## Milestone 2: Descriptive stats

My client/dataset is SportsStats. It is a sports analysis firm partnering with local news and elite personal trainers to provide "interesting" insights to help their partners. Insights could be patterns/trends highlighting certain groups/events/countries, etc. for the purpose of developing a news story or discovering key health insights.

- 1. Provide a summary of the different descriptive statistic you looked at and WHY. I did some descriptive statistic to answer the following questions:
  - What kind of athlete has the most chance of getting a medal?
    - To do so, I looked at the average values of Age, Height and Weight for the top winning countries in the Olympics to see if I can identify some key patterns.
  - Do more athletes mean more medals?
    - To do so, I looked for any correlation between the number of athletes in a team and the number of medals won.
- 2. Submit 2-3 key points you may have discovered about the data, e.g. new relationships? Did you come up with additional ideas for other things to review?

What kind of athlete has the most chance of getting a medal?

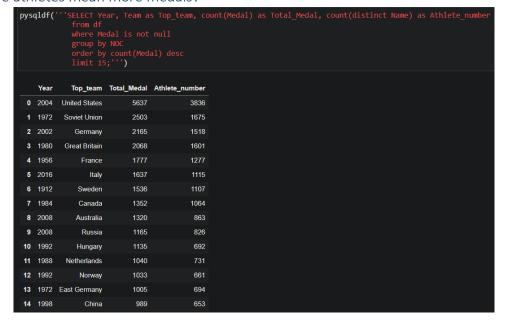
```
pysqldf('''SELECT NOC, Team, avg(Age), avg (Height), avg(Weight), count(Medal) as Total_Medal
from df
             group by NOC order by count(Medal) desc limit 10;''')
               Team avg(Age) avg (Height) avg(Weight) Total_Medal
0 USA United States 24.896638 179.519904 75.091918
            Germany 26.416510 178.682646 74.468635
                                                            2165
   GBR Great Britain 27.846310 178.585981
                                           74.552347
             France 27.982153 178.591992 74.383300
              Italy 26.956195 178.426520 75.703008
                                                            1637
            Sweden 28.066406 179.541457 76.492188
                                                            1536
6 SWE
            Canada 25.826677 176.171004
                                           73.541705
            Australia 24 819011 178 884459 74 095533
8 AUS
                                                            1320
             Russia 25.661790 177.277729 72.109692
                                                            1165
```

Here, we can see from the Top 10 country with the most medals that an athlete has a higher winning chance when:

- He/she is between 24 and 27 years old.
- He/she is between 176 cm and 179 cm.
- He/she weighs between 74 kg and 76 kg

Obviously, these stats depend on what sports is considered. These stats give an overall for all sports.

## Do more athletes mean more medals?



From here, we can the top country with most medals in a single competition. We can see that there is a correlation between the number of athlete and the number of medals won. The more athletes you have, the more medals you get.

Next, I can look at the influence of a given sport as you have individual sport (1 athlete per competition = 1 medal) and collective sport (X athletes per competition = X medals).

3. Did you prove or disprove any of your initial hypotheses? If so, which one and what do you plan to do next?

My hypotheses were:

- The more athletes your team have, the higher chances you have to a win a medal.
  - This hypothesis has been confirmed by my research. I planned to look deeper into it by filtering by sport.
- Young athletes in the prime of their career (around 25 years old) have a higher chance to win a medal.
  - This hypothesis has been confirmed by my research. I planned to look deeper into it by filtering by sport to see the influence of height and weight.

## 4. What additional questions are seeking to answer?

I need to find the results for a given sports to have a more precise understanding of my data. I will try to find the best athlete stats for swimming and basketball for example.