

DIC Phase 1

Report

Geetansh - 50607410
Sataakshi - 50607324
Rudraksh - 50604938

1. Problem Statement:-

A Deep Dive into Historical Olympic Performance Trends

This project aims to analyze the historical trends of Olympic performances, focusing on medal tallies, country-wise achievements, sport-wise achievements, and participation in various sports. By examining these patterns, we can discover insights such as the performance of certain countries in specific sports, shifts in performance over time, and variations in participation based on factors like gender and athlete age, etc. The goal is to identify and explore significant trends in the Olympics over the years. In addition, we might take this analysis a step further by predicting the 2024 Olympic medal tally, with a plan to compare these predictions against the actual results (data strategy yet to be defined).

Explain the potential of your project to contribute to your problem domain. Discuss why this contribution is crucial.

The project "A Deep Dive into Historical Olympic Performance Trends" provides valuable insights into sports analytics by analyzing historical Olympic data. This contribution is crucial as it helps identify strengths and weaknesses in countries' performances. It will be a guide into how funds have been and will be allocated after carefully analyzing the country's performance. It will also be a way to encourage the participation of underrepresented groups. It will also promote inclusivity in sports. Additionally, predicting the 2024 Olympic medal tally using historical data fosters a data-driven approach to future preparations. Overall, this project significantly enhances the understanding of Olympic performance trends, ultimately helping athletes and sports committees improve performance and participation. We are also finding trends based on the ages of athletes to understand sports careers better.

2. Questions Geetansh

First Question:-

What is the general trend in women participation country wise over the years? What countries are doing well and how do they compare to the best performing countries?

This general question can lead to multiple hypotheses including the need of promoting women empowerment, understanding and highlighting gender disparity, urgent need to change in policies and awareness, etc.

This is a significant question because in this day and age we should ideally have equal women participation in sports, especially in a major competition like the Olympics. This also helps us identify which countries need to focus on the gender disparity issue more.

Second Question :-

Are there any sports which are on the decline and losing popularity among participants? Also, are there some sports which have gained popularity over the recent years?

This general question can again lead to multiple hypotheses including the need to spread awareness about some particular sports, predict which sport to remove from the Olympics, what sports have previously suffered this fate, etc.

This is a significant question because we should be aware of which sports are losing popularity to save them from getting extinct. On the world stage we should be able to identify which sports are on the rise, it can be beneficial for marketing and branding (great opportunity for money making).

2. Questions Sataakshi

Question 1

How do the trends in medal counts for team sports compare to those for individual sports across different countries over the years, and what insights can be drawn from these comparisons regarding each country's performance in the Olympic Games?

What?

This analysis compares the trends in medal counts for team sports versus individual sports across different countries over the years. Using queries to extract data I have analysed how the number of medals have been won in team events vs in individual events for different countries. This study provides an analysis of how countries perform in team versus individual competitions.

Why?

The question of how trends in medal counts for team sports compare to those for individual sports across different countries is significant for several reasons. It enhances our understanding of national strengths, allowing us to identify which countries excel in team versus individual sports and informing national sports policies. This analysis can also guide investment in sports programs. So, this question provides valuable insights that can shape the future of sports.

Question 2

How has the participation of women athletes in various sports evolved, and what trends can be observed in terms of minimum and maximum participation levels across selected sports?

What?

The analysis of women's participation in various sports over time reveals significant trends in their involvement. By filtering the data for women athletes, we observe how participation levels have increased in these years in all sports.

Why?

The analysis aims to highlight advancements in women's participation in sports, demonstrating progress in inclusion and representation. By examining trends over time, we can assess the effectiveness of initiatives aimed at increasing female participation and identify sports where participation still lags. This information informs sports organizations and advocates about the current situation of women's sports. This will help in gender equity in olympics.

2. Questions Rudraksh

- 1) Is the average age of athletes increasing or not?

As advancements in technologies and therapies improve and preserve Athletes' health, is it getting reflected through average age or not?

Also, is the male athlete's average age more than the female athlete or not? Male athletes tend to have longer careers than female athletes. That's what we want to investigate as well.

We are showing min, max, and average values of age to do all our analysis.

- 2) We are seeing what sports have the shortest and longest career.

This, in general, can help people to decide which sports to select as a career if they are interested in being an Olympian and playing for the longest time in a sport.

We are showing an average of (max-min) values of age to do all our analysis.

