

**School of Computer Science and Engineering**  
(Computer Science & Engineering)  
Faculty of Engineering & Technology  
Jain Global Campus, Kanakapura Taluk - 562112  
Ramanagara District, Karnataka, India

**2023-2024**  
( IV Semester)

**A Project Report on**  
**“The Olympics: A Data Odyssey”**

**Submitted in partial fulfilment for the award of the degree of**

**BACHELOR OF TECHNOLOGY**  
**IN**  
**COMPUTER SCIENCE AND ENGINEERING**  
**(Artificial Intelligence & Data Engineering)**

**Submitted by**

**Ayushi Tawari, Mutta Datta Sai Vishnu Mohan, Niranjana J**

**22BTRAD008, 22BTRAD026, 22BTRAD027**

**Under the guidance of**

**Mr. Arnab Roy**  
**Project Practice Head and Mentor**  
**Futureense Technologies**

## Department of Computer Science and Engineering

School of Computer Science & Engineering

Faculty of Engineering & Technology

Jain Global Campus, Kanakapura Taluk - 562112

Ramanagara District, Karnataka, India

### CERTIFICATE

This is to certify that the project work titled “**The Olympics: A Data Odyssey - Exploring Performance and Trends**” is carried out by **Ayushi Tawari (22BTRAD008), Mutta Datta Sai Vishnu Mohan (22BTRAD026), Niranjana J (22BTRAD027)**, a bonafide student(s) of Bachelor of Technology at the School of Engineering & Technology, Faculty of Engineering & Technology, JAIN (Deemed-to-be University), Bangalore in partial fulfillment for the award of degree in Bachelor of Technology in Computer Science and Engineering, during the year **2023-2024**.

**Mr. Arnab Roy**

Project Practice Head  
and Mentor  
Futureense Technologies

Date: 10-04-2024

**Dr. Aditya Pai H ,**

Program Head,  
Computer Science and  
Engineering,  
School of Computer Science &  
Engineering(AI & DE)  
Faculty of Engineering &  
Technology  
JAIN (Deemed to-be University)

Date: 10-04-2024

**Dr. Geetha G**

Director,  
School of Computer Science  
& Engineering  
Faculty of Engineering &  
Technology  
JAIN (Deemed to-be  
University)

Date: 10-04-2024

Name of the Examiner

Signature of Examiner

## DECLARATION

We, **Ayushi Tawari (22BTRAD008)**, **Mutta Datta Sai Vishnu Mohan (22BTRAD026)** and **Niranjana J (22BTRAD027)**, students of CSE (AI & DE) IV semester B.Tech in **Computer Science and Engineering –Artificial Intelligence and Data Engineering**, at School of Engineering & Technology, Faculty of Engineering & Technology, **JAIN (Deemed to- be University)**, hereby declare that the internship work titled “**The Olympics: A Data Odyssey**” has been carried out by us and submitted in partial fulfilment for the award of degree in **Bachelor of Technology in Computer Science and Engineering -Artificial Intelligence and Data Engineering** during the academic year **2023-2024**. Further, the matter presented in the work has not been submitted previously by anybody for the award of any degree or any diploma to any other University, to the best of our knowledge and faith.

Name: Ayushi Tawari  
USN : 22BTRAD008

Signature

Name: Mutta Datta Sai Vishnu Mohan  
USN :22BTRAD026

Signature

Name: Niranjana J  
USN : 22BTRAD027

Signature

Place : Bangalore  
Date : 10-04-2024

## ACKNOWLEDGEMENT

*It is a great pleasure for us to acknowledge the assistance and support of a large number of individuals who have been responsible for the successful completion of this project work.*

*First, I take this opportunity to express my sincere gratitude to Faculty of Engineering & Technology, JAIN (Deemed to-be University) for providing us with a great opportunity to pursue our Bachelor Degree in this institution.*

*We are deeply thankful to several individuals whose invaluable contributions have made this project a reality. We wish to extend our heartfelt gratitude to **Dr. Chandraj Roy Chand, Chancellor**, for his tireless commitment to fostering excellence in teaching and research at Jain (Deemed-to-be-University). We are also profoundly grateful to the honorable **Vice Chancellor, Dr. Raj Singh, and Dr. Dinesh Nilkant, Pro Vice Chancellor**, for their unwavering support. Furthermore, we would like to express my sincere thanks to **Dr. Jitendra Kumar Mishra, Registrar**, whose guidance has imparted invaluable qualities and skills that will serve us well in our future endeavors.*

*We extend our sincere gratitude to **Dr. Hariprasad S A, Director** of the Faculty of Engineering & Technology, and **Dr. Geetha G, Director** of the School of Computer Science & Engineering within the Faculty of Engineering & Technology, for their constant encouragement and expert advice. Additionally, we would like to express my appreciation to **Dr. Krishnan Batri, Deputy Director (Course and Delivery)**, and **Dr. V. Vivek, Deputy Director (Students & Industry Relations)**, for their invaluable contributions and support throughout this project.*

*It is a matter of immense pleasure to express our sincere thanks to **Dr. Aditya Pai H, (Program Head), Computer Science and Engineering**, School of Computer Science & Engineering Faculty of Engineering & Technology for providing right academic guidance that made our task possible.*

*I would like to thank our guide **Mr. Arnab Roy**, Project Practice Head and Mentor at Futureense Technologies for sparing his valuable time to extend help in every step of our work, which paved the way for smooth progress and fruitful culmination of the project.*

*We would like to thank our Project Coordinator **Mr. Akash Das**, and all the staff members of Computer Science and Engineering for their support.*

*We are also grateful to our family and friends who provided us with every requirement throughout the course.*

*We would like to thank one and all who directly or indirectly helped us in completing the work successfully.*

*Signature of Students*

## ABSTRACT

The Olympics, a grand stage for the world's top athletes, offers a wealth of data waiting to be explored. In this analysis, we delve into a comprehensive dataset of Olympic athletes, uncovering intriguing patterns and trends across various dimensions. First, we examine gender disparities in Olympic participation and performance. Our findings reveal the evolution of gender representation and the sports where gender gaps are most pronounced. Next, we analyse the distribution of medals, identifying the most successful countries, athletes, and sports. This analysis offers insights into the historical dominance of certain nations and the potential factors contributing to their success. We then explored the relationship between age and athletic performance, revealing the age ranges where athletes tend to excel in their respective sports. This investigation highlights the delicate balance between experience and physical prowess in Olympic competition. Physique analysis follows, shedding light on the relationship between athlete's height, weight, and performance. By comparing these attributes across sports, we identify trends that may influence athletic success and inform recruitment strategies. Country analysis offers a geographical perspective, examining the distribution of Olympic talent and medals across the globe. This examination reveals regional trends and disparities, as well as the potential impact of socio-economic factors on athletic performance. We also investigated seasonal trends in Olympic performance, comparing summer and winter games. This analysis uncovers patterns in athletic dominance, sport popularity, and the evolving nature of the Olympic Games. In conclusion, this data odyssey through Olympic history provides valuable insights into the complex factors influencing athletic success. By examining trends in gender, medals, age, physique, country, and season, we contribute to a deeper understanding of the Olympic Games and the remarkable athletes who compete in them.

## TABLE OF CONTENTS

<b>Sl.</b>	<b>Title</b>	<b>Page no.</b>
<b>1.</b>	<b>Introduction</b>	<b>10</b>
1.1	Background and motivation	10
1.2	Overall objectives	11
1.3	Delimitation of research	11
1.4	Benefits of research	11
<b>2</b>	<b>Implementation</b>	<b>13</b>
2.1	Details of dataset	13
2.2	Data Preprocessing	14
<b>3</b>	<b>Insights and trends</b>	<b>15</b>
3.1	Country-wise analysis	15
3.2	Physique analysis	20
3.3	Participation analysis	27
3.4	Gender analysis	30
3.5	Seasonal analysis	31
3.6	India's sporting Tapestry	36
<b>4</b>	<b>Inferences</b>	<b>39</b>
<b>5</b>	<b>Conclusion</b>	<b>41</b>

## LIST OF FIGURES

Fig. No.	Description of the figure	Page No.
1.1	Olympics logo	10
3.1.1(a)	Dominant Country Each Year	15
3.1.2(a)	Dominant Country Each Year- Summer Olympics	16
3.1.2(b)	World map- Summer Olympics	16
3.1.2(c)	Participation by country and gender- Summer Olympics	17
3.1.2(d)	Medals by country- Summer Olympics	17
3.1.3(a)	Dominant Country Each Year- Winter Olympics	18
3.1.3(b)	World map- Winter Olympics	18
3.1.3(c)	Participation by country and gender- Winter Olympics	19
3.1.3(d)	Medals by country and gender- Winter Olympics	19
3.2.1(a)	Age of male Vs. female athletes	20
3.2.1(b)	Age Of Participants By Year	20
3.2.2(a)	Variation of Weight for Men Athletes over time	22
3.2.2(b)	Variation of Weight for Female Athletes over time	22
3.2.3(a)	Variation of Height for Men Athletes over time	23
3.2.3(b)	Height distribution - Men	23
3.2.3(c)	Variation of Height for Female Athletes over time	24
3.2.3(d)	Height distribution - Women	24
3.2.4(a)	Mean of physical features- top 5 sports	25
3.2.4(b)	Mean of physical features- men vs women	26
3.2.4(c)	Correlation matrix	27
3.2.4(d)	Mean of physical features- summer vs winter	28
3.3(a)	Athletes participation	28
3.3(b)	Participation Count In Every Sport Over The Years	29
3.3(c)	Top 10 athletes	30
3.4(a)	Gender analysis	31
3.4(b)	Variation in number of male athletes over time	31
3.4(c)	Variation in number of female athletes over time	32



3.5(a)	Athletes Participation In Olympics- Summer vs Winter	32
3.5(b)	Evolution of number of sports	33
3.5(c)	Popular Sports In Olympics – Summer	34
3.5(d)	Popular Sports In Olympics - Winter	34
3.5(e)	Top countries- Summer	35
3.5(f)	Top countries- Winter	35
3.5(g)	City – Summer Olympics	36
3.5(h)	City – Winter Olympics	36
3.6(a)	India's participation	37
3.6(b)	Sports With Max Popularity In India	37
3.6(c)	Top 5 Indian Female Sports Players	38
3.6(d)	Top 20 Indian Sports Players	38
3.6(e)	Distribution Of Medals Won By India Across Sports	38

# Chapter 1. Introduction

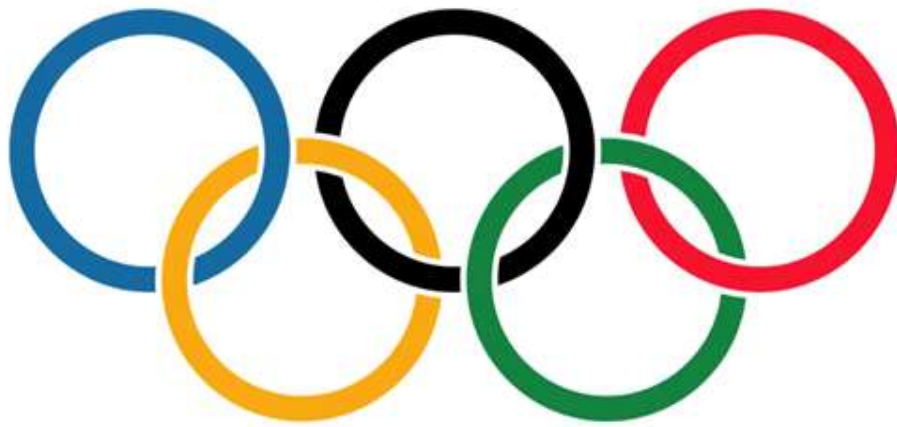


Fig1.1- Olympics logo

## 1.1 Background and Motivation

Analyzing the Olympics dataset can offer valuable insights into various aspects of the Olympic Games, spanning from historical trends and performances to socio-cultural dynamics and global participation.

1. **Historical Trends:** The Olympics dataset contains records spanning several decades, offering a rich historical perspective on the evolution of the Games. Analyzing this dataset allows us to identify trends in participant demographics, medal distributions, and the emergence of new sports over time.
2. **Socio-Cultural Dynamics:** The Olympics serve as a platform for global unity and cultural exchange. Analyzing participation trends, including the representation of different countries, genders, and age groups, can shed light on socio-cultural dynamics and inequalities within the world of sports.
3. **National Identity and Pride:** The Olympics are often a source of national pride, with countries vying for medals and recognition on the global stage. Analyzing the performance of athletes from different countries can provide insights into national sporting cultures and the significance of sports in shaping national identity.

4. Impact of Policy and Investment: Examination of medal counts and performance trends can highlight the impact of governmental policies, investment in sports infrastructure, and athlete development programs on a country's success in the Olympics. It can also reveal disparities in resources and support across nations.

## **1.2 Overall Objectives:**

- Athlete Performance: Analyze factors influencing medal acquisition (e.g., age, height, weight, specific sports/events).
- Trends over Time: Identify trends in participation, medal distribution, and dominance in different sports across various Olympic Games.
- Demographic Analysis: Understand the demographics of Olympic athletes (age, sex, height, weight) across different sports and countries.)
- National Performance: Compare performance of different countries based on medal count, participation rates, and athlete demographics.

