

Coursera Data Science Capstone Project Findings

Jacob Ryan

Overview of the Olympics

- The Olympics were originally held from 776 BC to 393 AD in Olympia, Greece.
- But the first modern olympics took place 1503 years later in 1896.
- With both the Summer and Winter games being held every 4 years there has been numerous athletes, events, and countries over the years.

Perspective, perspective

- With the modern Olympics being around for over 200 years, you, the audience, might have several question regarding with certain statistics from the events over the years.
- This presentation will answer three question I've asked the data set, from a non sports fan viewer's perspective, and then show the graphs of my findings.

My three questions

The questions I want to ask my data set are:

- Do certain regions excel better at certain sports and have a higher win average when compared to others?
- Which athletes had a better medal placement average in the years they competed?
- What is the average 'Age', 'Height', and 'Weight' for sports on athletes that won a medal?

Hypothesis for my data set

Before I started searching my data set the hypothesis I had were:

- If an athlete has appeared the most number of times in a sport then they should have a high medal win percentage.
- If a country gets more snow year around then they should yield better results in the winter events when compared to summer events.
- If a country has had more athletes throughout the events then they should have a higher medal count?

Do certain regions excel better at certain sports and have a higher win average when compared to others? (1)

- I first started by getting an overall medal count by each region of the top 20 countries with the most medals.

[Source query](#)

	Total_medals ▲	region ▲
1	5637	USA
2	3947	Russia
3	3756	Germany
4	2068	UK
5	1777	France
6	1637	Italy
7	1536	Sweden
8	1352	Canada
9	1349	Australia
10	1135	Hungary
11	1040	Netherlands
12	1033	Norway
13	993	China
14	913	Japan
15	900	Finland
16	691	Switzerland
17	653	Romania
18	644	Czech Republic
19	638	South Korea
20	597	Denmark

↓ 20 rows | 3.62 seconds runtime

Do certain regions excel better at certain sports and have a higher win average when compared to others? (2)

- I then separated the medals in to their respective placements, Gold, Silver, and Bronze.

[Source query](#)

	Gold_medal ▲	Silver_Medal ▲	Bronze_Medal ▲	Total_medals ▲	region ▲
1	2638	1641	1358	5637	USA
2	1599	1170	1178	3947	Russia
3	1301	1195	1260	3756	Germany
4	678	739	651	2068	UK
5	501	610	666	1777	France
6	575	531	531	1637	Italy
7	479	522	535	1536	Sweden
8	463	438	451	1352	Canada
9	368	459	522	1349	Australia
10	432	332	371	1135	Hungary
11	287	340	413	1040	Netherlands
12	378	361	294	1033	Norway
13	351	349	293	993	China
14	247	309	357	913	Japan
15	198	270	432	900	Finland
16	175	248	268	691	Switzerland
17	161	200	292	653	Romania
18	123	262	259	644	Czech Republic
19	221	232	185	638	South Korea
20	179	241	177	597	Denmark

Do certain regions excel better at certain sports and have a higher win average when compared to others? (3)

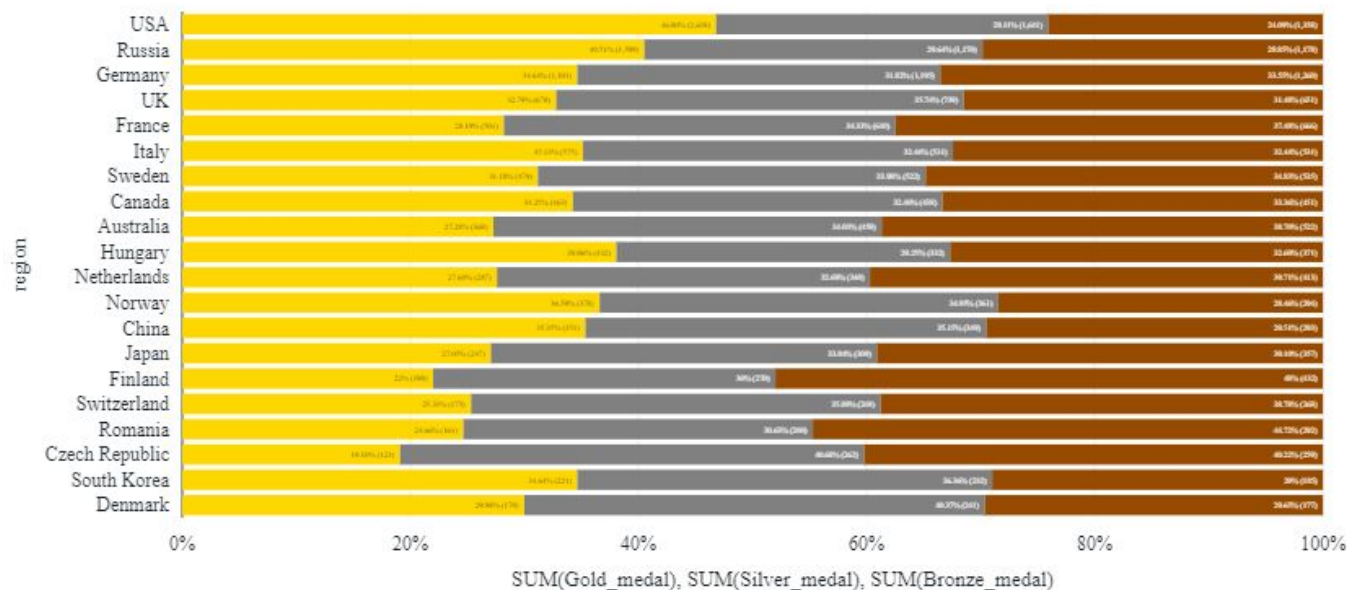
- Next I divided each country's medal type by the total medals to get each country's medal win percentage.

