Olympic Games Data Analysis

DATA ANALYSIS PROJECT FOR SPORTSSTATS

JULY 2025

MIA TROIANO

Proposal Description

This project will analyze data from 100+ years of the Olympic games. This analysis will include various trends on the data such as the teams with the most participating athletes, teams with the most medals, mean athlete age and more. This type of data may be interesting for those on the Olympic Committees, as it shows these trends over many years and can inform them on which countries are top performers. This data may also be used by sports commentators as it gives history of the games to discuss with the fans.

Questions to Answer

1. Which teams have the most representation and how does this change over time?

2. Which teams have the most medals and how does this change over time?

3. What are the mean demographics in each year of the games? How does the age and gender distribution change over time and by team?

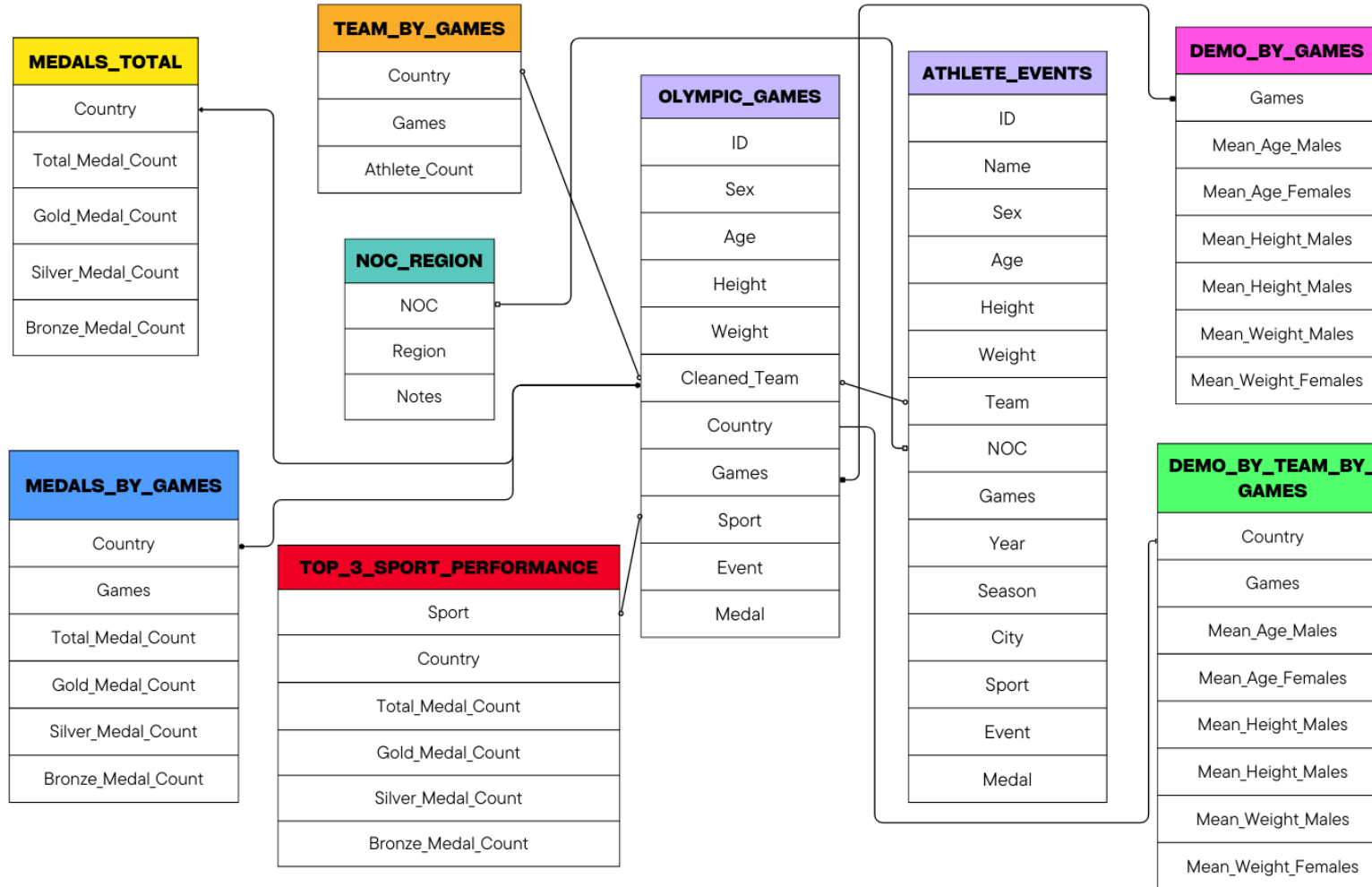4. Which teams are the top performers in each sport?

Initial Hypotheses

1. I think that European teams will have the most athlete representation per games in the early years of the games (early 1900s), however the United States, Russia and China will have more athlete representation in the later years and into the 2000s.

2. Similarly to before, I believe that European teams will have more medals in the early games with the Unites States and China having more medals in the later years. I think China will have the most medals overall within the years from this dataset.