By analyzing these different aspects, the dataset can reveal valuable information about the Olympics and the athletes who compete.

## **1.3 Delimitation of research**

Delimitations in research define the boundaries and limitations within which the study will be conducted. Here are some delimitations specific to analyzing the Olympics dataset:

- Temporal Scope: The analysis may focus on a specific range of years or Olympics editions due to data availability or relevance. The dataset contains information about the athletes and their participation from 1896 to 2016.
- Scope of Variables: The analysis may focus on specific variables within the dataset, such as athlete demographics, medal counts, or participation rates, while excluding other variables that are not directly relevant to the research objectives.
- Data Quality: Delimitations may be imposed based on the quality and reliability of the dataset. For example, the analysis may exclude data

with a high degree of missing values or inconsistencies that could compromise the validity of results.

- **Research Objectives:** Delimitations are often defined by the specific research questions or objectives of the study. The analysis may focus on addressing specific research questions while omitting broader or tangential topics.

## **1.4 Benefits of research**

Research on the analysis of Olympics datasets can yield several benefits, including:

- **Insights into Sporting Trends:** By examining historical data on athlete performances, medal counts, and participation rates, researchers can identify trends and patterns in sports over time. This information can help sports organizations, policymakers, and stakeholders make informed decisions about resource allocation, infrastructure development, and athlete training programs.
- **Understanding Societal Dynamics:** The Olympics serve as a reflection of broader societal trends and dynamics, including cultural shifts, gender equality, and globalization. Analyzing Olympics datasets can provide insights into societal attitudes towards sports, nationalism, and identity, as well as the impact of socio-political events on sports participation and performance.
- **Global Collaboration and Diplomacy:** The Olympics represent a platform for global cooperation and diplomacy, bringing together athletes and nations from around the world. Research on Olympics datasets can foster cross-cultural understanding, promote dialogue between nations, and contribute to the advancement of international cooperation in sports.
- **Inspiration and Motivation:** Research on Olympics datasets can inspire and motivate athletes, coaches, and sports enthusiasts by showcasing examples of perseverance, determination, and excellence in sports. It highlights the achievements of athletes and the transformative power of sports in uniting people and overcoming barriers.

# Chapter 2. Implementation

Two datasets were used in the data analysis. The datasets' characteristics and specifics are as follows:

## 2.1 Details of the dataset:

Athlete dataset and NOC regions dataset.

- Athletes – events dataset

Rows: 271117

Columns: 15

The dataset contains information about the athletes and their participation from 1896 to 2016.

- NOC\_regions dataset

Rows: 231

Columns: 3

The dataset contains information regarding different National Olympic Committees (NOC) and their respective regions.

Breakdown of the columns in Athlete dataset:

- ID: Unique identifier for each athlete.
- Name: Athlete's full name.
- Sex: Athlete's gender (male or female).
- Age: Athlete's age at the time of the competition.
- Height: Athlete's height (measurement in centimetres).
- Weight: Athlete's weight (measurement in pounds).
- Team: Country the athlete represents.
- NOC: National Olympic Committee code (unique 3-letter code for each country).
- Games: Combined field indicating year and season (e.g., "2012 Summer").
- Year: Year of the Olympic Games.
- Season: Season of the Olympic Games (Summer or Winter).
- City: Host city of the Olympic Games.
- Sport: Specific sport the athlete competed in (e.g., gymnastics, swimming).
- Event: Specific event within the sport (e.g., 100m freestyle, men's high jump).
- Medal: Medal awarded to the athlete (gold, silver, bronze, or "NA" for no medal).

## 2.2 Data Preprocessing

1. **Data Merging:** Initially, we merged two separate files into a single CSV file. This consolidation ensures that all relevant data are contained within a single dataset, facilitating easier analysis and manipulation.
2. **Duplicate Removal:** Next, we identified and removed any duplicate records present in the dataset. Duplicate records could skew analysis results and inflate counts, leading to inaccurate conclusions. By eliminating duplicates, we ensure that each observation in the dataset is unique.
3. **Null Value Imputation:**  
For string value columns: We addressed missing or null values by replacing them with the string value "unknown". This standardizes the representation of missing categorical data and allows for consistent handling during analysis.  
  
For numerical columns: We imputed missing or null values with the mean of the respective column. Imputation with the mean helps preserve the integrity of the data distribution and minimizes the impact of missing values on subsequent analyses.

By systematically performing these preprocessing steps, we ensure that the dataset is cleaned and standardized, laying a solid foundation for further analysis and modelling.

# Chapter 3. Insights and trends

Numerous insightful and educational discoveries and information were discovered following the analysis of the Olympic data. We are examining the key findings and patterns from this research under each objective in this chapter.

## 3.1 Country-wise Analysis

### 3.1.1 Overall analysis

Up to the Rio de Janeiro Olympics in 2016, 498 countries took part in Olympic Games. A total of 480 nations competed in the summer Olympics while 77 nations took part in the winter Olympics. Around 403 teams exclusively competed in the summer Olympics.

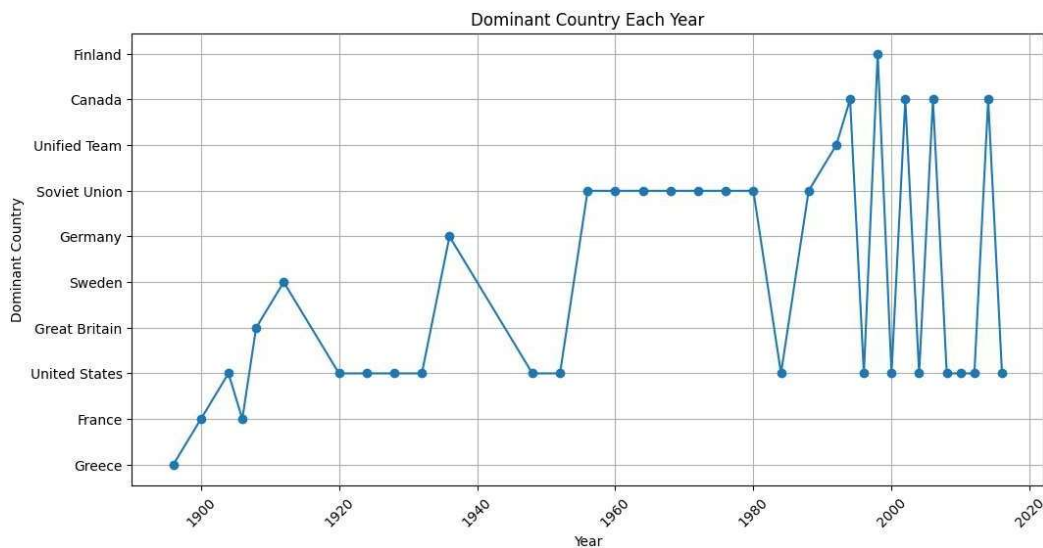


Fig 3.1.1(a) - Dominant country each year

- ☐ The 1896 summer, officially known as the Games of the I Olympiad and commonly known as Athens 1896 was the first international Olympic Games held in Athens, Greece.
- ☐ Greece was the first big star at the start of the Olympics in 1896.
- ☐ The United States has consistently secured the highest number of medals throughout Olympic history.
- ☐ The Soviet Union gained prominence from the 1950s to the 1990s but saw a decline in performance after the 2000s.
- ☐ Throughout the history of the Olympics, the United States of America

(USA) stands out as the country with the highest number of participants and the most medals won.

### 3.1.2 Summer Olympics-country analysis

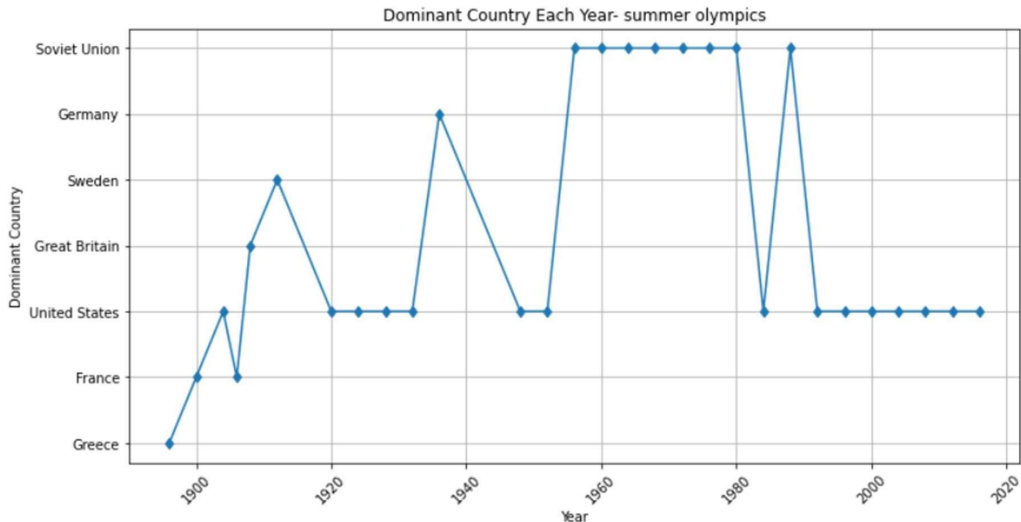


Fig 3.1.2(a) - Dominant Country Each Year- Summer Olympics

Total number of medals won by countries in summer olympics

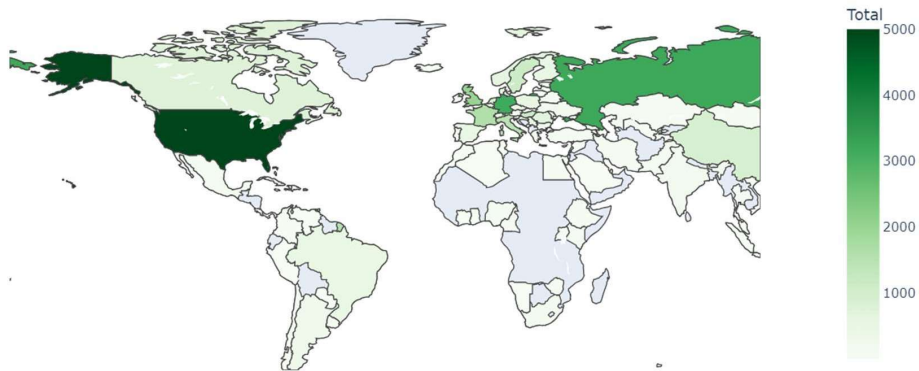


Fig 3.1.2(b) - World map- Summer Olympics

- ☐ The United States continues to have strong performances at the summer Olympics, particularly from the early years of the twenty-first century.
- ☐ The Soviet Union displayed its abilities between the 1950s and the 1980s, but it vanished as a result of its downfall.
- ☐ Certain parts of Asia and North America showcases good performance.



