Exploring the Relationship Between Age, Height, Weight and Olympic Performance

Eduardo Garcia

Introduction

For this project I will be utilizing the "Olympic Dataset - 120 Years of Data" which contains the records of athletes that participated in the modern Olympics up to 2016, including details such as the year, sport, age, height, weight and medal won (if applicable).

I plan to analyze the impact of age, height and weight on Olympic performance, as well as how these impacts vary between males and females and across the different types of events. The results of this study will be beneficial for personal trainers, fitness coaches, and health professionals.

Questions and Hypotheses

Initial Questions:

- How does age impact performance?
- How do height, weight, and BMI affect performance?
- How do these effects differ between men and women and across different events?

Initial Hypotheses:

- Athletes in their mid-to-late 20s will display superior performance.
- The BMI of the top athletes will generally fall within the "healthy" range (around 20-23).
- There will be a moderate to strong positive correlation between height and weight.
- Women will generally have lower height, weight, and BMI compared to men.
- Sports that primarily rely on strength (such as weightlifting and throwing) will favor athletes with higher height, weight, and BMI.
- Sports that primarily require speed and agility (like track and gymnastics) will favor athletes with lower height, weight, and BMI.

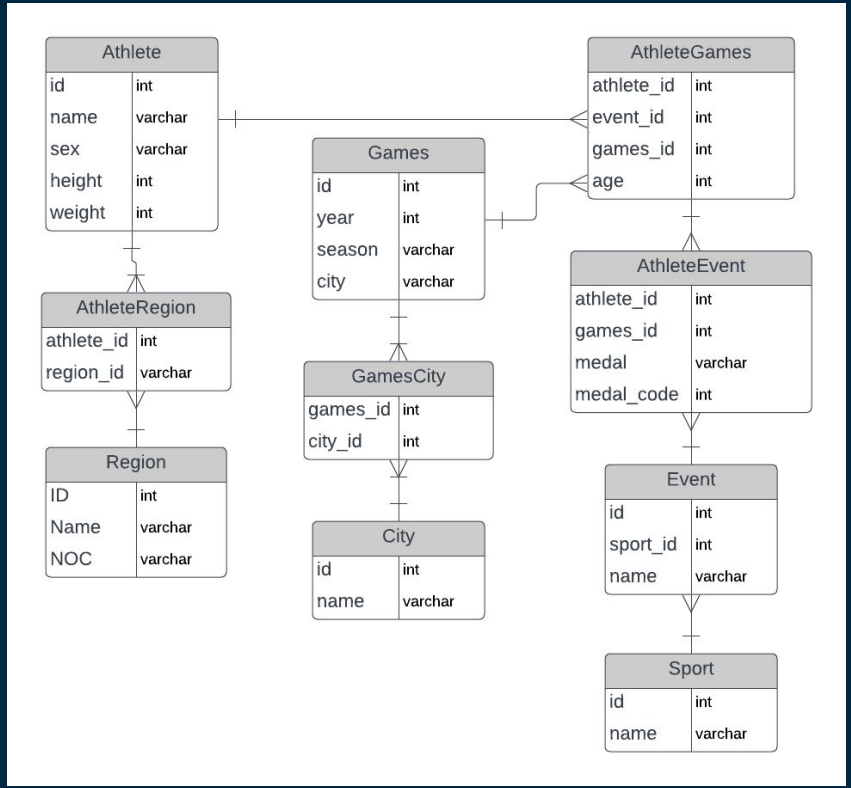
Format of Data

The data came in two files:

1. athlete_events.csv
    a. Categorical features: Name, Sex, Team, NOC (National Olympic Committee), Games, Season, City, Sport, Event and Medal.
        i. Medal = Gold, Silver, Bronze or no value if the athlete did not win a medal.
    b. Numerical Features: Age, Height, Weigh and Year.
        i. Age values change with respect to year, but Height and Weight values are the same for all athlete's records.
2. noc_regions.csv
    a. Categorical features: NOC, region, notes.
        i. 'NOC' and 'notes' reflect territories as they were defined at the time an athlete competed at an Olympics and 'region' indicates what territories these areas belong to as of 2016

The features that are applicable for my analysis are: Sex, Age, Height, Weight and Medal.

