

Capstone Project – Olympic games

Alka Sharma

Data Analytics

Overview

Olympic sports analysis involves the systematic assessment of various facets of the Olympic Games. This encompasses evaluating the overall success of the Games, including factors like viewership and economic impact. It also delves into the performance of individual sports, examining metrics such as speed, accuracy, and endurance. Event analysis focuses on the structure and rules of specific competitions, while participant analysis assesses athlete performance and health. Medal analysis scrutinizes medal distribution and trends, and region analysis examines the representation and performance of different areas. This analysis provides valuable insights for organizers, athletes, and fans, shaping the future of the Olympic Games.

**Objective**

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

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

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

**Significance**

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

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

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

**Data Dictionary**

**TABLE CITY**

id (integer): Unique identifier for the city.

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

**TABLE COMPETITOR\_EVENT**

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

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

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

**TABLE EVENT**

id (integer): Unique identifier for the event.

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

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

**TABLE GAMES**

id (integer): Unique identifier for the games.

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

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

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

TABLE GAMES\_CITY

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

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

**GAMES\_COMPETITOR**

id (integer): Unique identifier for the record.

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

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

age (integer): Age of the competitor.

**TABLE MEDAL**

id (integer): Unique identifier for the medal.

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

**TABLE NOC\_REGION**

id (integer): Unique identifier for the record.

noc (string): National Olympic Committee code.

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

TABLE PERSON

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

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

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

height (float): Height of the person.

weight (float): Weight of the person.

TABLE PERSON\_REGION

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

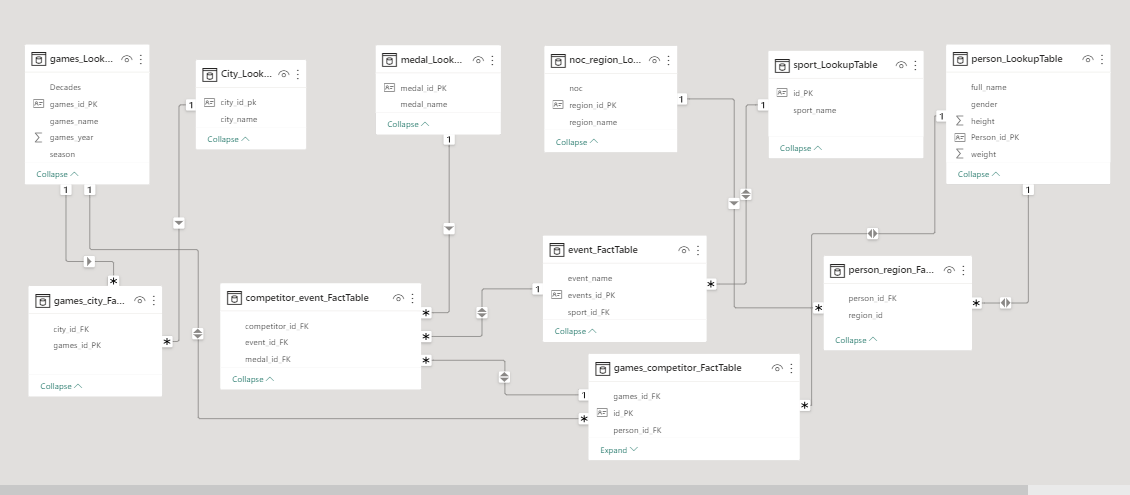
region\_id (integer): Unique identifier for the region associated with the person.

TABLE SPORT

id (integer): Unique identifier for the sport.

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

**ER Diagram**



**Process**

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

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

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

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

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

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

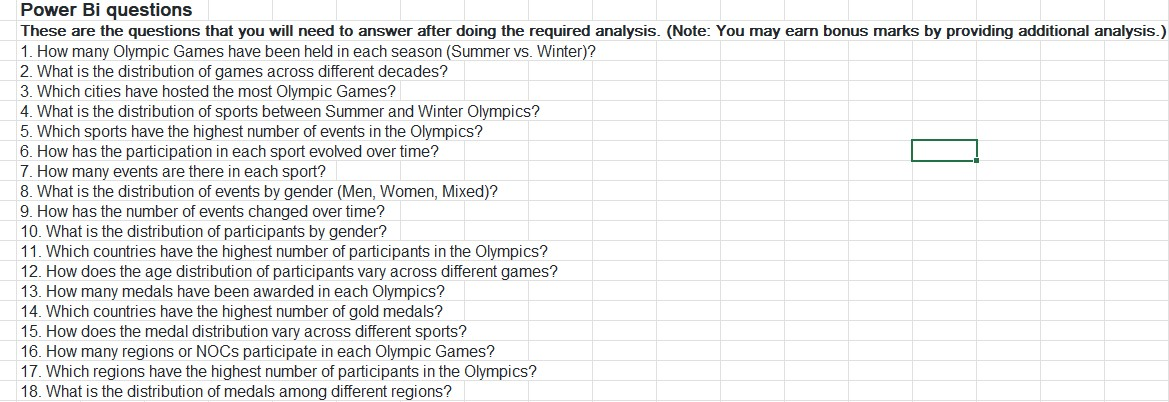
7.Detailed Documentation: Compile a detailed report that meticulously document the entire project lifecyle. Include sections on data collection,