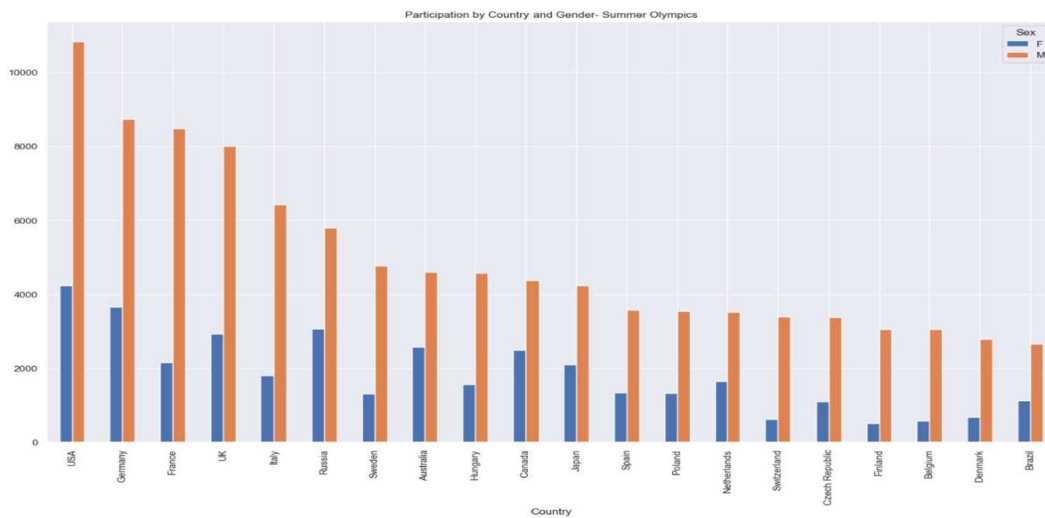


Fig 3.1.2(c) – Participation by country and gender- Summer Olympics

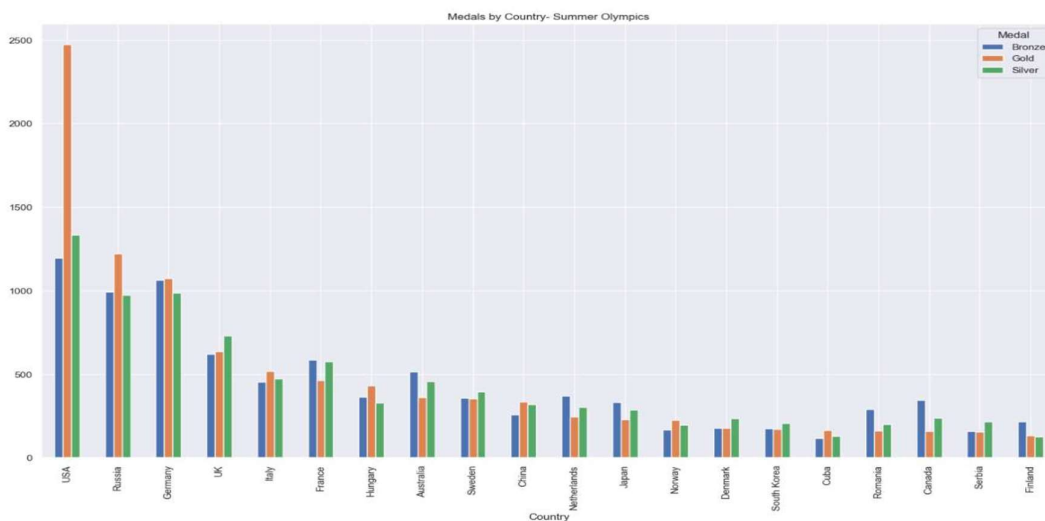


Fig 3.1.2(d) - Medals by country- Summer Olympics

- ☐ In the Summer Olympics (1896, 1900, 1904, 1906, 1908, 1912, 1920, 1924, 1928, 1932, 1936, 1948, 1952, 1956, 1960, 1964, 1968, 1972, 1976, 1980, 1984, 1988, 1992, 1996, 2000, 2004, 2008, 2012, 2016), the USA stood out as the frontrunner in terms of participation, boasting the highest numbers for both men and women, followed by Russia (Soviet Union) and Germany.
- ☐ On the flip side, Barbados showed the lowest male participation, while Ancora lagged behind in female representation.
- ☐ When it came to the medal tally, once again, the USA dominated, securing the most medals overall and Barbados secured the last position. However, in the race for gold, Germany had to yield to the prowess of Russia.

### 3.1.3 Winter Olympics-country analysis

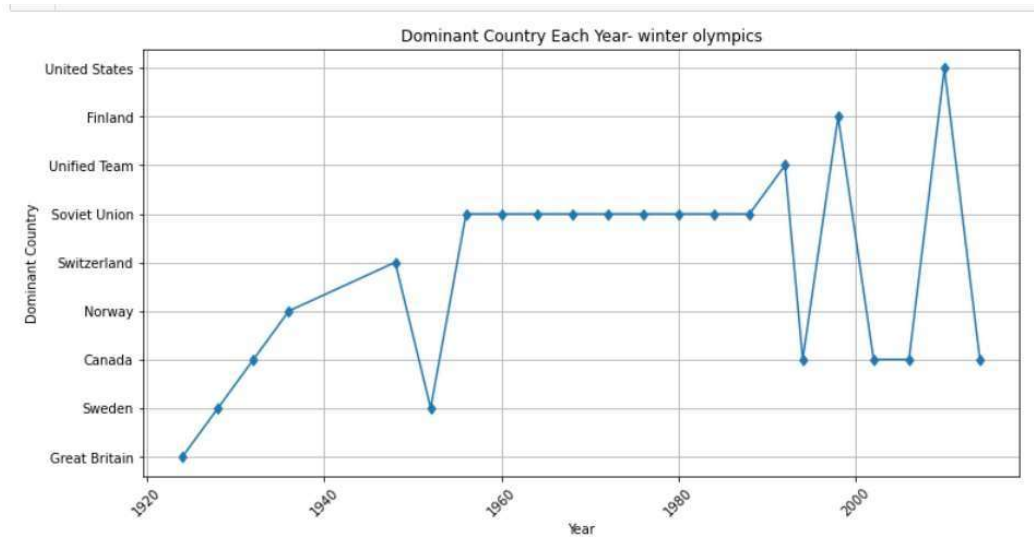


Fig 3.1.3(a) - Dominant Country Each Year- Winter Olympics

Total number of medals won by countries in winter olympics

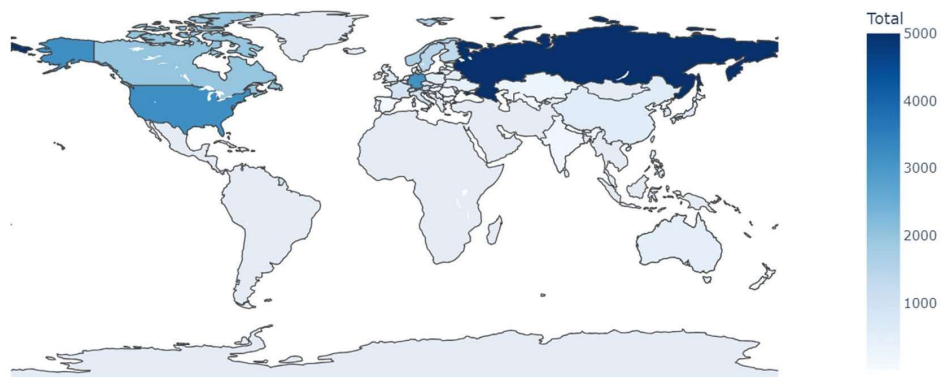


Fig 3.1.3(b) - World map - Winter Olympics

- Meanwhile in winter Olympics, countries like Russia (Soviet Union), Canada and other European countries shows a better performance.
- Factors such as frosty landscapes and a deep-rooted love for winter sports etc. give these countries a geographical and cultural edge in the Winter Olympics.

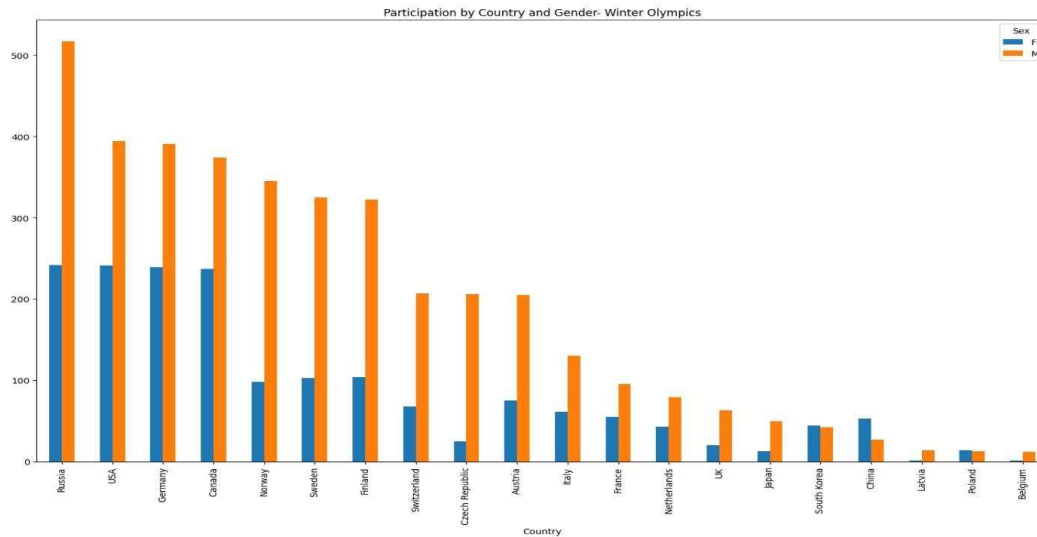


Fig 3.1.3(c) - Participation by country and gender- Winter Olympics

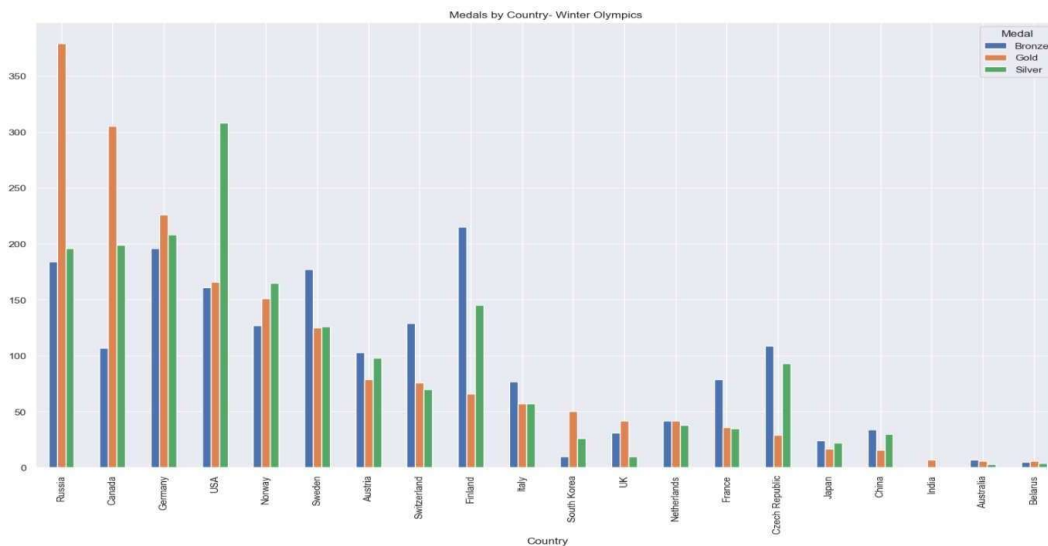


Fig 3.1.3(d) - Medals by country- Winter Olympics

- In the Winter Olympics (1924, 1928, 1932, 1936, 1948, 1952, 1956, 1960, 1964, 1968, 1972, 1976, 1980, 1984, 1988, 1992, 1994, 1998, 2002, 2006, 2010, 2014), Russia stood out as the frontrunner in terms of participation, boasting the highest numbers for both men and women, closely followed by USA.
- On the flip side, Bulgaria showed the lowest male participation, while Austria-2 lagged behind in female representation.
- When it came to the medal tally Russia dominated, securing the most medalsoverall and Nepal secured the last position. In the race for gold, Canada held second position, for silver USA held the second position and for bronze Finland held the second position.

## 3.2 Physique Analysis

An athlete's body plays a major role in their performance in the Olympics. A top performance physique is attained via years of training and diet, as each sport requires a certain build. Dominance is important, but having a strong physique also helps athletes avoid injuries, which is important for any Olympian going for gold. Here we are focusing on the age, height and weight of the athletes. Let's analyze the physique characteristics trends in different aspects:

### 3.2.1 Age analysis

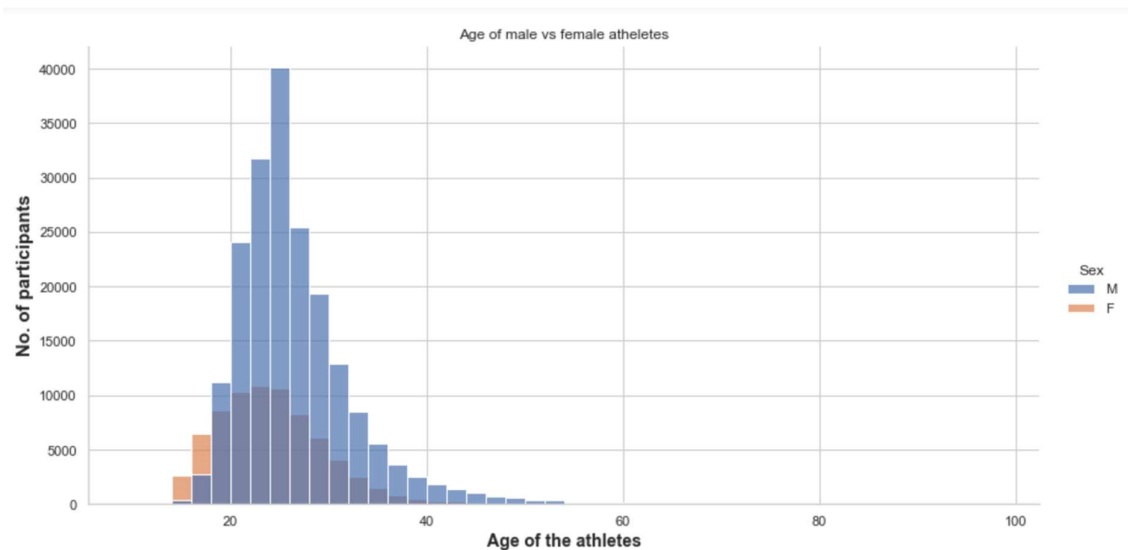


Fig 3.2.1(a) - Age of male Vs. female athletes

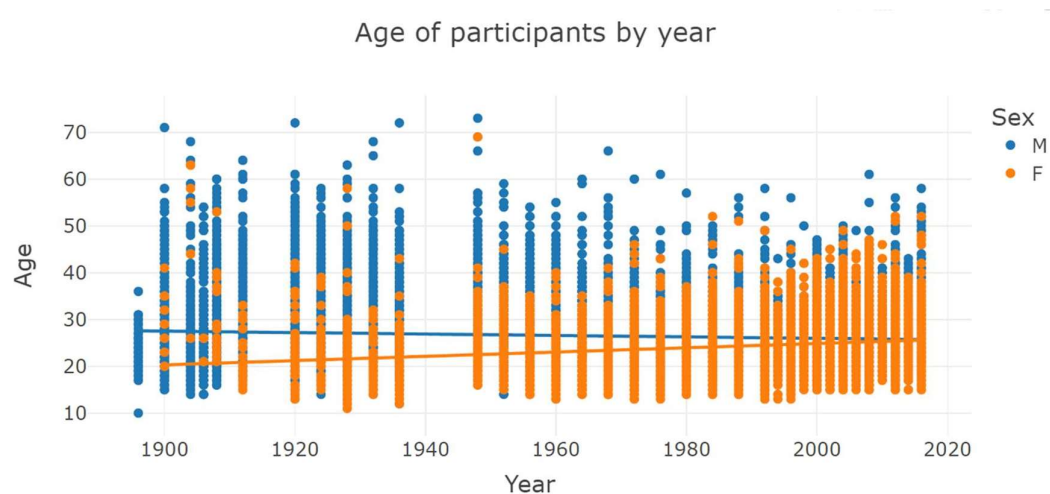


Fig 3.2.1(b) - Age of participants by year

- Throughout Olympic history, the average age of male participants has consistently hovered around 26, while their female counterparts have

maintained an average age of approximately 25.

