The Descriptive Statistics I looked at with the reason:

1. Which top 10 countries have won the highest number of medals in Olympics’ history?
2. What is the trend of year vs. medals won by top 5 countries found from point 1? Was there a slight change over the years? Was there a sudden rise and fall over the years? If yes, then what might be the reason?
3. The majority of countries’ medals are concentrated in a certain number of sports.
4. Which Age group the highest number of participants come from? This could answer the questions like: Was it due to awareness to this age group about the sport/Olympics or was it about physical capability needed for the particular sport?
5. Was there any sudden or high hike seen in the number of participants took part in Olympics? The number of participants in last 20 years have been increased more than ever in Olympics’ history. What might be the reason?
6. Is there any trend associated with the season and medal won by the country?

2-3 Key points about the data: New Relationships: Additional Ideas for other things to review:

1. There is a relationship between medals won by particular country and sports. It is common for all countries that majority of their medals are concentrated in a certain number of sports.
2. There is a proportional relationship between medals won by countries vs. total number of players participated in Olympics.
3. There could be a relationship between a player and medals won by him/her in different events of the same sport.
4. Additional Idea: One can check the relationship between medal-winning top 4 countries and season (winter/summer). Is there any relationship between medal-winning and season?

Prove/Disprove Hypotheses:

1. First hypothesis is True: I looked into the list of top 5 sports that have the most medals. I computed the percentage of these sports in relation to the total medals won by particular country. Based on the results, I came to the conclusion that all countries’ majority of medals are concentrated in a certain number of sports.
2. Second hypothesis is True: I studied the total number of participants of specific country in the Olympics. I looked into the percentage of the participants who won medals. Based on the results, I came to the conclusion that countries won more medals because they had higher number of participants than other countries.
3. Third hypothesis is False: I studied the participants who won the medals in different events. I grouped it according to sports. I have found different participants won in different events. This is because the requirement to win the sport is different for different events in the same sport.

Additional Questions:

1. Was there a sudden hike seen for any particular group of countries during the specific year? What was the reason behind it?
2. Which sport can be won by 35+ age group? Is there any specific group of sports/events which was won by the upper age group?

**Dive Deeper: Correlation/s I have discovered and What I learnt from those correlations.**

1. During initial data exploration, I thought to look into the relationship of top 10 countries in each Olympics event and then plot a graph of them to number of participants from those countries. But, obviously, top 10 countries are found to be different in each Olympics event. So, I decided to go with top 5 countries which have won the highest number of medals till now in Olympics history. I found ‘Soviet Union’ was also one of the countries out of top 5 countries. As it got resolved in later years, ‘Soviet Union’ was eliminated from analysis.
   1. I found the strong correlation between no. of medals won by the country and number of participants from that country.
   2. The Pearson correlation coefficients between no. of medals won by countries and number of participants from those countries are as following:

|  |  |
| --- | --- |
| **Country** | **Pearson Correlation Coefficient** |
| United States | 0.89791106 |
| Germany | 0.93401656 |
| Great Britain | 0.71517374 |
| France | 0.70679510 |

* 1. I learned that countries have won a greater number of medals because more participants came from those countries.

1. While doing gender-wise analysis, I found the correlation between no. of medals won by female sex and by male sex of particular country in top 10 sports. The list found was almost contained similar sports.
   1. I learned from this analysis that in their respective countries, the attention is being given equally to both sexes’ federation in terms of funding, training programs etc.

**Go Broader: Summary of other things I am considering in my analysis.**

1. Season-wise analysis: What impact does the season have on medals won by country?
   1. I found, in search for the answer to this question, that as there are a smaller number of sports played in Olympics in winter season, the medal-winning count was also found in a less number.
   2. Top 5 countries who won the highest number of medals in winter season are: Canada (575), United States (533), Norway (443), Sweden (428), Finland (426) and in summer season are: United States (4686), Soviet Union (2061), Germany (1687), Great Britain (1598), France (1408). The list of winter is totally differed from that of summer because winter sports might be more famous in cold countries compared to other countries.
2. Sports-wise analysis: Most of the medals are concentrated in specific sports for particular country. That is because basketball is more famous among people of United States, and cycling or fencing is more famous among people of France.
3. Age-wise analysis: No. of medals won in top 10 sports by any country was achieved by the age group of 20-27 years.

