**SQL for Data Science**

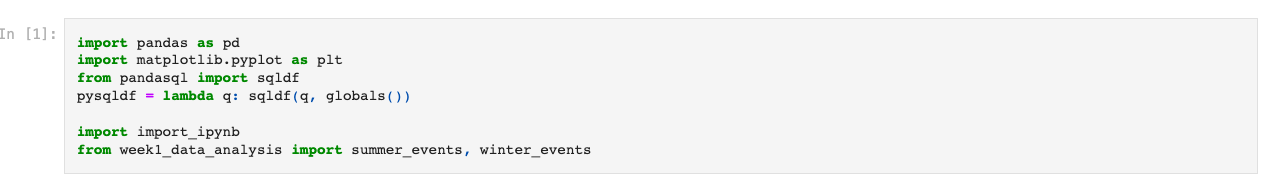
**Capstone Project**

**Milestone 2: Descriptive Stats**

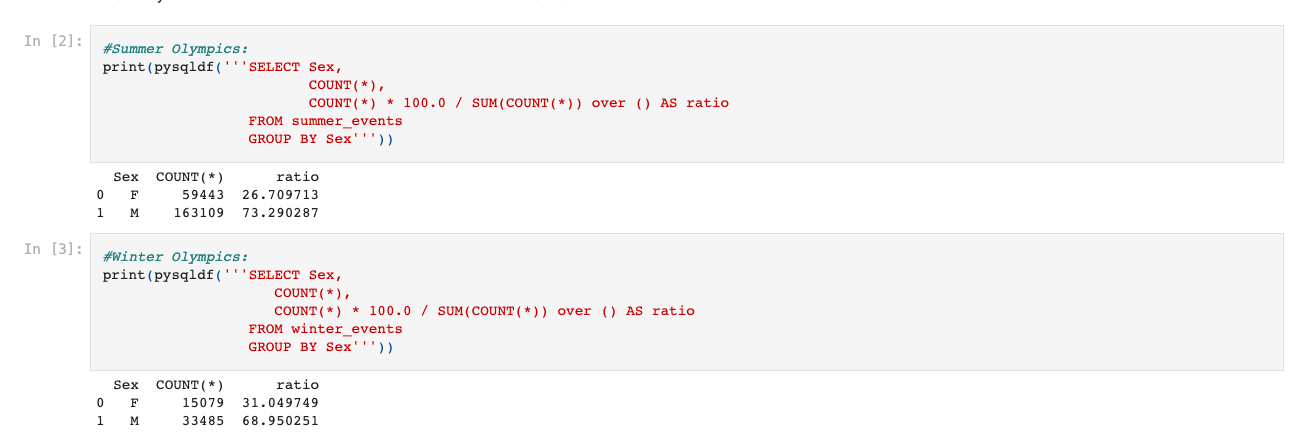
**Emi Bode**

1. *Provide a summary of the different descriptive statistics you looked at and WHY.*

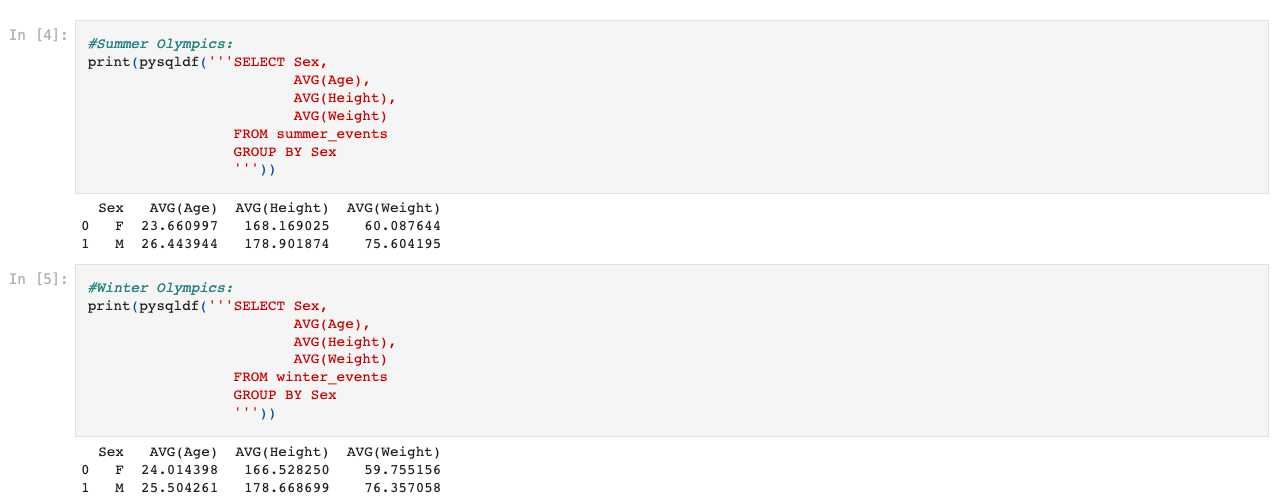
In this part of the project we will import our week1\_data\_analysis.ipynb notebook in Jupyter.