- Delving into the origins of the Games, we find that in the inaugural Olympics, men had an average age of 24, while women were slightly younger at 20, reflecting the youthful vigor of the event's early years.
- Fast forward to the turn of the millennium, and we witness a striking continuity in the age demographics. Despite the passage of time and the evolution of sports culture, the average age for both male and female participants remained steadfast at 26, indicating a remarkable stability in the age composition of athletes from the early 1900s to the 2000s.
- It was observed that the optimal age to win a gold, silver and bronze medal was in between 20 and 30.
- Sports like shooting, equestrianism, art competitions, sailing, fencing etc. has the most number of older participants. This shows that in these sports, the experience is more important than the physique. These sports focus on technique, strategy, and mental focus, areas that can sharpen with age. They also tend to be adaptable for various physical abilities and offer a social aspect, perfect for lifelong learning and camaraderie.
- While in sports like swimming, figure skating, gymnastics, rowing etc. the age of participants is lower than the average which in turn implies that physique plays a major role in the performance. Compared to older athletes, they gain from their resistance, flexibility, and capacity for recuperation. These disciplines demand a combination of flexibility, power, and agility that naturally peaks at a younger age. The lighter frames and explosive energy of younger athletes give them a distinct advantage in these sports, making peak performance more challenging to sustain as athletes mature.

### 3.2.2 Weight analysis

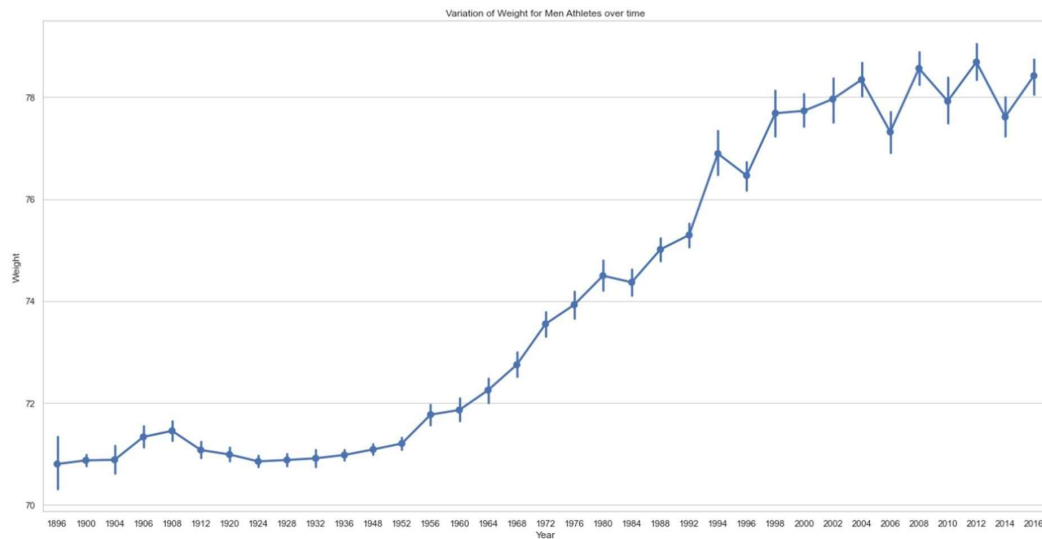


Fig 3.2.2(a) – Variation of weight for male athletes over time

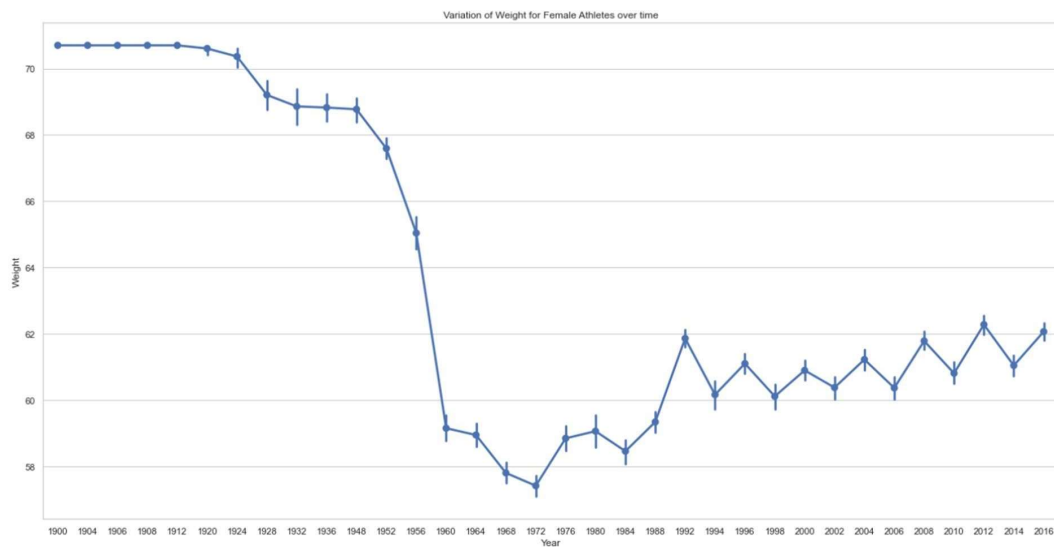


Fig 3.2.2(b) – Variation of weight for female athletes over time

- ☐ In the ancient arena of Olympic competition, where strength and skill collide in a breathtaking display of human potential, the quest for the optimal weight has witnessed a fascinating journey of evolution.
- ☐ Over time, the ideal weight for men to win a medal has gone up from about 71 to a sturdy 78 units.
- ☐ In a curious twist of fate, the optimal weight for female medallists has

embarked on a downward trajectory, descending from 71 units to 62. Yet, it is in the delicate balance of this weight reduction, particularly noted around the years 1955 and 1960, where the true essence of feminine strength and agility shines brightest.

### 3.2.3 Height analysis

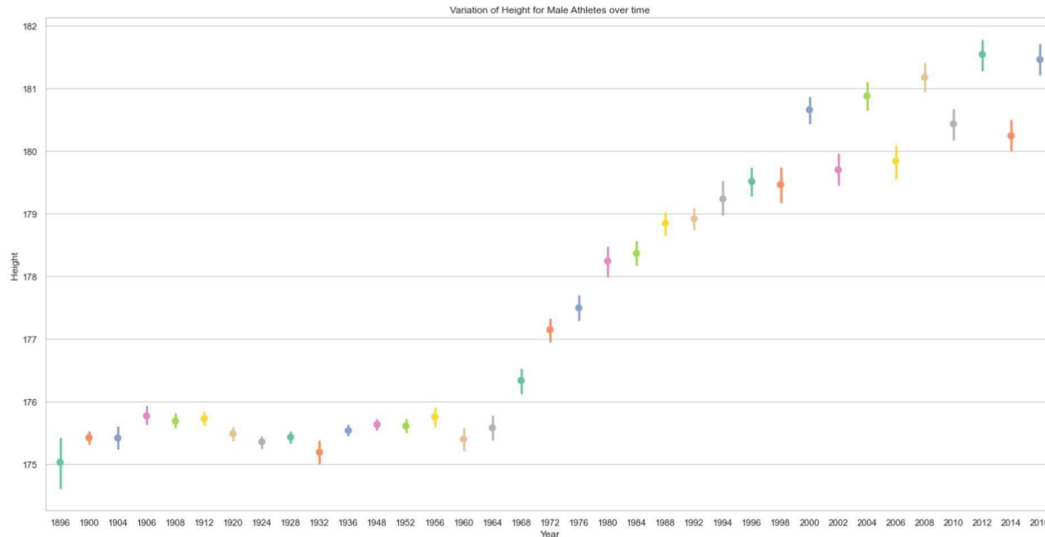


Fig 3.2.3(a) – Variation of height for male athletes over time

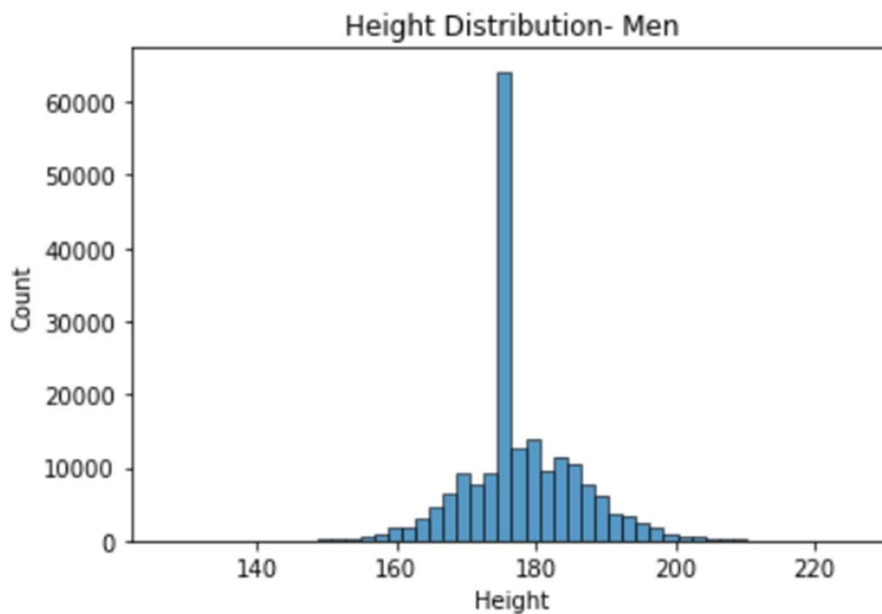


Fig 3.2.3(b) – Height distribution - Men

- In the towering realm of Olympic achievement, the bar for male athletes has steadily climbed from a starting point of 175cm to a range of 181-182cm.

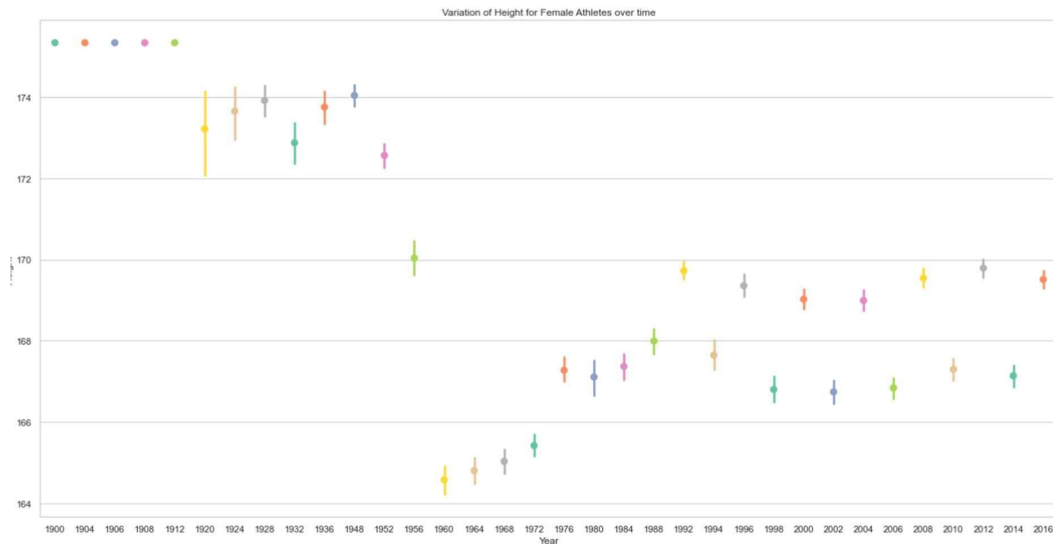


Fig 3.2.3(c) – Variation of height for female athletes over time

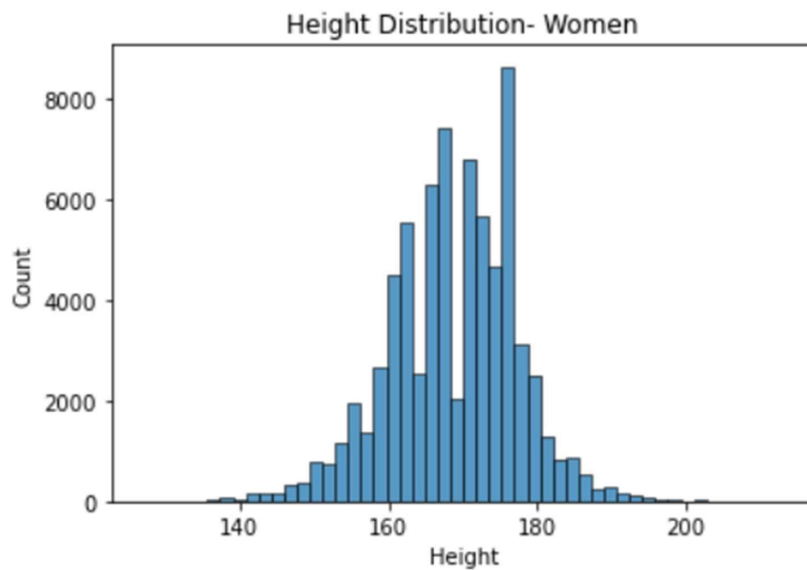


Fig 3.2.3(d) – Height distribution – Women

- Yet, for the graceful women standing tall at 170cm and above, they navigated a dip in requirements, descending to a lower threshold of 166cm and below around the 1950s. However they reclaimed their heights, surging back to a range between 166 and 170cm after the 1970s, and have remained steadfast ever since.