3. I think the age demographics will show a gradual decline in age of the athletes over the years, with an increase in height and a decrease in weight as athlete body type norms have evolved. I think this will stay pretty consistent by team with no noticeable differences.

4. I think that teams such as Russia, Canada and Scandanavian countries will dominate the events in the winter sports with a higher total medal count. I think teams such as the United States and China will dominate in summer Olympic sports with a higher total medal count.

Data Analysis Approach

- I will be mostly examining the teams, games, medal and sport columns.

- I will be using aggregates to examine descriptive statistics of the demographics of the athletes.

- I will be using aggregates to count various variables, such as athletes and medals.

- I will utilize the chart feature on Snowflake and SourceTable to create a simple visualization with the year of the games on the Y-axis to demonstrate how a variable has changed over the years.

Entity Relationship Diagram

MEDALS_TOTAL

Country
Total_Medal_Count
Gold_Medal_Count
Silver_Medal_Count
Bronze_Medal_Count

TEAM_BY_GAMES

Country
Games
Athlete_Count

NOC_REGION

NOC
Region
Notes

OLYMPIC_GAMES

ID
Sex
Age
Height
Weight
Cleaned_Team
Country
Games
Sport
Event
Medal

ATHLETE_EVENTS

ID
Name
Sex
Age
Height
Weight
Team
NOC
Games
Year
Season
City
Sport
Event
Medal

DEMO_BY_GAMES

Games
Mean_Age_Males
Mean_Age_Females
Mean_Height_Males
Mean_Height_Males
Mean_Weight_Males
Mean_Weight_Females

MEDALS_BY_GAMES

Country
Games
Total_Medal_Count
Gold_Medal_Count
Silver_Medal_Count
Bronze_Medal_Count

TOP_3_SPORT_PERFORMANCE

Sport
Country
Total_Medal_Count
Gold_Medal_Count
Silver_Medal_Count
Bronze_Medal_Count

DEMO_BY_TEAM_BY_GAMES

| |
|---|
| Country |
| Games |
| Mean_Age_Males |
| Mean_Age_Females |
| Mean_Height_Males |
| Mean_Height_Males |
| Mean_Weight_Males |
| Mean_Weight_Females |

ERD

Technical Difficulties

- A difficulty that I ran into was that the 'team' column had to be extensively cleaned, as there were different formats to the same countries and I wanted to make sure all versions were accounted for in the analysis

- Other difficulties arose with multiple columns containing either 'null' or 'n/a', which needed to be screened out for any mathematical analysis

# Initial Findings – Team Representation

Top 20 Countries by Total Olympic Athlete Count
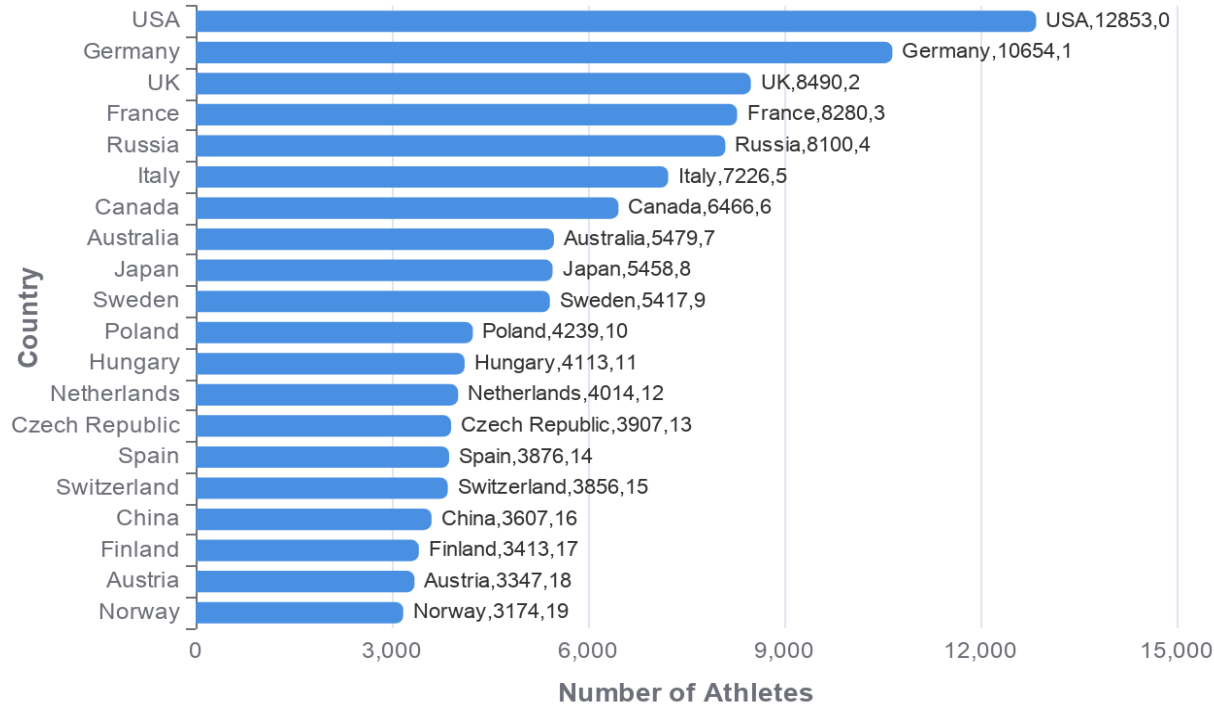
- USA, Germany and UK dominated representation in the Olympics over the span from 1896 to 2016, including 51 different games