Then we will analyse the data by using “Sex” as a parameter in order to examine the male-female differences

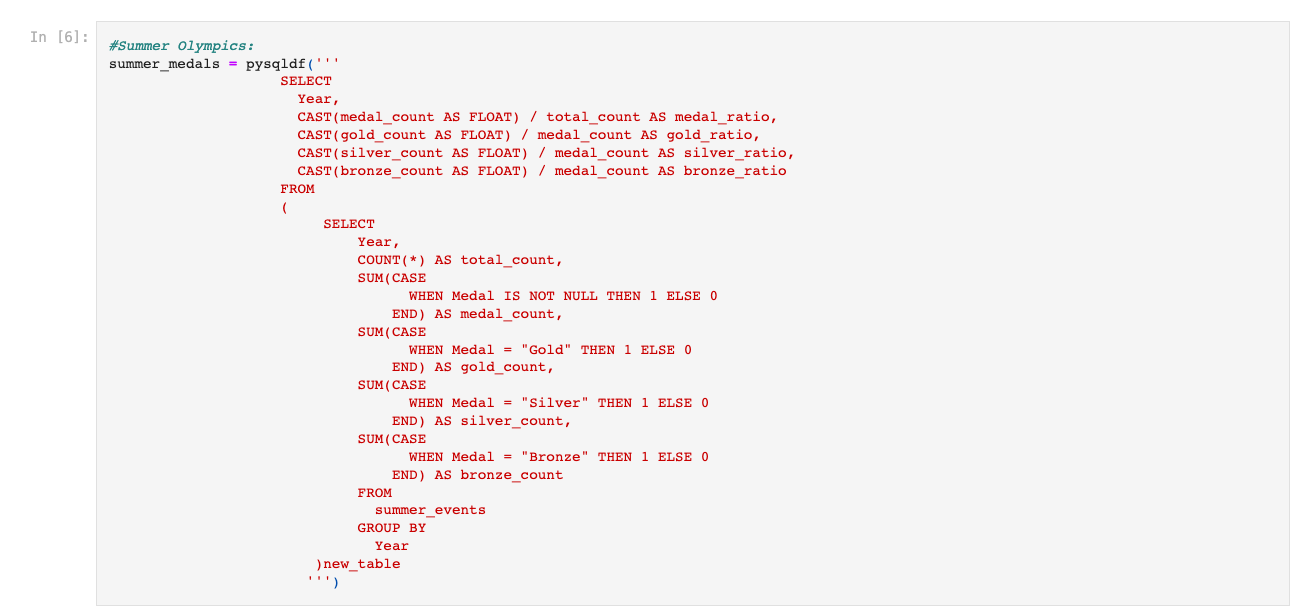


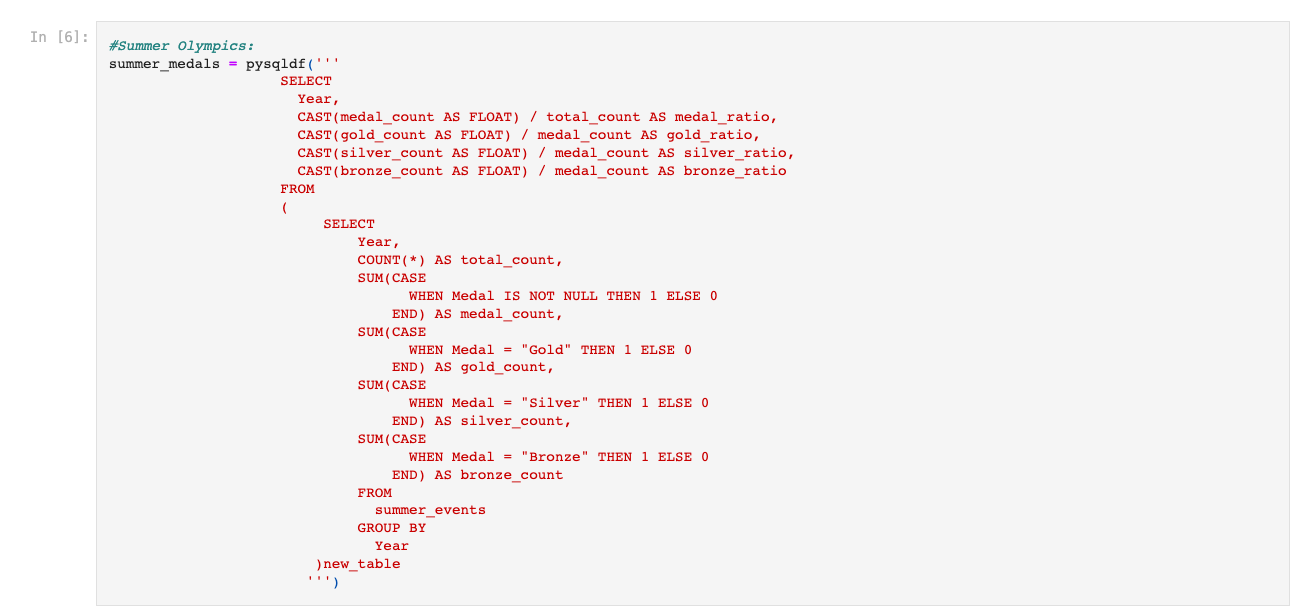
As we can see from the results, men are still more dominant even though the ratio between the Summer Olympics and the Winter Olympics is different. From what we can tell from the results the ratio of women to men has increased over time.