[Source query](#)

Table Visualization 1 +							
	region	Total_medals	Gold_medal	Gold_Percent	Silver_medal	Silver_Percent	Bronze_medal
1	USA	5637	2638	0.47	1641	0.29	1358
2	Russia	3947	1599	0.41	1170	0.3	1178
3	Germany	3756	1301	0.35	1195	0.32	1260
4	UK	2068	678	0.33	739	0.36	651
5	France	1777	501	0.28	610	0.34	666
6	Italy	1637	575	0.35	531	0.32	531
7	Sweden	1536	479	0.31	522	0.34	535
8	Canada	1352	463	0.34	438	0.32	451
9	Australia	1349	368	0.27	459	0.34	522
10	Hungary	1135	432	0.38	332	0.29	371
11	Netherlands	1040	287	0.28	340	0.33	413
12	Norway	1033	378	0.37	361	0.35	294
13	China	993	351	0.35	349	0.35	293
14	Japan	913	247	0.27	309	0.34	357
15	Finland	900	198	0.22	270	0.3	432
16	Switzerland	691	175	0.25	248	0.36	268
17	Romania	653	161	0.25	200	0.31	292
18	Czech Republic	644	123	0.19	262	0.41	259
19	South Korea	638	221	0.35	232	0.36	185
20	Denmark	597	179	0.3	241	0.4	177

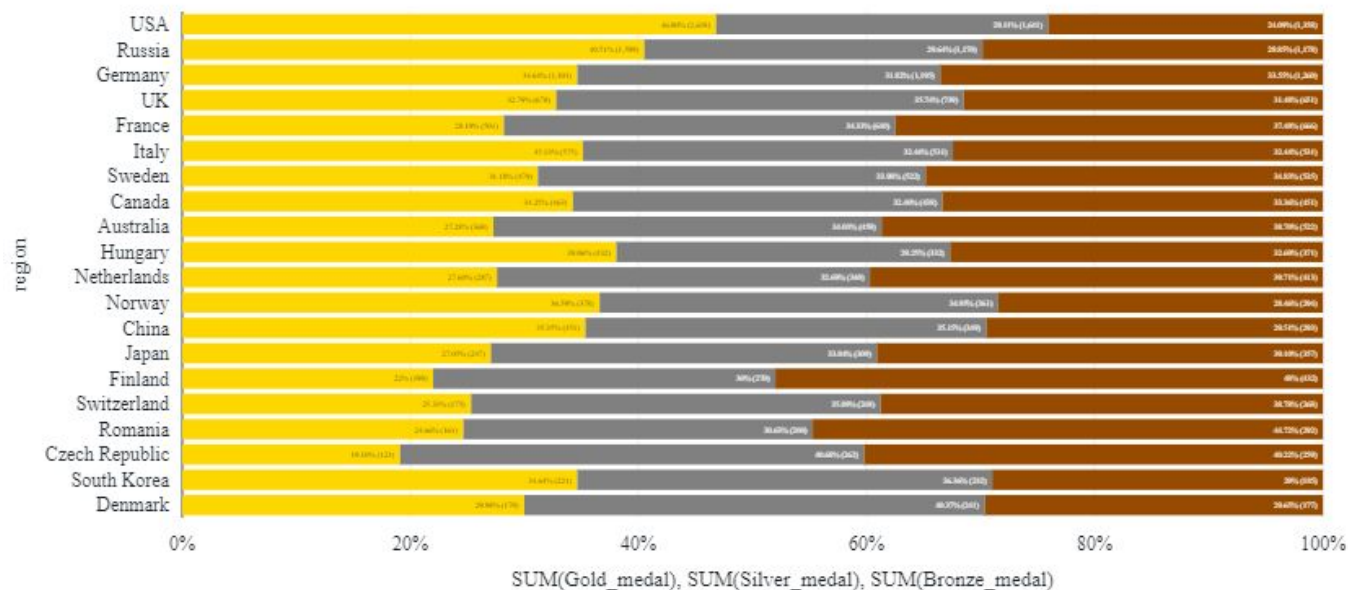
Do certain regions excel better at certain sports and have a higher win average when compared to others? (4)

- Then I went and put my findings in a bar graph for visualization.