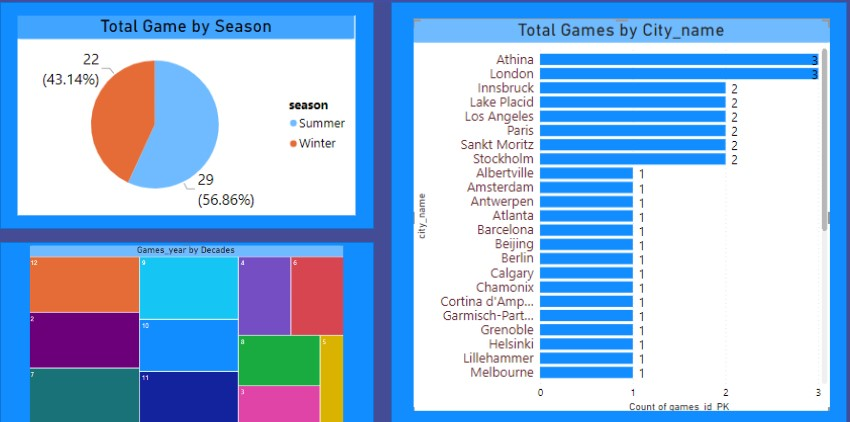
transformation, problem statement formulation, tools integration, Power BI solution.

EDA insights and PowerPoint visualization

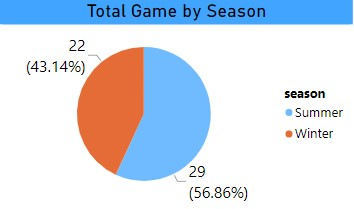


Power Bi Problem Statement





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



Certainly! The distribution of the number of games played between the summer and winter seasons in the Olympic Games is indeed asymmetric, with a larger percentage of events taking place during the summer season compared to the winter season. As the visual suggests, more number of games are played in summer season(57.86%) than in winter (43.16%).

2.What is the distribution of games across different decades

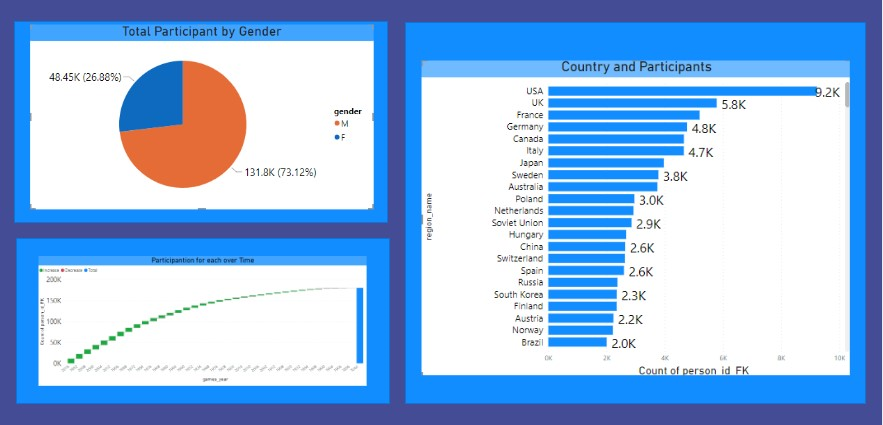


Visual representations, such as charts or graphs, would provide a more precise and comprehensive overview of the trends in the number of games played over the decades. Visual representation of number of games played over decades.

