SQL CAPSTONE PROJECT

WEEK 1

**Step 1: Preparing for Your Proposal**

1. **Which client/dataset did you select and why?**

Client 3: Sports Stats (Olympics Dataset - 120 years of data)

SportsStats 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.

I chose this dataset to gain key insights from the data and which physical characteristics increase the likeliness to win a medal

1. **Describe the steps you took to import and clean the data.**
2. I imported the data into a pandas Data Frame into a Jupyter Notebook. There were two csv files - athlete\_events.csv and noc\_regions.csv
3. I observed the null values present in each column. I saw that the null values of the following columns must be removed.

Age 9474

Height 60171

Weight 62875

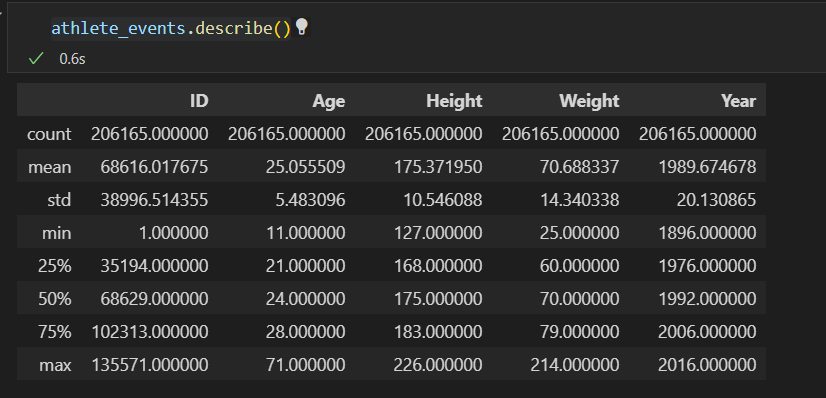
**The null values of the medals column were required in order to judge whether an athlete has won a medal or not. So they were not removed**

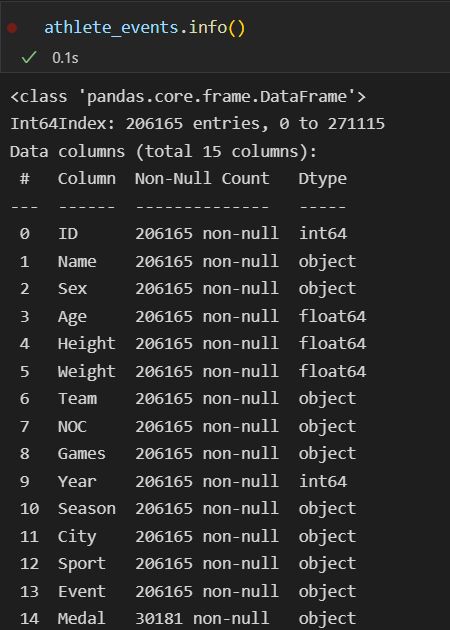
Similarly I removed the Null values of region in the NOC table.

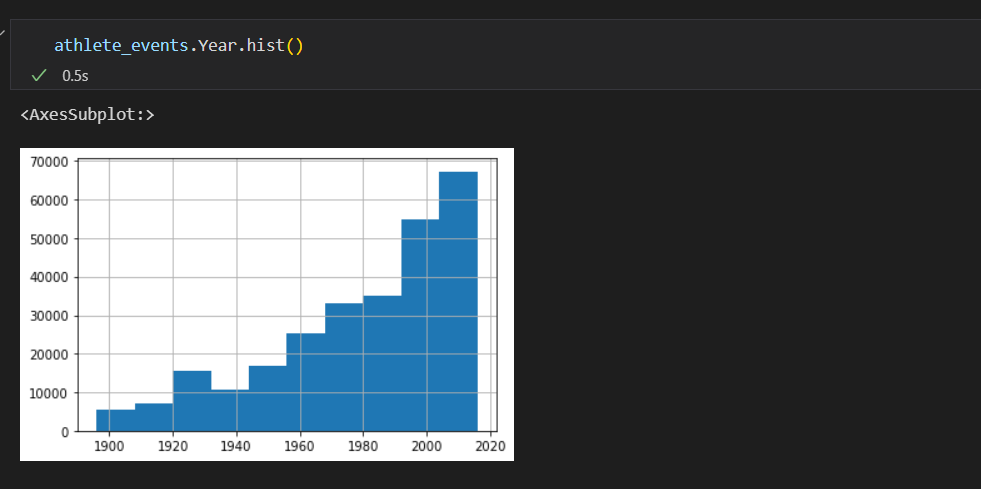
**The null values of the notes column were required since that was additional information so it was not removed.**

1. **Perform initial exploration of data and provide some screenshots or display some stats of the data you are looking at.**

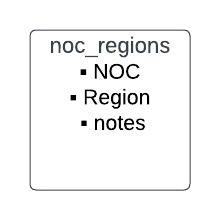
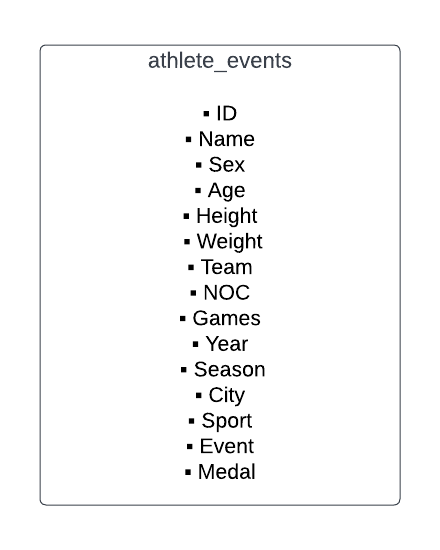
**Basic Info of the Dataset:**

****

****

**This is a histogram of the years of the dataset**

1. **Create an ERD or proposed ERD to show the relationships of the data you are exploring.**



**Step 2: Develop Project Proposal**

**Description**

I would like to find out which physical body characteristics are necessary for winning a medal

**Questions**

I would like to find the answers to the following questions in the data

1. The average height and weight characteristics for medal winners in each game

This will help me discover the ideal body type for a certain game

1. What is the average age of medal winners in each game

This will help me discover the ideal age in order to win a game

1. I wish to know which Country has won the most medals in each game

**Hypothesis**

1. The average BMI value of the medal willing athletes matches with the ideal values
2. Developed countries have won more medals due to high infrastructure of sports.

**Approach**

Applying SQL queries to the dataset.

Using Where and GROUP BY clauses will help me find the answer.