### 3.2.4 Descriptive analysis on physique

Here is the mean age, height and weight of gold medalists in the top 5 sports in Olympics. Considering this analysis is done on the gold medalist's data, it can be concluded that these findings indicate the features of an optimal physique.

---

Mean Age, Weight, and Height for Gold Medalists in Athletics :  
Mean Age: 25.003734129947723  
Mean Weight: 71.74756054443593  
Mean Height: 177.83015309127222

Mean Age, Weight, and Height for Gold Medalists in Swimming :  
Mean Age: 20.799818016378527  
Mean Weight: 73.46226635111096  
Mean Height: 180.74211859049518

Mean Age, Weight, and Height for Gold Medalists in Rowing :  
Mean Age: 25.328220858895705  
Mean Weight: 78.99726114634544  
Mean Height: 182.65886056924208

Mean Age, Weight, and Height for Gold Medalists in Gymnastics :  
Mean Age: 23.562579013906447  
Mean Weight: 62.743751794438694  
Mean Height: 167.87497783068713

Mean Age, Weight, and Height for Gold Medalists in Fencing :  
Mean Age: 28.78787878787879  
Mean Weight: 71.41866901679634  
Mean Height: 176.925945425514

Fig 3.2.4(a) – Mean of physical features- top 5 sports

- ☐ The average age of an Olympic gold winner in the top 5 sports is between 20 and 25 years old, with fencing having the highest average age at 28.7.
- ☐ The average weight is generally between 70 and 75 kg, with gymnastics having a mean weight of 62 kg and rowing having a mean weight of 79 kg, both of which are outside of the average range of 70-75 kg.
- ☐ The typical range for height is 175–185 cm, with gymnastics having a lower mean height of 162 cm.

Mean Age, Weight, and Height for female  
Mean Age: 25.0  
Mean Weight: 64.0  
Mean Height: 171.0

Mean Age, Weight, and Height for male  
Mean Age: 26.0  
Mean Weight: 77.0  
Mean Height: 180.0

Fig 3.2.4(b) – Mean of physical features- men vs women

- Mean age of a female athlete is 25 while a male athlete has a mean age of 26.
- The average height for a female and male athlete is 171 cm and 180 cm respectively.
- Average weight of a female athlete is 64 kg, while that of a male athlete is 77 kg.
- As we can see, male athletes compete in the Olympics at a higher age than women do. Because women attain biological maturity sooner than males, which enable them to compete in the Olympic Games earlier, this age distribution may be explained.

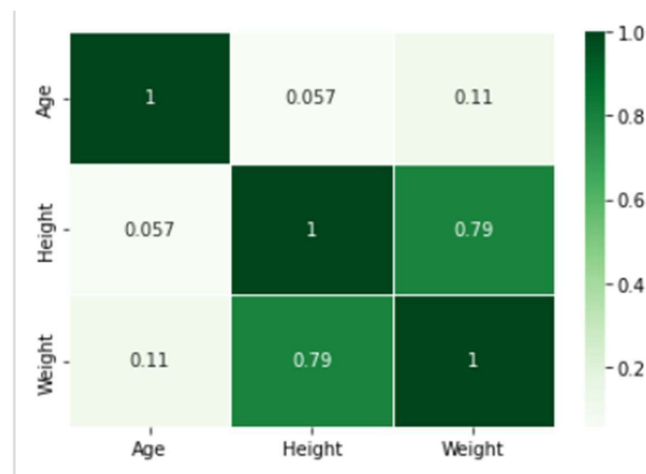


Fig 3.2.4(c) – Correlation matrix

Now let's take a look at the correlation matrix (heatmap) between age, height and weight.

- As one might anticipate, there is a positive link between "Height" and "Weight," but not between "Age" and the remaining variables.
- There's no one-size-fits-all - each sport demands a physique optimized for its unique performance needs.

		Age		
		min	max	mean
Season	Sex			
Summer	F	11.0	69.0	24.380110
	M	10.0	73.0	26.455206
Winter	F	13.0	46.0	25.186637
	M	14.0	58.0	26.469619

Fig 3.2.4(d) – Mean of physical features- summer vs winter

- ☐ Because the cold impairs performance more severely in older athletes, the maximum age for Winter Sports Athletes is considerably lower than for Summer Olympic Athletes.
- ☐ The mean age for both summer and winter Olympics falls under the same range of 24-27.

### 3.3 Participation Analysis

- ☐ As for the athletes, there have been 74,522 female athletes and 196,594 male competitors that have competed in the Olympics thus far. The majority of the participants were from the US, France, and the UK.
- ☐ With 13821 participants, the 2000 Summer Olympics drew the largest crowd, closely followed by the 1996 Summer Olympics with 13780 participants.
- ☐ Over the years, 48564 people participated in the winter Olympics while 221167 people attended the summer Olympics.

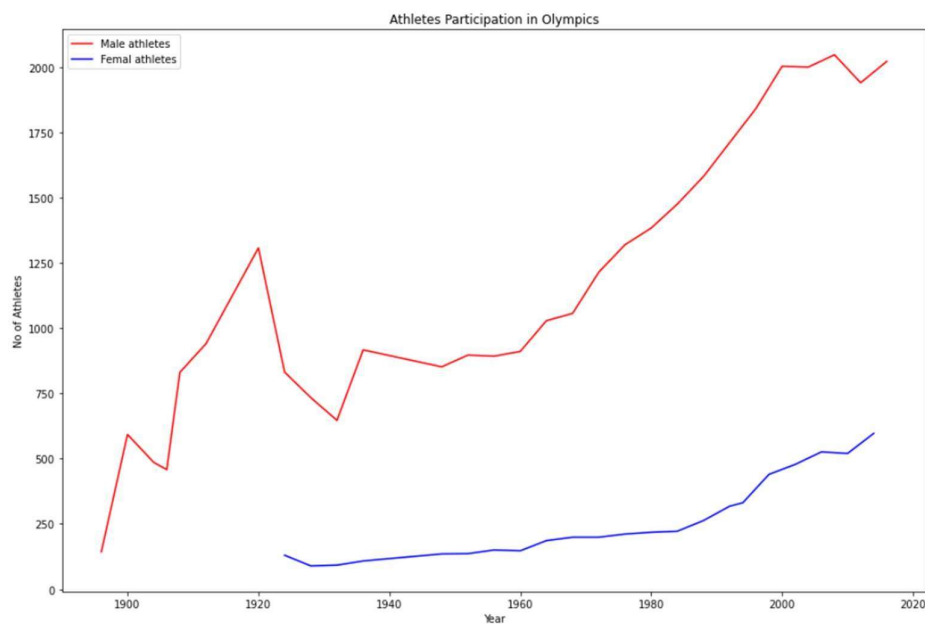
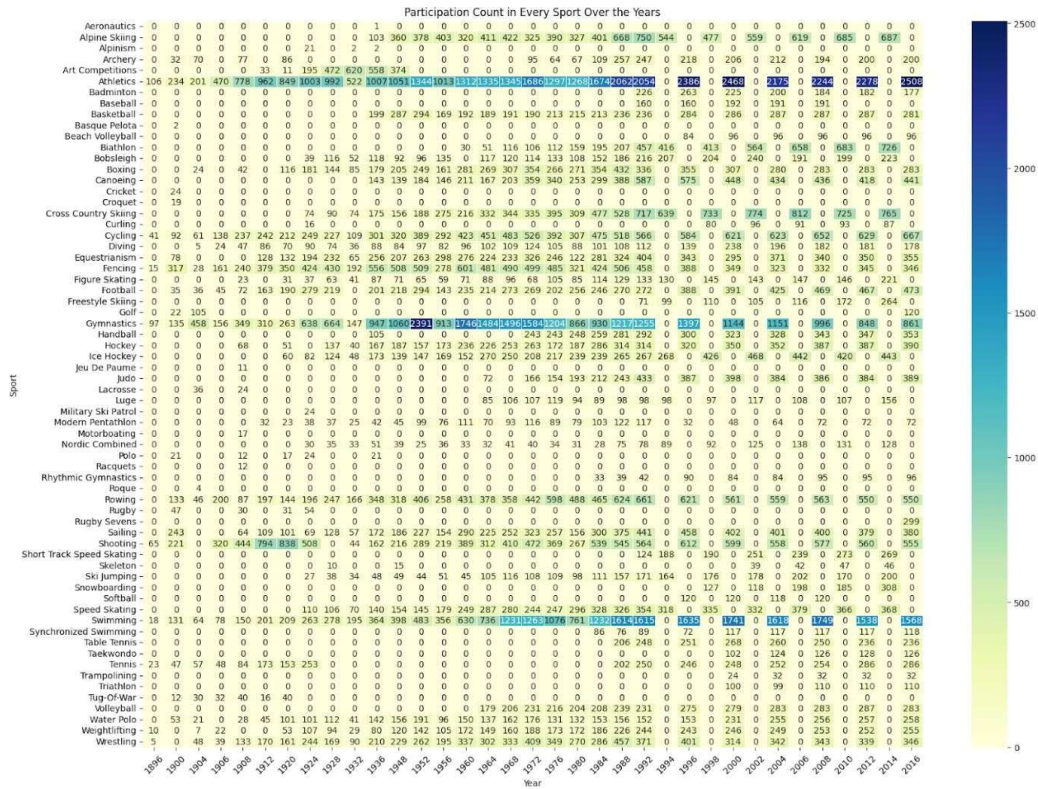


Fig 3.3(a) – Athletes participation

- Only 17.2% of the athletes win a medal.



- Sports like skeleton, ski jumping, snowboarding, softball, speed skating, synchronized swimming, table tennis, taekwondo, trampolining, triathlon, volleyball, and others are newly added activities that have grown in popularity over time. Other sports that follow the same trend include boxing, canoeing, cross-country skiing, curling,

freestyle skiing, Judo, Luge, Modern pentathlon, Nordic combined, and Rhythmic Gymnastics.

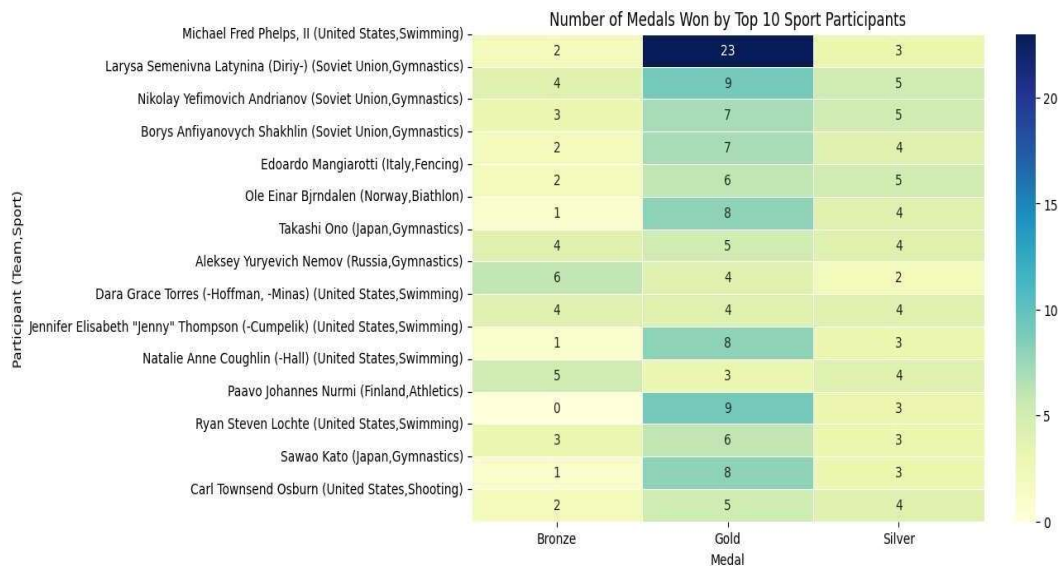


Fig 3.3(c) – Top 10 athletes

- Michael Phelps, the legendary American swimmer, tops the list of athletes with most number of medals in Olympics history with 23 gold, 3 silver and 2 bronze medals.
- Soviet Union Gymnasts Larysa Latynina, Nikolay Andrianov, and Borys Shakhlin, comes in the next positions, announcing their quality of craft and talent.