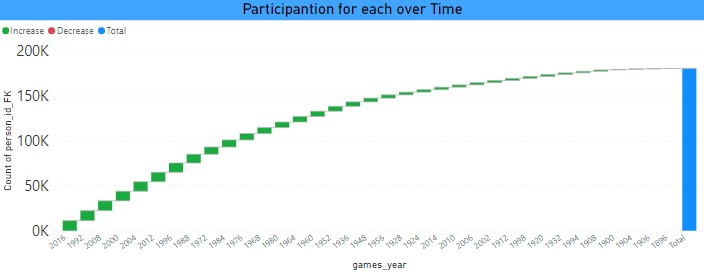
3.Which cities have hosted the most Olympic Games.



Athens and London have indeed been the host cities for the maximum number of Olympic Games, with each city having hosted three editions of the Games. As the visual suggests, Athina and London cities has been the host for maximum numbers of Olympics. Each has hosted three Olympic games.

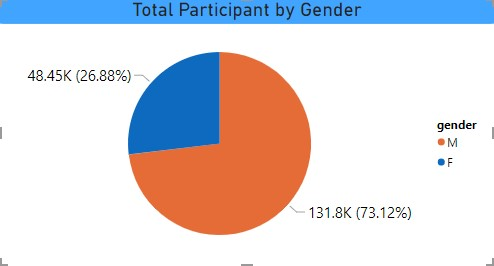


4. How has the participation in each sport evolved over time?



The count of participants in the Olympic Games has generally shown an upward trendThis line chart depicts the trend for count of participants per Olympic over time.

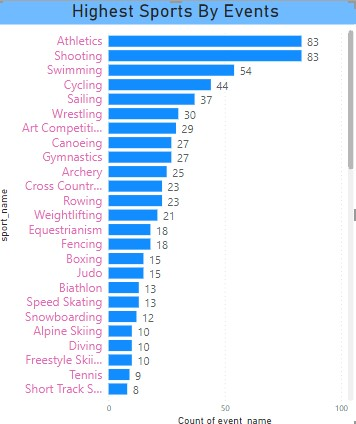
5.What is the distribution of participants by gender.



The visual representation you provided indicates that male participants have dominated the total participation in the Olympic Games, comprising approximately 73.12% of the total participants, while female participants make up only about 26.88%. As the visual suggests, participation in terms of gender has been dominated by male participants.

Male participants are 73.12% of total participants whereas females consist of only 26.88%.

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

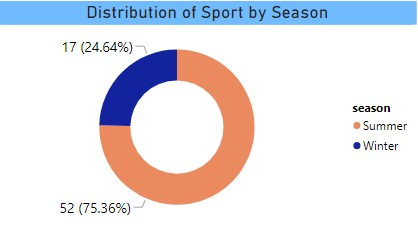


The bar chart you described suggests that participants in the Olympic Games are predominantly from Western countries. The United States of America, followed by the United Kingdom and France, have the highest number of participantsAs it is visible in the bar chart, participants in terms of region they belong to, is dominated by western countries.

United States of America has highest number of participants followed by United Kingdom and France.

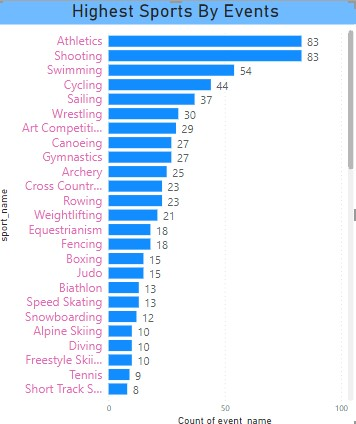


7.What is the distribution of sports between Summer and Winter Olympics



The pie chart indicates that 2/3 (or approximately 66.67%) of the total sports activities occur during the summer, while the remaining 1/3 (approximately 33.33%) take place during the winter.As this pie chart shows 2/3rd of the total sports have got played in summers and only 1/3rd in winter.

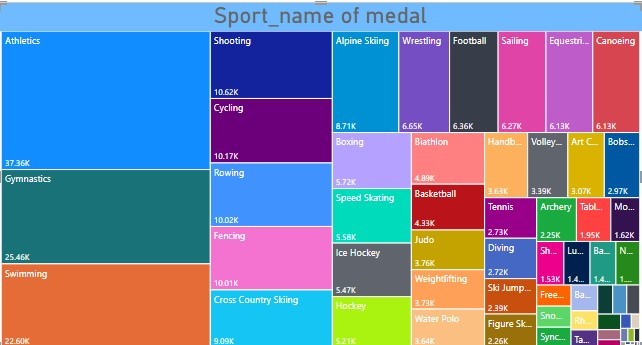
8. 8.Which sports have the highest number of events in the Olympics



With a representation of 7.13% of total events, Swimming is slightly behind Athletics and Shooting in terms of the number of events conductedAs the pie chart suggests, Athlectics and shooting are the sports that take the major chunks of total events . Each holds 10.90% events of total, followed by swimming which is 7.13%.