Entity-Relationship Diagram



Cleaning Data and Introducing Fresh Metrics

Data Cleaning

- I removed the rows with missing age, height and weight values

New Metrics

- I introduced BMI which is $BMI = \frac{Weight}{Height^2}$
- Medal Code: Gold = 1, Silver = 2, Bronze =3, No Medal = 4
- Categorized Sports based on objectives

Approaches and Technical Hurdles

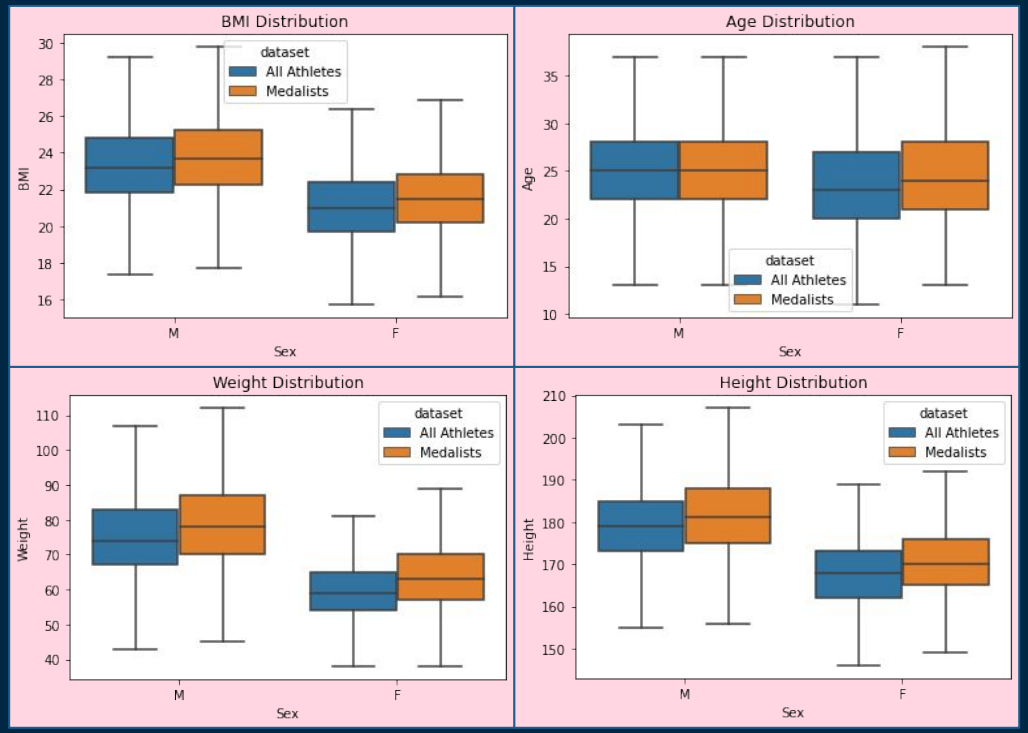
The data frames used for the deep analysis contain the following sport properties, grouped by sport category:

- Across all athletes and medalists, average metrics are compared, and the difference between these averages is reported.
- Additionally, average metrics are compared between men and women, and the difference between these averages is reported.
- Finally, average height, actual average weight, and predicted average weight (based on height using linear regression) are compared, and the difference between actual and predicted weight is reported.

To address the challenges, the following intermediate steps were taken

- Records were filtered to include only medalists or athletes of either men or women, as needed for the comparison.
- Metrics were grouped by sport for each of the two subsets of athletes.
    - The final data frames were created by joining the two subsets on the "sport" column.

Distributions of Olympic Athletes and Medalist: A Comparative Analysis



Men

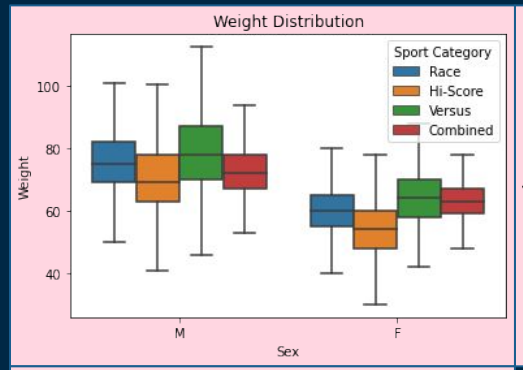
	dataset	AVG(Age)	AVG(Height)	AVG(Weight)	AVG(BMI)
0	All Athletes	25.667887	178.967548	75.795053	23.545946
1	Medalists	25.863130	181.353775	79.252080	23.964186