3. Data Retrieval

For this project, we have utilized the following datasets in Kaggle to analyze historical Olympic data and make predictions for the upcoming Paris 2024 Olympics. We have uploaded them on Google drive, connected our notebook to Google drive and then loaded them:

- 1). 126 Years of Historical Olympic Dataset: This dataset provides a rich history of the Olympic Games from 1896 to 2022. We used it to analyze trends in athlete participation, country performance, and the evolution of different sports over time.
[\[https://www.kaggle.com/datasets/muhammadehsan02/126-years-of-historical-olympic-dataset\]](https://www.kaggle.com/datasets/muhammadehsan02/126-years-of-historical-olympic-dataset)

This datasets has multiples files :-

```
Athlete_Events_Details = pd.read_csv('Olympic_Athlete_Event_Details.csv')
Event_Results = pd.read_csv('Olympic_Event_Results.csv')
Athlete_Biography = pd.read_csv('Olympic_Athlete_Biography.csv')
Medal_Tally = pd.read_csv('Olympic_Medal_Tally_History.csv')
Games_Summary = pd.read_csv('Olympic_Games_Summary.csv')
Country_Profile = pd.read_csv('Olympic_Country_Profiles.csv')
```

- 2). World Population 1960-2018: This dataset contains population data from 1960 to 2018 for various countries. We utilized it to analyze population growth trends and their potential impact on global sports participation.

[\[https://www.kaggle.com/datasets/imdevskp/world-population-19602018?select=population_total_long.csv\]](https://www.kaggle.com/datasets/imdevskp/world-population-19602018?select=population_total_long.csv)

```
Population_Total = pd.read_csv('population_total_long.csv')
```

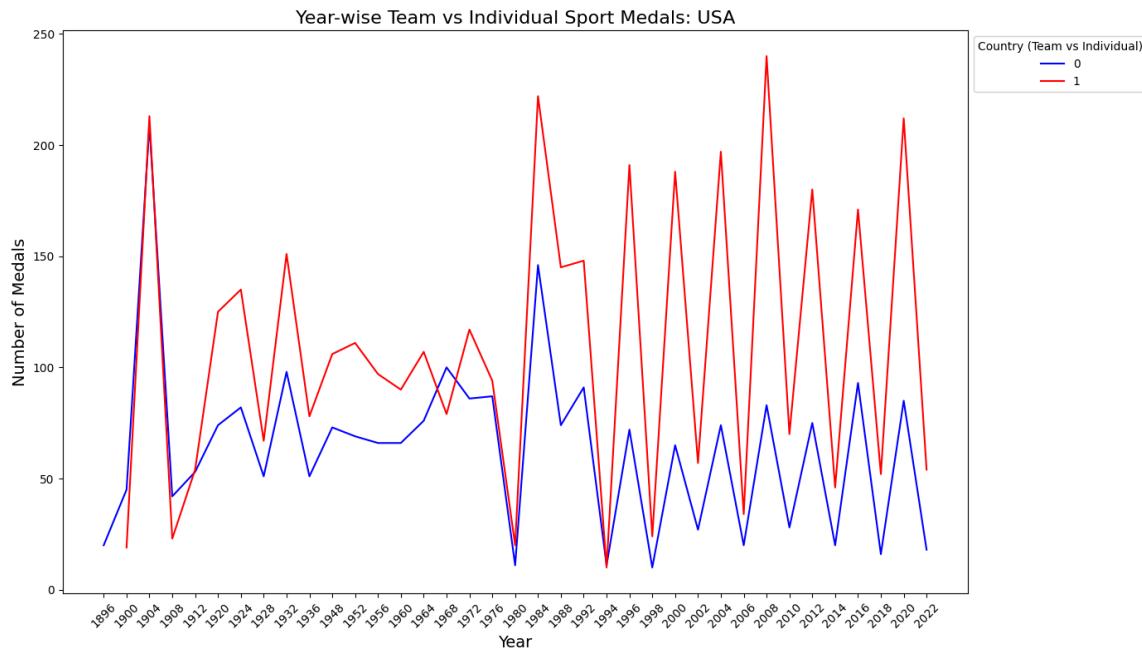
4. Data Cleaning

We have kept a more detailed explanation in the Ipython notebook.