- European countries dominate the top 20 list, with 14 in the top 20

Initial Findings – Team Representation

Olympic Athlete Count Trends Over Time (1896-2016)
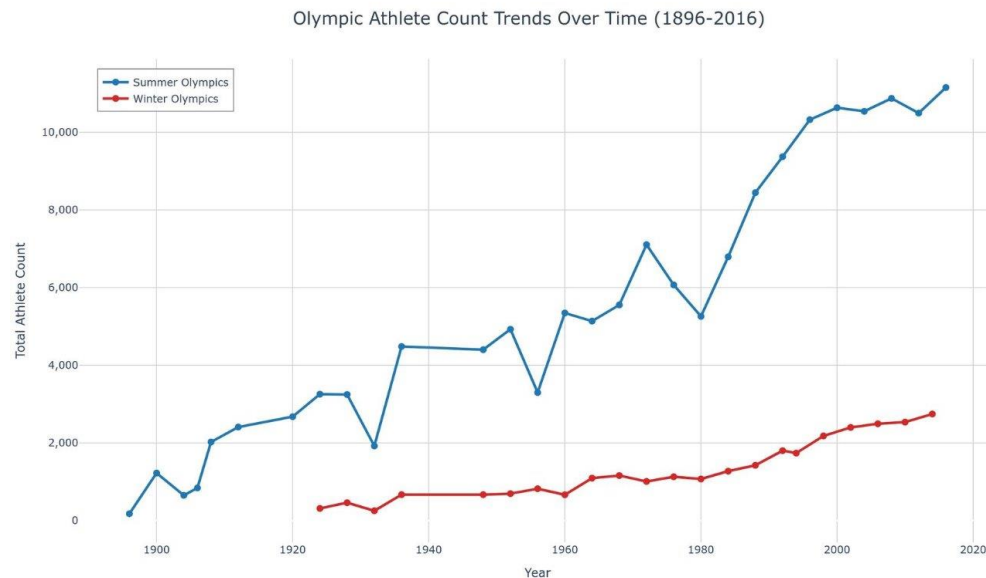
- Growth in representation is seen in both the Winter and Summer Olympics over the years, with a greater growth rate seen in the Summer games

- Major dips in participation can be seen around major historical events, such as the Great Depression, post-war time and boycotts

- The greatest representation was seen in the Summer 2016 Olympic games

Team Representation by Games
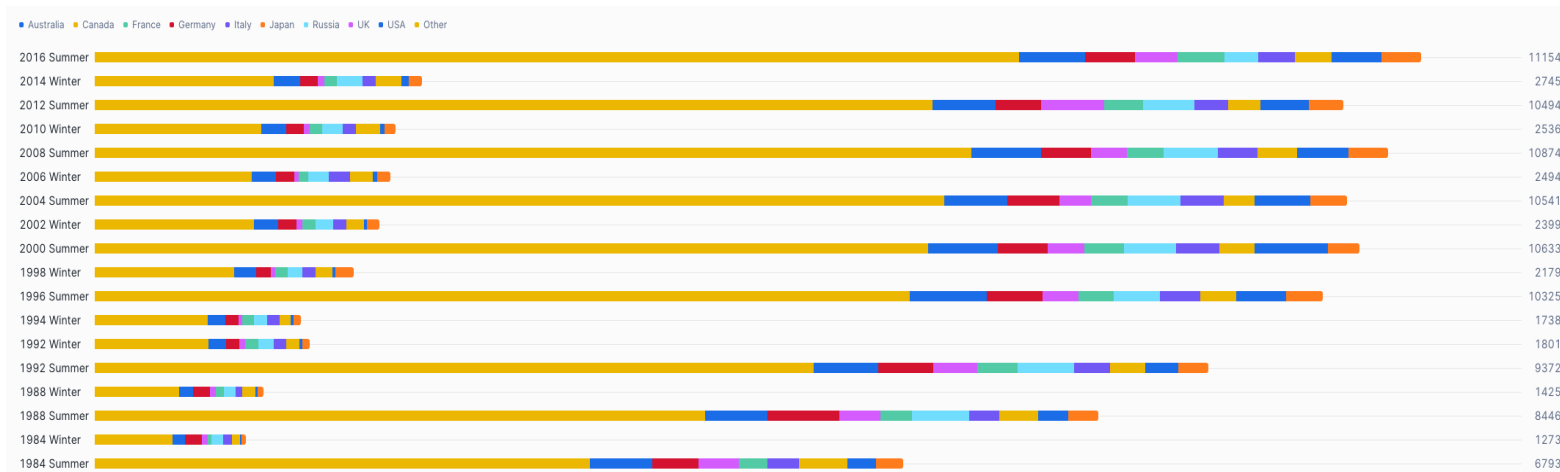
	COUNTRY	GAMES	ATHLETE_COUNT
1	UK	1908 Summer	735
2	Germany	1972 Summer	721
3	France	1900 Summer	720
4	USA	1996 Summer	648
5	Australia	2000 Summer	617
6	Germany	1988 Summer	606
7	China	2008 Summer	599
8	USA	2008 Summer	588
9	USA	2000 Summer	586
10	Germany	1976 Summer	557