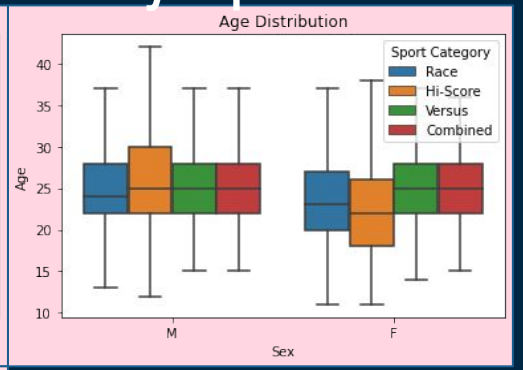
Women

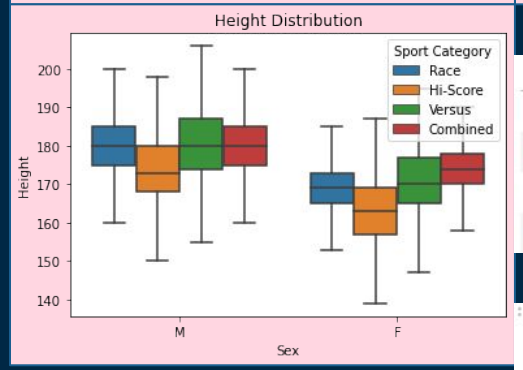
	dataset	AVG(Age)	AVG(Height)	AVG(Weight)	AVG(BMI)
0	All Athletes	23.781552	167.867261	60.027223	21.193275
1	Medalists	24.596152	170.538238	63.222663	21.631776

- The majority of male athletes fall within the age range of 22-28 years old, with a height between 175-190 cm, a weight ranging from 67-87 kg, and a BMI falling within 22-25.
- Similarly, female athletes typically range from 20-27 years old, with a height varying between 160-172 cm, a weight ranging from 55-70 kg, and a BMI falling within 20-23.
- It is worth noting that, with the exception of age, all other distributions appear slightly elevated among medalists compared to the overall population of athletes.

Distribution of Olympic Athletes by Sport Category







Overall, the distributions of height and weight appear to be lowest for high-score sports and highest for versus sports. This may be due to the following factors:

- High-score sports: Certain high-score events, such as gymnastics and figure skating, require flexibility, while others, such as archery and shooting, are more about strategy and accuracy than athletic ability.
- Races: In races, competitors are "active" for the entire race, requiring a constant supply of strength and/or endurance.
- Versus sports: Many versus sports involve physical struggle with an opponent, such as in fighting sports and rugby. Additionally, versus sports often require quick reactions based on the opponent's actions.

