PRESENTING MY FINDINGS

(Olympics Dataset)

- by AKASH CHOVATIYA

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.

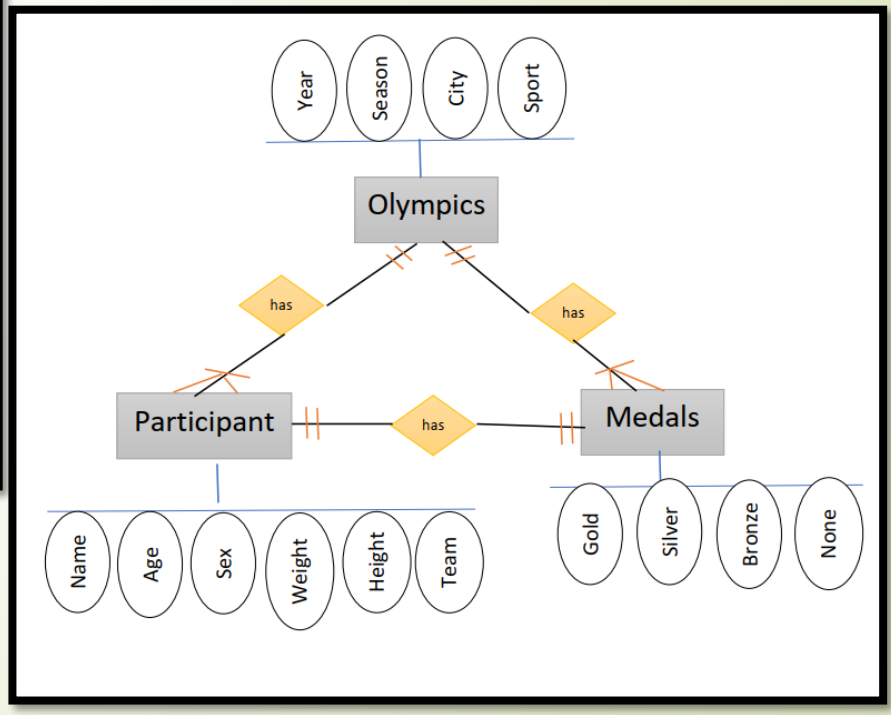
Entity Relationship Diagram (ERD) of the data

Olympics	
Year	int
Season	str
City	str
Sport_Event	str
Prize	List[Medals]
Participants	List[Athletes]

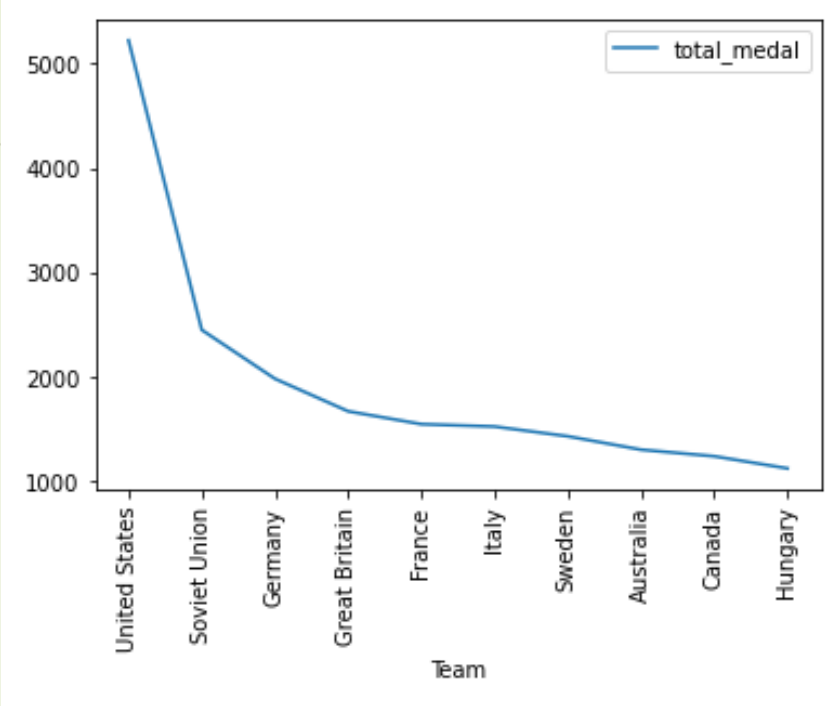
Medals	
Gold	str
Silver	str
Bronze	str
None	str

Athlete	
Name	str
Age	int
Weight	float
Height	float
Sex	str
Team	List[Country]