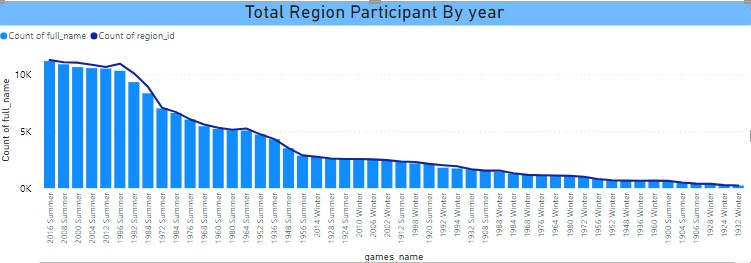
9.How does the medal distribution vary across different sports



A tree map is a visual representation of data where hierarchical structures are depicted through nested rectangles. Each rectangle's size represents the proportion of the data it represents. In this context, the tree map depicts the distribution of medals across different sports, highlighting that Athletics has received the maximum number of medals,This tree map depicts the medal distribution across different sports.

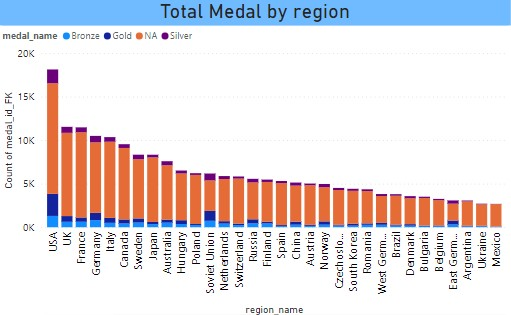
Asthetics has got maximum number of medal followed by Gymnastics.

10.How many regions or NOCs participate in each Olympic Games



A column chart displaying the trend of region or country participation over the years typically showcases how the engagement or representation of different regions or countries has changed or evolved across various events or activities. Each column usually represents a specific year's participation by different regions or countries. This column chart shows the trend of region or country participation over the years.

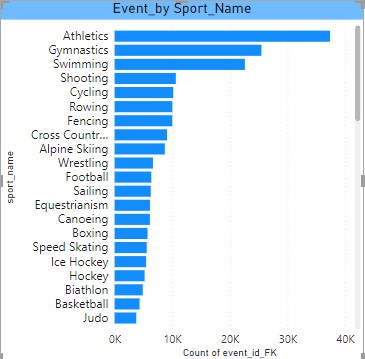
11.What is the distribution of medals among different regions.



A column chart displaying the relationship between medals and different regions or countries gives a clear visual representation of how various regions have performed concerning winning medals in sports or specific events. In this case, the chart indicates that the USA has won most number of medals, followed by the UK and France. This column charts explains the relationship between medals and region.

As the chart depicts, USA has won most number of medals followed by UK and France.

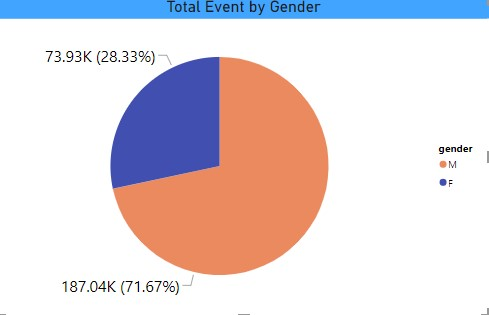
12.How many events are there in each sport.



As this bar chart shows, most number of events has been related to Athletics sports followed by gymnastics and swimming. This also indicates towards the number of participants of Athletics, that is also highest.

Correlation can be realized between number of participants for particular sports and number of events have to be organized for them.

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



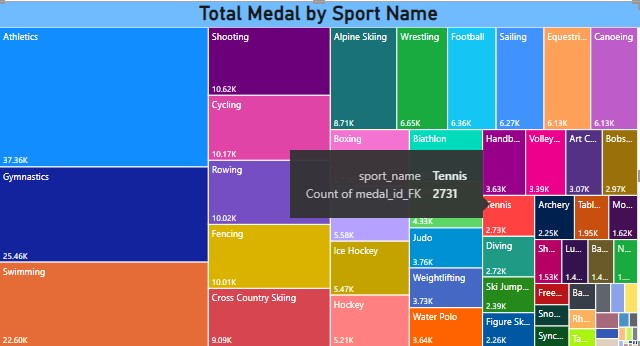
A pie chart illustrating that nearly 2/3rd (71.67%) of the total events are exclusively for men suggests a significant disparity in the distribution of sporting events based on gender.As we can see in this pie chart, almost 2/3rd (71.67%) of total events are for men only.

