Prompt:

Dive Deeper

Look deeper into the features you are investigating, consider:

- Relationships / Correlation, Pearson Correlation
- Linear Regression for future prediction (if the relationship is linear)
- Textual Analysis for TF-IDF (Term Frequency-Inverse Document Frequency; Row-based and column-based, stop-word removal?

Specify 1-2 correlations you discovered. List the fields that you found to be correlated and describe what you learned from these correlations.

Go Broader

Expand the features you are investigating. Look for connections/relationships that you may have initially missed.

1. What jumps out at you now?
2. Use the descriptive stats to point you to features that you may now want to consider.

What key terms did you discover in any text analysis, for whom? Any themes? If you are not analyzing text, summarize what other things you are considering in your analysis?

New Metric

Create 1 or 2 new metrics to track relationships of data you discovered. Explain why you created them.

Response:

Last week, I created avg_age_diff, avg_height_diff, and avg_weight_diff to represent the difference between the average age, height, and weight of all athletes in a particular sex/sport vs that of medalists only. This week, I was interested in exploring how these metrics differed between men and women in the same sport. Therefore, I created a new table with similarly named metrics, except this time showing the differences between men and women.

I found that in most sports, men are about 12-18 kg heavier than women, and about 9-13 cm taller than women. Surprisingly, though, there was one sport in which the differences were very low:

boxing. This is likely due to the presence of the same weight classes for both men and women (https://www.unorthodoxx.co.uk/post/the-difference-in-rules-of-mens-and-womens-boxing).

In addition, I was interested to see how trends of height, weight, and age might differ between different types of sports. Therefore, I created a new feature, Sport Category, assigning one of the following three categories to sports:

1. Race (sports in which the object is to get the fastest time): Rowing, Bobsleigh, Speed Skating, Short Track Speed Skating, Swimming, Triathlon, Athletics (Track), Cycling, Sailing, Canoeing, Cross Country Skiing, Alpine Skiing, , Biathlon, Skeleton, Luge, Motorboating
2. Hi-Score (sports in which athletes take separate turns competing, and whomever gets the best score, distance, etc. wins):Synchronized Swimming, Rhythmic Gymnastics, Gymnastics, Athletics (Field), Snowboarding, Shooting, Weightlifting, Diving, Figure Skating, Freestyle Skiing, Trampolining, Golf, Equestrianism, Ski Jumping, Archery
3. Versus (sports in which two athletes/teams compete directly against each other): Basketball, Softball, Hockey, Curling, Football, Volleyball, Baseball, Water Polo, Handball, Ice Hockey, Rugby Sevens, Tug-Of War, Rugby, Lacrosse, Judo, Boxing, Wrestling, Badminton, Fencing, Table Tennis, Tennis, Taekwondo, Beach Volleyball
4. Combined (disciplines combining sports from more than one of the above categories): Athletics (All-Around), Modern Pentathlon, Nordic Combined

Looking across the different categories, I found that athletes in hi-score events tend to have the least height/weight. Races were in the middle, and “versus” sports had the tallest/heaviest athletes - they included the sports with the top five average heights. I feel this might be the case due to the fact that certain hi-score events require flexibility (Gymnastics, Figure Skating, Diving, etc.) and others require athletes to be active for only a short amount of time during each turn. In races, competitors are “active” for the whole race, requiring a constant supply of strength throughout. But it makes sense that “versus” events are at the top, since many involve physically struggling with an opponent (fighting sports, Rugby, etc), and they often require quick reactions based on the actions of the opponent, which can be tougher to anticipate than say, the starting gun to a track race. The one hypothesis I had not yet conducted comprehensive analysis of was #3: the association between height and weight. This week, I looked into this association, both overall, with men and women separately, and among the different categories of sports I established above. The Pearson correlation coefficients between height and weight for these combinations of genders and event categories are:

- All athletes
    - All events: 0.87

    - Races: 0.92
    - Hi-score events: 0.74
    - Versus events: 0.89
- Men
    - All events: 0.73
    - Races: 0.66
    - Hi-score events: 0.43
    - Versus events: 0.78
- Women
    - All events: 0.71
    - Races: 0.78
    - Hi-score events: 0.44
    - Versus events: 0.78

  For the most part, height and weight are moderately to strongly positively correlated. The sports that appeared most frequently as outliers were: Big:

- F Rhythmic Gymnastics: avg_height=167.8 cm, avg_weight=48.8 kg (~10 kg less than expected according to best fit line among all athletes/events)
- F Weightlifting: avg_height = 160.5 cm, avg_weight=67.8 kg (~15 kg more than expected)
- M Weightlifting: avg_height = 169.4 cm, avg_weight = 81.8 kg (~15 kg more than expected)

Moderate:

- M Bobsleigh: avg_height = 182.2 cm, avg_weight = 90.3 kg (~10 kg more than expected)
- M Judo: avg_height = 177.5 cm, avg_weight = 83.7 kg (~1- kg more than expected)
- M Rugby Sevens: avg_height = 182.8 cm, avg_weight = 91 kg (~10 kg more than expected)

It seems that for the most part, Olympic sports require well-balanced skill sets and training practices that do not deviate much from the naturally occurring height-weight correlation. To explain some of the exceptions (listed above), here is what I found:

- F Rhythmic Gymnasts are much taller and thinner than the average for their age. Reasons for this include:
- According to a study conducted by the Journal of Clinical Endocrinology & Metabolism (https://en.as.com/en/2021/08/06/olympic_games/1628240586_789360.html), rhythmic gymnasts h

ave strict diets and put their bodies through extreme intensity, which causes height velocity to stop later in puberty for them - around the age of 18, rather than 15 for most women. Additionally, due to rhythmic gymnastics routines involving throwing and catching (balls, hoops, etc.), they are made easier by longer limbs.

- M/F Weightlifters are very short considering their weight. This is mainly due to the fact that the taller you are, the less you can lift as a percentage of your weight (https://www.inverse.com/mind-body/leg-day-observer-weightlifting-height).
- M Bobsledders' high weight given their height makes sense due to physics: the more weight is in a bobsled, the faster it can go down the track (https://bobskeleton.org.uk/?p=73). In fact, competitors with larger weights were consistently outperforming the rest of the competition until weight limits were introduced in the 1950s.