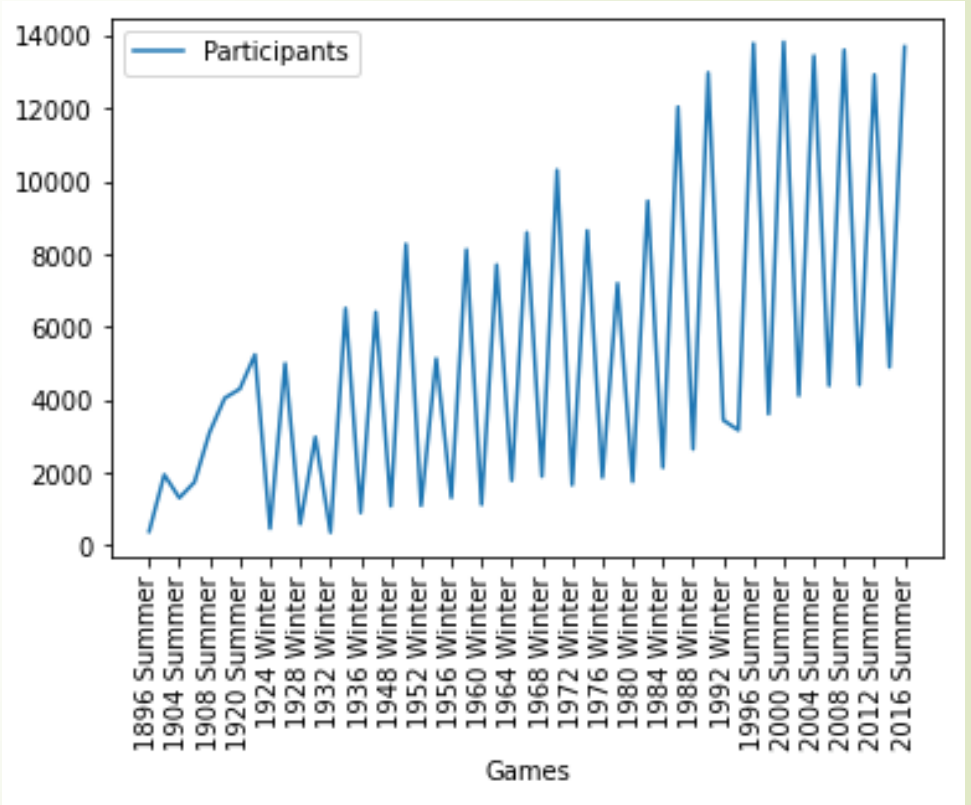
Country	
NOC	str
Region	str
Notes	str



Initial Data Exploration

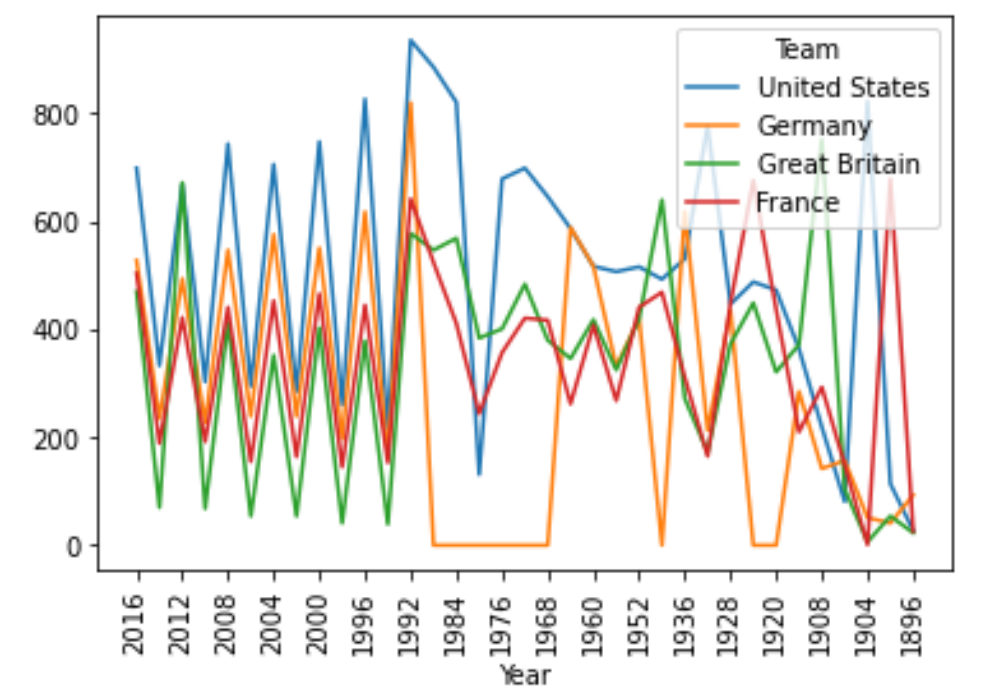


Top 10 countries vs. Total medals won by them

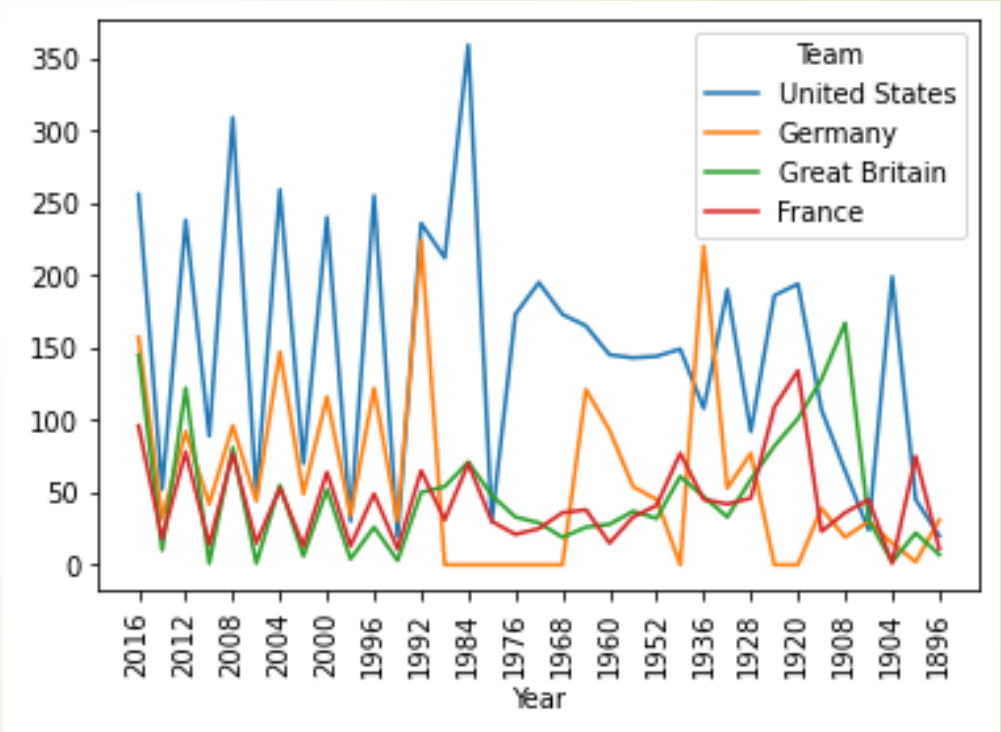


Games vs. Total No. Of Participants

Initial Data Exploration



Participants over the years from the top 4 countries



Medals won over the years by the top 4 countries

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

	sport_percent	sport_USA	cumulative_percent
0	20.521173	Athletics	20.521173
1	20.425369	Swimming	40.946541
2	6.533819	Basketball	47.480360
3	6.380533	Rowing	53.860893
4	5.288369	Ice Hockey	59.149262
5	3.698026	Shooting	62.847289
6	3.391454	Gymnastics	66.238743
7	2.682506	Diving	68.921249
8	2.529220	Equestrianism	71.450469
9	2.471738	Water Polo	73.922207

United States

	sport_percent	sport_Germany	cumulative_percent
0	10.282258	Rowing	10.282258
1	9.173387	Hockey	19.455645
2	8.014113	Canoeing	27.469758
3	7.762097	Athletics	35.231855
4	7.661290	Swimming	42.893145
5	7.056452	Equestrianism	49.949597
6	5.241935	Football	55.191532
7	4.737903	Cycling	59.929435
8	4.032258	Biathlon	63.961694
9	3.881048	Gymnastics	67.842742

Germany

HYPOTHESIS 1: Medals are concentrated in a few sports

	sport_percent	sport_UK	cumulative_percent
0	19.964136	Athletics	19.964136
1	12.970711	Rowing	32.934848
2	10.101614	Cycling	43.036461
3	7.591154	Swimming	50.627615
4	7.053198	Hockey	57.680813
5	5.200239	Shooting	62.881052
6	4.961148	Equestrianism	67.842200
7	3.705918	Sailing	71.548117
8	3.347280	Boxing	74.895397
9	3.048416	Gymnastics	77.943814

