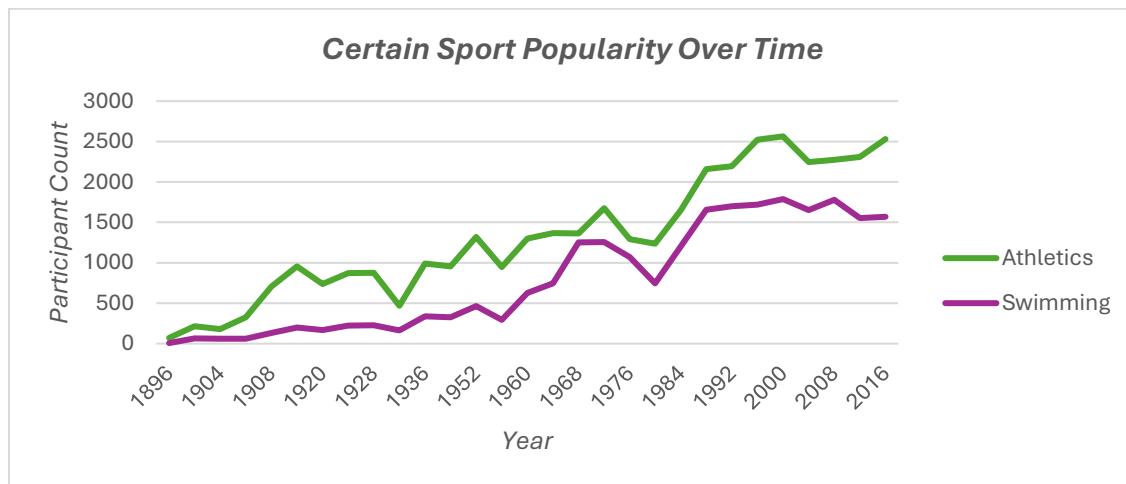
### Approach

To examine the historical popularity of selected sports in the Olympics, the following SQL query was used. It calculates the number of participants in **Athletics** and **Swimming** for each Olympic year:

```
SELECT sport_name, games_year, COUNT(games_competitor_id) AS popularity  
FROM consolidated_fact  
WHERE sport_name IN ('Athletics', 'Swimming')  
GROUP BY sport_name, games_year  
ORDER BY sport_name, games_year;
```

The results were exported to Excel and visualized as a multi-line chart showing how the participant count for each sport evolved over time.

### Visualization



### Conclusion

The analysis reveals important trends in the popularity of Olympic sports:

- **Athletics** has consistently maintained the highest participation levels across all Olympic editions. A marked increase is evident after the **1960s**, with peaks around **1996–2008**.
- **Swimming** also shows a **steady upward trajectory**, particularly gaining momentum from the **1950s** onward. However, its participant numbers have consistently remained below those of Athletics.
- Both sports demonstrate a strong **long-term growth trend**, indicating their continued importance and relevance in the Olympic Games.
- The findings underscore how **traditional core sports** not only retain their foundational place in the Olympics but also continue to **attract growing global interest and athlete participation**.

### *Sheet 6 – Region-Specific Sports*

EDA Question 6: Are there any sports that are specific to a particular region or culture?

#### Approach

To investigate whether certain Olympic sports are predominantly represented by athletes from specific regions, a two-step **Common Table Expression (CTE)** SQL approach was employed:

```
WITH total_competitor AS (
  SELECT sport_name,
    COUNT(DISTINCT games_competitor_id) AS total_athletes
  FROM consolidated_fact
  GROUP BY sport_name
),
sport_by_region AS (
```

```

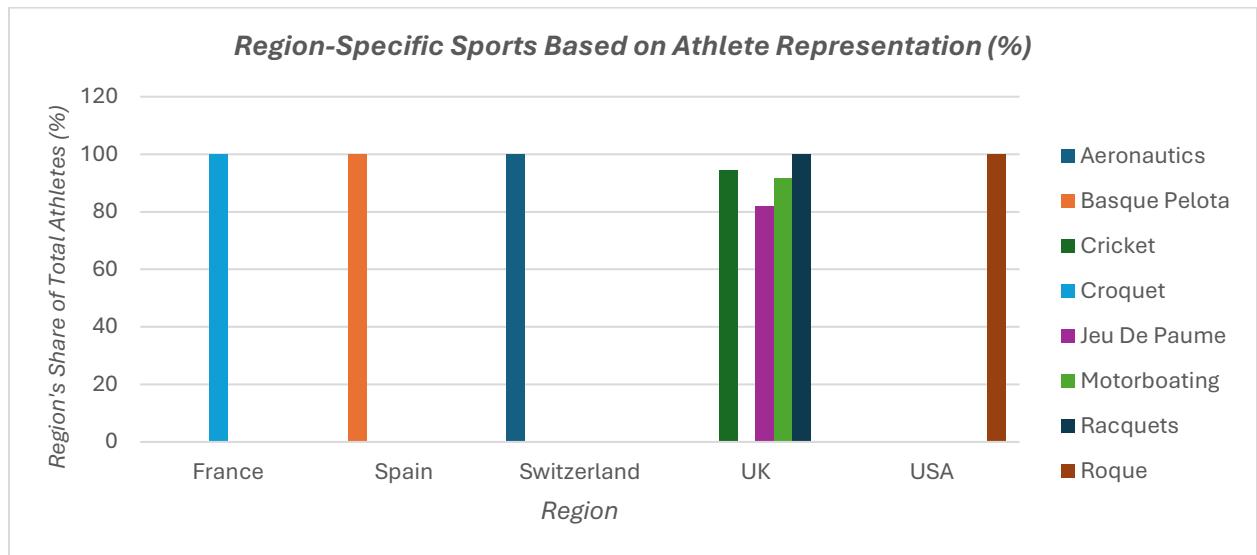
SELECT sport_name,
       region_name,
       COUNT(DISTINCT games_competitor_id) AS athletes_count
  FROM consolidated_fact
 GROUP BY sport_name, region_name
)

SELECT
       s.sport_name,
       s.region_name,
       s.athletes_count,
       t.total_athletes,
       ROUND((s.athletes_count * 1.0 / t.total_athletes) * 100, 2) AS region_share_percent
  FROM sport_by_region s
 JOIN total_competitor t ON s.sport_name = t.sport_name
 WHERE (s.athletes_count * 1.0 / t.total_athletes) > 0.8
 ORDER BY region_share_percent DESC;

```

This query identified sports where **more than 80%** of the participants came from a single region—indicating potential **regional or cultural specialization**. The data was then visualized using a clustered column chart in Excel.