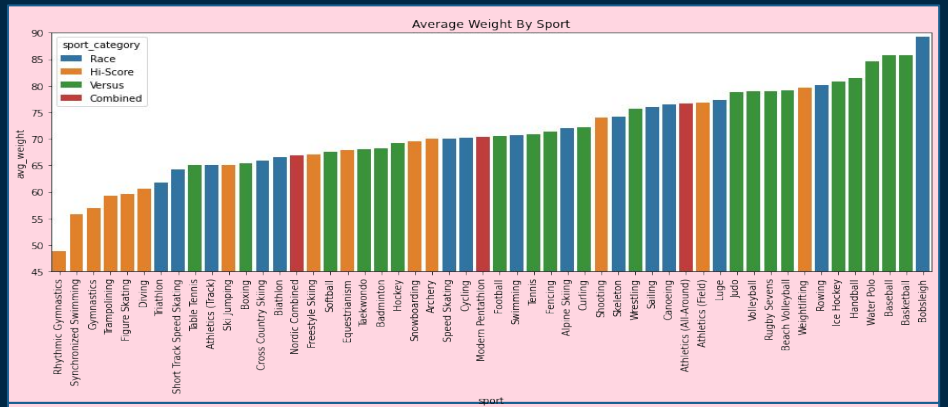
Men

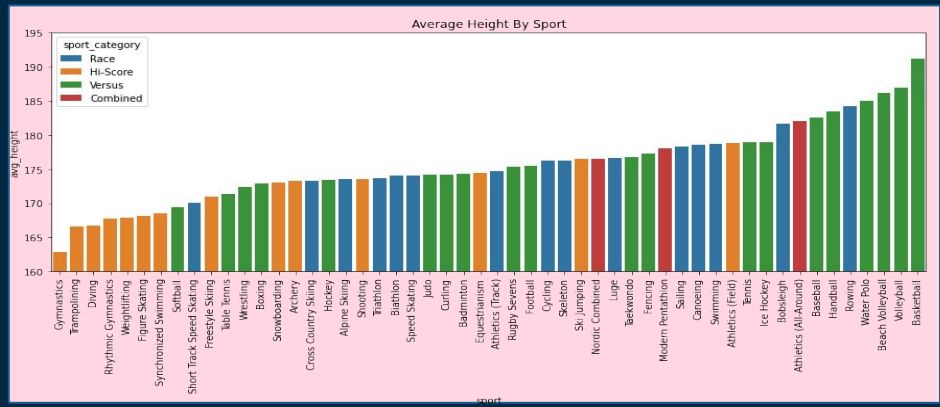
	Sport Category	AVG(Age)	AVG(Height)	AVG(Weight)	AVG(BMI)
0	Combined	25.283363	179.886865	72.973380	22.484926
1	Hi-Score	26.961565	174.303191	72.555794	23.722686
2	Race	24.991181	180.527449	75.855142	23.208377
3	Versus	25.612952	180.582244	78.774666	23.996591

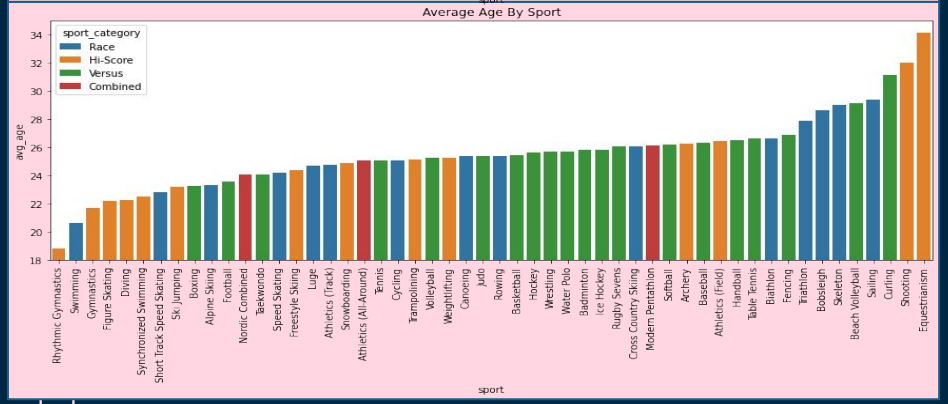
Women

	Sport Category	AVG(Age)	AVG(Height)	AVG(Weight)	AVG(BMI)
0	Combined	25.206597	173.342014	62.713542	20.843834
1	Hi-Score	23.077226	162.981613	55.549977	20.729912
2	Race	23.454155	169.153359	60.542570	21.116008
3	Versus	25.393187	171.408947	64.821920	22.001204

