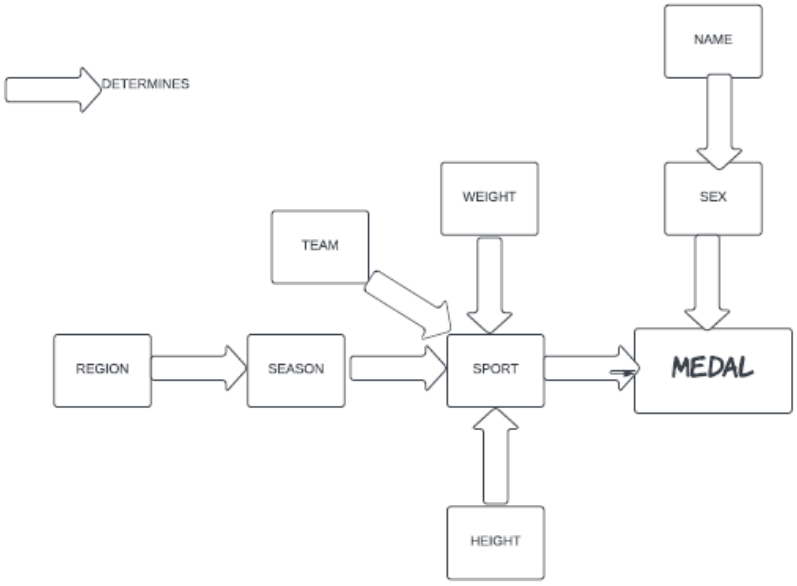
CAPSTONE : SportsStats' insights to develop news story and provide better health

E.R.D Diagram

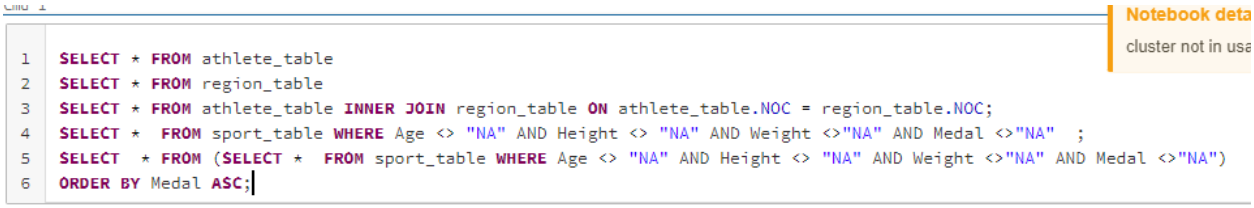


INTRODUCTION

Being a huge fan of sports myself, I decided to work with SportsStats firm as a data scientist to find trends, patterns to help them reach their goal: to develop news story and find insights on key health. They are partnering with local news and personal trainers. My audience would be countries, individuals who are interested in having a better health to maximize their chances of winning a competition, but also who are interested in news about their favorite sport, teams, games, and events. Being a champion and winning a gold medal is not an easy task. A lot of discipline and hard work comes with it. I will be finding patterns to give advice on how to better prepare for the competition.

Cleaning/Importing Data

I used databricks community edition to import my data. I first created a cluster. Then I went to the data section, created a table where I uploaded the dataset previously downloaded. To clean the data, I first joined both datasets by region since both datasets had a region column called NOC. Then I deleted all na values from dataset. The dataset had a column called notes which had mostly null values, I deleted the column.



Questions that I want to answer with the data:

- What does it take to win a gold medal?
- How do you choose a sport? What attributes do you need?
- Can a sport be played anywhere and anytime?

Assumptions:

- The age plays a role in winning a medal due to your physical and moral condition
- Height and weight can determine which sport you might have more chances of being good at
- Season and region determine what sport is being played. You notice that some countries play some specific sport on the daily basis which is mainly due to their climate

Approach:

- I am going to look at the age column and try to identify if the younger you are, the more chance you must win a medal.
- The height, weight can have direct impact on how good you perform.
- I would like to explore a relationship that exist between season, region, and sport. Does the climate/season affect what sport can be played? I am going o look at season, region columns.

Descriptive stats, relationships, and hypothesis:

- I calculated the average age and compared with the medal won per team and the conclusion proves that the countries that have won more medals such as Norway, US, and Russia have young athletes that compete for them between 22 to 25. The exception was Italy and Finland who have both an average of 27, 28 respectively. Italy has about 100% of his athlete competing for the sport rowing which is best known as low impact exercise compared to the other sports such as gymnastic, football…

- After checking some healthy weight for different height using average height and average weight, I can conclude that US, Norway, and Russia fit in the recommended weight for their height which means that they are healthy enough hence their great performances at the competition.
- Indeed, you notice that no sports have been played in two different seasons which confirms that the season affects the sport being played. Some regions might play different sports because they might have a change of climate during that year.

Additional questions?

Does the gender have an impact on winning a medal?