## Visualization



## Conclusion

The analysis clearly highlights sports that are **regionally concentrated or culturally unique**:

- **Aeronautics** – 100% of athletes from **Switzerland**
- **Basque Pelota** – 100% from **Spain**
- **Croquet** – exclusively represented by **France**
- **Roque** – 100% from the **USA**
- **Racquets** – 100% from the **UK**

Additionally, other sports with **strong regional skew** include:

- **Cricket** – 94.44% of participants from the **UK**
- **Motorboating** – 91.67% from the **UK**
- **Jeu De Paume** – 81.82% from the **UK**

These results indicate that such sports are either:

- **Culturally rooted** in their respective regions,
- Historically played in limited geographies,
- Or are **discontinued events** that never gained widespread adoption.

This insight reinforces how the **Olympic Games reflect not only global athleticism but also the cultural identities and sporting traditions of specific nations.**

### *Sheet 7 – Gender Disparity in Events*

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

#### Approach

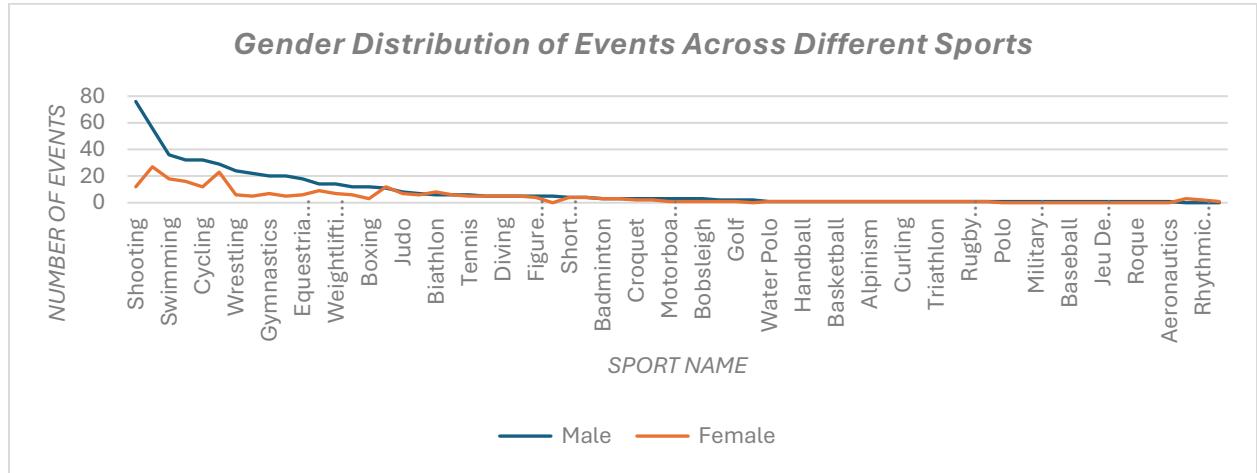
To analyze gender disparities in Olympic sports events, the following SQL query was executed. It first identified **distinct sport–event–gender combinations**, then aggregated the number of events held for **males (M)** and **females (F)** across all sports:

```
SELECT sport_name,  
       SUM(CASE WHEN gender = 'M' THEN 1 ELSE 0 END) AS Male,  
       SUM(CASE WHEN gender = 'F' THEN 1 ELSE 0 END) AS Female  
FROM (  
       SELECT DISTINCT sport_name, gender, event_name  
       FROM consolidated_fact  
       JOIN person ON consolidated_fact.person_id = person.id  
) AS unique_event  
GROUP BY sport_name
```

**ORDER BY Male DESC, Female DESC;**

This dataset was exported to Excel and used to create a multi-line chart comparing the number of male and female events per sport.

### Visualization



### Conclusion

This analysis reveals **significant gender-based disparities** in Olympic event availability:

- Sports like **Shooting, Cycling, Wrestling**, and **Canoeing** have historically had **many more male-specific events** than female ones.
- In contrast, **Rhythmic Gymnastics** and **Synchronized Swimming** are exclusively **female**, showcasing the opposite imbalance.
- Some modern and neutral sports have shown **better gender parity**, but overall, the data illustrates that **male athletes have had broader representation in terms of available events**.
- These disparities reflect historical biases and **evolving gender roles in sports**, though recent editions of the Olympics have shown gradual efforts toward equal representation.

### *Sheet 8 – New Events in Recent Editions*

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

#### Approach

To determine whether any **new sports** were introduced in the **recent Olympic Games (2016 or later)**, a SQL query was written using **Common Table Expressions (CTEs)** to compare the distinct sport names in two time periods: before and after 2016.

```
WITH past_year_sports AS (
    SELECT DISTINCT sport_name
    FROM consolidated_fact
    WHERE games_year < 2016
),
present_year_sports AS (
    SELECT DISTINCT sport_name
    FROM consolidated_fact
    WHERE games_year >= 2016
)
SELECT present.sport_name
FROM present_year_sports present
LEFT JOIN past_year_sports past ON present.sport_name = past.sport_name
WHERE past.sport_name IS NULL;
```

This query identifies **any sport that appears for the first time** in or after the year **2016**, and is absent in all prior editions.

## Visualization

*This question produces a single-result outcome rather than a dataset suitable for charting. Therefore, no chart was generated for this analysis.*

## Conclusion

The query result reveals that:

- **Rugby Sevens** is the **only new sport** introduced in or after the **2016 Olympic Games**.
- No other completely new sports were added during this period.

This indicates that the **International Olympic Committee (IOC)** has maintained a relatively **stable core set of sports** in recent years. Rather than frequently introducing new sports, the IOC appears to have **focused on evolving the format, structure, and gender balance within existing disciplines**.

The inclusion of Rugby Sevens represents an effort to modernize the Olympic program with sports that are **dynamic, globally relevant**, and appealing to **younger audiences**, aligning with the IOC's strategic goal to remain contemporary and engaging.

## *Sheet 9 – Discontinued Events*

EDA Question 9: Are there any events that have been discontinued or removed from the Olympics?

## Approach

To identify sports that were once part of the Olympic Games but are no longer included in recent editions, a SQL query was used to find the **last recorded year** each sport appeared in the dataset:

```
WITH last_appearance AS (
    SELECT sport_name, MAX(games_year) AS last_year
    FROM consolidated_fact
```

GROUP BY sport\_name

)

SELECT sport\_name, last\_year

FROM last\_appearance

WHERE last\_year < 2016;

This query returns a list of sports whose **last Olympic appearance was before 2016**, suggesting they have since been **discontinued or removed** from the official program.