- a. Deleted duplicate values.
- b. Removed trailing spaces.
- c. Converted strings to lower.
- d. Dropped columns that were not needed.
- e. One hot encoded categorical data.
- f. Extracted numerical data from texts and created new columns.
- g. Merged some reference tables to have complete info.
- h. Dropped columns where essential data was null.
- i. Normalized the data where values were too large or too small.
- j. Missing values have been imputed.

5. EDA & Hypothesis Sataakshi

EDA 1: Country-wise Comparison of Team vs. Individual Sport Medals Over the Years



Inference: The graph depicting year-wise medal counts for team versus individual sports for the USA, we see:

The graph indicates that the USA has consistently won more medals in team sports (shown in red) compared to individual sports (shown in blue) across the years. There is a strong participation in team-oriented events reflecting the country's competitive advantage.

In summary, the graph highlights a strong national performance in team sports for the USA, while also providing valuable information about the potential for growth in individual sports through strategic investments and focused training programs.

EDA 2: Sport-wise Trend of Women's Participation Over the Years

	sport	year	women_participation_count
0	artistic gymnastics	1900	1
1	golf	1900	12
2	tennis	1900	15
3	archery	1904	17
4	figure skating	1908	6
..
585	speed skating	2022	184
586	ice hockey	2022	235
587	alpine skiing	2022	276
588	biathlon	2022	348
589	cross country skiing	2022	445

[590 rows x 3 columns]

We start by filtering the data sports wise to aggregate the participation of women in different sports over the years followed by visualizing them.



We can infer from this graph that alpine skating has seen a significant increase in women's participation when compared from 1936 vs 2022. We can see that a sport that had less than 50 participants has over 300 participants at some point too. This clearly signs of the gender equality aspect. This number might be still very less when compared to the male participants but it is a start.

This EDA hence shows that all the sports will see even greater participation in the future too.

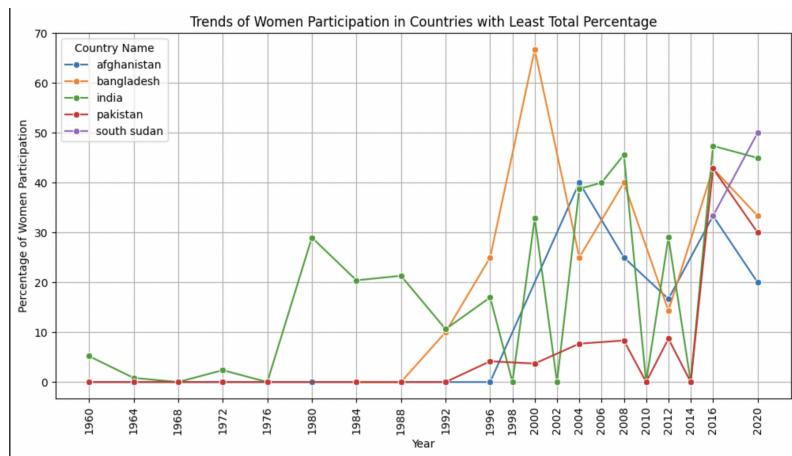
5. EDA and Hypothesis Geetansh

More detailed analysis is written in the Python notebook.

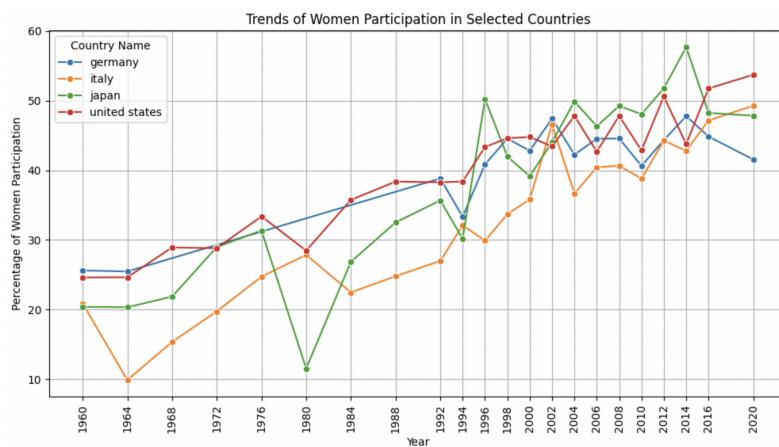
EDA 1 :- Percentage Women participation per country per olympic

Insights :- There are a lot of countries where women participation is significantly low as compared to men. These countries need to pay attention and target women participation in sports.