### 3.4 Gender Analysis

The gender gap is narrowing in the Olympics, but disparities remain. Traditionally male-dominated sports are seeing increasing female participation, while some sports favor specific physiques that tend to skew participation towards one gender or another. However, advancements in equipment and training methods are constantly pushing the boundaries for what's considered the "ideal" physique in each sport, paving the way for a more balanced future. These are our major findings in this aspect:

- Men have access to many more teams and events than do women. Not so much as the quantity of sports played by both sexes together.
- This is explained by the fact that women were not allowed to participate in as many contests as men until London 2012 and that women do not play as many sports as men do elsewhere, which drastically lowers the number of teams that may compete.

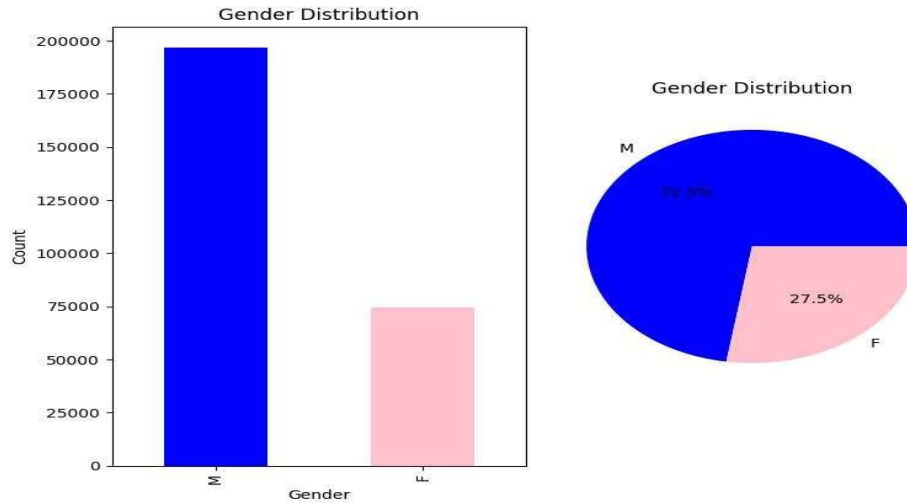


Fig 3.4(a) – Gender distribution

- Throughout the annals of Olympic history, female participation has remained a minority, comprising less than one-third of the total athletes.

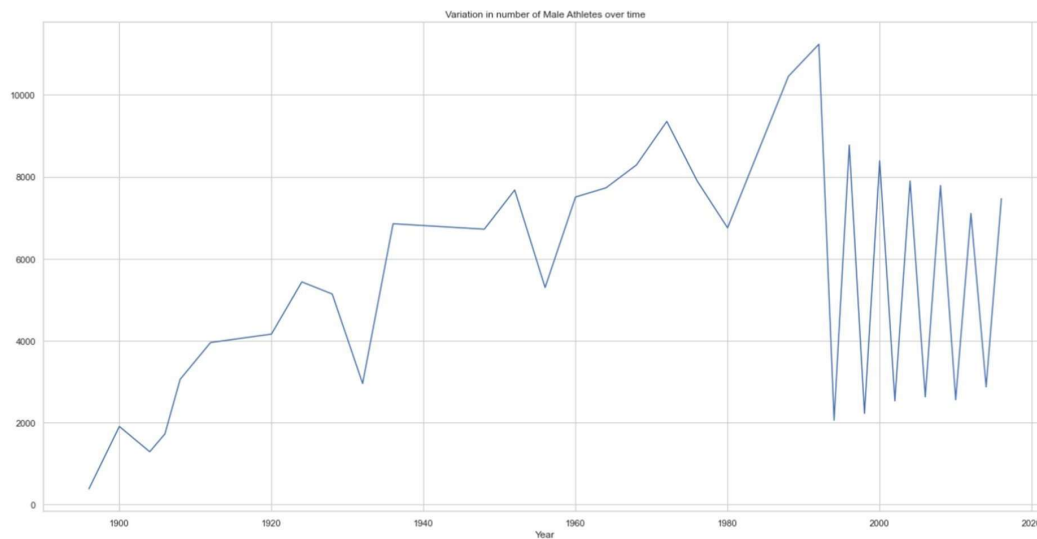


Fig 3.4 (b) – Variation in number of male athletes over time

- Male participation peaked around the 1990s, reaching its highest point.

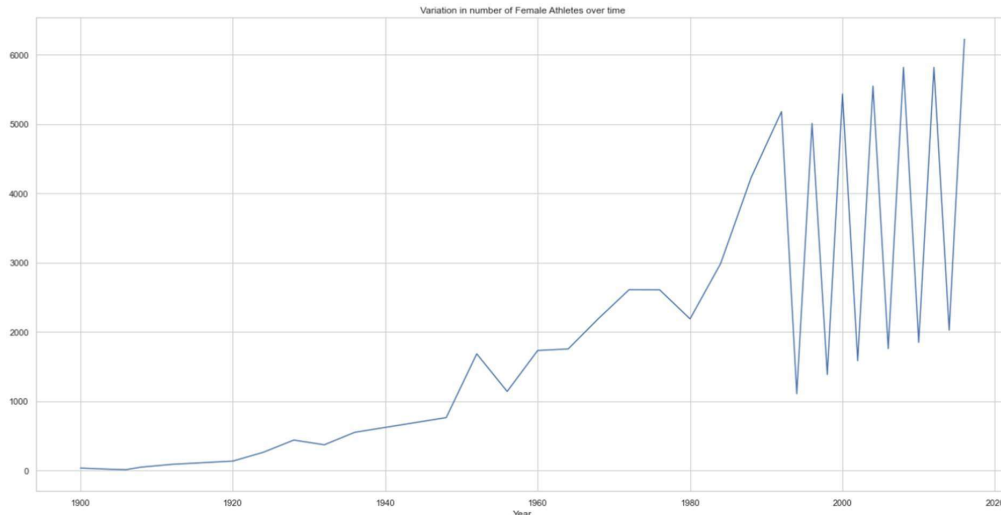


Fig 3.4(c) – Variation in number of female athletes over time

- ☐ Female participation didn't hit its peak until after 2015.
- ☐ Over the years, male participation experienced significant growth, whereas female participation showed a slower progression.

### 3.5 Seasonal Analysis

There are summer and winter versions of Olympic games. Here are the key insights from our season analysis:

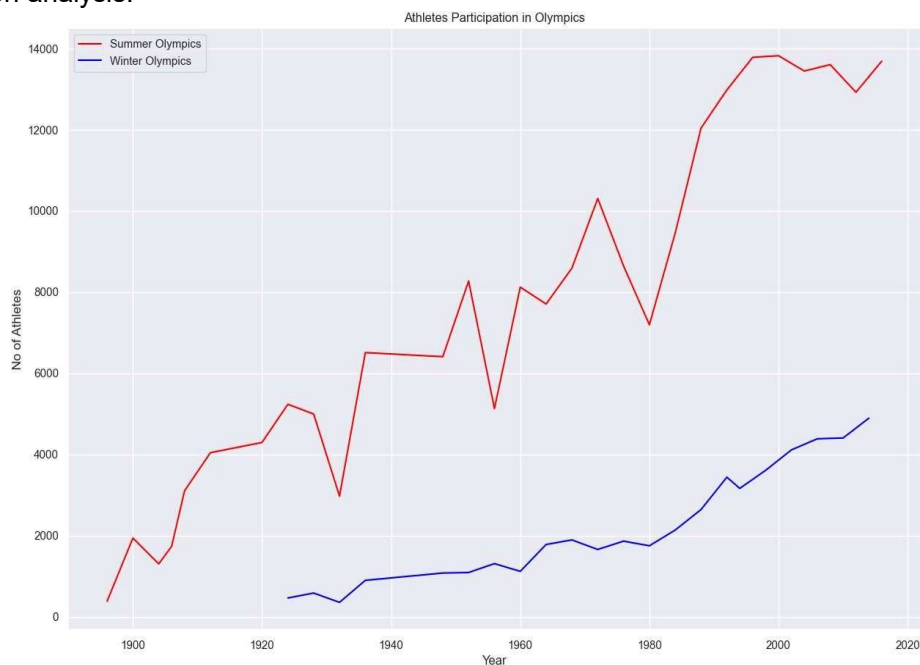


Fig 3.5(a) – Athlete participation - Summer vs Winter

- The Summer Olympics have historically enjoyed greater popularity and participation compared to their Winter counterpart.
- A few factors contribute to the popularity of the Summer Olympics. First and foremost, diversity is essential. With over 30 sports to Winter's 15 offerings, Summer has something for everyone. Second, accessibility is really important. Compared to winter sports that require specific equipment and weather conditions, many summer sports are well-known year-round pastimes that attract a wider fan base. Lastly, summer sports have long-standing fan favorites like track and field, which have rich histories and renowned players that appeal to spectators more than certain winter sports.
- Both events were held in the same year until 1992, with notable years being 1924, 1928, 1932, and so forth. The inaugural Olympics in 1896, hosted in Greece, marked the beginning of the Summer Games tradition.
- The Winter Olympics emerged around 1924, coinciding with the Summer Games for several decades.
- However, from 1994 onwards, they were staggered, with a two-year gap between each event. For instance, the Winter Olympics of 1994 preceded the Summer Olympics of 1996. This scheduling pattern persisted thereafter, allowing each event to command its own spotlight and attention.

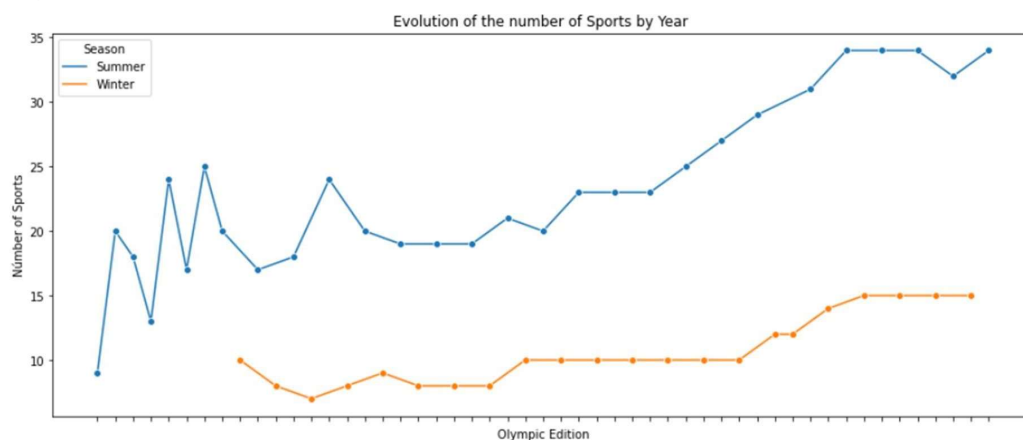


Fig 3.5(b) – Evolution of number of sports

- The line plot shows that, except for the Winter Olympic Games, the number of sports has increased virtually every edition of the Olympic Games.