### Visualization

*As this query returns a direct list of discontinued sports rather than time-series or quantitative data suited for graphical presentation, no chart was generated for this question.*

### Conclusion

The analysis confirms that **several sports have been removed from the Olympic program over time:**

- Notable discontinued sports include:
  - **Tug-of-War**
  - **Croquet**
  - **Roque**
  - **Jeu de Paume**
  - **Motorboating**

These events were likely discontinued due to factors such as:

- Low global popularity or participation
- Limited competitive appeal
- Logistical and infrastructure challenges

- IOC prioritization of sports with broader appeal

Some sports like **Baseball** and **Softball** were removed after **2008** but later **reintroduced in the 2020 Tokyo Olympics**, indicating that **event inclusion is sometimes reversible**, often depending on the **host country's interests** and global trends.

Certain sports that appear to end around **2014** (e.g., Luge, Figure Skating) are likely still active but may be missing from the dataset due to **data cutoff or limitations**, particularly for Winter Games.

This highlights the **dynamic nature of the Olympic sports roster**, reflecting changes in global interest, competitiveness, and the strategic direction of the **International Olympic Committee (IOC)**.

#### *Sheet 10 – Athlete Height and Weight Trends*

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

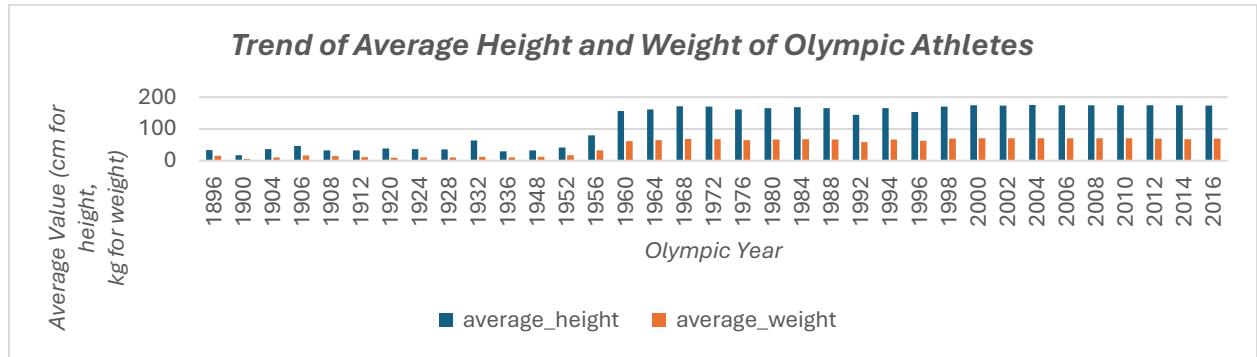
#### Approach

To analyze changes in the **physical characteristics** of Olympic athletes over time, the following SQL query was used to compute the **average height and weight per Olympic year**:

```
SELECT games_year,
       AVG(person_height) AS average_height,
       AVG(person_weight) AS average_weight
  FROM consolidated_fact
 WHERE person_height IS NOT NULL AND person_weight IS NOT NULL
 GROUP BY games_year
 ORDER BY games_year;
```

This query ensures that only valid height and weight data is considered, and it aggregates the data by year to identify historical trends. The results were visualized in a clustered bar chart to compare average height and weight across Olympic years.

### Visualization



This visual enables a clear comparison between height and weight progression over time.

### Conclusion

The data reveals **significant upward trends** in both **average height** and **weight** of Olympic participants, particularly after the **1950s**, with the most noticeable jump occurring **after 1960**:

- **Average height** steadily increased and has stabilized around **175–180 cm** in recent decades.
- **Average weight** also rose in parallel, stabilizing around **70 kg**.

This shift is likely due to:

- **Enhanced training methodologies**
- **Improved global nutrition and health**
- **More competitive and scientifically optimized athlete selection processes**

The early Olympic years display greater variance, likely due to **incomplete data or lower participant volume**. The clear trends in later years underscore the **evolution of athleticism and physical standards** required to compete at the Olympic level.

## *Sheet 11 – Dominance in Sports by Country*

EDA Question 11: Are there any dominant countries or regions in specific sports or events?

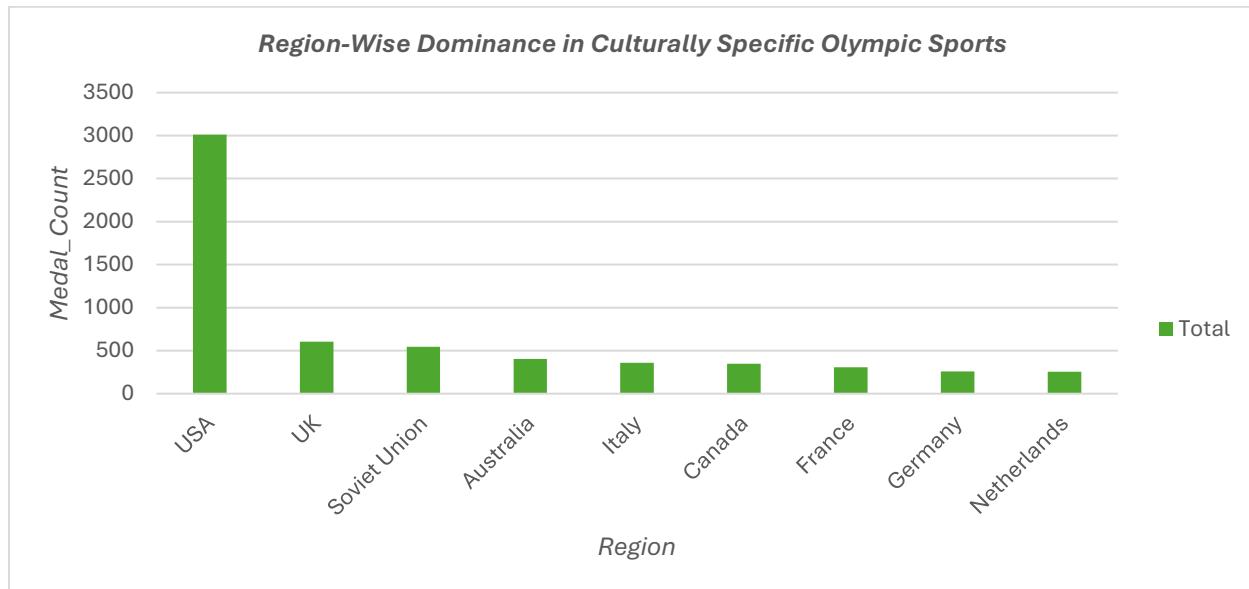
### Approach

To explore **regional dominance** in Olympic events, especially in sports with strong cultural or national affiliations, the following SQL query was executed:

```
SELECT region_name, sport_name,  
       SUM(CASE WHEN medal_name = 'Gold' THEN 1 ELSE 0 END) AS Gold,  
       SUM(CASE WHEN medal_name = 'Silver' THEN 1 ELSE 0 END) AS Silver,  
       SUM(CASE WHEN medal_name = 'Bronze' THEN 1 ELSE 0 END) AS Bronze,  
       COUNT(medal_name) AS Total_Medal_Count  
FROM consolidated_fact  
WHERE medal_name IN ('Gold', 'Silver', 'Bronze')  
GROUP BY region_name, sport_name  
ORDER BY Total_Medal_Count DESC  
LIMIT 15;
```

This query calculates total medals earned by each region in each sport, helping identify sports where countries demonstrate clear dominance.