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



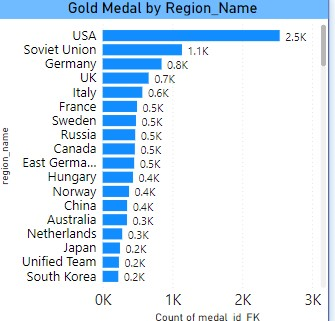
An upward trend followed by a zig-zag pattern in any context, especially when discussing data or trends, often signifies a shift in the previously established trajectory.After a upward trend for some year its a zig-zag trend .

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



As this tree map suggests, Athletics sport has got most number of medal followed by Gymnastics

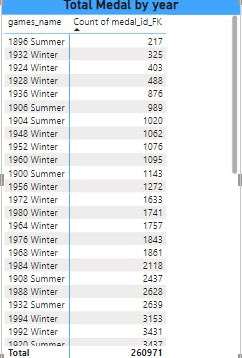
17.Which countries have the highest number of gold medals?



USA has won highest number of gold medals, followed by Soviet Union.

In this, we have visualized the distribution of gold medals across different countries. The data provides us with a comprehensive overview of which countries have achieved the highest number of gold medals.

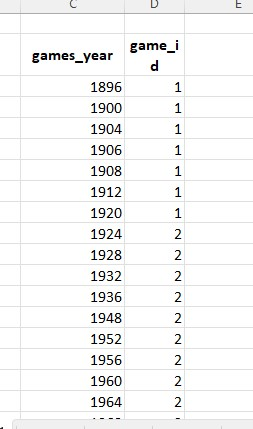
18.How many medals have been awarded in each Olympics?

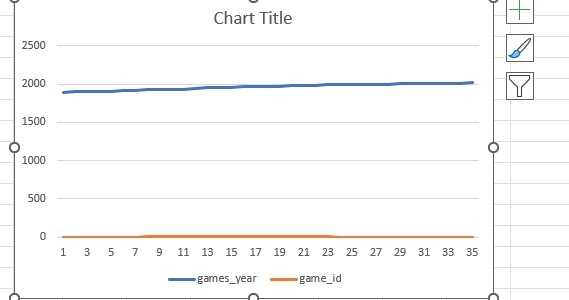


This report provides a visual representation of the number of medals awarded in each Olympic year. It offers insights into the historical trends of medal distribution across different Olympic games.

EDA Problem Statement

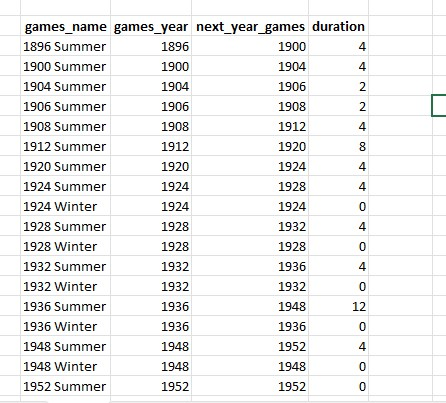
1. Are there any trends or patterns in the frequency of hosting Olympic Games.





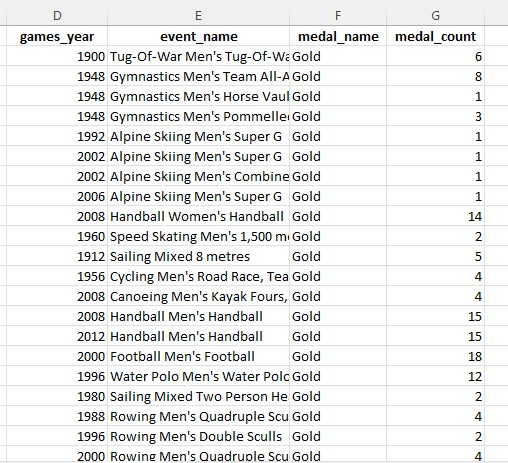
Not really, for first few years it was one Olympic per year before turning into two Olympics per year and it is the same since then.