Below are the trends for countries where women participation is the least



Now, let's compare this with leading countries in sports.



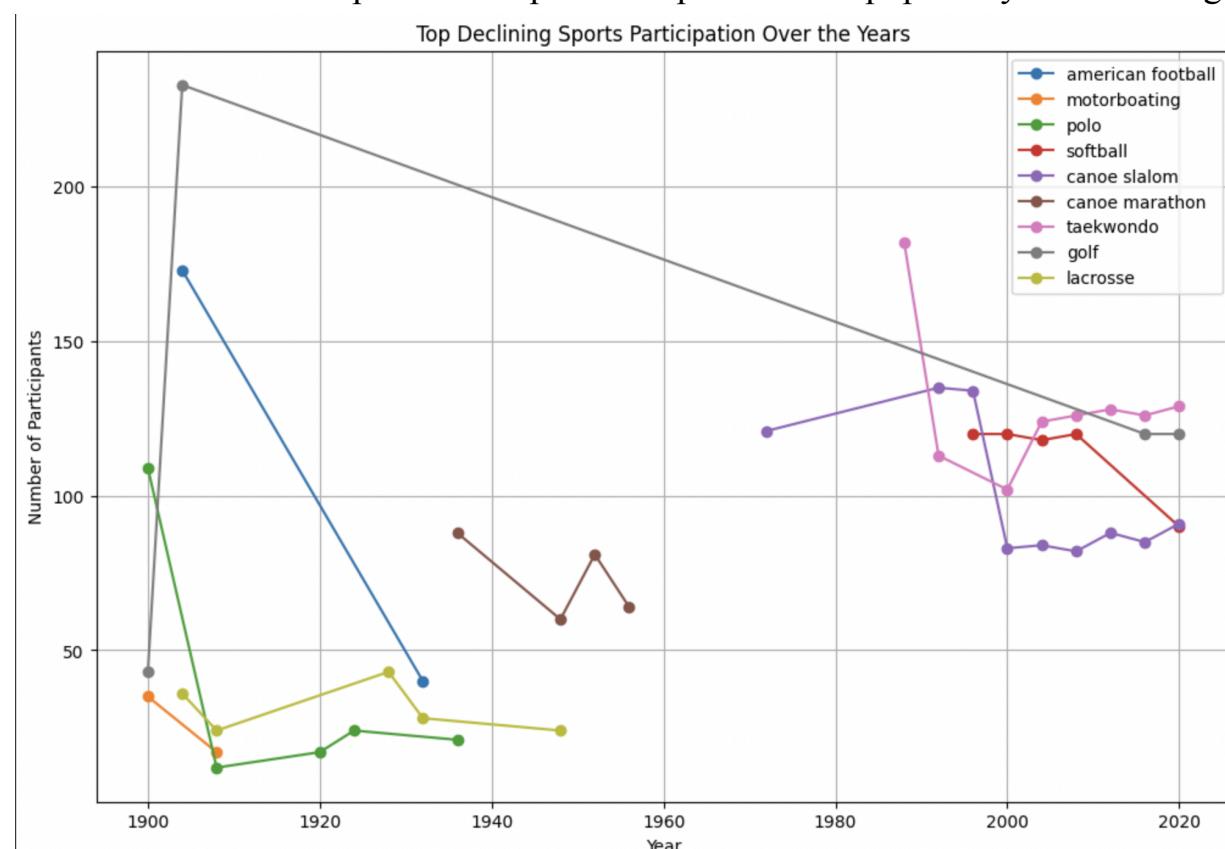
We clearly see the difference between the rise in women's participation in these two graphs.

This data can be passed downstream and used for predicting the future participation of women in the Olympics. Also, this is a significant indicator of women's empowerment in each country and can be very useful to highlight these issues.