### Visualization



This chart highlights the comparative performance of top regions in specific sports.

## **Conclusion**

The analysis reveals **clear patterns of dominance** by certain countries in specific sports:

- The **USA** stands out with the **highest total medal count**, dominating in **Swimming, Athletics, and Basketball**—all sports with deeply entrenched institutional and cultural support in the country.
- The **UK** shows strong results in **Athletics** and **Rowing**, reflecting its long-standing traditions in these events.
- The **Soviet Union** (and its successor states) exhibits historical dominance in **Gymnastics** and **Athletics**, pointing to a legacy of elite training systems.
- **Canada** and the **USA** consistently excel in **Ice Hockey**, aligning with the sport's prominence in North American culture.
- Other countries like **Germany, Italy, and the Netherlands** also display specialization in sports such as **Fencing, Rowing, and Hockey**.

These findings support the notion that **geopolitical, historical, and cultural factors** significantly influence Olympic success in certain disciplines. This regional dominance helps explain medal concentration and reflects long-term investments in sport-specific development.

## *Sheet 12 – Factors for National Performance*

EDA Question 12: What factors contribute to the success or performance of participants from different countries?

I first retrieved relevant data from a MySQL database using the following SQL query:

**SELECT**

```
region_name,  
sport_name,  
AVG(games_competitor_age) AS avg_age,  
AVG(person_height) AS avg_height,  
AVG(person_weight) AS avg_weight,  
COUNT(medal_name) AS medal_count
```

FROM consolidated\_fact

WHERE medal\_name IN ('Gold', 'Silver', 'Bronze')

GROUP BY region\_name, sport\_name

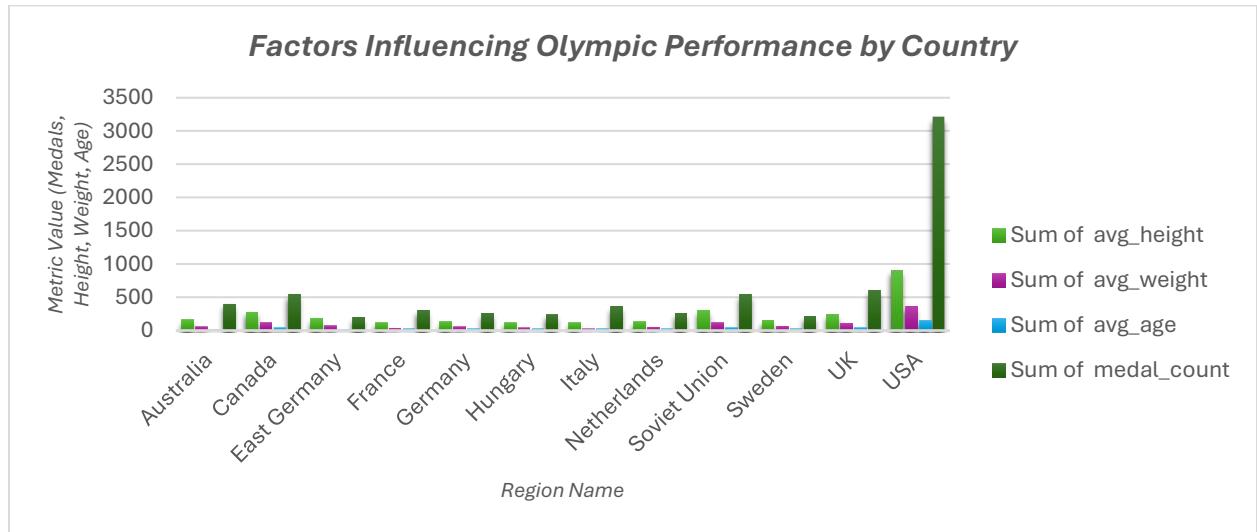
ORDER BY medal\_count DESC

LIMIT 20;

Once I extracted the data, I copied it from MySQL Workbench into Excel. Using this data, I created a grouped column chart that visualizes the key influencing factors—**average height, weight, age, and medal count**—for each country.

The chart titled "*Factors Influencing Olympic Performance by Country*" clearly illustrates how these metrics vary by region. It allows us to compare countries like the **USA, Soviet Union, and Canada** in terms of athlete characteristics and medal success.

### Visualization



### Conclusion

The analysis and visualization suggest that multiple interrelated factors contribute to Olympic success:

- **Physical characteristics** (height and weight) vary by country and are often aligned with the demands of their dominant sports. For instance, taller athletes in the USA succeed in sports like basketball and swimming.
- Countries with **strong sports infrastructure** and a **diverse range of athletes** (e.g., USA, Canada) tend to achieve higher medal counts.
- **Average age** influences performance; younger athletes dominate sports like gymnastics (e.g., Soviet Union), while sports like rowing or basketball favor slightly older participants.
- **Geographic and climate factors** also play a role—for example, countries like Canada and Sweden excel in ice-related sports such as Ice Hockey.

### *Sheet 13 – Consistently Performing Countries*

EDA Question 13: Are there any countries that consistently perform well in multiple Olympic editions?