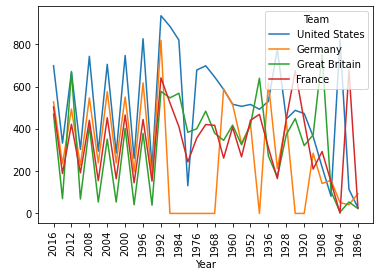
**New Metric: New metric/s to track the relationship of data I have discovered.**

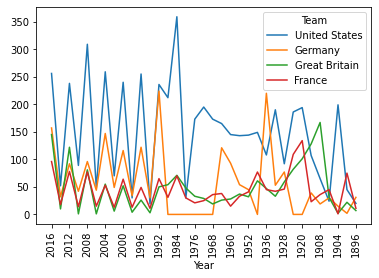
1. A metric to track relationship between no. of medals won by country vs. no. of participants participated from that country is Percentage Evaluation ((No. of medal/No. of participants) \* 100). The reason I am using percentage evaluation is because I want to know that countries are winning a greater no. of medals just because they are sending a greater number of participants to Olympics.
2. To track whether a country is gender-neutral in sports area, a medal-winning count by both genders is the necessary metric.

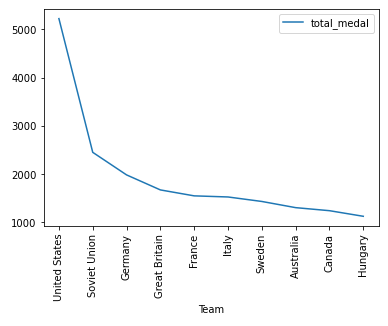
**About Dataset and Client and Target Audience :**

* The Olympics dataset, which has 120 years of data of participants in Olympics, has been chosen for analysis. This dataset consists of two tables with details of names of participants, sports they participated, countries they belong to, medals if they won any, etc.
* My purpose of this analysis is concentrated to top performing countries in Olympics history to find possible reasons for their success.
* My client is any firm who is working with government’s sport authority or any news media. My analysis might be helpful these clients in order to establish any strategies while preparing participants or creating news stories to aware local public.
* My target audience is government itself, and elite personal trainers.

**Initial Data Exploration :**







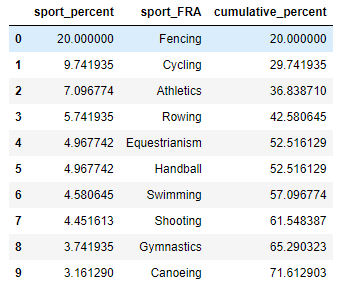
**The BIG PICTURE from initial data exploration – Question-to-Answer :**

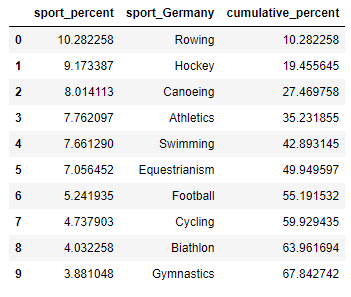
* Soviet Union was one of the top 5 countries who won the highest number of medals in Olympics history. However, there is no data of Soviet Union after the year 1990 because it got dissolved in year 1990. So, Soviet Union was not considered in data analysis part. The United States, Germany, France and Great Britain are cosidered for analysis purpose.
* What is the trend of year vs. medals won by top 4 countries? Was there a slight change over the years or it Was a sudden rise and fall over the years? What might be the reason?
* Is there any relation between number of participants from one country and medals won by that country?
* Is there any relationship between age group and sport? For example, participants from the specific age group predominantly win in a specific sport?
* The number of participants in last 20 years have been increased more than ever in Olympics history? What might be the reason?
* Is there any trend associated with season and medals won by the country?
* Do men and women shine in different sports?

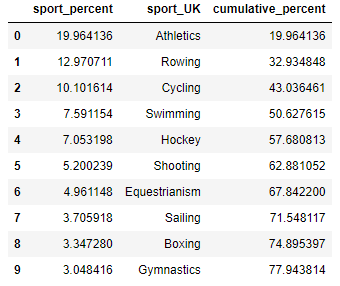
**Initial Hypotheses :**

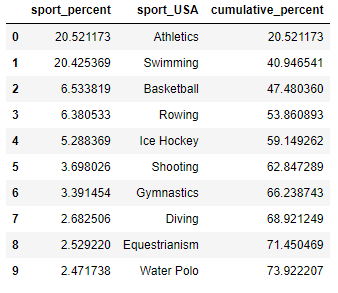
* It is common for all countries that majority of their medals are concentrated in a certain number of sports.
* The country/countries won more number of medals because it/they have sent greater number of participants.
* At most of the times, a player wins a medal in different events of the same sport.