•The UK had the greatest athlete representation to date, sending 735 athletes to the 1908 Summer Olympics

Total Medal Count

Total Olympic Medals by Country (Top 25 Countries)
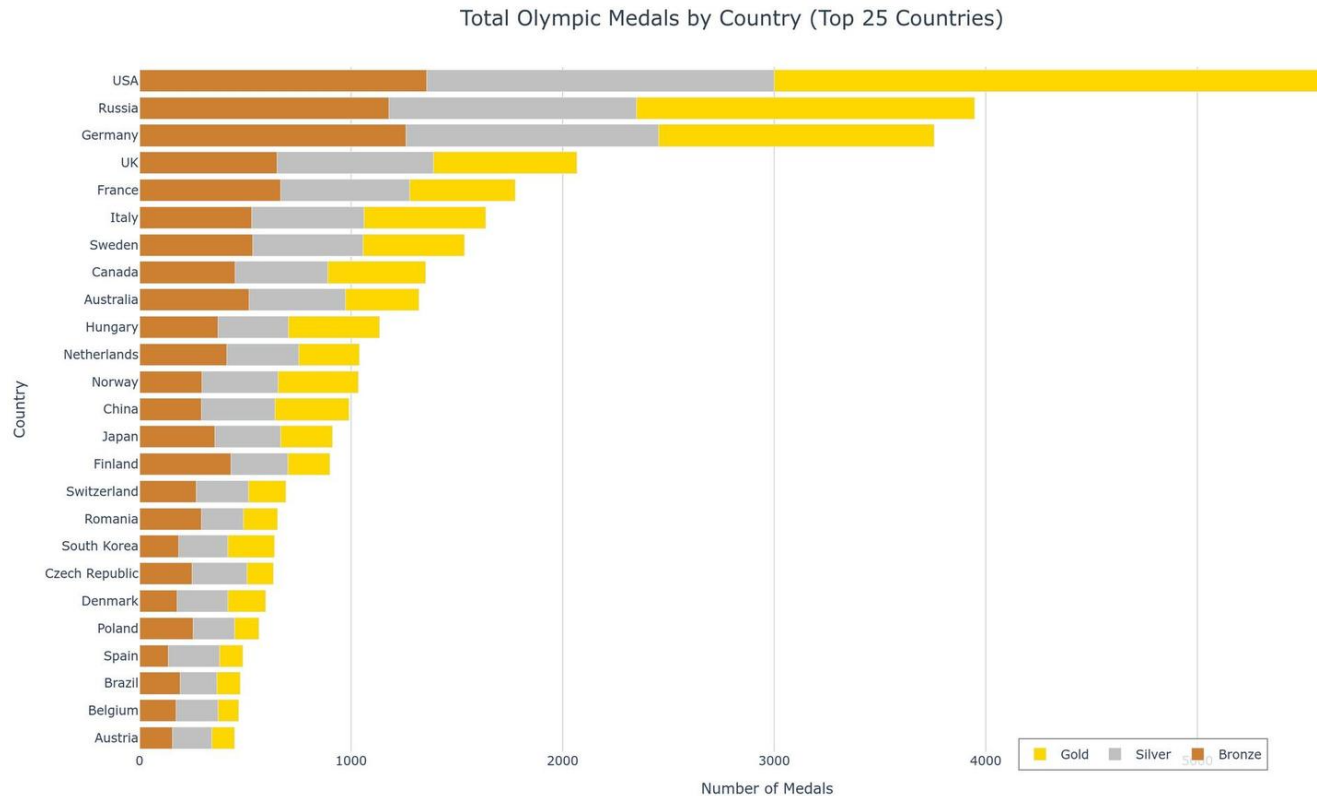
- USA dominates the Olympic Medal rankings, with 5,637 total medals won

- European countries make up 15 of the top 25 countries with the most medals won

- It should be noted that this is total medals, not broken down by Winter and Summer Games

Medal Count By Games

COUNTRY	GAMES	TOTAL_MEDAL_COUNT	GOLD_MEDAL_COUNT	SILVER_MEDAL_COUNT	BRONZE_MEDAL_COUNT
Russia	1980 Summer	442	187	129	126
USA	1904 Summer	394	128	141	125
UK	1908 Summer	368	147	131	90
USA	1984 Summer	352	186	116	50
USA	2008 Summer	317	127	110	80
Russia	1988 Summer	300	134	67	99
Germany	1988 Summer	296	111	91	94
Russia	1976 Summer	286	114	95	77
Germany	1976 Summer	273	123	76	74
Germany	1980 Summer	264	115	88	61
USA	2016 Summer	264	139	54	71
USA	2004 Summer	263	117	75	71
USA	1996 Summer	259	159	48	52
Germany	1972 Summer	253	74	83	96
USA	2012 Summer	248	145	57	46
USA	2000 Summer	242	130	61	51
France	1900 Summer	235	52	101	82
USA	1992 Summer	224	89	50	85
Germany	1936 Summer	224	93	70	61
Russia	1992 Summer	220	92	61	67