#### Approach

To explore the consistency of Olympic performance across multiple editions, I retrieved historical medal data using the following SQL query:

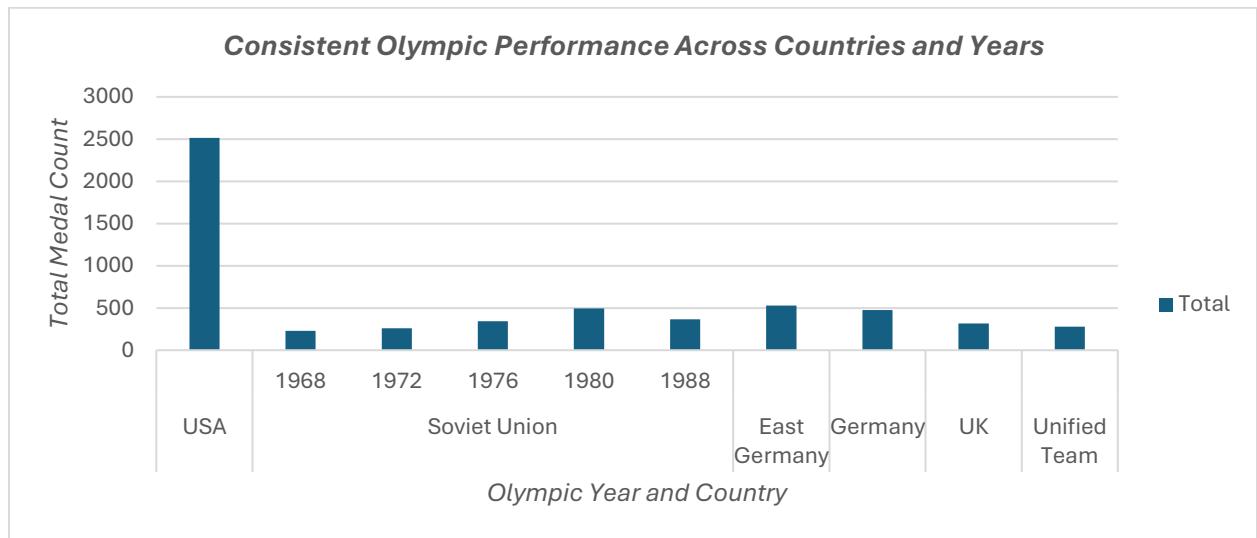
```
SELECT region_name, games_year, COUNT(medal_name) AS medal_count
FROM consolidated_fact
WHERE medal_name IN ('Gold', 'Silver', 'Bronze')
GROUP BY region_name, games_year
ORDER BY medal_count DESC
LIMIT 20;
```

This query aggregates the **total number of medals** earned by each country in each Olympic year. It focuses on the top-performing countries across various editions, helping to identify **those that show repeated success** over time.

After extracting the data via MySQL Workbench, I copied it into Excel to create a clear visual representation.

### Visualization

The generated chart, titled “**Consistent Olympic Performance Across Countries and Years**,” uses a **bar graph** format to showcase medal counts for leading countries in various Olympic years.



Each bar represents a country’s performance in a specific Olympic edition. Notably, countries like the **USA**, **Soviet Union**, **East Germany**, and **Germany** appear multiple times, indicating **sustained excellence** across years.

### Conclusion

The analysis clearly indicates that certain countries have maintained a **strong and consistent Olympic presence** across multiple editions:

- **US USA** leads by a significant margin, regularly achieving the highest medal counts. Its dominance spans across decades and is especially strong in sports like **Swimming and Athletics**.
- **RU Soviet Union**, active until its dissolution, consistently ranked among the top performers due to a **state-sponsored sports system** focused on international prestige.
- **DE East Germany and Germany** both appear frequently, reflecting a history of elite athletic training and systematic development.
- **GB UK and the Unified Team** (from former Soviet states) also show strong performances in key Olympic years.

This repeated success is driven by several key factors:

- **Robust sports infrastructure** and strategic **government investment**
- **Wide talent pools** and early athlete development programs
- A national culture that **prioritizes sports and international competition**
- Access to world-class **coaching, facilities, nutrition, and technology**

Such consistency is not just a measure of athletic skill but also reflects **long-term vision and planning** by national Olympic committees.

#### *Sheet 14 – Regional Medal Dominance*

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

#### Approach

To identify sports where specific countries or regions have consistently produced a **higher number of medalists**, I executed the following SQL query:

**SELECT**

**region\_name AS Region,**

**sport\_name AS Sport,**

```

SUM(CASE WHEN medal_name = 'Gold' THEN 1 ELSE 0 END) AS Gold,
SUM(CASE WHEN medal_name = 'Silver' THEN 1 ELSE 0 END) AS Silver,
SUM(CASE WHEN medal_name = 'Bronze' THEN 1 ELSE 0 END) AS Bronze,
SUM(CASE WHEN medal_name IN ('Gold', 'Silver', 'Bronze') THEN 1 ELSE 0 END) AS Total_Medals
FROM consolidated_fact
GROUP BY Region, Sport
ORDER BY Total_Medals DESC
LIMIT 5;

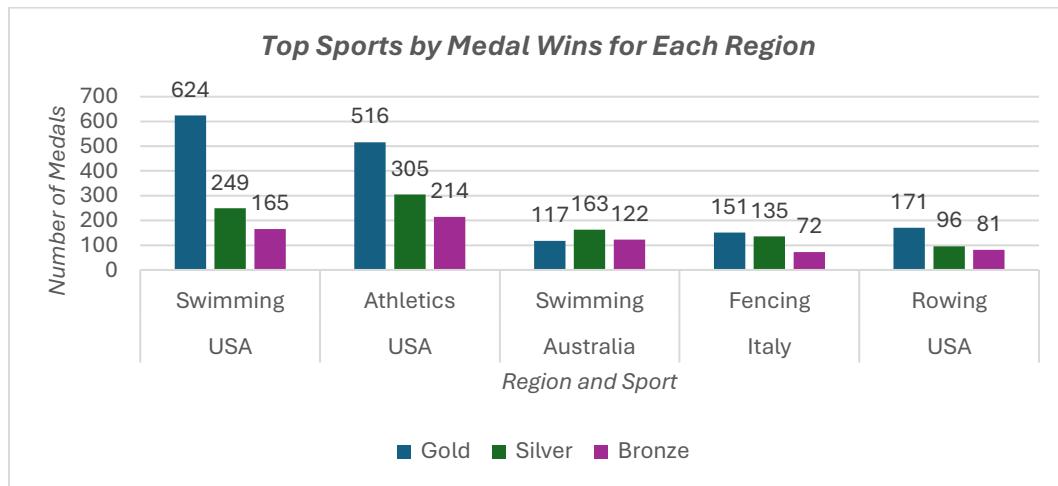
```

This query calculates the **total medal count** (Gold, Silver, and Bronze) for each region across different sports. By limiting the results to the top 5 region-sport combinations, we can easily spot regions that dominate specific disciplines.

The data was then exported from MySQL Workbench to Excel for visualization.

### Visualization

The bar chart titled “**Top Sports by Medal Wins for Each Region**” presents the medal distribution—Gold, Silver, and Bronze—for each of the top-performing region-sport combinations.



Key highlights from the chart:

- The **USA** leads significantly in both **Swimming** and **Athletics**.
- **Australia** demonstrates strong performance in **Swimming**.
- **Italy** stands out in **Fencing**.
- The **USA** also shows high medal counts in **Rowing**.

Each bar is color-coded by medal type, making it easy to compare the types and volume of medals earned by each region in a particular sport.