**HYPOTHESIS 1: Medals are concentrated in a few sports** :



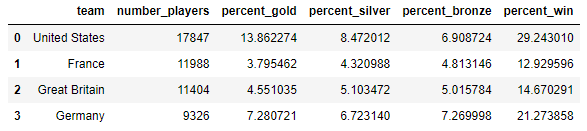






* The tables that are presented above show that most of the medals won by top 4 best performing countries are concentrated in a few sports. Above tables show that even though there are 66 categories of sports, majority of medals i.e. More than 67% are won in top 10 sports.
* Based on the above facts, the first hypothesis can be considered true.

**HYPOTHESIS 2: More participants means more medals** :



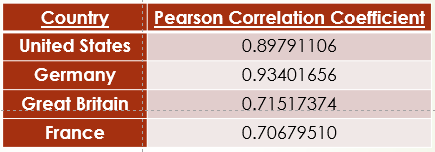
* The table presented above denotes percentage of participants who won a medal out of the total participants from respective country in the Olympics.
* The United States won the highest number of medals because it sent a greater number of participants to the Olympics. However, this is not true in case of other three countries as Germany have the higher win percentage even though it sent less number of participants compared to France and Great Britain.
* Based on the above facts, the second hypothesis is false. But we can change the second hypothesis little bit, which can be proved.
* Changed Hypothesis: The United States won the highest no. of medals because it sent a greater no. of participants.

**HYPOTHESIS 3: A player wins a medal in different events of the same sport :**

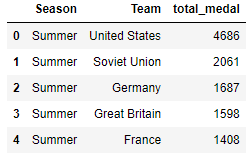
* After looking into the participants in each sport and whether they won medals in different events of the same sport, the third hypothesis can be concluded false.
* The reason is: even in the ‚Athletics‘ category sport, there are different events which need different physical abilities to win in a event.

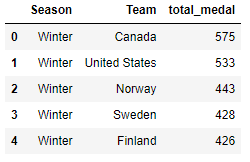
**Insights Discovered :**

* The Pearson correlation coefficients between no. of medals won by countries and number of participants from those countries are as following:



* No. of medals won in sports was not only concentrated to age group till 35 years. This can be proved by looking in maximum age of the participants who won medals.
* While doing gender-wise analysis, I found the correlation between no. of medals won by female sex and by male sex of particular country in top 10 sports. The list found was almost contained similar sports. I learned from this analysis that in their respective countries, the attention is being given equally to both sexes’ federation in terms of funding, training programs etc.
* What impact does the season have on medals won by country?
* Top 5 countries who won the highest number of medals in winter and summer season are: Canada, United States, Norway, Sweden, Finland and in summer season are: United States, Soviet Union, Germany, Great Britain, France. The list of winter is totally differed from that of summer (except United States) because winter sports might be more famous in cold countries compared to other countries. The up-down curve in medal history graph for top 4 best performing countries is understood by the above fact. The countries’ overall best performance is subjected to more sports played in summer compared to winter.





**New Metric: New metric/s to track the relationship of data I have discovered.**

* A metric to track relationship between no. of medals won by country vs. no. of participants participated from that country is Percentage Evaluation ((No. of medal/No. of participants) \* 100). The reason I am using percentage evaluation is because I want to know that countries are winning a greater no. of medals just because they are sending a greater number of participants to Olympics.
* To track whether a country is gender-neutral in sports area, a medal-winning count by both genders is the necessary metric.

**Summary of Insights :**

* HYPOTHESIS 1: It is common for all countries that majority of their medals are concentrated in a certain number of sports.
* HYPOTHESIS 2: The country/countries won more number of medals because it/they have sent greater number of participants.
* Changed HYPOTHESIS 2: The United States won the highest no. of medals because it sent a greater no. of participants.
* HYPOTHESIS 3: At most of the times, a player wins a medal in different events of the same sport.
* ***Discoveries about data:***
* No. of medals won in sports is not only concentrated to age group till 35 years.
* The list of the top 10 sports in which medals won by both genders are almost similar.
* The countries’ overall best performance is subjected to more sports played in summer compared to winter. Difference in seasonal games and participation is the main reason for up-down curve in the history graph.

**Recommendations and Next Steps:.....**

* Age-group analysis could be done deeper. For example, which age-group tends to be most favourite to win a specific event/sport.
* Is country winning a sport that has more popularity in their region?
* Countries‘ popular players has not been analyzed. Analyzing which can give interesting incidents, statistics etc.
* Physical attributes such as height and width could be responsible for winning some sports/events, which can be analyzed.