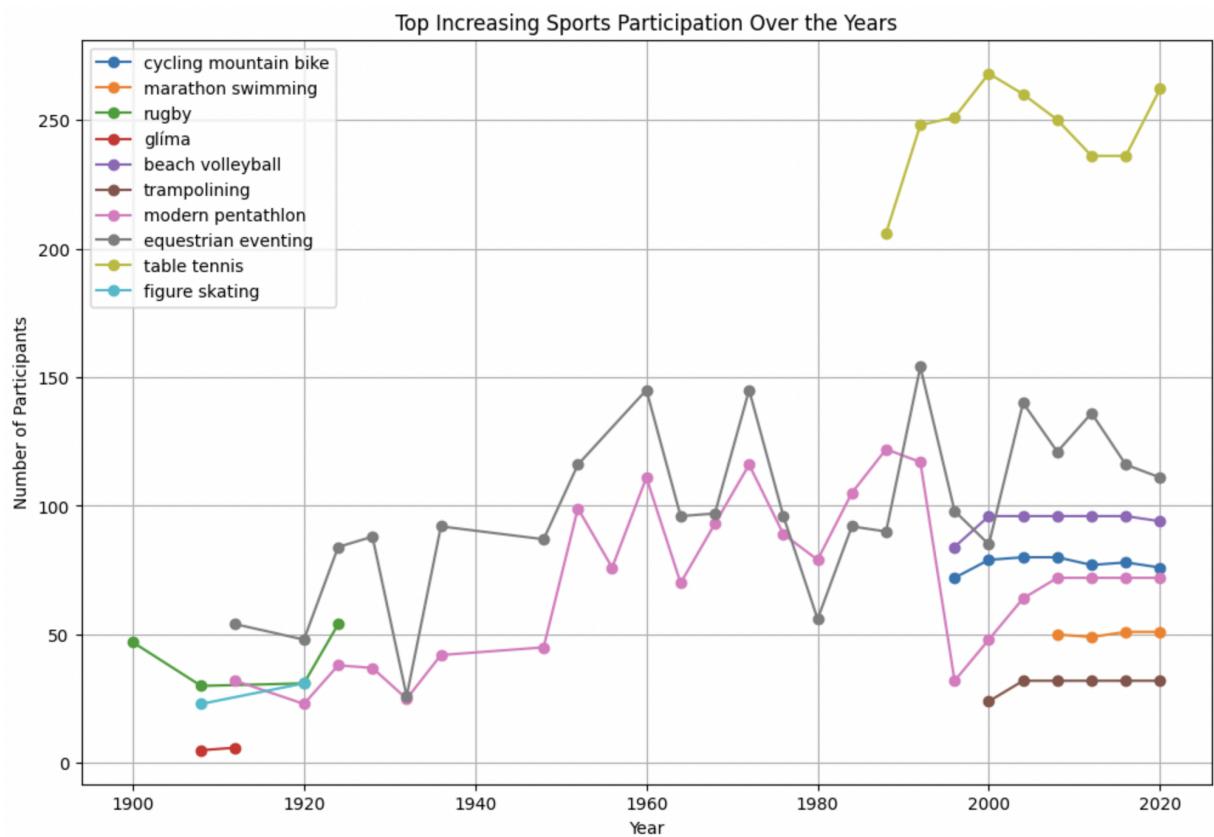
EDA 2:- Identifying Declining and On The Rise Sports

Insights:- There are a few sports whose trends are declining. In the past, the declining sports were removed from the Olympics altogether. The participation slowly decreases and then vanishes. Some sports have been gaining more popularity in the recent past.

Below we see extinct sports in the past and sports whose popularity is decreasing.