- Despite having the second most overall medals, Russia has won the most medals, along with most gold medals, at a single games: the 1980 Summer Olympics
- Russia, USA, UK and Germany make up nearly all of the top 20 of the most medals won in a single games

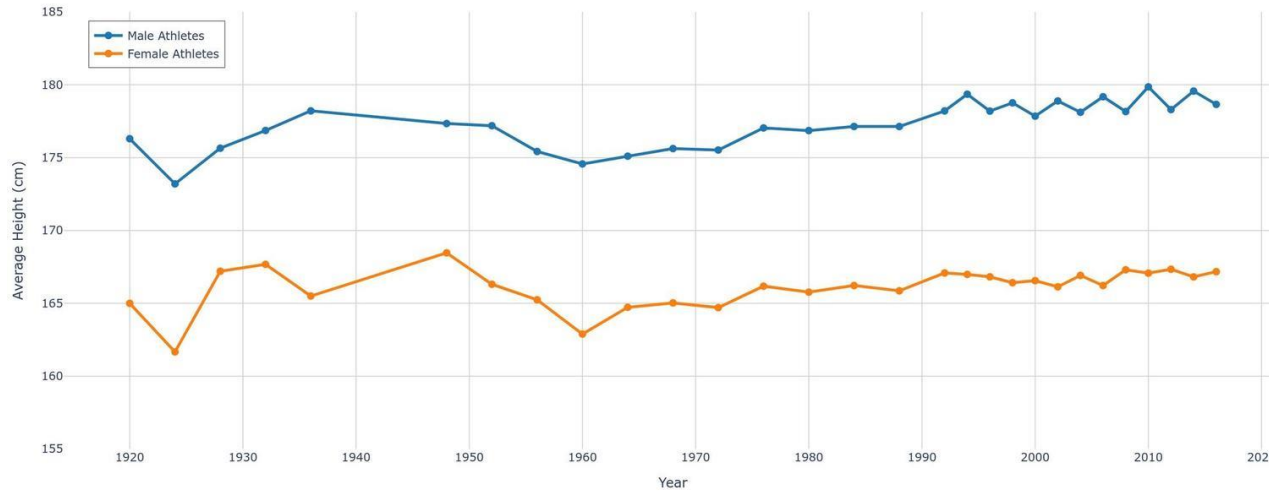Olympic Athletes Age Analysis Over Time (1900-2016)
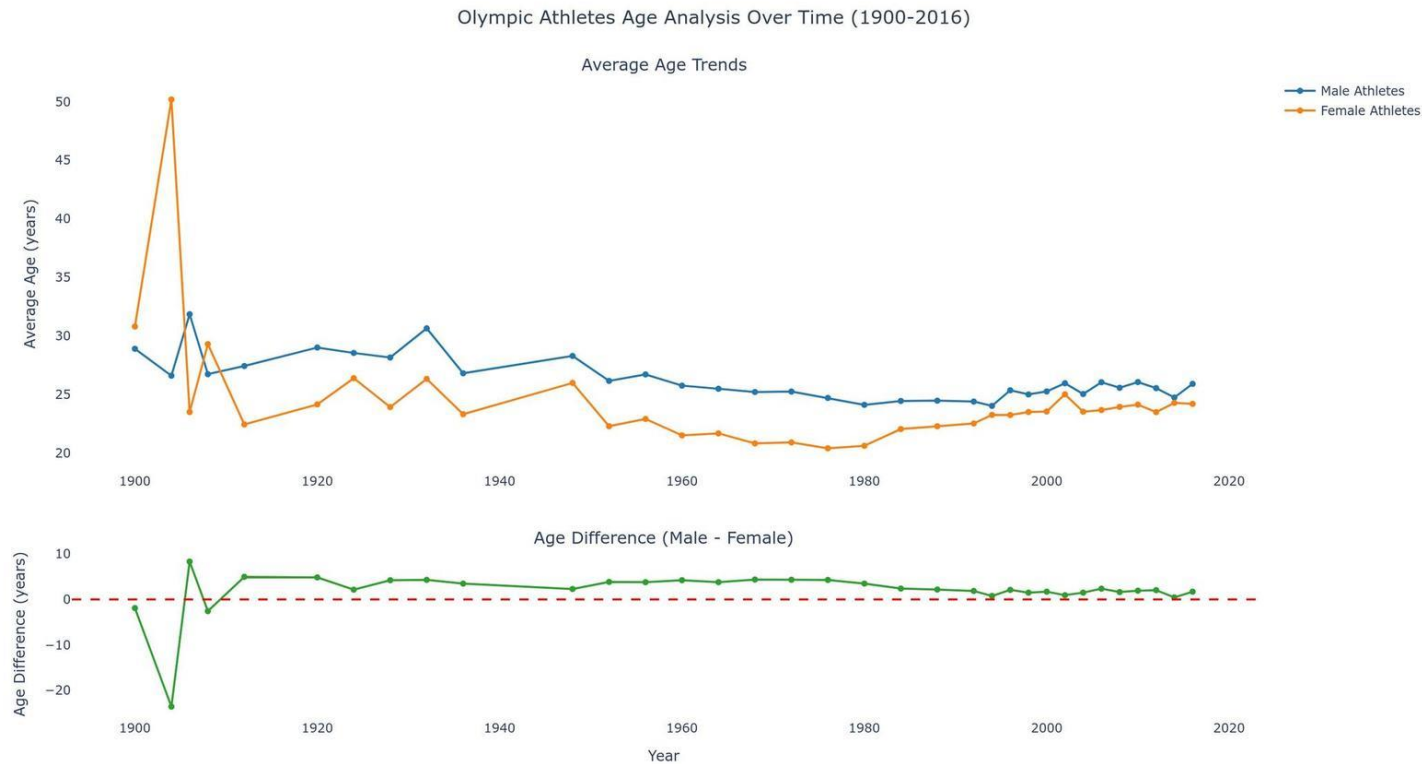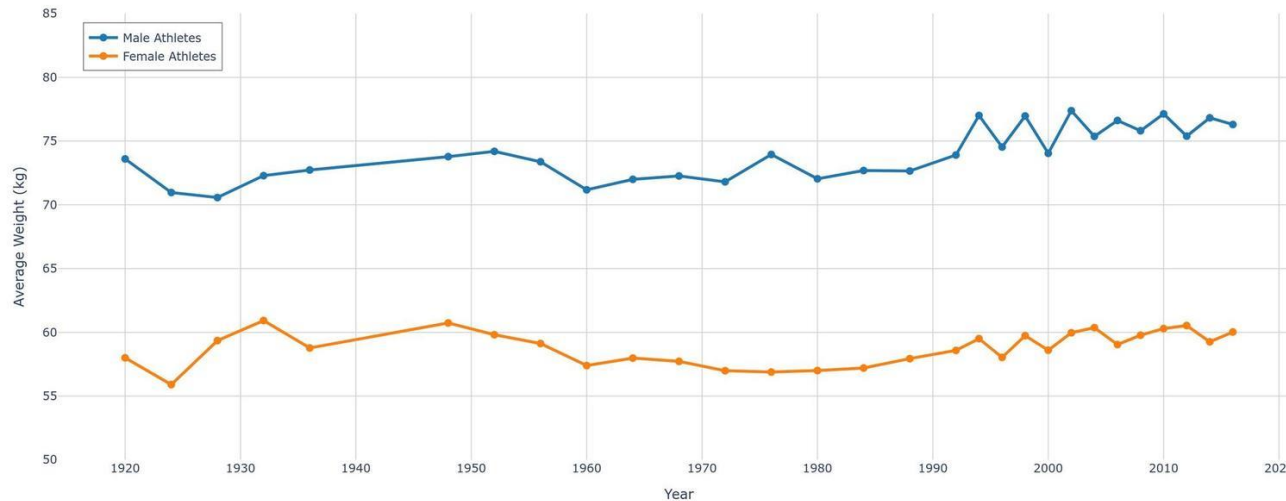
Athlete Demographics - Age

