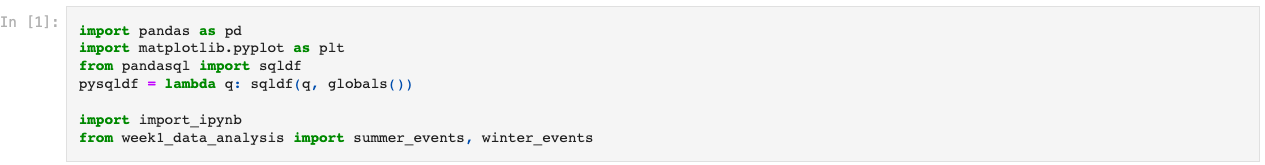
**Milestone 3: Beyond Descriptive Stats**

**SQL for Data Science**

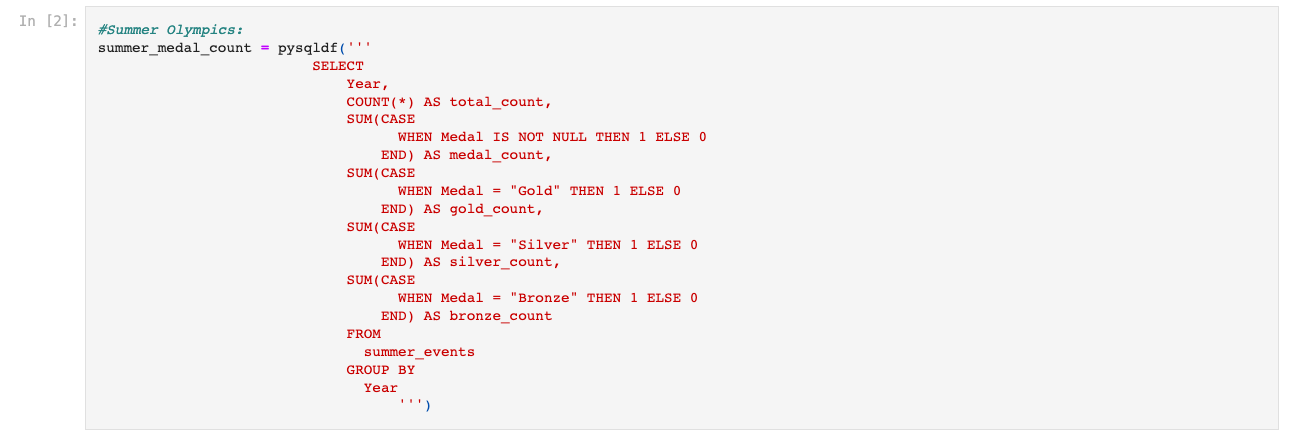