Average Athletic Metrics by Sport

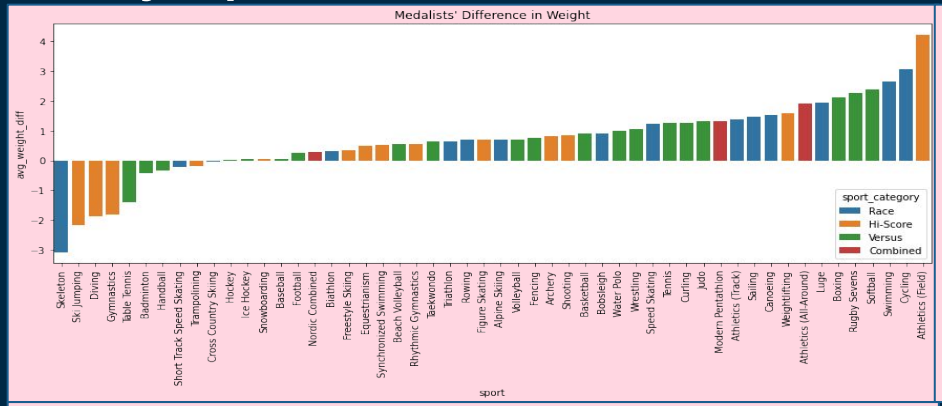


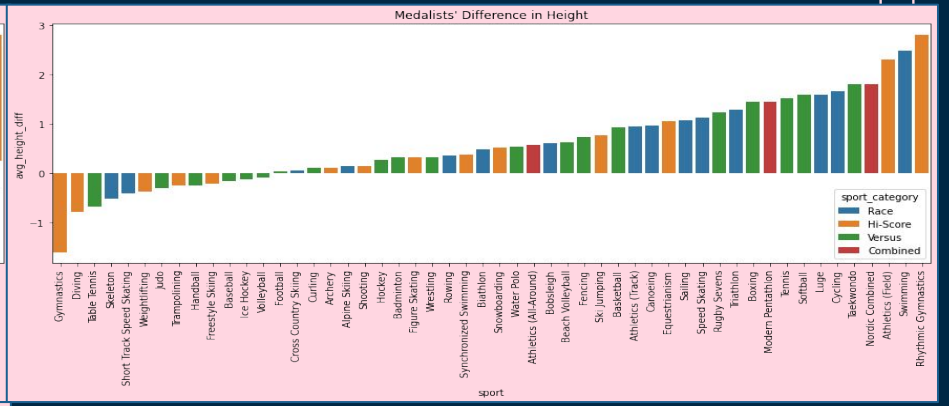


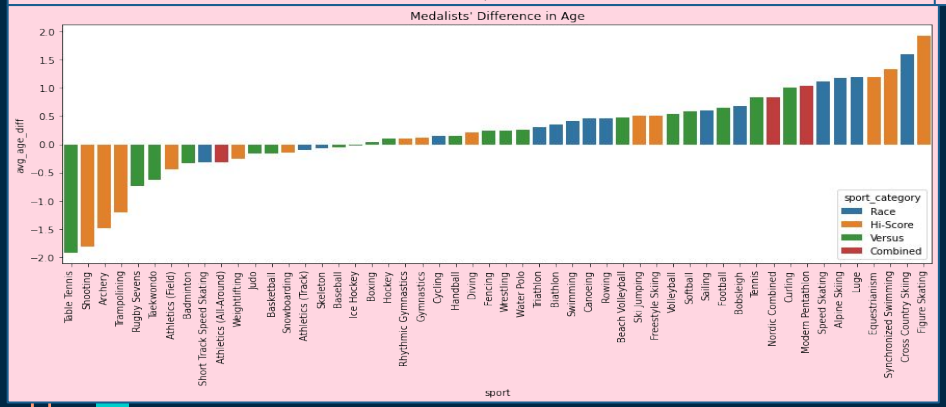


- Most of the sports with the tallest and heaviest athletes are team versus sports, where height is a key advantage.
- Sports with the shortest and lightest athletes include many hi-score sports that demand flexibility.
- Equestrian, shooting, curling, and sailing athletes are significantly older than the rest.
- Bobsleigh athletes are the heaviest, while weightlifting athletes are among the lightest.