- Average male athlete age is slightly older (26.3 years) than female athletes (24.4)

- Age differences are converging for male and female athletes, showing ages are becoming more similar over time

- Male athletes are getting slightly younger while female athletes are getting slightly older

# Athlete Demographics - Height and Weight

- Male athletes tend to be an average of 11.2cm taller and 15.2kg heavier than female athletes

- Both genders are getting taller and heavier over time, with height increasing more gradual than weight

Top Performers By Sporting Event

Olympic Medals Heatmap: Top 10 Sports by Country
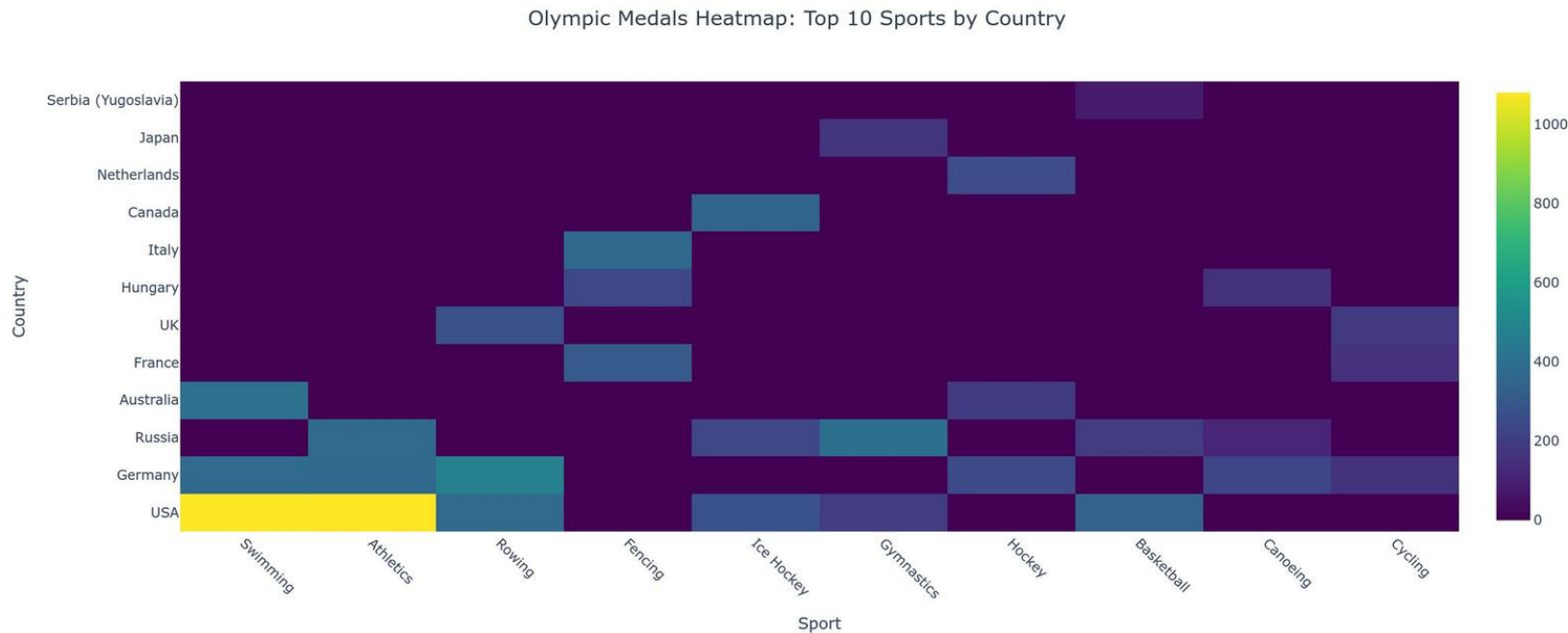
- USA dominates in swimming, athletics and basketball

- Germany dominates in rowing and canoeing

- Italy dominates in fencing

- Canada dominates in ice hockey

- Russia dominates in gymnastics

Top Performers By Sporting Event
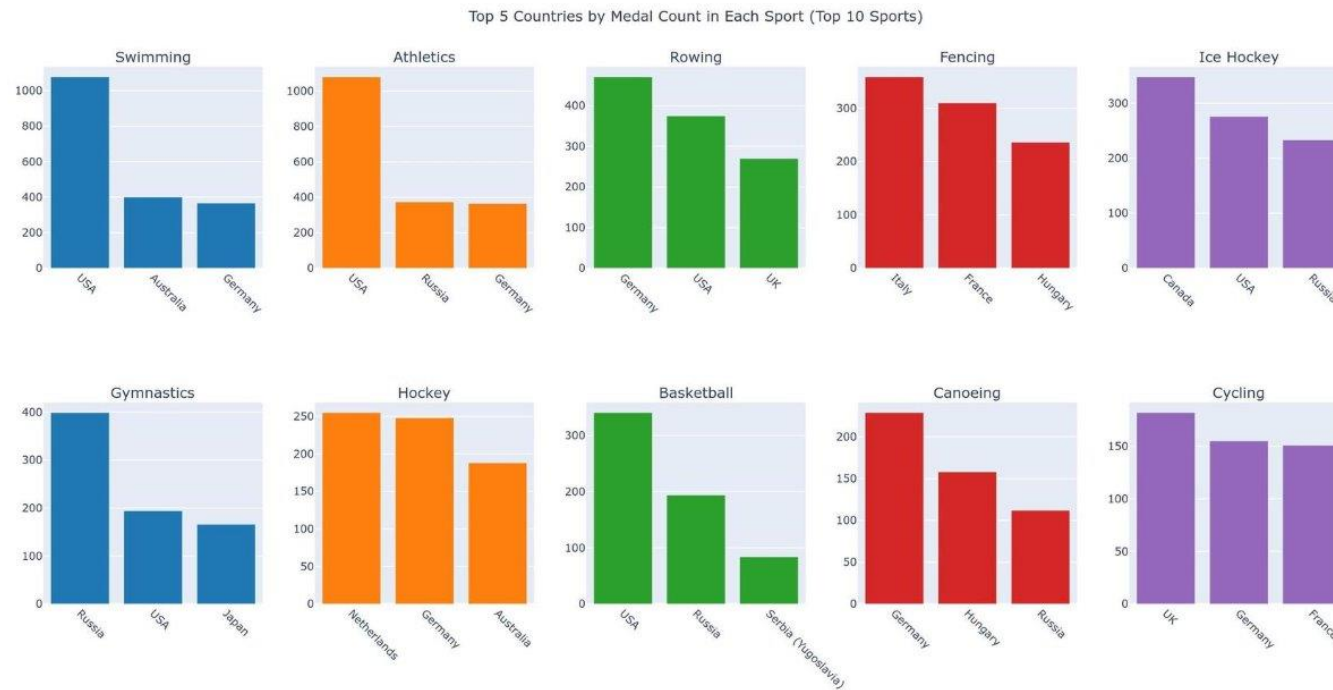
- USA and Germany are amongst the top 3 medal count for 6 of the 10 top sports