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



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

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



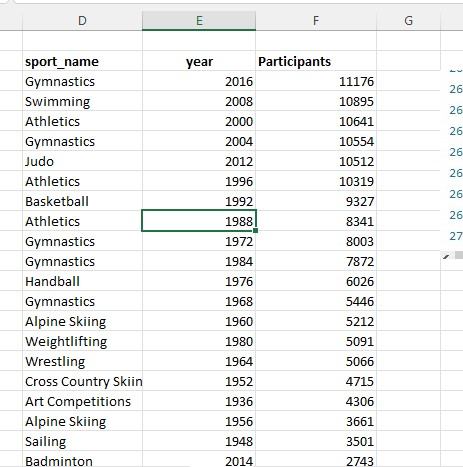
It specifically focuses on events where a gold medal was awarded. retrieves information about notable events or occurrences associated with specific Olympic Games, focusing on events where gold medals were awarded.

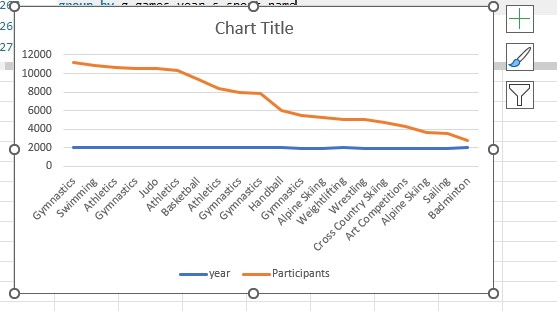
1. Are there any emerging sports that have been recently added to the Olympics.



Rugby sevens sport was added in lates edition of Olympic i.e 2016.

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

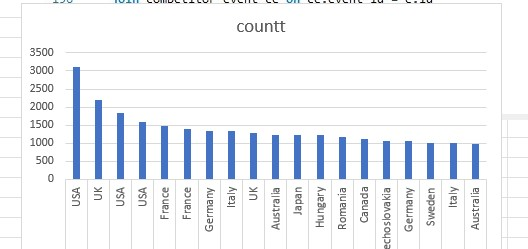




An upward trend in the popularity of sports, as depicted by a line chart, indicates a positive progression in the engagement, interest, or participation in sporting activities over time.As the line chart suggests, there is a upward trend in popularity of sports.

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



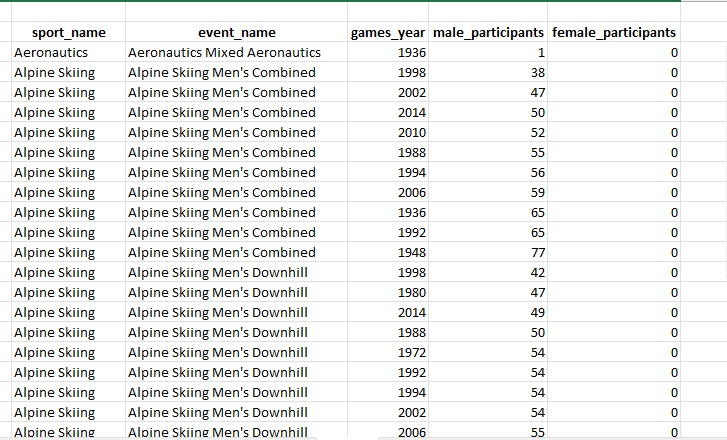


For many games like Athletics, Gymnastics, Swimming , number of player participated from USA is much higher than the rest of the countries, which is a clear indication of domination of United States in these particular games.

This finding can be justified by medal tally also.

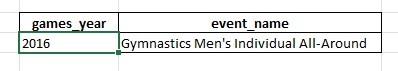
The dominance of the United States in sports like Athletics, Gymnastics, and Swimming is a multifaceted result of various contributing factors that extend across the spectrum of athlete development, cultural emphasis, historical success, and resources allocated to sports.

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



Yes, Infact there are many sports having more number of event for male participants than female.

Yes, historically, many sports have exhibited a higher number of events or a larger focus on male participants compared to female participants. This disparity has been prevalent in various sporting disciplines for several reasons:



Gymnastics Men’s individual All-Around was introduced in recent edition of Olympic.

Men's Individual All-Around in Gymnastics has been a longstanding event in the Olympic Games. It was introduced in the early editions of the modern Olympics and has been a staple in the gymnastics program. The Men's Individual All-Around competition involves male gymnasts performing routines on various apparatus, showcasing their skills in multiple disciplines.