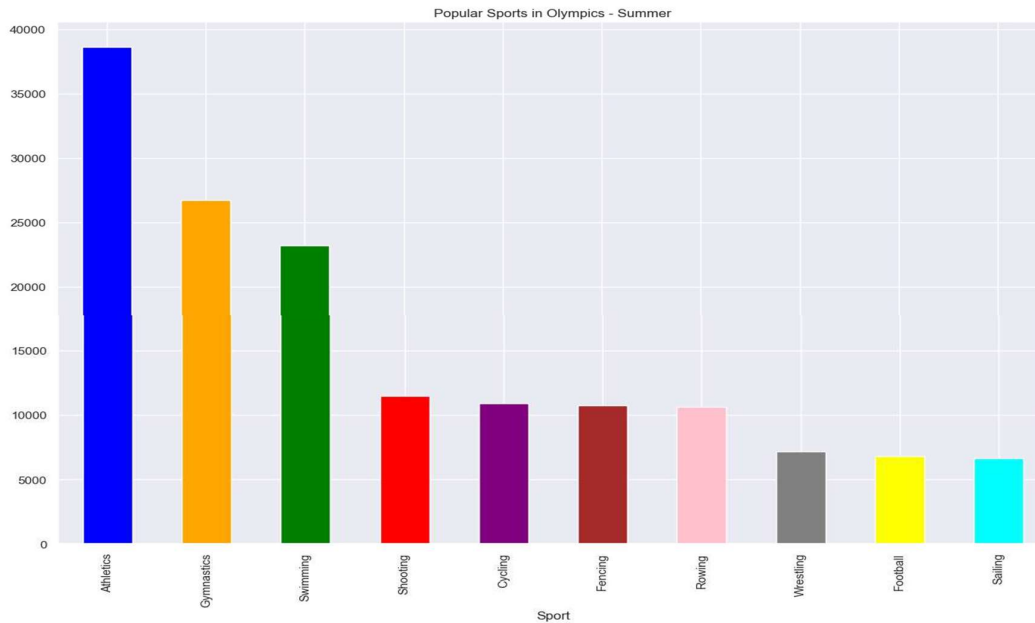


Fig 3.5(c) – Popular sports - summer

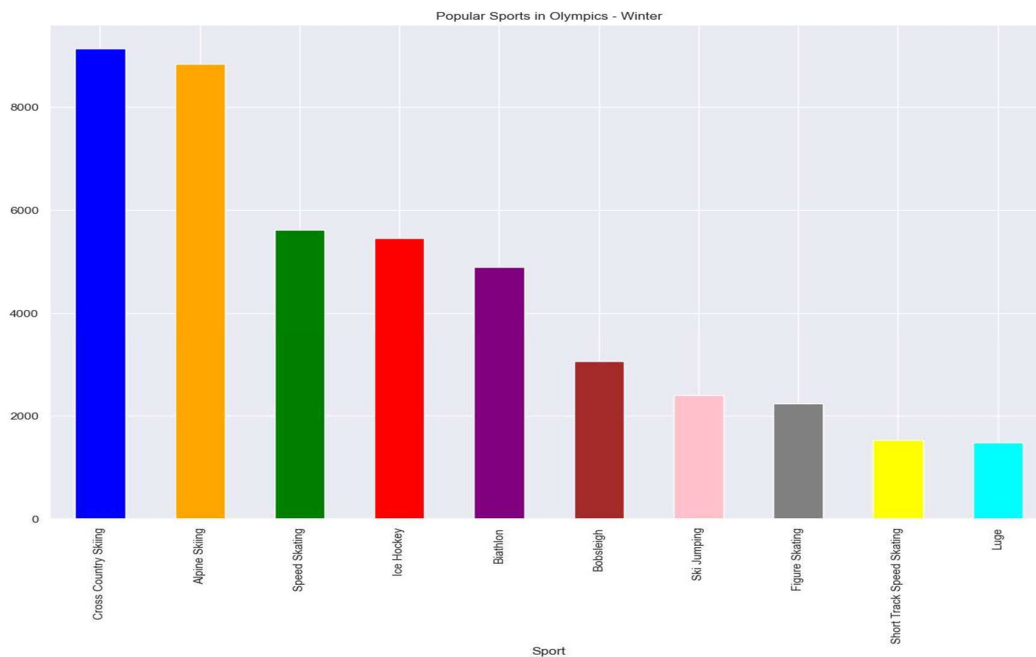


Fig 3.5(d) – Popular sports- winter

- The Summer Olympics boasted a triumvirate of beloved sports: athletics, gymnastics, and swimming, captivating audiences worldwide with their blend of athleticism, grace, and aquatic prowess.
- Meanwhile, the Winter Olympics showcased the thrilling spectacle of cross-country skiing, alpine skiing, and speed skating, each event carving its own niche in the icy terrain

- The popularity of sports in the Summer and Winter Olympics is influenced by a variety of factors, including geographic considerations, climatic considerations.

#### Summer Season

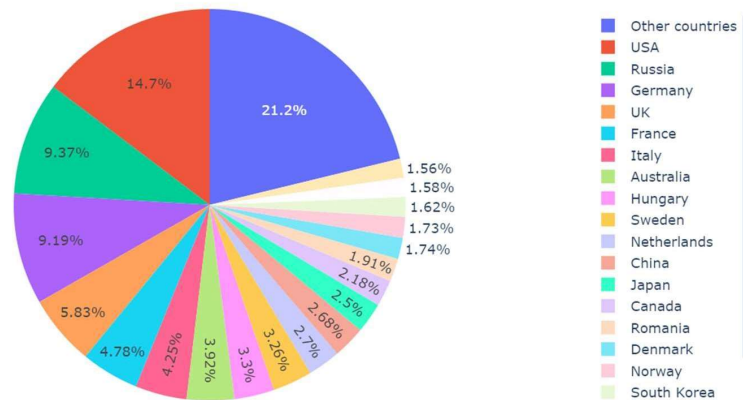


Fig 3.5(e) – Top countries- summer

#### Winter Season

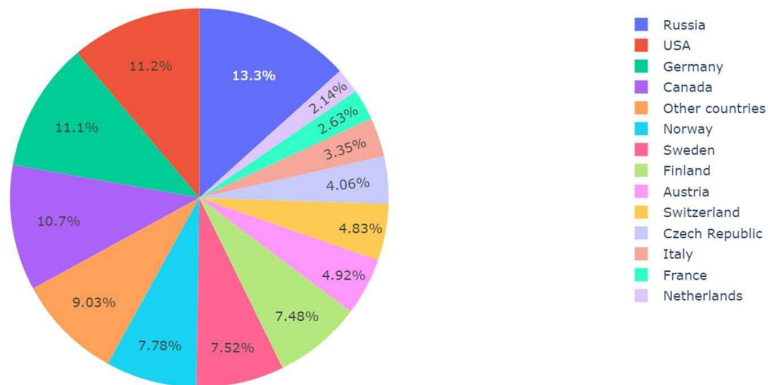


Fig 3.5(f) – Top countries- winter

- Most medals were won by USA in Summer Olympics, while in Winter Olympics Russia ranked highest.

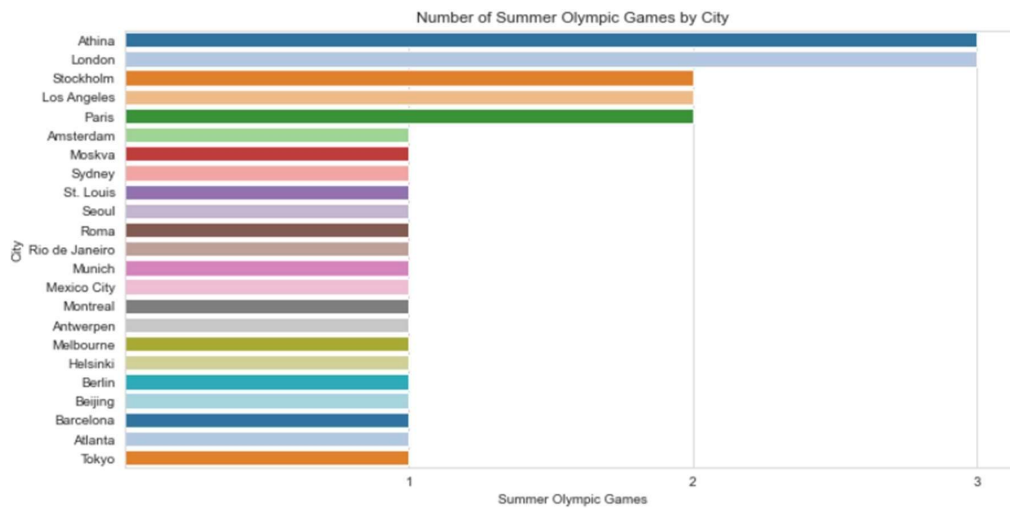


Fig 3.5(g) – City – Summer Olympics

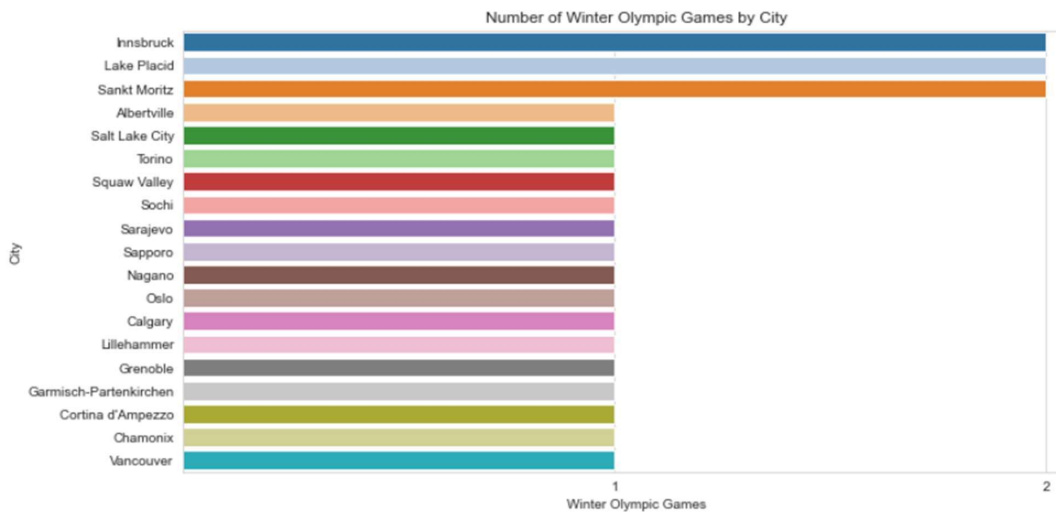


Fig 3.5(h) – City – Winter Olympics

- The cities that have played host to this sporting event are listed along with the number of times they have done so. Upon initial observation, it appears that European cities are more common, which might be explained by the fact that the historic sporting culture was initially popularized and resurrected on this continent.

### 3.6 India's Sporting Tapestry

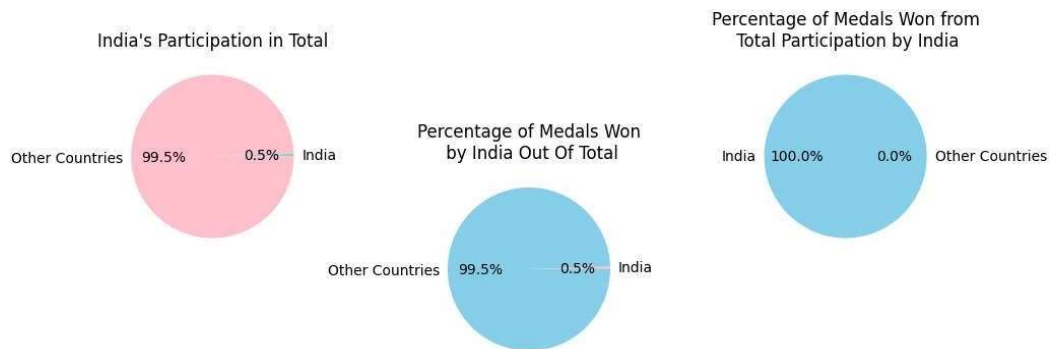


Fig 3.6(a) –India's participation

- ☐ Compared to other countries, India has had a diminished presence in total Olympics participation.
- ☐ However, what sets India apart is its remarkable consistency in securing atleast one medal in every sport in which it has participated.

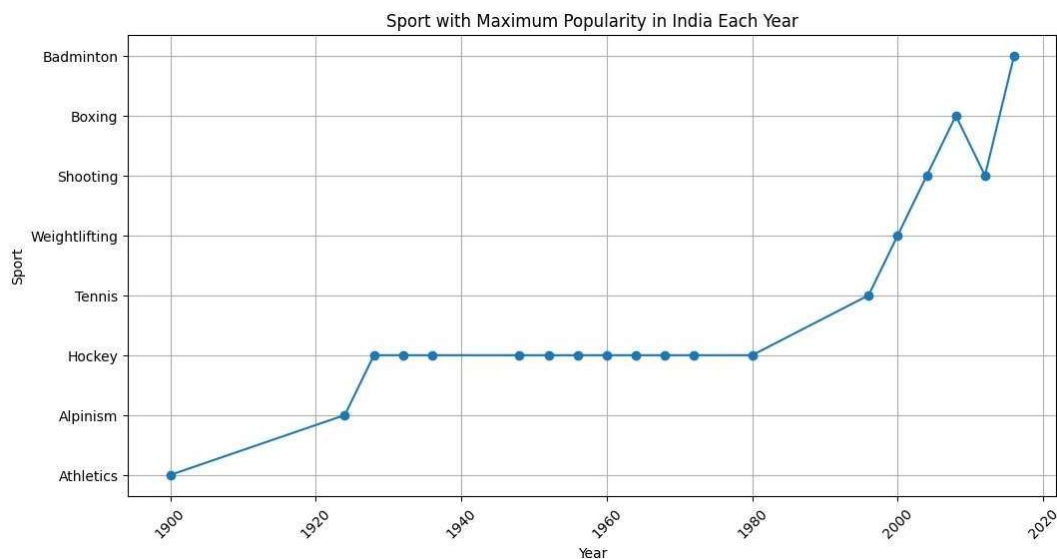


Fig 3.6(b) – Sports with maximum popularity in India each year

- ☐ In the early years of the Olympics, athletics held a prominent place in India's sporting landscape.
- ☐ However, by 1920, the rise of hockey saw it surpass athletics as the nation's most beloved sport.
- ☐ Fast forward to the turn of the millennium, disciplines such as shooting, boxing, and badminton emerged as the new darlings of the sporting scene, capturing the hearts and minds of a new generation of athletes and enthusiasts.