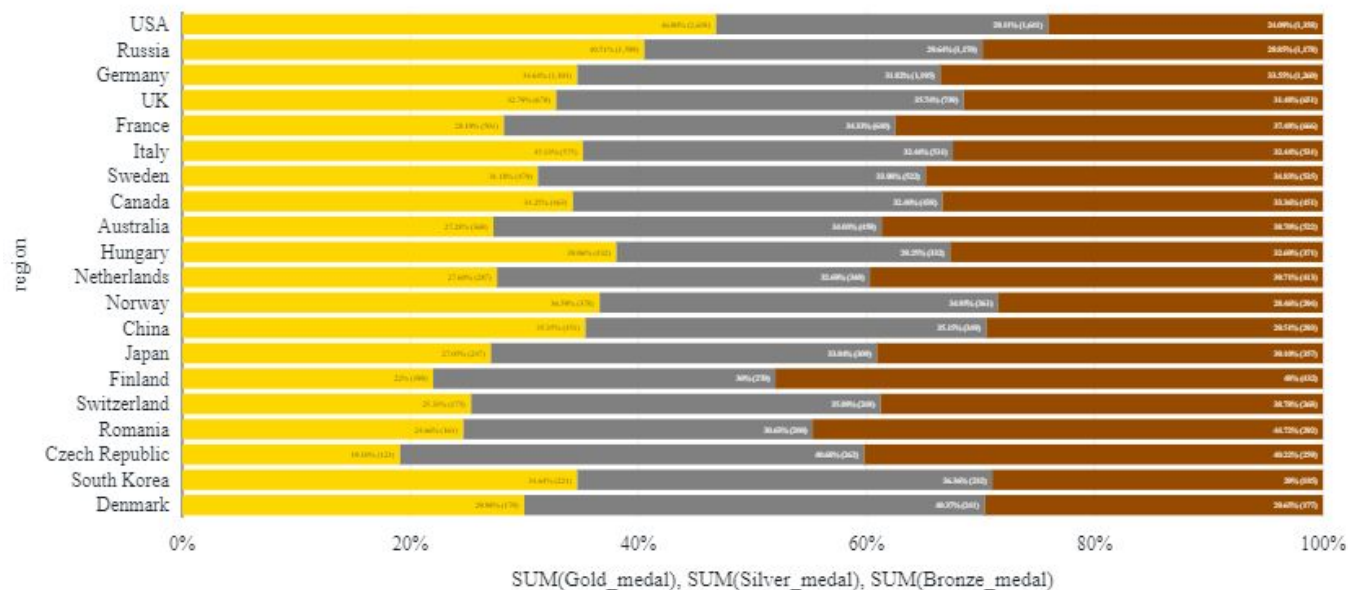
Do certain regions excel better at certain sports and have a higher win average when compared to others? (5)

- USA's medals 46.80% of medal wins were gold medals at 2638 medals out of a total of 5637.



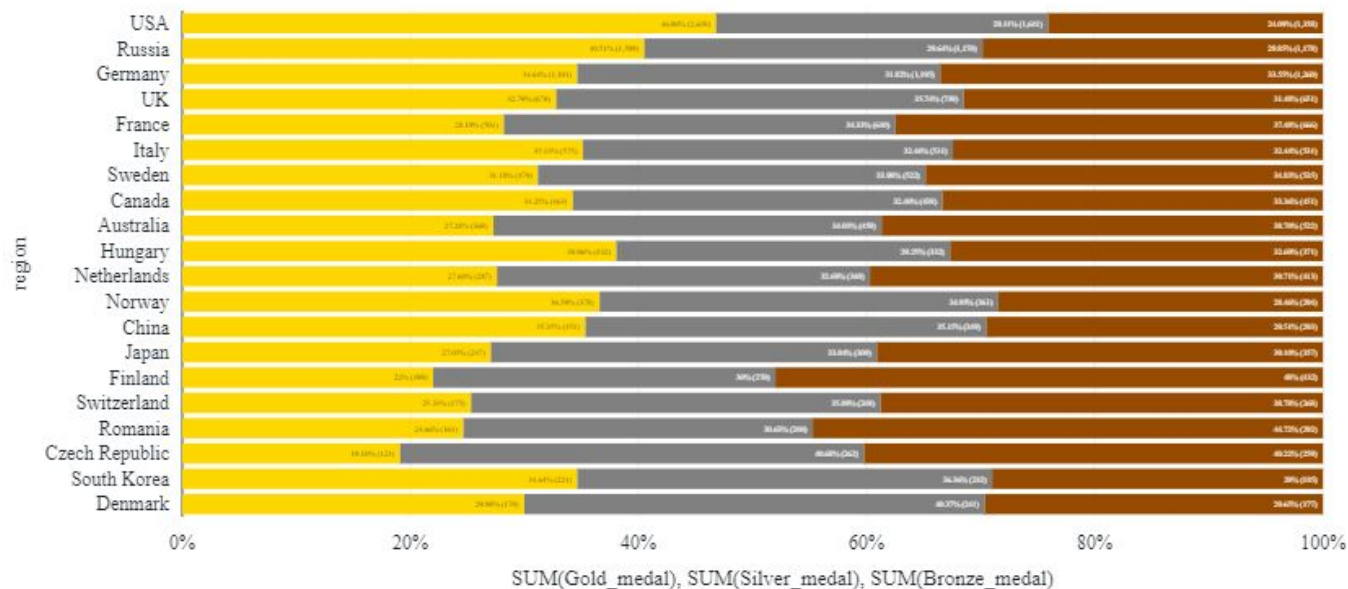
Do certain regions excel better at certain sports and have a higher win average when compared to others? (6)

- The Czech Republic has the highest silver medal percentage at 40.68% of medal wins at 262 silver medal wins out of 644 total medals.



Do certain regions excel better at certain sports and have a higher win average when compared to others? (7)

- And Finland has the highest Bronze medal percentage at 48% of medal wins at 432 bronze medals out of 900 total medals.



Do certain regions excel better at certain sports and have a higher win average when compared to others? (8)

- With USA having a total of 5637 medals and a 40.8% Gold medal win percentage the USA has the Gold medal count when compared to the other countries.
- While Russia is in second via total medal counts at 3947 total medals, the Czech Republic and Finland have higher Silver medal and Bronze medal win percentage. With Czech having a 40.68% Silver wins and Finland's 48% Bronze medal wins.

Do certain regions excel better at certain sports and have a higher win average when compared to others? (9)

- Now to see which countries perform better in either the Summer or Winter events I separated the medal win into the Summer games...