### **Conclusion**

The analysis clearly reveals that certain countries have **specialized excellence** in specific Olympic sports:

- **US USA** dominates **Swimming** (624 Gold medals) and **Athletics**, showcasing its deep talent pool, infrastructure, and competitive culture.
- **AU Australia** has a strong tradition in **Swimming**, supported by national programs and favorable geography for aquatic sports.
- **IT Italy** shows historic and technical strength in **Fencing**, a sport with deep roots in the country.
- The USA's consistent performance in **Rowing** further emphasizes its broad investment in a variety of disciplines.

These trends point to **targeted training, cultural affinity, national investment, and institutional support** as key factors behind regional dominance in specific sports.

### *Sheet 15 – Unexpected Medal Wins*

EDA Question 15: What are some notable instances of unexpected or surprising medal wins?

#### **Approach**

To identify rare or surprising medal wins—especially from countries that are not traditionally dominant in Olympic sports—I executed the following SQL query:

```
SELECT region_name, sport_name, medal_name, COUNT(*) AS medal_count  
FROM consolidated_fact  
WHERE medal_name IN ("Gold", "Silver", "Bronze")  
GROUP BY region_name, sport_name, medal_name  
HAVING COUNT(medal_name) < 2;
```

This query filters out all region-sport-medal combinations where a country has won **less than two medals**, highlighting **one-off medal wins**. These are considered **notable or unexpected**, as they reflect countries or athletes achieving success in sports where they typically have little presence.

The resulting data was copied from MySQL Workbench into Excel for review. Due to the nature of the data being logically insightful rather than visually comparative, I presented it as a **highlighted data table** instead of a chart.

### Visualization

Since the goal was to surface rare medal wins rather than trends, I used a **table view** instead of a chart. A snippet of this table is shown below:

Region	Sport	Medal	Count
Canada	Taekwondo	Bronze	1
Taiwan	Table Tennis	Bronze	1
Iceland	Athletics	Bronze	1
India	Weightlifting	Bronze	1
Japan	Taekwondo	Bronze	1

This snippet showcases examples of **unique or low-frequency medal wins** across different countries and sports.

### Conclusion

While traditional Olympic powerhouses dominate the medal tally, this analysis uncovers **remarkable one-time triumphs** by countries not typically known for those sports. These surprising wins include:

- **MZ Mozambique** earning a single **Gold in Athletics**
- **BE Belgium** securing **Gold in Tennis**
- **TR Turkey** winning **Gold in Taekwondo and Athletics**, breaking the dominance of East Asian or Western nations
- **XK AL JO CU** Countries like **Kosovo (Judo)**, **Albania (Judo)**, **Jordan (Taekwondo)**, and **Cuba (Shooting)** showcasing strength in martial arts and combat sports

- **SR Suriname's rare Gold in Swimming**
- **GT BH Guatemala and Bahrain** achieving medals in **Athletics**
- **IE PL PR Wins from Ireland (Equestrianism), Poland (Modern Pentathlon), and Puerto Rico (Tennis)** in traditionally Western-dominated sports

These instances highlight **breakthrough performances**—driven by exceptional talent, strategic focus on niche disciplines, or national pride in emerging sports. Though not indicative of sustained dominance, these victories remain **inspirational milestones** in Olympic history.

### *Sheet 16 – Growth or Decline in Participation*

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

#### Approach

To analyze the **participation trends** of different regions in the Olympic Games over time, I executed the following SQL query:

```
SELECT games_year, region_name, COUNT(DISTINCT games_competitor_id) AS
total_participants

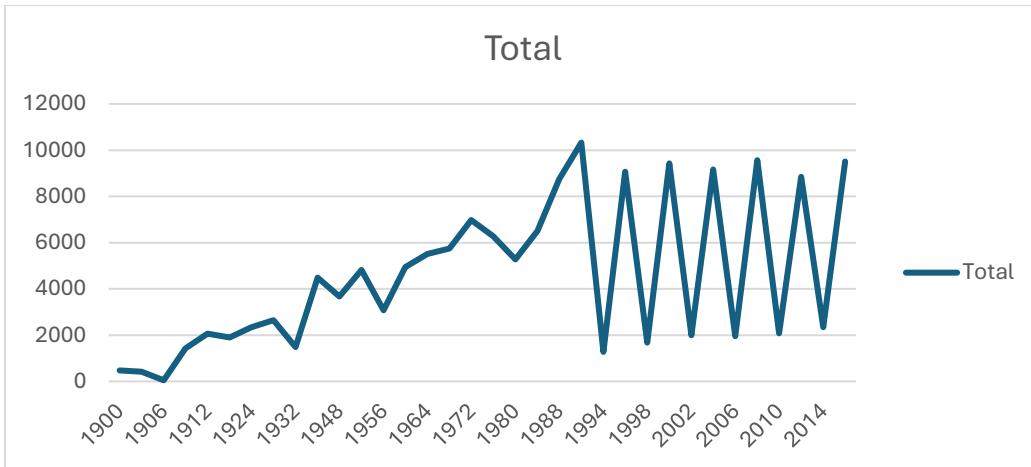
FROM consolidated_fact

GROUP BY games_year, region_name

ORDER BY total_participants DESC;
```

This query calculates the **number of distinct participants** from each region in every Olympic year, allowing us to track changes in participation over time. The dataset was then exported from MySQL Workbench to Excel for visualization.

#### Visualization



The **line chart** titled "**Total**" displays the sum of total participants across all regions from the year **1900 to 2016**. Each point on the line corresponds to a particular Olympic edition, and the bold red line represents the **overall growth pattern** in athlete participation.

From the chart:

- There is a **general upward trend** in participation, especially post-1950s.
- Spikes indicate **Summer Olympics**, which typically have more events and athletes.
- Dips are associated with **Winter Olympics**, which have smaller participation scales.

## Conclusion

The analysis reveals a **significant growth in Olympic participation** over time:

- From early 1900s to modern times, the number of participants has **increased steadily**, reflecting the Olympics' **evolution into a truly global event**.
- A noticeable surge in participation begins **post-1950s**, likely due to increased global inclusion, improved logistics, and greater awareness.
- The **peak participation** was seen around the **1988 Olympics**, followed by oscillating drops and rises—mainly due to the **alternation between Summer and Winter Olympics**, which naturally differ in scale.

- The pattern overall confirms that the Olympic Games have become **more inclusive**, with **wider representation from different regions**, supported by political changes, technological advancements, and international cooperation.