**Capstone Project**

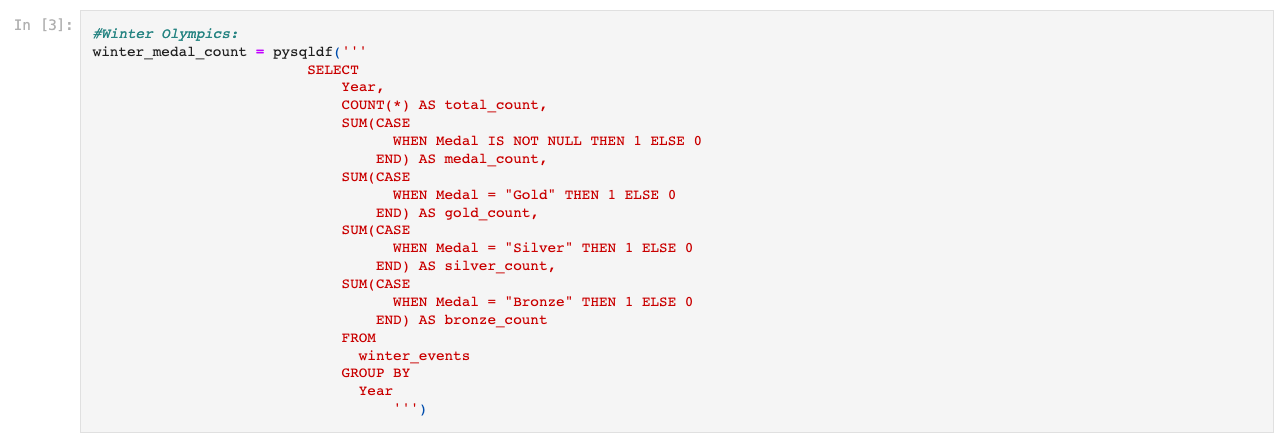
Emi Bode

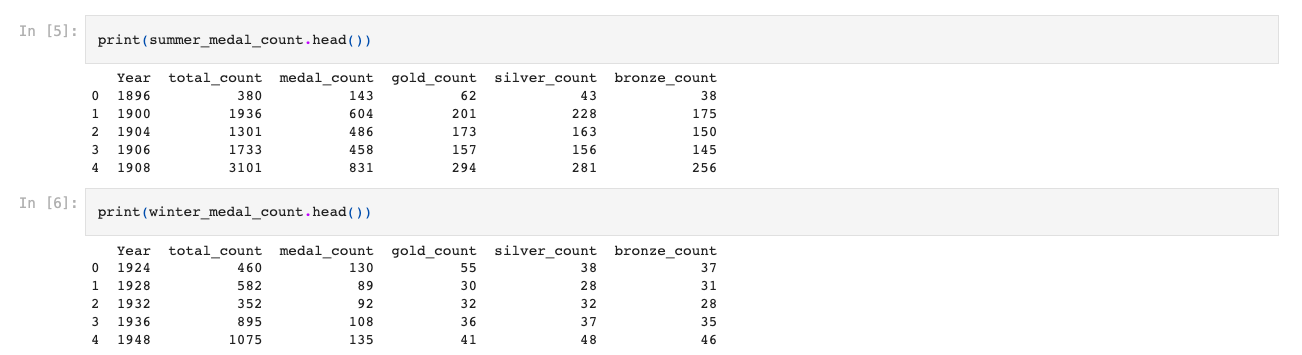
First we will import week1\_data\_analysis.ipynb notebook to Jupyter.