Athletic Differences Between Medalists and All Athletes by Sport

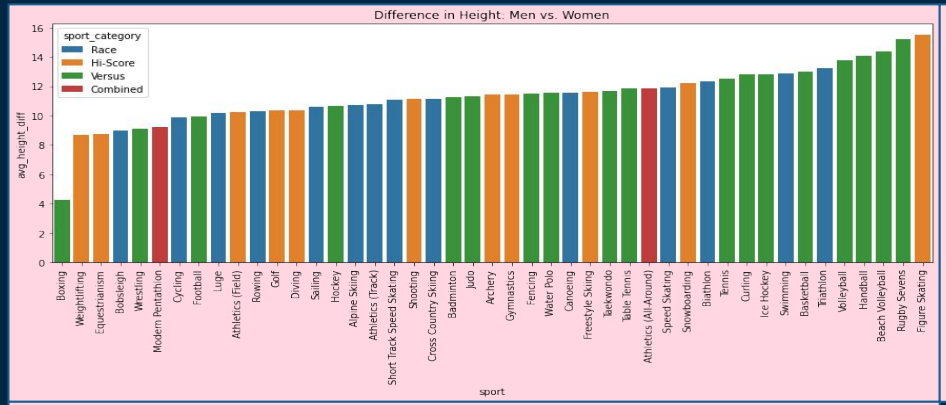


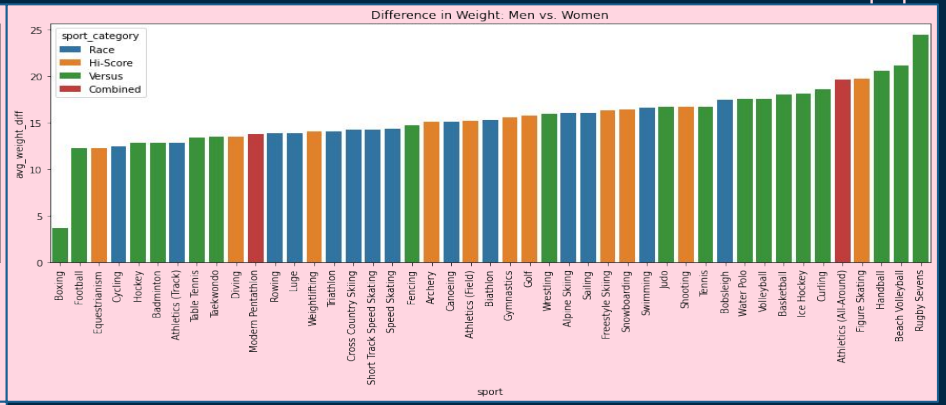


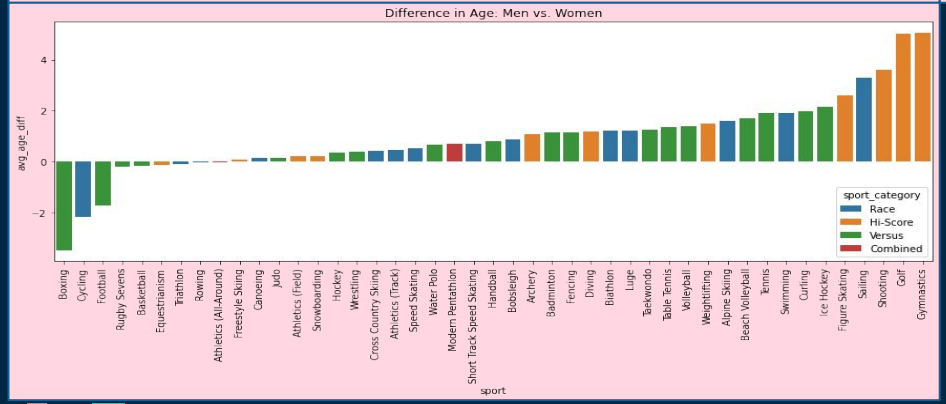


- Medalists in athletics (field) are the heaviest athletes compared to the average athlete.
- Medalists in shooting and archery are the second and third youngest athletes compared to the average athlete, respectively.
- Medalists in gymnastics are the shortest and fourth lightest athletes compared to the average athlete, but they are also the tallest athletes in rhythmic gymnastics.
- Medalists in table tennis are the youngest, third shortest, and fifth lightest athletes compared to the average athlete.
- Medalists in swimming are the second tallest and third heaviest athletes compared to the average athlete.