[Source query](#)

Table		Visualization 1		+				
	region	season	Total_medals	Gold_medal	Gold_Percent	Silver_medal	Silver_Percent	Bro
1	USA	Summer	5002	2472	0.49	1333	0.27	119
2	Russia	Summer	3188	1220	0.38	974	0.31	994
3	Germany	Summer	3126	1075	0.34	987	0.32	106
4	UK	Summer	1985	636	0.32	729	0.37	620
5	France	Summer	1627	465	0.29	575	0.35	587
6	Italy	Summer	1446	518	0.36	474	0.33	454
7	Australia	Summer	1333	362	0.27	456	0.34	515
8	Hungary	Summer	1123	432	0.38	328	0.29	363
9	Sweden	Summer	1108	354	0.32	396	0.36	358
10	Netherlands	Summer	918	245	0.27	302	0.33	371
11	China	Summer	913	335	0.37	319	0.35	259
12	Japan	Summer	850	230	0.27	287	0.34	333
13	Canada	Summer	741	158	0.21	239	0.32	344
14	Romania	Summer	651	161	0.25	200	0.31	290
15	Denmark	Summer	592	179	0.3	236	0.4	177
16	Norway	Summer	590	227	0.38	196	0.33	167
17	South Korea	Summer	552	171	0.31	206	0.37	175
18	Poland	Summer	538	111	0.21	185	0.34	242
19	Serbia	Summer	532	157	0.3	216	0.41	159
20	Spain	Summer	487	109	0.22	243	0.5	135

Do certain regions excel better at certain sports and have a higher win average when compared to others? (10)

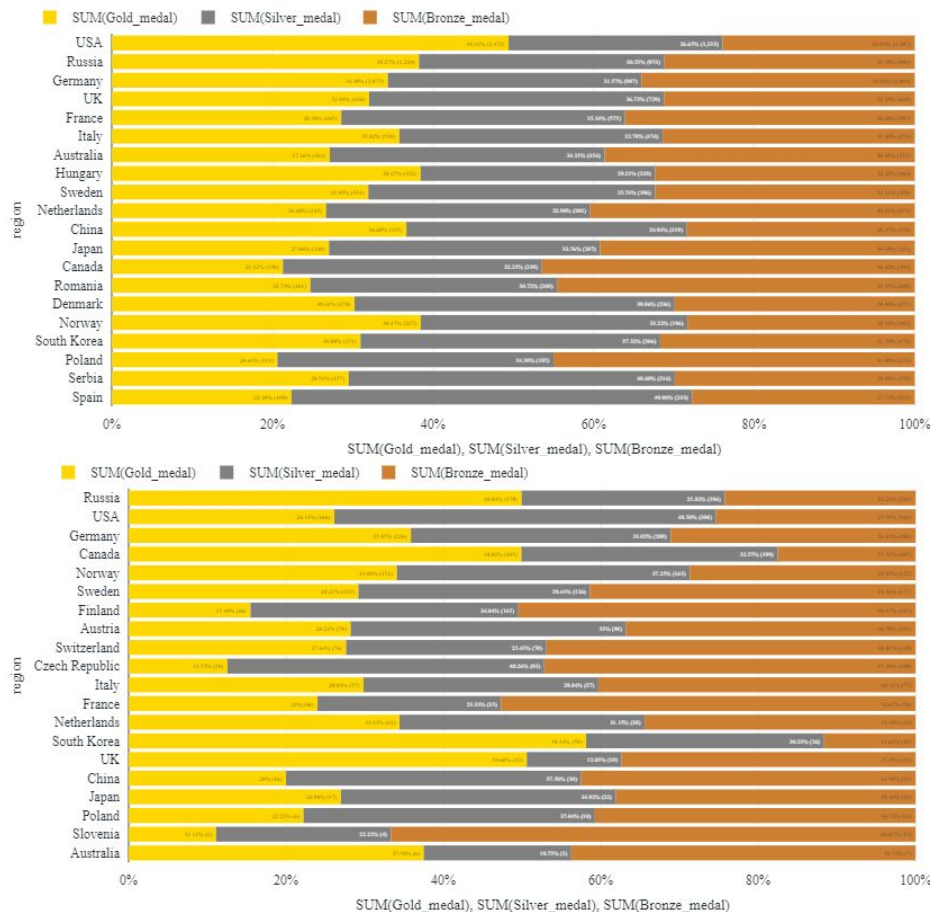
- And into Winter games...

[Source query](#)

	region	season	Total_medals	Gold_medal	Gold_Percent	Silver_medal	Silver_Percent
1	Russia	Winter	759	379	0.5	196	0.26
2	USA	Winter	635	166	0.26	308	0.49
3	Germany	Winter	630	226	0.36	208	0.33
4	Canada	Winter	611	305	0.5	199	0.33
5	Norway	Winter	443	151	0.34	165	0.37
6	Sweden	Winter	428	125	0.29	126	0.29
7	Finland	Winter	426	66	0.15	145	0.34
8	Austria	Winter	280	79	0.28	98	0.35
9	Switzerland	Winter	275	76	0.28	70	0.25
10	Czech Republic	Winter	231	29	0.13	93	0.4
11	Italy	Winter	191	57	0.3	57	0.3
12	France	Winter	150	36	0.24	35	0.23
13	Netherlands	Winter	122	42	0.34	38	0.31
14	South Korea	Winter	86	50	0.58	26	0.3
15	UK	Winter	83	42	0.51	10	0.12
16	China	Winter	80	16	0.2	30	0.38
17	Japan	Winter	63	17	0.27	22	0.35
18	Poland	Winter	27	6	0.22	10	0.37
19	Slovenia	Winter	18	2	0.11	4	0.22
20	Australia	Winter	16	6	0.38	3	0.19

Do certain regions excel better at certain sports and have a higher win average when compared to others? (11)

- Even though USA has a higher medal count overall, Russia has a higher gold medal count than USA with half of Russia's Winter game medals begin gold.
- Notice how Canada is 4th in medal count won from the Winter games when compared to it's 14th position in Summer game medals



Which athletes had a better medal placement average in the years they competed? (1)

- I first started by getting an overall athlete count by each region of the top 20 countries with the most athletes.