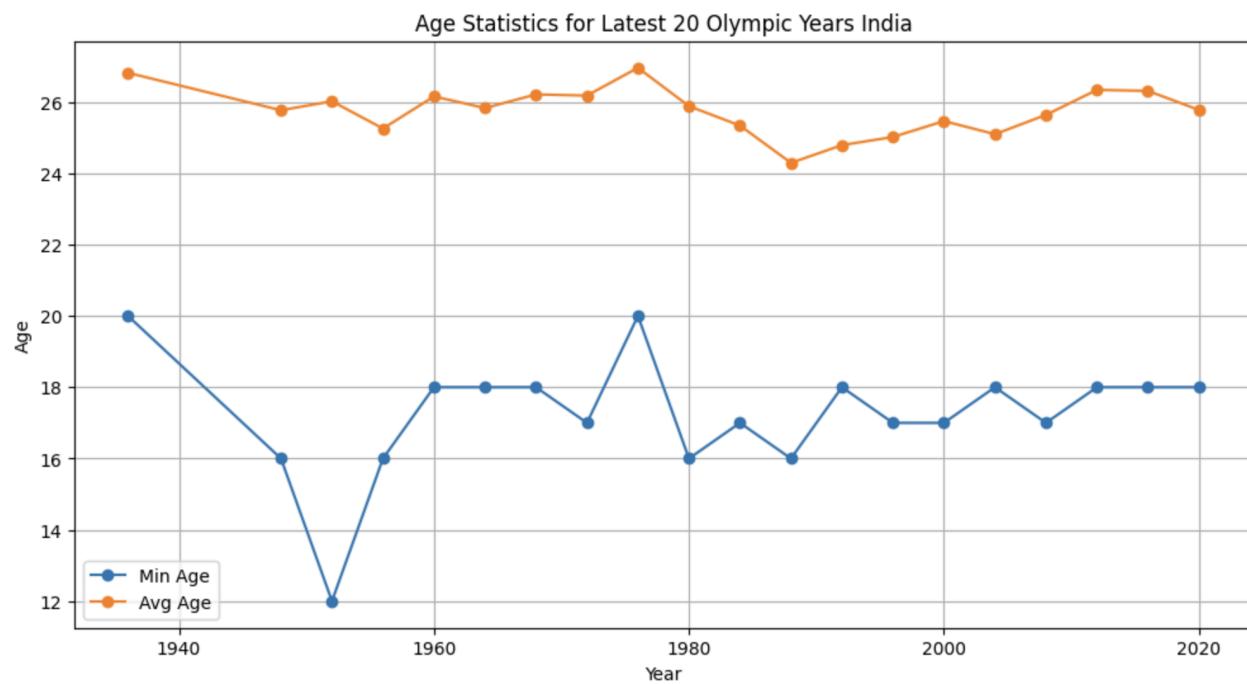
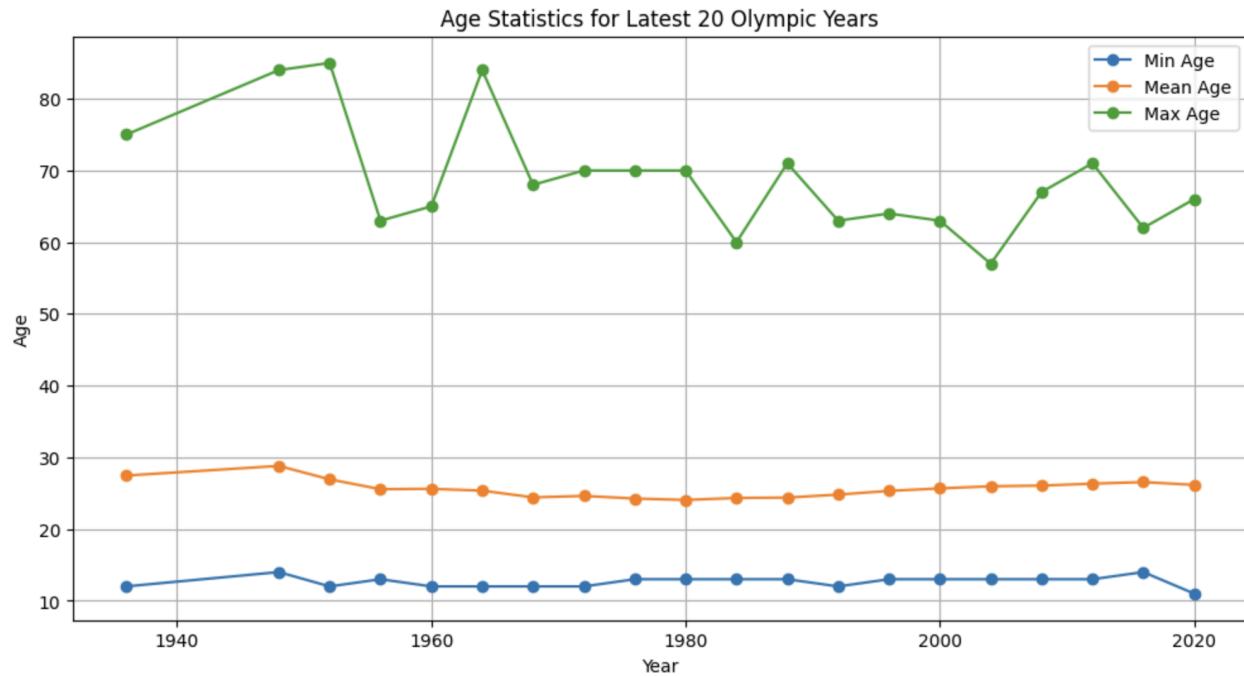
Below are the sports which are on the rise.