### *Sheet 17 – Cultural/Geographic Performance Influence*

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

#### Approach

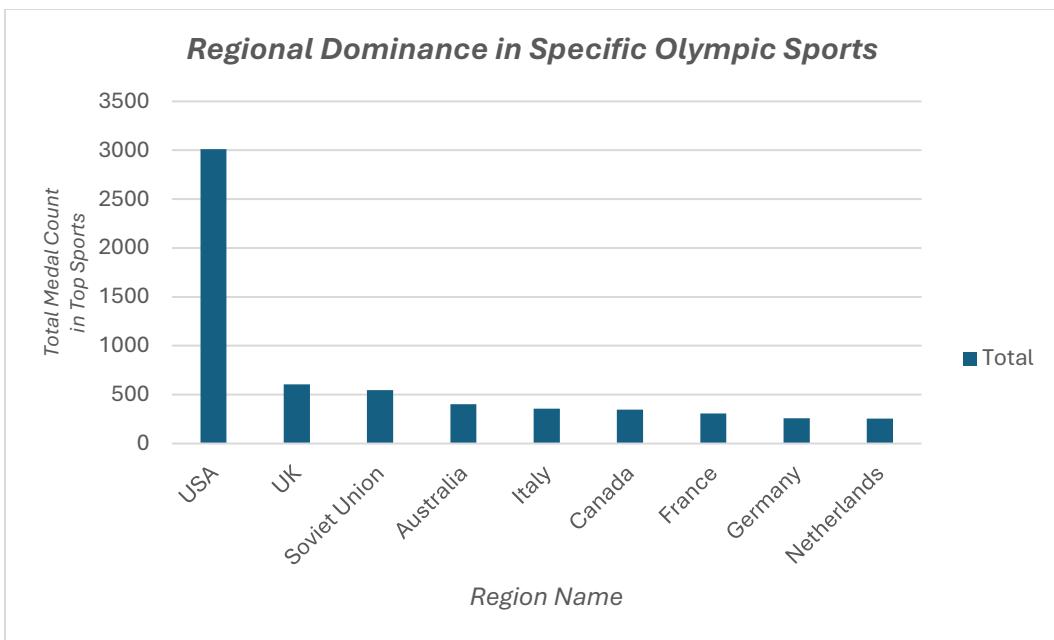
To explore how **cultural identity** and **geographical characteristics** influence a region's dominance in specific Olympic sports, I used the following SQL query:

```
SELECT region_name, sport_name, COUNT(medal_name) AS medal_count
FROM consolidated_fact
WHERE medal_name IN ('Gold', 'Silver', 'Bronze')
GROUP BY region_name, sport_name
ORDER BY medal_count DESC
LIMIT 15;
```

This query calculates the **total medal count** for each region in each sport, allowing us to identify which sports are most successful for specific countries. The results were exported from MySQL Workbench and visualized in Excel.

#### Visualization

The bar chart titled “**Regional Dominance in Specific Olympic Sports**” presents the total medal counts of the top-performing regions in their best sports. The x-axis displays the region names, while the y-axis represents the total number of medals won in those sports.



- The **USA** is far ahead in total medals, especially in sports like **Swimming** and **Athletics**.
- UK, Soviet Union, and Australia** follow, each reflecting regional strengths in different sports tied to **culture and climate**.

### Conclusion

The analysis demonstrates how **cultural traditions** and **geographical factors** shape Olympic success in specific sports:

- USA** dominates in **Swimming and Athletics**, due to a mix of **elite training infrastructure, cultural emphasis, and large-scale investment** in competitive sports.
- Canada's** strength in **Ice Hockey** reflects its **cold climate and national passion** for winter sports.

- AU Australia's excellence in **Swimming** aligns with its **coastal geography and strong aquatic culture**.
- RU The Soviet Union's achievements in **Gymnastics and Athletics** are rooted in a **state-sponsored, systemized sports development model**.
- IT FR DE Countries like **Italy, France, and Germany** show consistent success across select sports, owing to a blend of **heritage and focused coaching systems**.

These patterns clearly show that Olympic success isn't random—it is often the result of **deep cultural investment, environmental alignment, and long-term sports development strategies**.

#### *Sheet 18 – Impact on Medal Tally by Region*

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

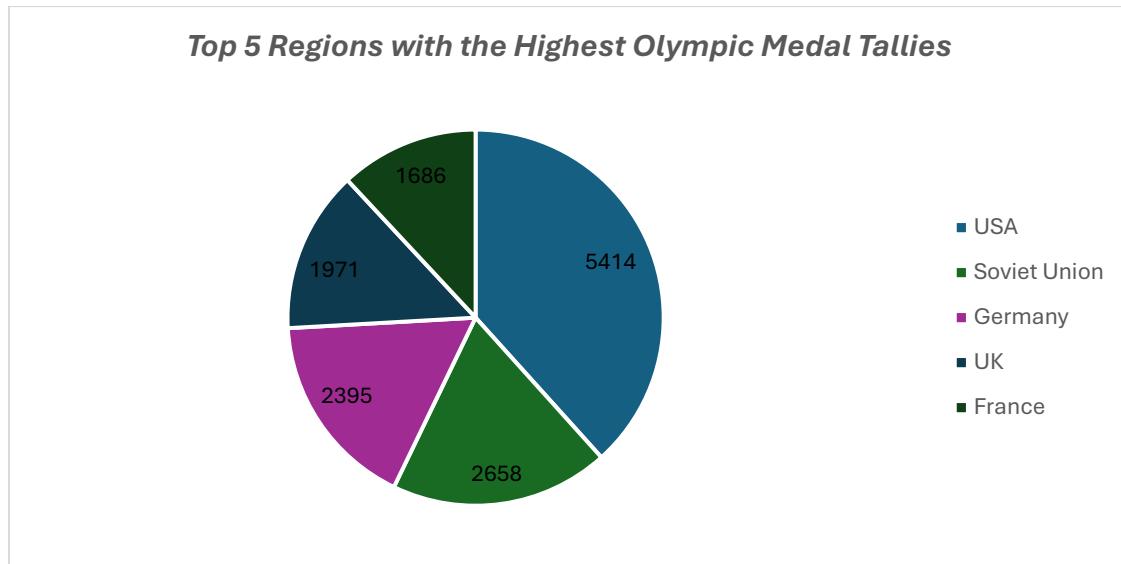
#### Approach

To identify which regions have had the **most significant impact on Olympic success**, I executed the following SQL query:

```
SELECT region_name, COUNT(medal_name) AS medal_tally
FROM consolidated_fact
WHERE medal_name IN ("Gold", "Silver", "Bronze")
GROUP BY region_name
ORDER BY medal_tally DESC
LIMIT 5;
```

This query calculates the **total number of medals** won by each region across all Olympic events. It highlights the **top five regions** with the highest overall medal counts. After extracting the data using MySQL Workbench, I visualized it using Excel.