[Source query](#)

	total_athletes ▲	region ▲
1	9652	USA
2	7541	Germany
3	6273	UK
4	6161	France
5	5597	Russia
6	4921	Italy
7	4810	Canada
8	4036	Japan
9	3868	Australia
10	3782	Sweden
11	2964	Poland
12	2937	Netherlands
13	2868	Switzerland
14	2815	China
15	2741	Czech Republic
16	2681	Hungary
17	2637	Spain
18	2346	Finland
19	2334	Austria
20	2289	South Korea

Which athletes had a better medal placement average in the years they competed? (2)

- Then I take those top 20 regions and I filter it down to each athlete, to what sport they competed in, and how many medals they won total.

[Source query](#)

	name	sport	Total_medals	region
1	Michael Fred Phelps, II	Swimming	28	USA
2	Larysa Semenivna Latynina (Diriy-)	Gymnastics	18	Russia
3	Nikolay Yefimovich Andrianov	Gymnastics	15	Russia
4	Takashi Ono	Gymnastics	13	Japan
5	Ole Einar Bjrndalen	Biathlon	13	Norway
6	Edoardo Mangiarotti	Fencing	13	Italy
7	Borys Anfiyanovych Shakhlin	Gymnastics	13	Russia
8	Dara Grace Torres (-Hoffman, -Minas)	Swimming	12	USA
9	Natalie Anne Coughlin (-Hall)	Swimming	12	USA
10	Birgit Fischer-Schmidt	Canoeing	12	Germany
11	Aleksey Yuryevich Nemov	Gymnastics	12	Russia
12	Paavo Johannes Nurmi	Athletics	12	Finland
13	Jennifer Elisabeth "Jenny" Thompson (-Cumpelik)	Swimming	12	USA
14	Sawao Kato	Gymnastics	12	Japan
15	Ryan Steven Lochte	Swimming	12	USA
16	Matthew Nicholas "Matt" Biondi	Swimming	11	USA
17	Carl Townsend Osburn	Shooting	11	USA
18	Viktor Ivanovych Chukarin	Gymnastics	11	Russia
19	Vra slavsk (-Odloilov)	Gymnastics	11	Czech Re
20	Mark Andrew Spitz	Swimming	11	USA
21	Polina Hryhorivna Astakhova	Gymnastics	10	Russia
22	Raisa Petrovna Smetanina	Cross Country Skiing	10	Russia

Which athletes had a better medal placement average in the years they competed? (3)

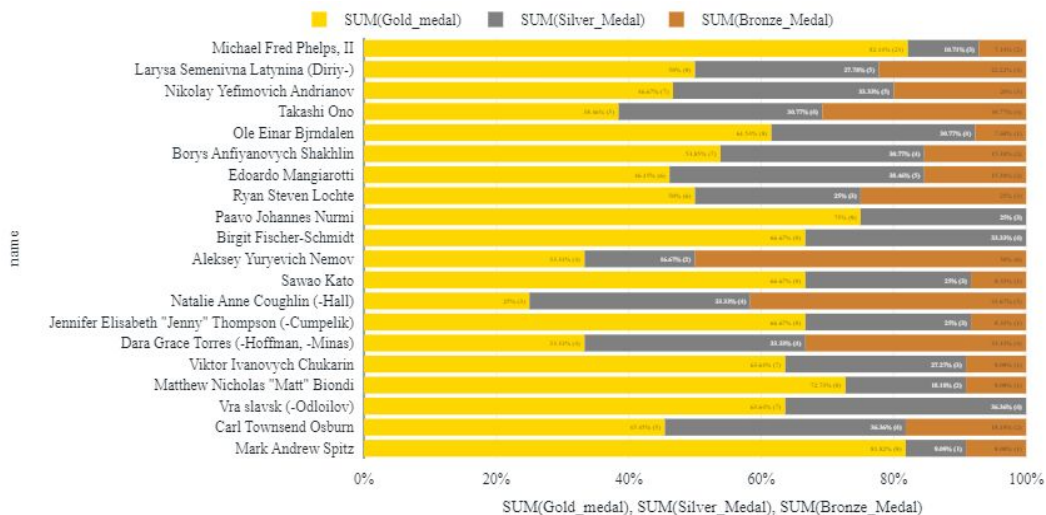
- Limiting to the top 20 athletes I then added to the query the athlete's minimum and maximum year, their total medal count, as well as type of medal.

[Source query](#)

	name	region	sport	min(year)	max(year)	Total_me
1	Michael Fred Phelps, II	USA	Swimming	2000	2016	28
2	Larysa Semenivna Latynina (Diriy-)	Russia	Gymnastics	1956	1964	18
3	Nikolay Yefimovich Andrianov	Russia	Gymnastics	1972	1980	15
4	Takashi Ono	Japan	Gymnastics	1952	1964	13
5	Ole Einar Bjørndalen	Norway	Biathlon	1994	2014	13
6	Borys Anfiyanovich Shakhlin	Russia	Gymnastics	1956	1964	13
7	Edoardo Mangiarotti	Italy	Fencing	1936	1960	13
8	Ryan Steven Lochte	USA	Swimming	2004	2016	12
9	Paavo Johannes Nurmi	Finland	Athletics	1920	1928	12
10	Birgit Fischer-Schmidt	Germany	Canoeing	1980	2004	12
11	Aleksey Yuryevich Nemov	Russia	Gymnastics	1996	2004	12
12	Saiwao Kato	Japan	Gymnastics	1968	1976	12
13	Natalie Anne Coughlin (-Hall)	USA	Swimming	2004	2012	12
14	Jennifer Elisabeth "Jenny" Thompson (-Cumpelik)	USA	Swimming	1992	2004	12
15	Dara Grace Torres (-Hoffman, -Minas)	USA	Swimming	1984	2008	12
16	Viktor Ivanovich Chukarin	Russia	Gymnastics	1952	1956	11
17	Matthew Nicholas "Matt" Biondi	USA	Swimming	1984	1992	11
18	Vra slavsk (-Odloillov)	Czech Republic	Gymnastics	1960	1968	11
19	Carl Townsend Osburn	USA	Shooting	1912	1924	11
20	Mark Andrew Spitz	USA	Swimming	1968	1972	11