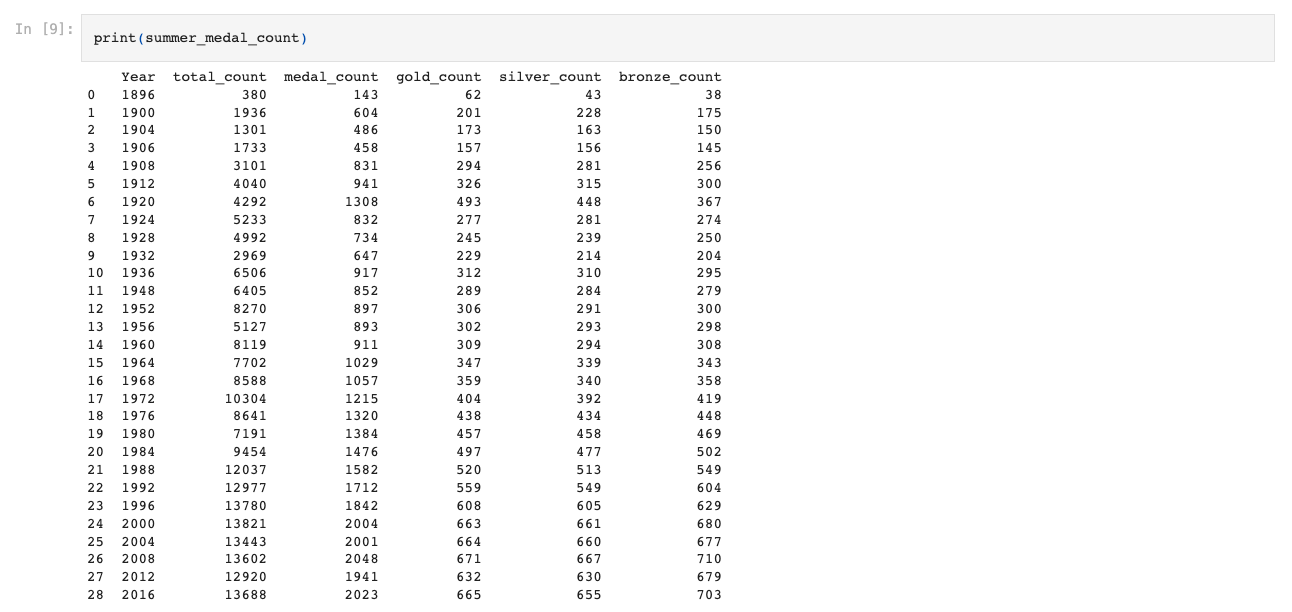


Then we will create two tables to count the total number of medals in the Summer and Winter Olympics. We will calculate the Peason correlation coefficient between the total number of medals in the Summer and Winter Olympics.



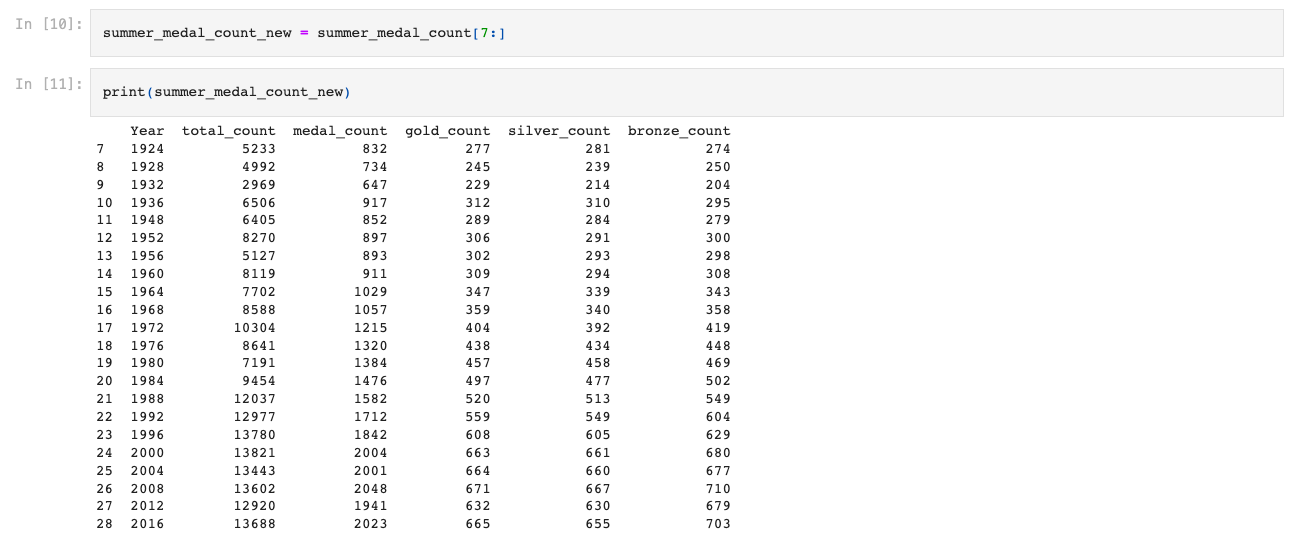






As we can see from the results, the length of the array of the number of medal counts in the Winter and Summer Olympics are different because the Winter Olympics started in 1924, but the Summer Olympics started in 1896.

Now we will create a shortened table of the Summer Olympics that started in 1924 to match the length of the Winter Olympics.



Next we will calculate the Pearson correlation coefficient between the total number of medals in the Winter and Summer Olympics from 1924 to 2016.



By analysing the results we can say that the Pearson correlation coefficient is positive with a value of 0.94 and the performance of a country in Winter Olympics is highly correlated to that in Summer Olympics.

Now we will calculate the Standard Deviation in country performance through years.



From the results we can see that from 1924 to 2016, the standard deviation in the summer olympics is about 3 times the standard deviation in the Winter Olympics.

Country performance by year changes more in the Summer Olympics.