Sport-Specific Sex Differences in Athletic Performance

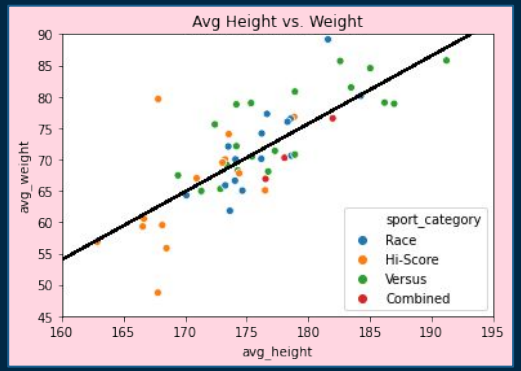


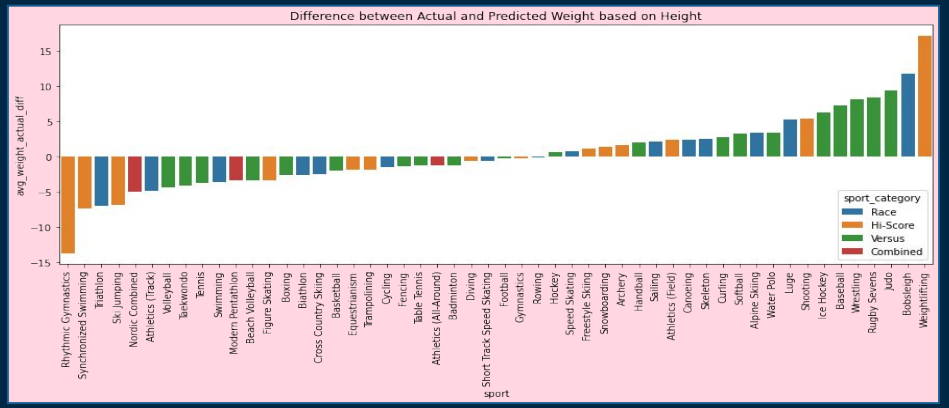




- Male athletes are consistently older, taller, and heavier than female athletes within the same sports, with the exception of outliers caused by non-physical factors, such as boxing, golf, football, and gymnastics, where men are approximately 5 years older than women.
- Male athletes also tend to be the tallest and heaviest athletes compared to female athletes in figure skating and team sports.

Correlation of Height and Weight





A moderate to strong positive correlation (r = 0.80) was observed between height and weight. Several outliers were identified, including:

- Weightlifting athletes: approximately 16 kg heavier than expected.
- Rhythmic gymnastics athletes: approximately 13 kg lighter than expected.
- Bobsleigh athletes: approximately 11 kg heavier than expected.
- Judo, rugby sevens, and wrestling athletes: approximately 8-9 kg heavier than expected.

Hypothesis Review

1. Athletes in their mid-to-late 20s will perform the best.
   a. Partially supported. The interquartile range of medalists (both male and female) also included athletes in their early-to-mid 20s.
2. The BMI of the best athletes will generally be near the middle of the healthy range (20-23).
   a. Supported for women, but not for men, whose BMI interquartile range is closer to 22-25.
3. Height and weight will be moderately to strongly positively correlated.
   a. Supported, with a regression model yielding a coefficient of $r = 0.80$.
4. Height, weight, and BMI in women will be less than those in men.
   a. Supported.
5. In events requiring mostly strength (e.g., weightlifting, throwing), successful athletes will have higher height, weight, and BMI values. In events requiring mostly speed/agility (e.g., track, gymnastics), they will have lower values.
   a. Mixed support. Sports such as gymnastics and athletics (field) met these expectations, but others such as weightlifting (where athletes are short and heavy) and athletics (track) surprised me.

Hypothesis Review: Later Questions

1. How do the effects of age, height, and weight differ between different categories of events?
    a. Generally, hi-score events have the youngest, shortest, and lightest athletes. Versus events have the oldest, tallest, and heaviest athletes, with race events in the middle.
2. Which sports have the biggest differences between sexes or categories and/or deviate the most from the general trend, and why?
    a. I uncovered many insights, including the following notable ones:
        i. Men are older than women in gymnastics and golf. Women are older than men in boxing, curling, and football.
        ii. In general, men are considerably taller and heavier than women in all sports, although not as much as usual in boxing.
        iii. Team sport athletes are particularly tall and heavy.
        iv. Athletes in sports demanding flexibility are short and light.
        v. Gymnasts are particularly young, short, and light, but the most successful artistic gymnast is the shortest and the most successful rhythmic gymnast is the tallest.
        vi. Weightlifters and bobsledders are particularly heavy considering their height.
    b. Most of these differences are due to physical factors and/or the nature of particular sports, although a few are caused by restrictions or classifications of competitors.

Conclusion

Recommendations:

- Personal trainers and Olympic coaches should adapt their training regimes to the specific needs of each athlete, taking into account their age, height, weight, and other relevant factors.
- This data can also be used to more effectively advise those who have not yet committed to a particular sport, helping them to choose a sport that is well-suited to their physical characteristics and abilities.

Next steps for analysis:

- Analyzing trends over time: Have athletes gotten younger/older, shorter/taller, lighter/heavier in general and in given sports?
- Investigating country trends: Which nations are most successful in general and in given sports? What factors could be responsible for these differences?