- Russia is amongst the top 3 medal count for 5 of the 10 top sports

New Metric – Body Mass Index (BMI)

Average BMI of Olympic Athletes Over Time (1896-2016)
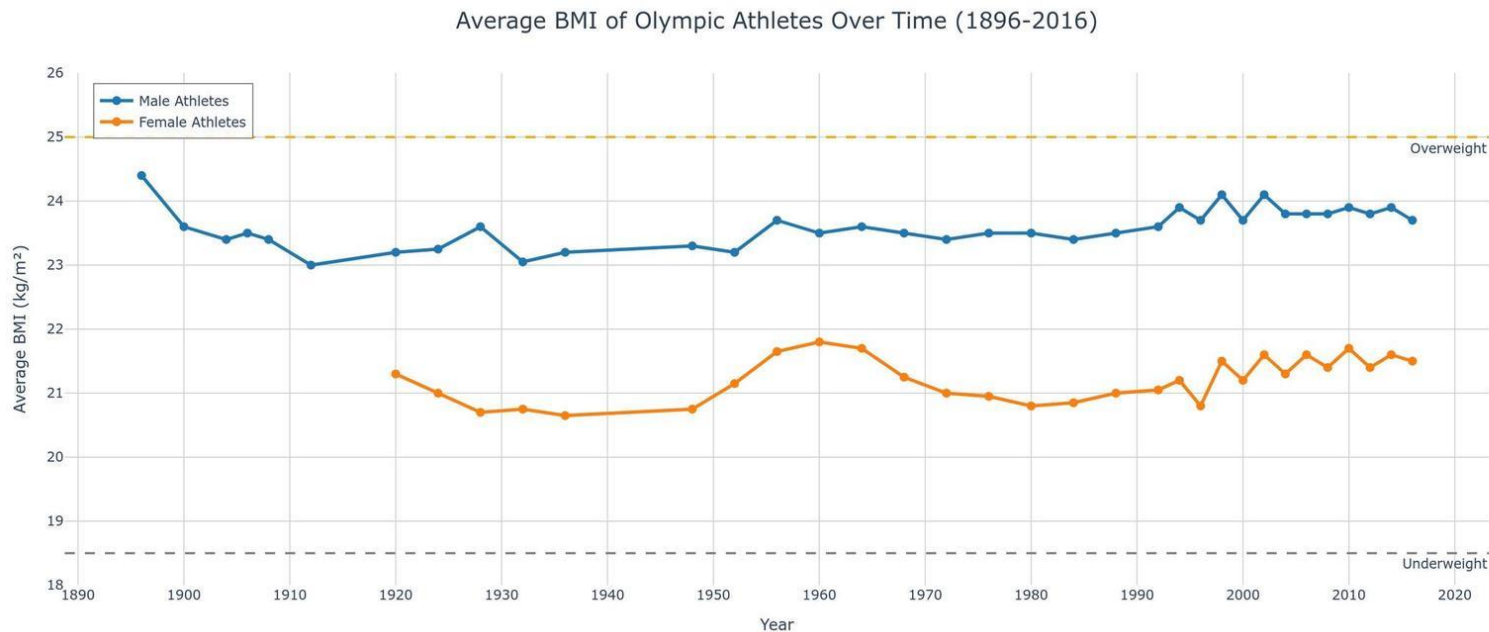
- Average BMI was used to examine population level body composition, with the assumption that BMI would decrease over the years as a leaner body composition was preferable for athletes

- The trend shows that there isn't much change in the average BMI over the years of the games

- Males tend to have a higher BMI than females, with both average BMIs falling well within healthy normal standards

Final Findings

1. The United States, Germany, the United Kingdom, and France were among the leaders in total representation at the Olympics, as well as among the leaders in representation at a single Olympic games

2. Despite being 5th in total representation at the Olympics, Russia has the second most medals won; second to the United States. My hypothesis that China would have the most medals won over the years was incorrect, as they ranked 13th in most medals won

3. Average athlete demographics have changed throughout the years, with age for males getting younger and age for females getting older. Height and weight for both males and females have been increasing, which disproves my initial hypothesis

4. The United States, Germany and Russia are consistently good at the top sporting events in the Olympics, whether it is the Summer or Winter sporting events

Further Directions

- This analysis focused on exploratory data analysis with descriptive statistics

- Further analysis can focus on trends to help predict future athlete participation, possible medal placement, athlete demographics and more, which can be used for a variety of planning or betting opportunities
  - Sports betting companies can use these predictive trends as a guideline for how to wager their bets for the Olympics

- Analysis can also look into relationships such as a country's success in the games when they are the host nation, which can be used to drive countries to place bids to host the Olympics