

## SQL FOR DATA SCIENCE CAPSTONE PROJECT

### Assignment 3: Week 3/ Milestone 3: Descriptive stats

**Name: Okonkwo Maureen**

**Client: Sports Stats**

**Data set: Olympics Dataset**

**Date: April 2022**

- 1) **Specify 1-2 correlations you discovered. List the fields that you found to be correlated and describe what you learned from those correlations**
  - a) **Age\_groups and medals**

I calculated the different age group and the medals won, using the Pearson chi square to determine the statistical significance between the variables, I concluded that age between 20 and 25 has an influence over the number of medals that can be won

#### **Number of Participants and Number of medals Variable**

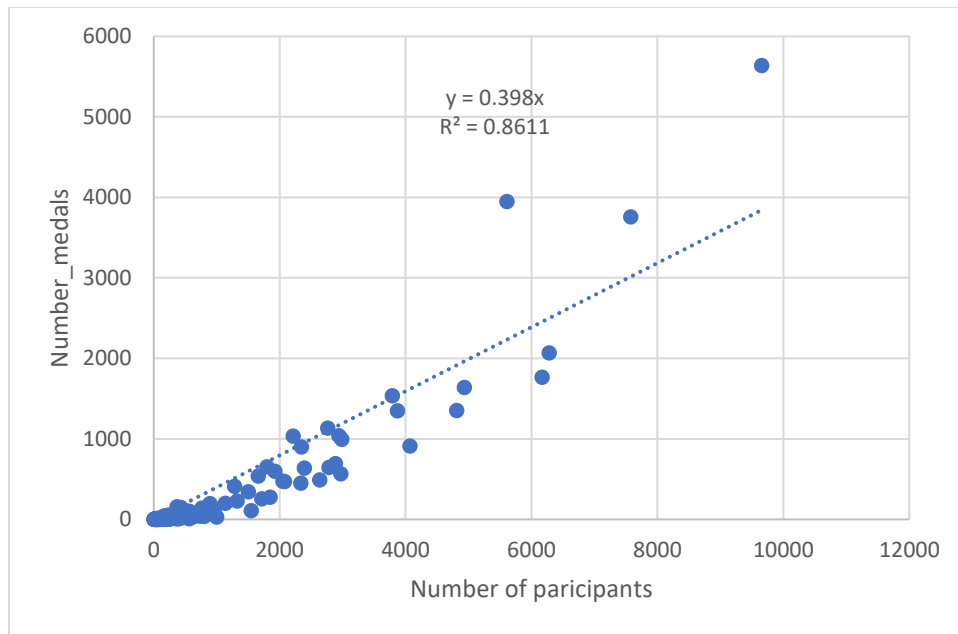
I calculated the number of participants represented by each country and the number of medals won by each participant in each country. To achieve this, I used Pearson Correlation which calculates the relationship between the two variables. The  $r = 0.929214$  shows that the two variables are positively correlated with P-Value ( $4.21685E-90$ ) which is  $< 0.0001$ . Therefore P-value =  $< 0.0001$

### **The Pearson correlation procedure**

<b>2 Variables</b>	<b>Num_medals</b>	<b>Num_participants</b>
--------------------	-------------------	-------------------------

Pearson Correlation	T-Statistics	DF	P-Value
0.92914	35.8939	204	4.21685E-90

Variable	N	Sum	Std Dev	Mean	Minimum	Maximum
Num_participants	206	136137	1376.05	660.86	3	9653
Num_medals	206	39763	625.73	193.02	0	5637



## 2) Create 1 or 2 new metrics to track relationship of data you discovered.

### Explain why you created them:

I created the medal and age group metrics where I analyzed the number of medals won by each age group and the group 20\_25 has won the highest medal in Gold, Bronze and Silver.

The age was the main variable I used for my metrics, where analyzing the age with sport I discovered that Gymnastic has an average age of 22, minimum age of 10 and maximum age of 49 while group 20\_25 was the highest represented age in the gymnastic sport.

The Art Competition is the sport that has the oldest participant of 97 years old

The total number of medals won in the Olympics is 946 with the season summer having the highest medal won which is 683.

The highest medal ever won in the Olympics was in the year 1980 with 385 medals won.