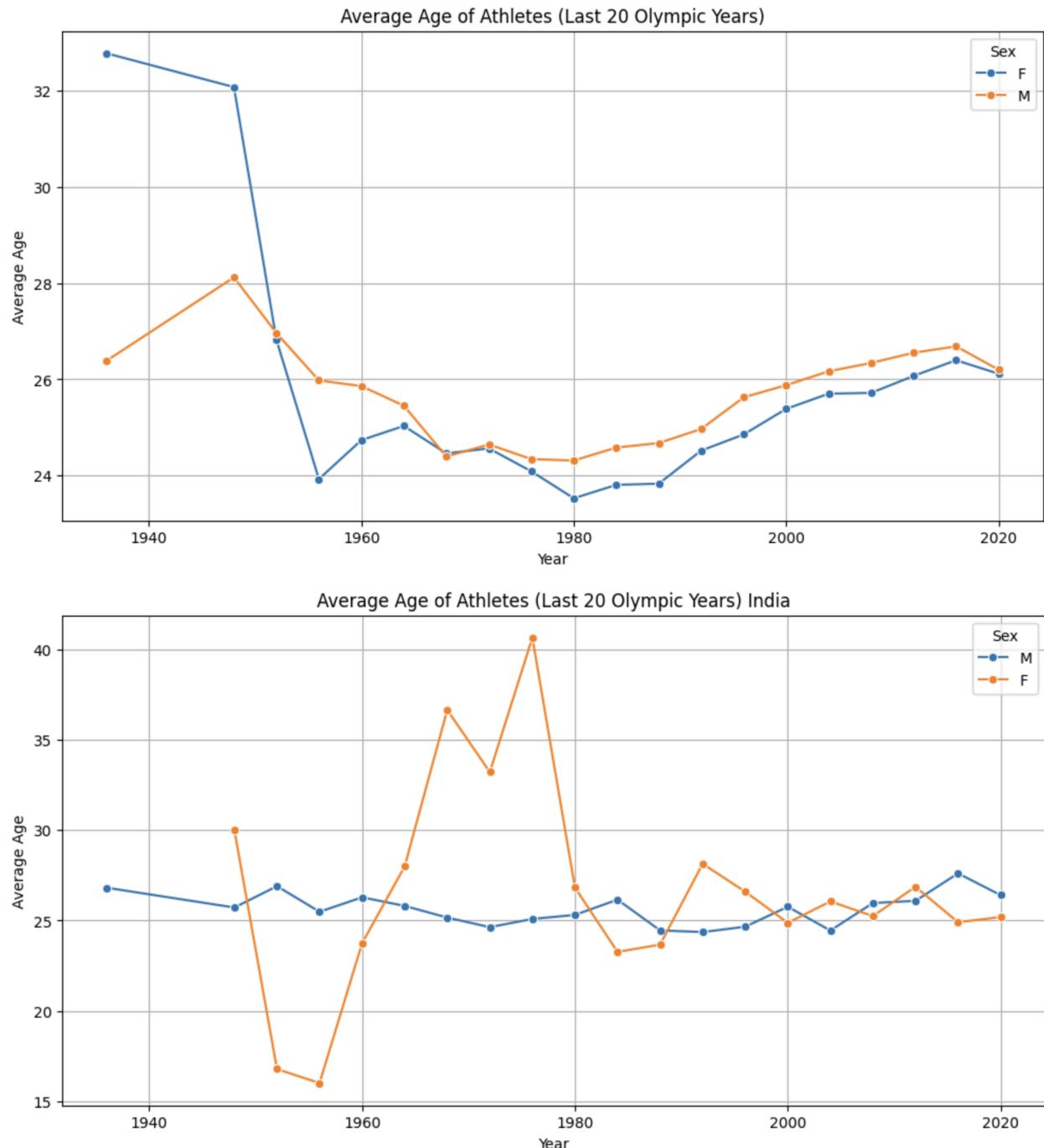
In the above two graphs, we see trends of increase and decline in sports. This data can be sent downstream, and we can predict which sport might get removed after how many years. We can also try to use this for marketing opportunities for the sports which are on the rise.