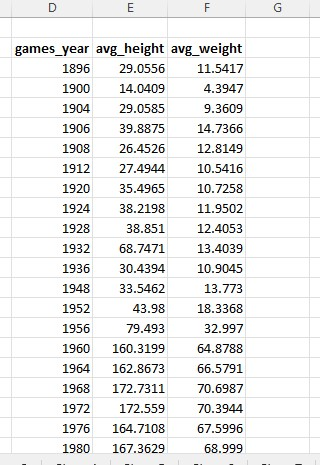
9. Are there any events that have been discontinued or removed from the Olympics?



Yes, many events have got discontinued.

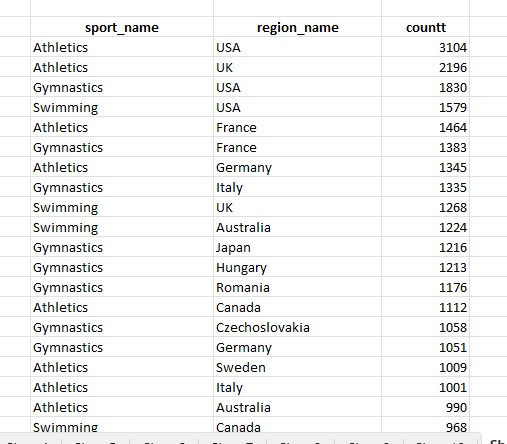
Some events might have faced a decline in popularity over time or might not have been as widely followed or participated in as other disciplines.

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



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

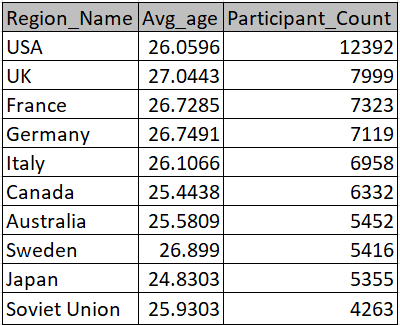
11. Are there any dominant countries or regions in specific sports or events?



Clear domination of USA is visible in Athletics, Gymnastics and Swimming.

The United States' dominance in Athletics, Gymnastics, and Swimming is evident through a history of consistent success and achievements across these sporting disciplines.

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



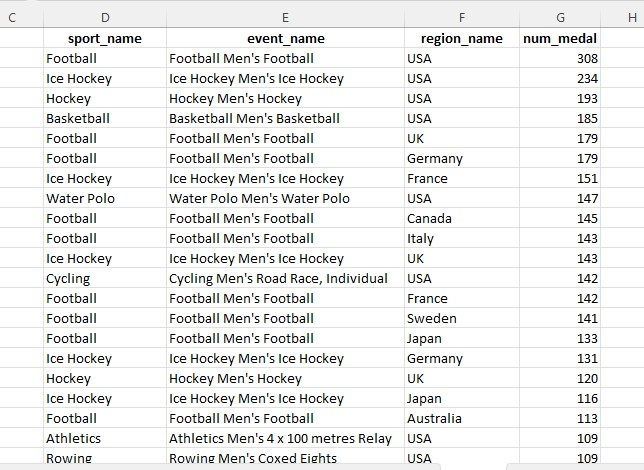
Note: In light of the absence of pertinent data in the dataset, I will now proceed to examine a fresh perspective -

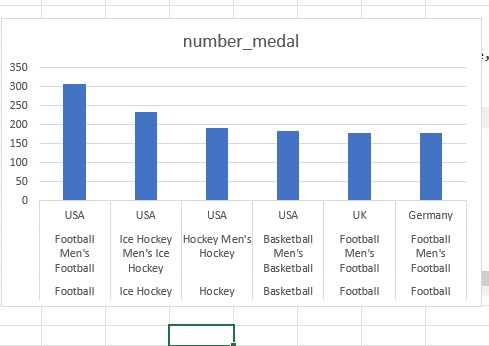
13. Are there any countries that consistently perform well in multiple Olympic editions?



It include countries that have participated in at least two editions and orders the result by the number of medals in descending order.