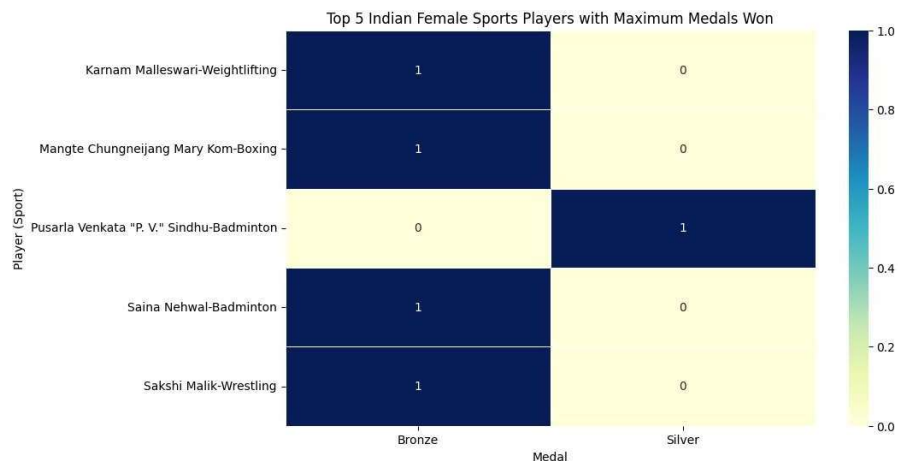


Fig 3.6(c) – Top 5 female Indian athletes

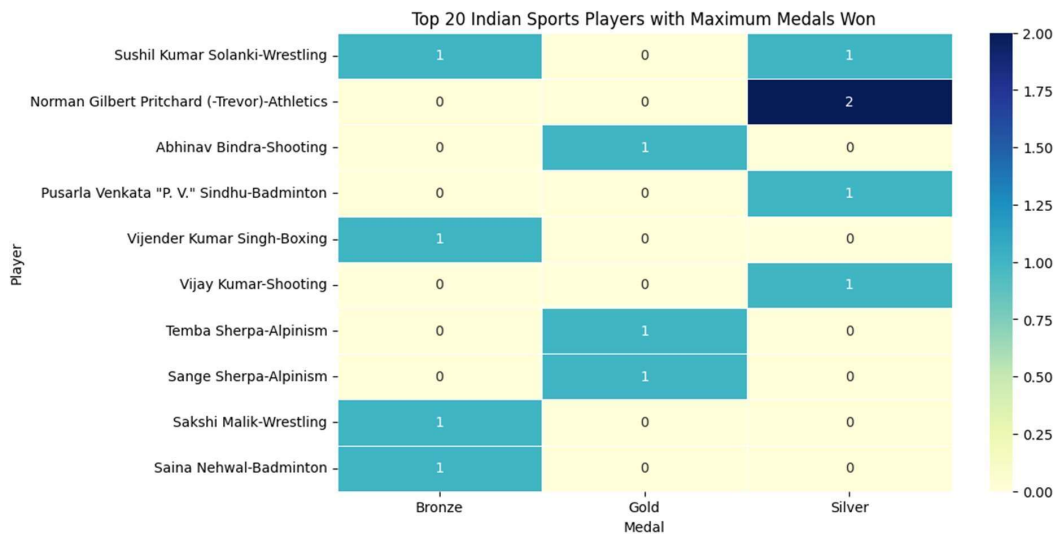


Fig 3.6(d) – Top 20 Indian Athletes

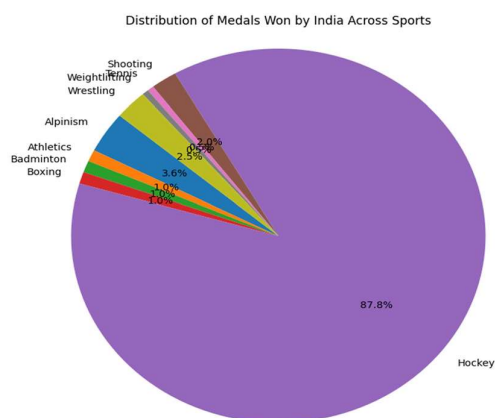


Fig 3.6(e) – India's medal distribution

- The Indian men's field hockey team won its first Olympic gold medal in 1928 and went on to win six consecutive gold medals from 1928 to 1956.
- Field hockey holds a special place in Indian culture and history. The sport was introduced to India during British colonial rule, and it quickly became popular among the masses. Hockey became deeply ingrained in the Indian sporting psyche and was embraced as a national passion.
- Hockey boasted the highest medal count, followed by alpinism.
- Weight-lifting and Tennis were popular sports before 2000.

# Chapter 4. Inferences

## 4.1 Female as minority

The underrepresentation of female athletes throughout Olympic history can be attributed to a combination of historical, cultural, and societal factors.

- Historically, sports have been predominantly male-dominated, with limited opportunities for women to participate due to societal norms and expectations. Women often faced barriers such as societal perceptions of femininity, restrictive gender roles, and limited access to athletic training and facilities.
- Culturally, many societies have held traditional beliefs and attitudes that prioritize and encourage male participation in sports over female involvement. This cultural bias has perpetuated the notion that certain sports are more suitable or appropriate for men, leading to fewer opportunities and support for female athletes.
- Institutionalized sexism within sports organizations and governing bodies has historically favoured male athletes in terms of funding, resources, and recognition. This disparity in support has hindered the development and progression of women's sports, further perpetuating their minority status in the Olympic arena.

While progress has been made in recent decades towards gender equality in sports, with initiatives aimed at increasing female participation and representation, there is still work to be done to address the systemic barriers and biases that continue to impact women in sports.

## 4.2 Factors influencing participation and dominance among nations

Several factors contribute to the discrepancy in participation and dominance among countries in the Olympics:

- **Historical Development:** Countries like the United States and Russia have invested heavily in sports infrastructure, talent development programs, and coaching resources over many decades. This long-term investment has contributed to their consistent presence and success in the Olympics.

Population and Resources: Countries with larger populations and greater financial resources often have an advantage in sports participation and performance. The United States, with its large population and strong economy, has a vast talent pool and resources to support athletes in various sports. Similarly, Russia benefits from a rich sporting tradition and government support for athletic development.

- Sports Culture: The culture and tradition surrounding sports play a significant role in a country's Olympic success. In some countries, certain sports are deeply ingrained in the cultural fabric and enjoy widespread popularity and support from the government, media, and public. This can lead to greater participation and success in those sports. For example, ice hockey is deeply rooted in the culture of Russia.
- Geographical and Climate Factors: The availability of suitable training facilities and natural resources can impact a country's ability to excel in certain sports. Countries with favourable climates or geographical features may have an advantage in sports like skiing, swimming, or track and field events. This could influence their participation and success in those sports compared to countries with less conducive environments.
- Government Policies and Investment: Government policies, funding allocations, and investment in sports infrastructure can have a significant impact on a country's Olympic performance. Countries that prioritize sports development and allocate resources accordingly are likely to have greater success in the Olympics. In contrast, countries with limited government support for sports may struggle to compete at the highest levels.



## Chapter 5. Conclusion

To sum up, this data analysis provides an engrossing look into the rich past and fascinating currents of the Olympic Games. The data shows an interesting pattern, starting with the first-ever Athens Games in 1896, where Greece was the early favorite, and ending with the Summer Olympics and Athletics' ongoing appeal. It's interesting to note that the research emphasizes the persistent gender disparity in participation, with women continuing to be underrepresented throughout Olympic history. Furthermore, climate and geographic variables have a big impact on how popular certain sports are in both summer and winter versions.

Furthermore, the analysis reveals intriguing information about the host cities is also revealed by the investigation, with London having hosted the most Olympic Games to date. This initiative acts as a launchpad for more research. Subsequent research endeavours may conduct more comprehensive examinations of the variables impacting participation patterns, including the continuous endeavor for gender parity and the influence of technology progress on sports performance. Examining these changing facets will help us comprehend the Olympics' future and how it may continue to motivate competitors and spectators throughout the globe for many years to come.

## REFERENCES

- [1]. Joshi A. P, Patel B. V. Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process. Orient.J. Comp. Sci. and Technol; 13(2,3). Available from: <https://bit.ly/39pJDXm>
- [2]. <https://www.datacamp.com/blog/data-sets-and-where-to-find-them-navigating-the-landscape-of-information>
- [3]. <https://www.simplilearn.com/exploratory-data-analysis-article>
- [4]. Zhang Z. Predictive analytics in the era of big data: opportunities and challenges. Ann Transl Med. 2020 Feb;8(4):68. doi: 10.21037/atm.2019.10.97. PMID: 32175361; PMCID: PMC7049053.
- [5]. <https://medium.com/@kaanerdenn/exploratory-data-analysis-eda-bf98e8586e6f>
- [6]. <https://medium.com/@jinalg.tagline/mastering-data-visualization-in-python-a-comprehensive-guide-0164a8b61493#:~:text=Python%20boasts%20a%20wide%20array,creating%20a%20variety%20of%20visualizations.>

## APPENDIX - I

### SOURCE

### CODE

Find out more about or project here:

**[https://github.com/niranjana628/Futureense\\_internship\\_capstone\\_project](https://github.com/niranjana628/Futureense_internship_capstone_project)**

## APPENDIX II- DATASHEETS

The Olympics data analysis was done on 2 datasets.

### 1. Athlete\_events csv file:

1	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
2	1	A Diliang	M	24	180	80	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NA
3	2	A Lamusi	M	23	170	60	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NA
4	3	Gunnar Nielsen Aaby	M	24	NA	NA	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NA
5	4	Edgar Lindenau Aabye	M	34	NA	NA	Denmark/Swede	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
6	5	Christine Jacoba Aaftink	F	21	185	82	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NA
7	5	Christine Jacoba Aaftink	F	21	185	82	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 1,000 metres	NA
8	5	Christine Jacoba Aaftink	F	25	185	82	Netherlands	NED	1992 Winter	1992	Winter	Albertville	Speed Skating	Speed Skating Women's 500 metres	NA
9	5	Christine Jacoba Aaftink	F	25	185	82	Netherlands	NED	1992 Winter	1992	Winter	Albertville	Speed Skating	Speed Skating Women's 1,000 metres	NA
10	5	Christine Jacoba Aaftink	F	27	185	82	Netherlands	NED	1994 Winter	1994	Winter	Lillehammer	Speed Skating	Speed Skating Women's 500 metres	NA
11	5	Christine Jacoba Aaftink	F	27	185	82	Netherlands	NED	1994 Winter	1994	Winter	Lillehammer	Speed Skating	Speed Skating Women's 1,000 metres	NA
12	6	Per Knut Aaland	M	31	188	75	United States	USA	1992 Winter	1992	Winter	Albertville	Cross Country S	Cross Country Skiing Men's 10 kilometres	NA
13	6	Per Knut Aaland	M	31	188	75	United States	USA	1992 Winter	1992	Winter	Albertville	Cross Country S	Cross Country Skiing Men's 50 kilometres	NA
14	6	Per Knut Aaland	M	31	188	75	United States	USA	1992 Winter	1992	Winter	Albertville	Cross Country S	Cross Country Skiing Men's 10/15 kilometres Pursuit	NA
15	6	Per Knut Aaland	M	31	188	75	United States	USA	1992 Winter	1992	Winter	Albertville	Cross Countrv S	Cross Countrv Skiing Men's 4 x 10 kilometres Relay	NA

Total: 271117 rows, 15 columns

### 2. noc\_regions csv file

1	NOC	region	notes
2	AFG	Afghanistan	
3	AHO	Curacao	Netherlands Antilles
4	ALB	Albania	
5	ALG	Algeria	
6	AND	Andorra	
7	ANG	Angola	
8	ANT	Antigua	Antigua and Barbuda
9	ANZ	Australia	Australasia
10	ARG	Argentina	
11	ARM	Armenia	
12	ARU	Aruba	
13	ASA	American Samoa	
14	AUS	Australia	
15	AUT	Austria	

Total: 231 rows, 3 columns

## INFORMATION REGARDING STUDENTS

STUDENT NAME	EMAIL ID	PERMANENT ADDRESS	PHONE NUMBER
AYUSHI TAWARI	AYUSHIT AWARI03 @GMAIL. COM	YASORAM INDRADHANUSH APARTAMENT TD ROAD ERNAKULAM KERALA	9995951128
MUTTA DATTA SAI VISHNU MOHAN	VISHNU.J OJO26@G MAIL.CO M	DR NGR'S GANDHI DAS APARTMENT,VIJAY AWADA,ANDHRA PRADESH	9391452963
NIRANJANA J	NIRANJA NAJITHEN DRAN@G MAIL.CO M	SARANGI (H), THAZHVARAMROAD, KAVUNGAL BYPASS, UPHILL (P.O), MALAPPURAM, KERALA	8891800705