Great Britain

	sport_percent	sport_FRA	cumulative_percent
0	20.000000	Fencing	20.000000
1	9.741935	Cycling	29.741935
2	7.096774	Athletics	36.838710
3	5.741935	Rowing	42.580645
4	4.967742	Equestrianism	52.516129
5	4.967742	Handball	52.516129
6	4.580645	Swimming	57.096774
7	4.451613	Shooting	61.548387
8	3.741935	Gymnastics	65.290323
9	3.161290	Canoeing	71.612903

France

HYPOTHESIS 1: Medals are concentrated in a few sports

- The tables that are presented in previous two slides shows that most of the medals won by top 4 best performing countries are concentrated in a few sports.

Country	% medals_concentration in top 10 sports
United States	73.922207
Germany	67.842742
Great Britain	77.943814
France	71.612903

- Above table shows 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

	team	number_players	percent_gold	percent_silver	percent_bronze	percent_win
0	United States	17847	13.862274	8.472012	6.908724	29.243010
1	France	11988	3.795462	4.320988	4.813146	12.929596
2	Great Britain	11404	4.551035	5.103472	5.015784	14.670291
3	Germany	9326	7.280721	6.723140	7.269998	21.273858

- 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 has 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 can be found in the table shown at the right side. The tables shows that even in the ‚Athletics' category sport, there are different events which need different physical abilities to win in a event.

	Name	Sport	Event
0	Adam McCright Nelson	Athletics	Athletics Men's Shot Put
1	Allyson Michelle Felix	Athletics	Athletics Women's 200 metres
2	Andrew Rock	Athletics	Athletics Men's 4 x 400 metres Relay
3	Bernard Rollen Williams, III	Athletics	Athletics Men's 200 metres
4	Bryan Ezra Tsumoru Clay	Athletics	Athletics Men's Decathlon
5	Christine Arron	Athletics	Athletics Women's 4 x 100 metres Relay
6	Coby Miller	Athletics	Athletics Men's 4 x 100 metres Relay
7	Darold Andre Williamson	Athletics	Athletics Men's 4 x 400 metres Relay
8	Darren Andrew Campbell	Athletics	Athletics Men's 4 x 100 metres Relay
9	Darvis Darell Patton	Athletics	Athletics Men's 4 x 100 metres Relay
10	De'Hashia Tonnek "DeeDee" Trotter	Athletics	Athletics Women's 4 x 400 metres Relay
11	Deena Michelle Drossin-Kastor	Athletics	Athletics Women's Marathon
12	Derrick Keith Brew	Athletics	Athletics Men's 400 metres
13	Derrick Keith Brew	Athletics	Athletics Men's 4 x 400 metres Relay
14	Dwight D. Phillips	Athletics	Athletics Men's Long Jump

Analysis from diving deeper and going broader

- 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

- 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.

	Sport	MIN(age)	MAX(age)
0	Alpine Skiing	18.0	32.0
1	Alpinism	22.0	57.0
2	Archery	18.0	53.0
3	Art Competitions	20.0	60.0
4	Athletics	15.0	41.0
5	Baseball	20.0	37.0
6	Basketball	19.0	37.0
7	Biathlon	19.0	35.0
8	Bobsleigh	31.0	41.0
9	Boxing	16.0	37.0

Analysis from diving deeper and going broader: Gender-wise Analysis

- 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.

	Sex	Sport	medals_won
0	F	Swimming	625
1	F	Athletics	389
2	F	Rowing	178
3	F	Football	147
4	F	Basketball	131
5	F	Equestrianism	122
6	F	Gymnastics	120
7	F	Hockey	113
8	F	Ice Hockey	101
9	F	Fencing	78

	Sex	Sport	medals_won
0	M	Athletics	1280
1	M	Swimming	791
2	M	Rowing	665
3	M	Cycling	421
4	M	Fencing	402
5	M	Shooting	355
6	M	Equestrianism	310
7	M	Basketball	248
8	M	Gymnastics	243
9	M	Hockey	217

Analysis from diving deeper and going broader: Season-wise Analysis

	Season	Team	total_medal
0	Summer	United States	4686
1	Summer	Soviet Union	2061
2	Summer	Germany	1687
3	Summer	Great Britain	1598
4	Summer	France	1408

	Season	Team	total_medal
0	Winter	Canada	575
1	Winter	United States	533
2	Winter	Norway	443
3	Winter	Sweden	428
4	Winter	Finland	426

- What impact does the season have on medals won by country?
- The above tables shows that 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.

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.

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.