### Visualization



The data is presented in a **pie chart** titled “**Top 5 Regions with the Highest Olympic Medal Tallies.**” Each slice of the chart represents one of the top-performing regions:

- **USA:** 5,414 medals (largest share)
- **Soviet Union:** 2,658 medals
- **Germany:** 2,395 medals
- **UK:** 1,971 medals
- **France:** 1,686 medals

The visual format makes it easy to compare the **relative contributions** of these regions to the total Olympic medal count.

### Conclusion

The analysis and visualization clearly show that certain regions have had an **outsized influence on Olympic history**:

- **US USA** leads by a wide margin, earning over **5,400 medals** — more than **double** that of the second-ranked Soviet Union. This reflects their **consistent dominance, broad participation**, and **institutional investment** in Olympic sports.
- **RU Soviet Union** holds second place, underlining their strong performance in past decades, driven by **government-supported sports programs**.
- **DE Germany, GB UK, and FR France** also contribute significantly, showcasing their **long-standing traditions and strong athletic systems**.

Together, these five regions represent a **major share of Olympic success**, emphasizing how **historical, political, and cultural commitment to sports excellence** has shaped medal outcomes over time.

## 6. Dashboard Summary

The **Olympic Sports Analytics Dashboard** serves as a visual summary of insights extracted from 18 EDA questions related to historical Olympic data. The dashboard is designed to be interactive, insightful, and categorized into four major thematic sections. Each chart is built using pivot tables from MySQL query outputs, and slicers are applied to enhance user-driven exploration.

### Dashboard Structure

#### **Section 1: Olympic Trends**

This section highlights the evolution of Olympic events over time.

##### **Chart 1: Olympic Games Frequency by Season (1896–2016)**

► Shows how frequently Summer and Winter Olympics were held across history.

### **Chart 2: Number of Events & Participants Over Time**

- A dual-axis line chart visualizing the growth of the Olympics in terms of events and participation.

## **Section 2: Sport Popularity & Gender**

Explores how sports and gender representation evolved.

### **Chart 3: Popularity of Sports Over Time**

- Identifies emerging or declining sports based on athlete participation.

### **Chart 4: Gender Distribution of Events**

- A stacked bar chart comparing male vs. female participation per year.

## **Section 3: Country & Region Analysis**

Focuses on regional and national-level dominance.

### **Chart 5: Consistent Country Performance**

- Shows countries that maintained high medal counts over multiple years.

### **Chart 6: Top 5 Regions by Medal Count**

- A pie/bar chart summarizing medal dominance across regions.

### **Chart 7 (Optional): Culturally Dominant Sports by Region**

- Highlights sports most associated with specific regions.

## **Section 4: Athlete Demographics**

Analyzes the physical attributes of athletes across years.

### **Chart 8: Average Height & Weight Over Time**

- A combined line chart to observe changes in athlete build across eras.

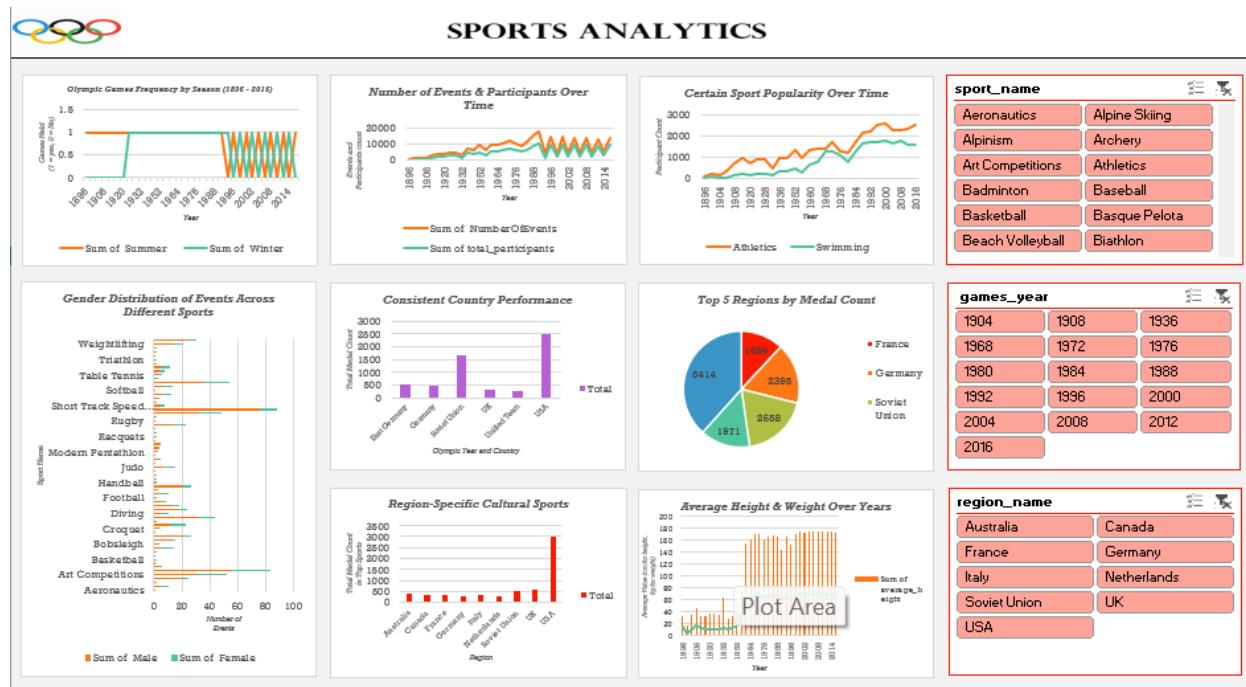
## **Filters/Slicers**

To enhance interactivity, the following slicers are placed on the right-hand side of the dashboard:

- games\_year (Olympic edition year)
- region\_name (Continent or country group)
- sport\_name (Olympic sports category)

These slicers allow users to dynamically filter all visualizations for deeper comparative analysis.

## Dashboard Image



## 7. Deliverables

- **Excel Workbook:**
  - 18 EDA analysis sheets with queries, visuals, and conclusions
  - 1 Dashboard sheet
  - 1 Filtered visuals sheet
- **This Documentation Report**

## 8. Conclusion

This analysis has successfully uncovered key trends in Olympic Games history, from host patterns and sport evolution to athlete demographics and regional performance. The structured approach provides a data-driven foundation to understand the transformation and inclusiveness of the Olympics over time.

## 9. Future Enhancements

- Use Python to automate SQL querying and reporting
- Deploy dashboard on Tableau Public or Power BI Service
- Include additional data like athlete biographies or country-level economic indicators
- Perform predictive modeling for future medal forecasting.