5. EDA and Hypothesis Rudraksh

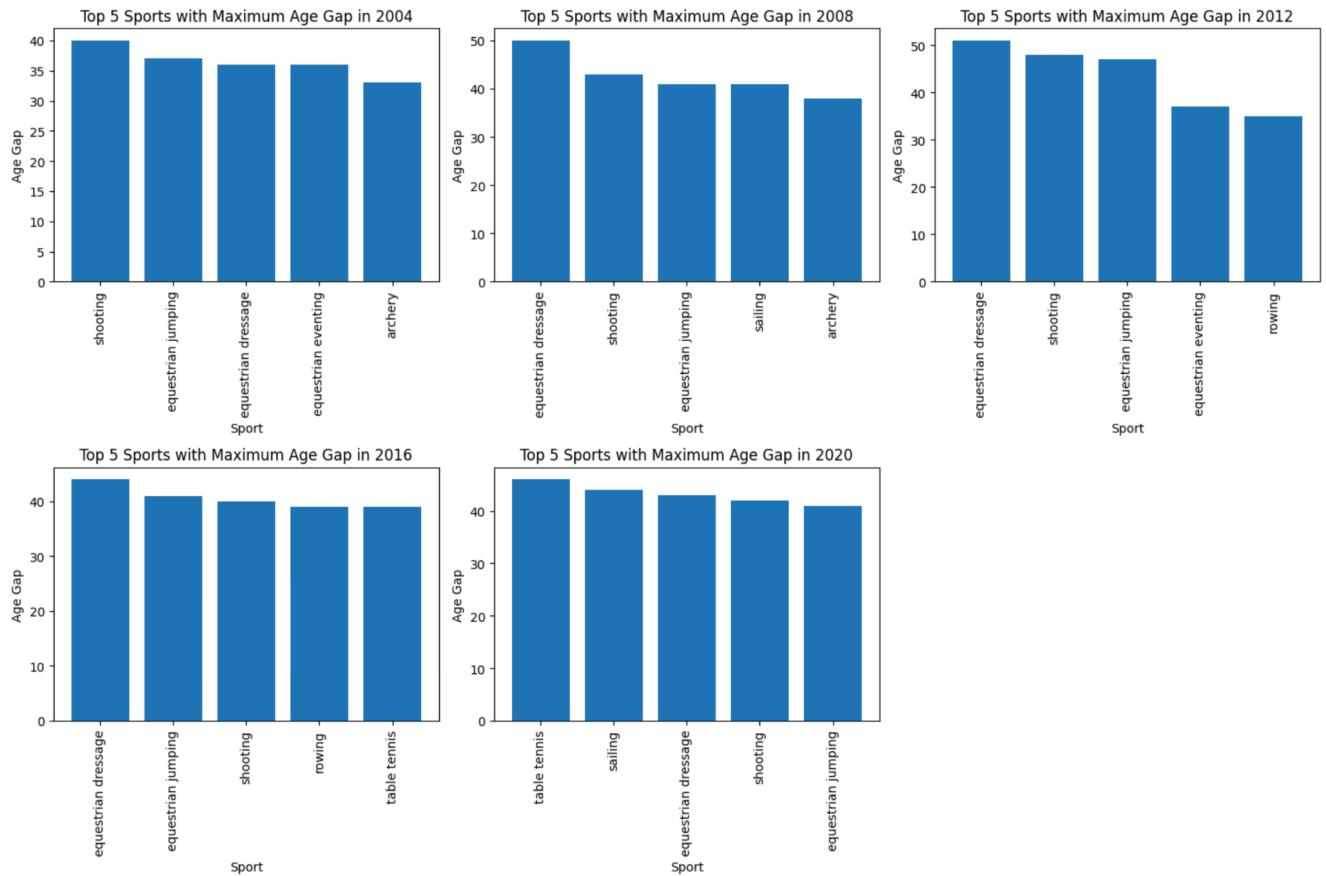
- 1) Through the graphs made below, we can see that there is not much change in the average age of the athletes in the world. Even the minimum and maximum ages being observed are more or less in the same band of ages.



There is a clear difference in the average age of male and female athletes in the world. For males, it is consistently greater than that of females. For India, it looks different there is no clear conclusion that can be made.



- 2) As can be seen, there are sports like shooting, equestrian jumping/dressage/eventing, archery, and table tennis, which have the longest careers spanning over 40 years. (12 Olympics)



As can be seen below graph, there are sports like boxing, triathlon, cycling, and gymnastics in which the career lasts for around 15 years (3-4 Olympics).