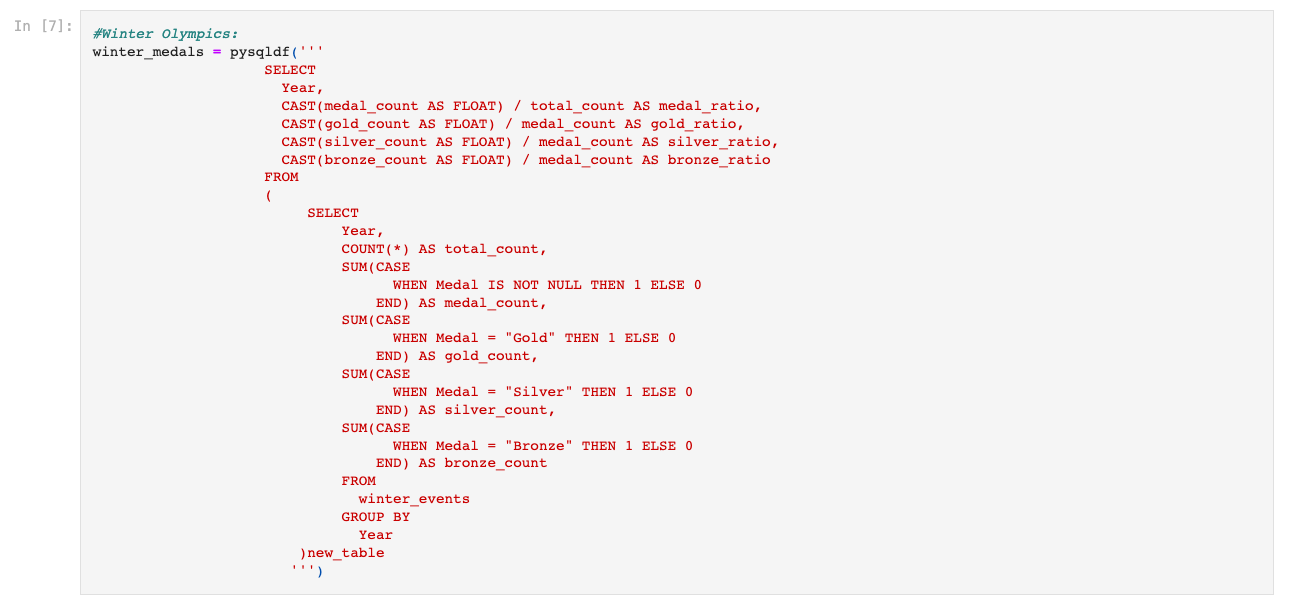


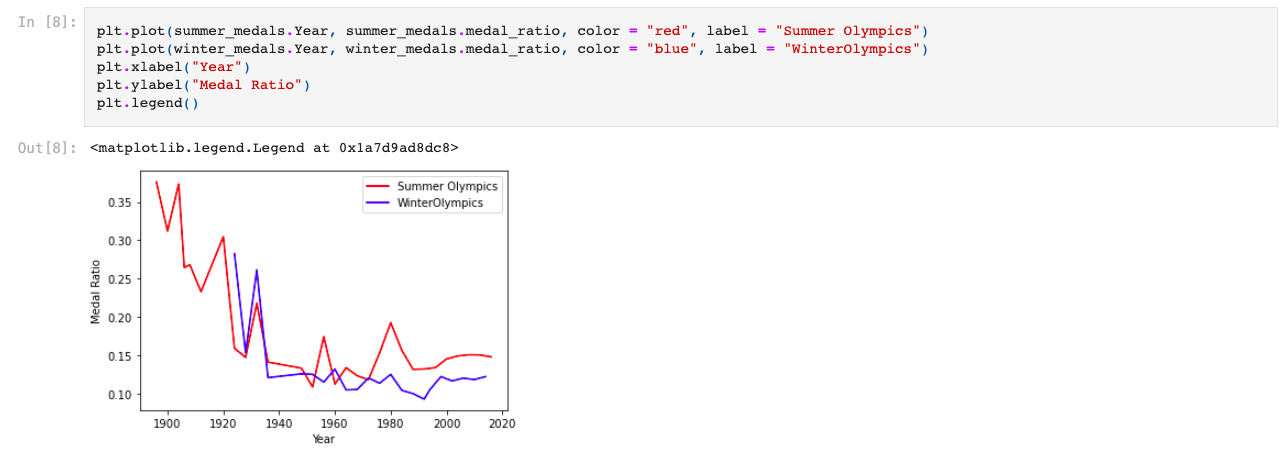
As we analyse the data we can see that there is a significant difference between male and females not just in terms of height and weight, but also in terms of age. We will have to consider social factors to better understand our data.