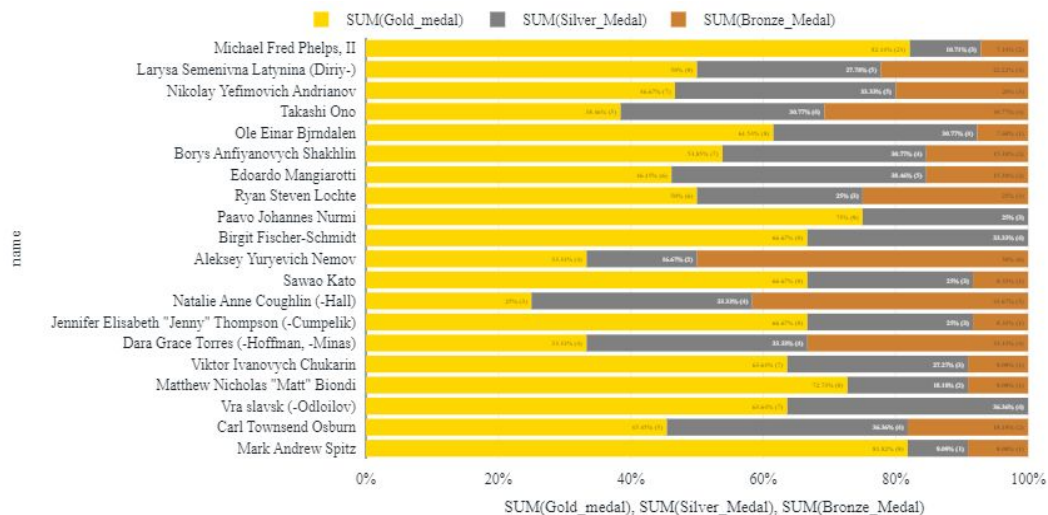
Which athletes had a better medal placement average in the years they competed? (4)

- By sorting my findings into a bar graph we can see how the top 20 athletes' medals are distributed amongst their placement.
- Michael Phelps has won a total of 28 medals in his 16 years of swimming for the Olympics, averaging 7 medal wins at every Olympics since they only happen every 4 years.
- Winning an astonishing 23 gold medals out of his 28 total, Phelps has a 82.14% gold medal win rate when compared to his silver and bronze placements.



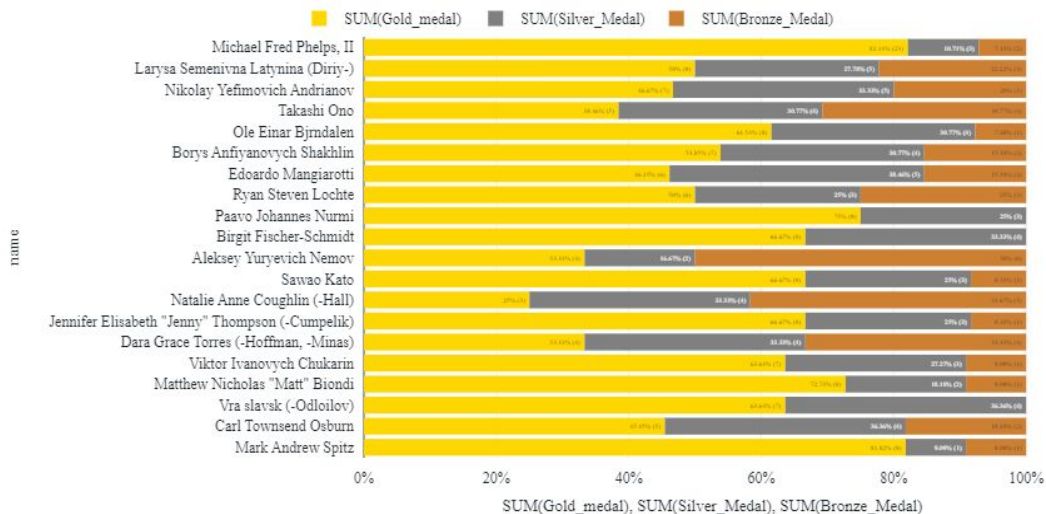
Which athletes had a better medal placement average in the years they competed? (4)

- By sorting my findings into a bar graph we can see how the top 20 athletes' medals are distributed amongst their placement.
- Michael Phelps has won a total of 28 medals in his 16 years of swimming for the Olympics, averaging 7 medal wins at every Olympics since they only happen every 4 years.
- Winning an astonishing 23 gold medals out of his 28 total, Phelps has a 82.14% gold medal win rate when compared to his silver and bronze placements.