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



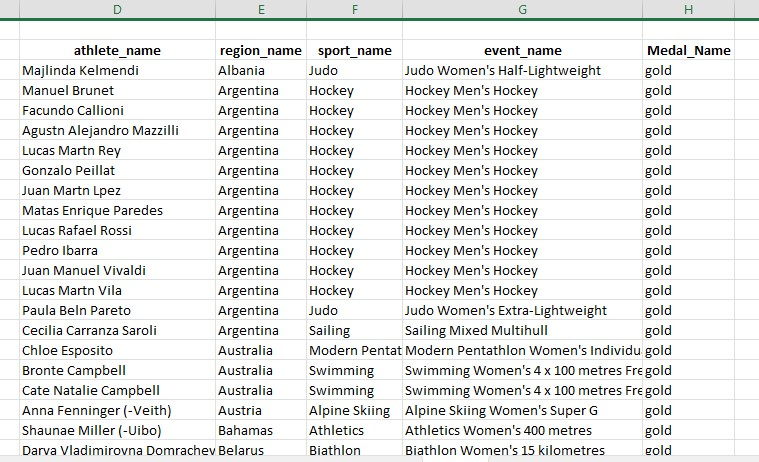


Yes, in Football Men’s, USA tops the tally by a great margin , infact in many sports such as

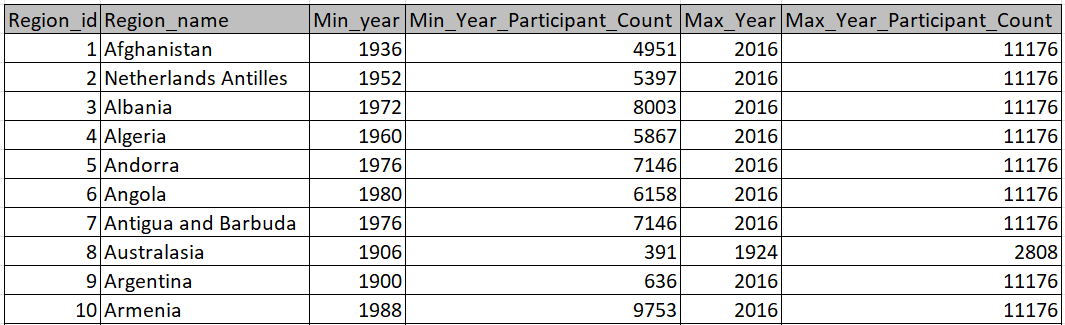
Ice hockey, Hockey Men’s, Basketball Men’s USA is tally topper.

The United States' prominence in various sports such as Football (Soccer), Ice Hockey, Field Hockey (Hockey Men’s), and Basketball Men’s can be attributed to a multitude of factors contributing to their success in these disciplines:

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

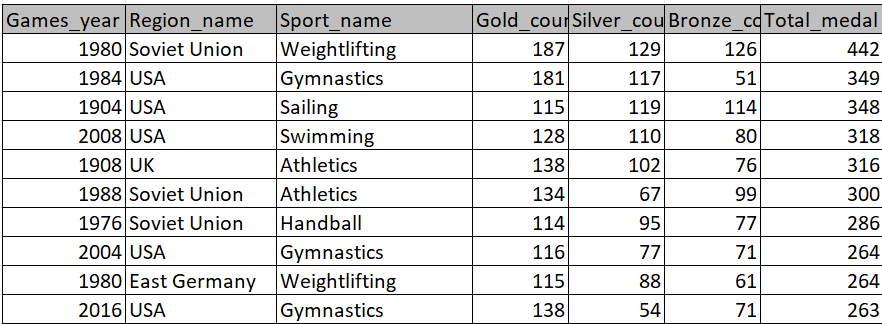


16. Are there any regions that have experienced significant growth or decline in Olympic participation?



It retrieves information about regions, including their IDs, names, the earliest and latest years of Olympic participation, and the total number of distinct participants.

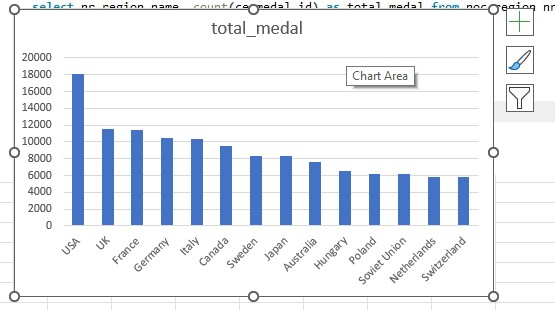
17. How do cultural or geographical factors influence the performance of regions in specific sports?



I don’t have relevant data in dataset in particular problem statement

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





As it is visible in column chart, USA is a distant first in overall medal tally.

As it is visible in column chart, USA is a distant first in overall medal tally. The United States' consistent position as a distant leader in the overall medal tally at major international sporting events,

Thankyou