Another interesting analysis is the ratio of medal winners.









From the graphic above we can see that in the last century the medal ratio fluctuated in the two competitions, but then it stabilised.



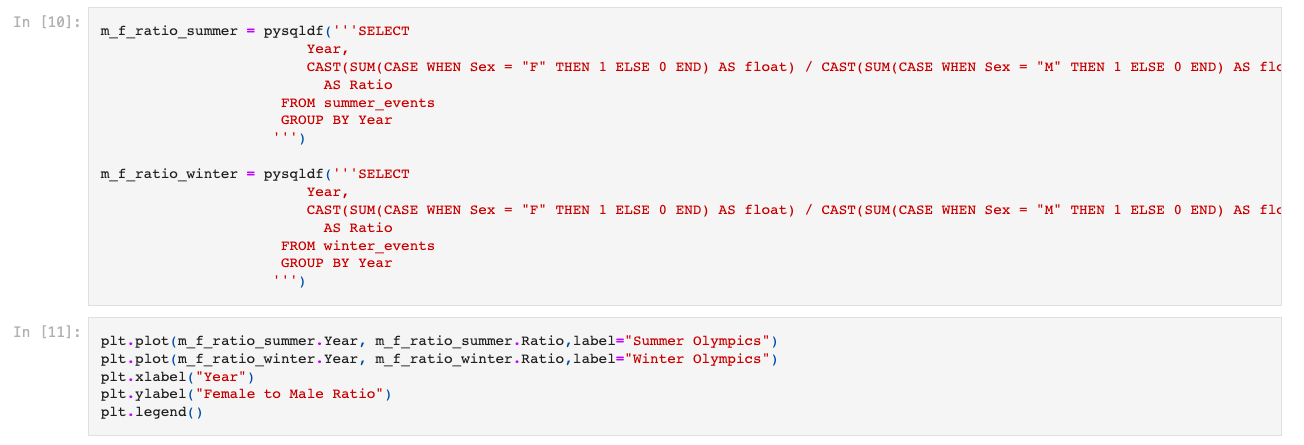
From the graphics we can tell that the relative percentage of gold, silver and bronze medals have also stabilised due to the reasons mentioned above.

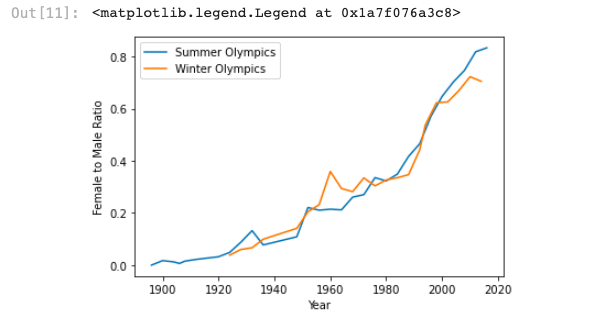
1. *Submit 2-3 key points you may have discovered about the data, e.g. new relationships? Aha's! Did you come up with additional ideas for other things to review?*

* By analysing the results we can say that the age gap between male and female in the Winter Olympics is smaller than in the Summer Olympics
* Another thing that is visible is the fact that in the recent Olympics, the ratio of medals and the percentage of participants who won medals has been stabilised
* There are differences between the number of participants in the Summer and Winter Olympics that must be analysed separately.

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

My first assumption is that the ratio of female to male has increased over time. We will create a year-by-year Female to Male ratio, then we will draw a line graph to see their relationship.





From what we can see the assumption is correct. Over time, the ratio of females to male has increased. There is an interesting detail: during World War II, the proportion of the Summer Olympics dropped sharply, but then it resumed its growth momentum.

1. *What additional questions are you seeking to answer?*

I will use A/B testing to test my other two hypotheses:

1. In the Winter Olympics, the annual performance of different countries varies greatly
2. A country’s performance in the WInter Olympics is related to its performance in the Summer Olympics