Which athletes had a better medal placement average in the years they competed? (5)

- But Phelps isn't the only athlete that has a gold medal placement that's over 50%.
- Looking at Mark Spitz, he's won 11 medals in his 4 years of swimming for the Olympics averaging 5.5 medal wins each Olympics. Winning 9 gold medals and a single medal in silver and bronze, Spitz's gold medal average is an 81.82% win rate in gold medals.
- Looking at more athletes like Paavo Nurmi, Matthew Biondi, and Birgit Fischer-Schmidt they all have a gold medal placement average that's over 50% of their total medal wins.



What is the average 'Age', 'Height', and 'Weight' for sports on athletes that won a medal? (1)

- First thing I did was I sorted my data into two sets per se, Men's events and Women's events.

Source Query

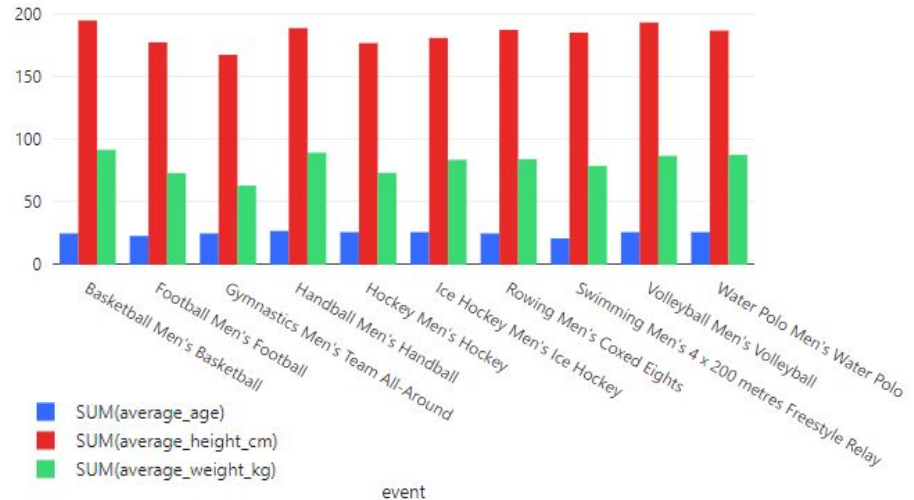
	event	Total_medals
1	Football Men's Football	1269
2	Ice Hockey Men's Ice Hockey	1230
3	Hockey Men's Hockey	1050
4	Water Polo Men's Water Polo	866
5	Rowing Men's Coxed Eights	730
6	Gymnastics Men's Team All-Around	713
7	Basketball Men's Basketball	687
8	Handball Men's Handball	588
9	Volleyball Men's Volleyball	495
10	Swimming Men's 4 x 200 metres Freestyle Relay	395

	event	Total_medals
1	Hockey Women's Hockey	478
2	Volleyball Women's Volleyball	474
3	Handball Women's Handball	472
4	Basketball Women's Basketball	393
5	Gymnastics Women's Team All-Around	390
6	Swimming Women's 4 x 100 metres Freestyle Relay	362
7	Football Women's Football	302
8	Ice Hockey Women's Ice Hockey	300
9	Rowing Women's Coxed Eights	299
10	Swimming Women's 4 x 100 metres Medley Relay	272

What is the average 'Age', 'Height', and 'Weight' for sports on athletes that won a medal? (2)

- After finding the top 10 events with the most medals won for both Men's and Women's, I queried to find the average 'Age', 'Height', and 'Weight' for athletes that won a medal in those events.
- With men's football, ice hockey, and hockey each having over 1000 athletes that won a medal there is no shortage of enough information to calculate the average. With men's Basketball and Volleyball having a similar average was expected.

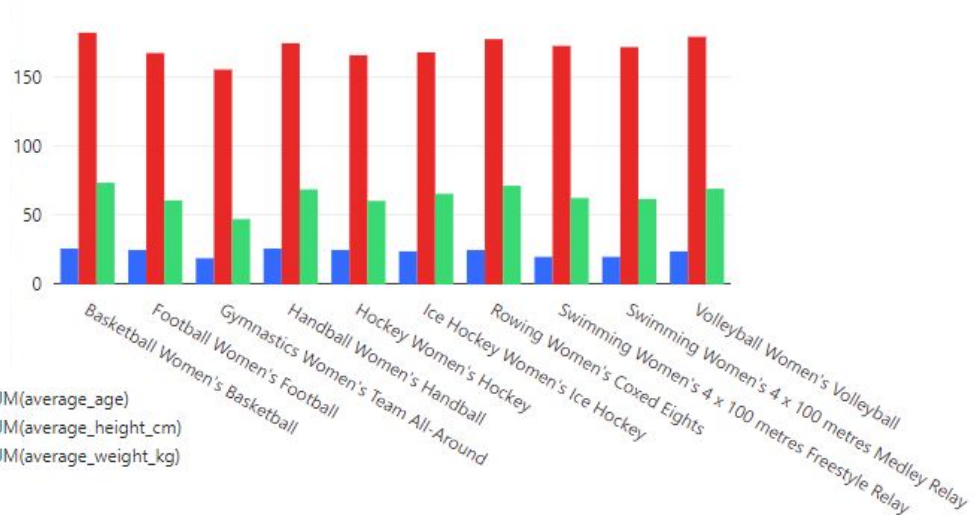
	event	Total_medals	average_age	average_height	average_weight
1	Football Men's Football	1269	23	177.48 cm	73.09 kg
2	Ice Hockey Men's Ice Hockey	1230	26	180.99 cm	83.78 kg
3	Hockey Men's Hockey	1050	26	176.85 cm	73.34 kg
4	Water Polo Men's Water Polo	866	26	186.80 cm	87.71 kg
5	Rowing Men's Coxed Eights	730	25	187.38 cm	84.18 kg
6	Gymnastics Men's Team All-Around	713	25	167.54 cm	63.13 kg
7	Basketball Men's Basketball	687	25	194.87 cm	91.68 kg
8	Handball Men's Handball	588	27	188.78 cm	89.39 kg
9	Volleyball Men's Volleyball	495	26	193.27 cm	86.93 kg
10	Swimming Men's 4 x 200 metres Freestyle Relay	395	21	185.22 cm	78.87 kg



