**Milestone 3: Beyond Descriptive Stats**

1. **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.

**Answer:**

I have used logistic regression to try to predict the medal that a ticket will win based on the variables Weight, Height and Age.

Next I will describe the process that I have carried out.

1. **Go Broader**

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

What jumps out at you now?

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?

**Answer:**

I was ignoring the NOC column that had a small error, there were RUS and URS values ​​that referred to Russia. This made my metric turn out wrong. So what I did was update all URS values ​​by RUS in both tables (athletes, regions).

So far I have discovered the following statements regarding the data sets.

• The age of most athletes in games is between 24 and 26 years old.

• The height of the athletes is 171 to 178 cm.

• Weight is 64 to 72 kg.

Also, regarding new data relationships. I believe that physical characteristics such as weight and age of an athlete influence whether or not to obtain medals.

On the other hand, another analysis that I am thinking of doing and it catches my attention is a machine learning model to predict which medals will be won the most in future events of the Olympic games.

1. **New Metric**

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