What is the average 'Age', 'Height', and 'Weight' for sports on athletes that won a medal? (2)

- While the women's events of athletes that won a medal aren't as large when compared to the men's, having nearly 500 total medals in hockey, volleyball, and handball is still more than enough data to pull from.

	event	Total_medals	average_age	average_height	average_weight
1	Hockey Women's Hockey	478	25	166.13 cm	60.53 kg
2	Volleyball Women's Volleyball	474	24	179.49 cm	69.33 kg
3	Handball Women's Handball	472	26	174.84 cm	68.88 kg
4	Basketball Women's Basketball	393	26	182.45 cm	73.69 kg
5	Gymnastics Women's Team All-Around	390	19	155.72 cm	47.32 kg
6	Swimming Women's 4 x 100 metres Freestyle Relay	362	20	172.89 cm	62.60 kg
7	Football Women's Football	302	25	167.68 cm	60.93 kg
8	Ice Hockey Women's Ice Hockey	300	24	168.21 cm	65.71 kg
9	Rowing Women's Coxed Eights	299	25	177.77 cm	71.50 kg
10	Swimming Women's 4 x 100 metres Medley Relay	272	20	172.01 cm	61.95 kg



[Source Query](#)

Revisiting my hypothesis

“If an athlete has appeared the most number of times in a sport then they should have a high medal win percentage.”

This hypothesis was proven to be true because when querying the athletes and looking at the years they competed you can calculate that they've participated in more than 2 olympics

“If a country gets more snow year around then they should yield better results in the winter events when compared to summer events.”

This hypothesis was also proven to be true with countries like Canada, Russia, and Norway yielding more wins in the winter events.

“If a country has had more athletes throughout the events then they should have a higher medal count?”

My final hypothesis was also proven to be true. When querying the number of athletes by country we saw countries like USA, Germany, and UK with the highest athlete count and those countries were constantly in the top 10 when comparing countries.

Metric tracking

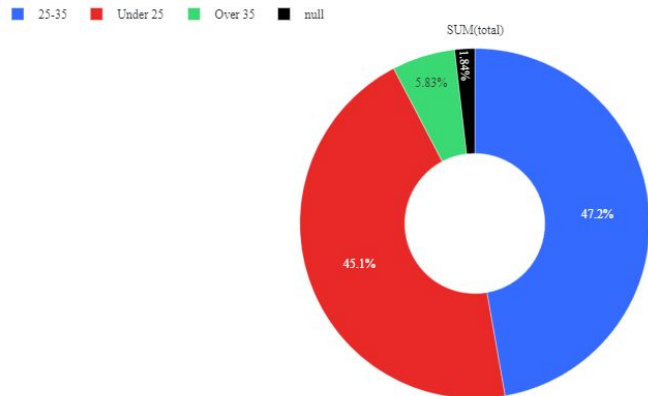
As I was querying my dataset I wanted to track down the age group of the athletes who've won medals.

I grouped them into 'Under 25', '25-53' and 'over 35' and then counted the medals each age group won. My results returned showed that the '25-35' group has the most medals at 18791 total while the 'Under 25' age group is not far behind at 17939 medals and the 'Over 35' has 2321 total medals.

Grouping the age groups by their total medal counts we can see that the age group of '25-53' has 47.2% of medals won while 'Under 25' has 54.1% of medals won and 'Over 35' and ages that weren't recorded make up the rest at 7%.

	age	total	Gold_medal	Silver_Medal	Bronze_Medal
1	25-35	18791	6317	6227	6247
2	Under 25	17939	6129	5806	6004
3	Over 35	2321	778	788	755
4	null	732	148	295	289

[Query Source](#)



Conclusion

- Working with the Olympic dataset and querying all my questions was an educational adventure.
- I would like to see the dataset expanded upon and see the athlete's times/scores but with how there's numerous events and how they're scored that can be quite difficult to accomplish with the current form the dataset is in.
- Changes I would make to the dataset is in my medal queries when an athlete didn't win a medal their medal entry was marked as 'NA' instead of being NULL so I had to make a case statement when querying to change the NAs to NULL to get the answers I was looking for.

Query sources

Here is the Databricks link for all of my queries: [CapstoneQueryies](#)

If the link doesn't work then I have provided screen shots of my query searches in the rest of this presentation

Question 1 Source Query 1

```
select
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
    end) as Total_medals,
  n.region
from
  athlete_events_1_csv as a
inner join
  noc_regions_3_csv as n
on
  a.NOC = n.NOC
group by
  n.region
order by
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
    end) desc
limit
  20
```

Table ▾ +			
	Total_medals ▲	region ▲	
1	5637	USA	
2	3947	Russia	
3	3756	Germany	
4	2068	UK	
5	1777	France	
6	1637	Italy	
7	1536	Sweden	
8	1352	Canada	
9	1349	Australia	
10	1135	Hungary	
11	1040	Netherlands	
12	1033	Norway	
13	993	China	
14	913	Japan	
15	900	Finland	
16	691	Switzerland	
17	653	Romania	
18	644	Czech Republic	
19	638	South Korea	
20	597	Denmark	
⬇ 20 rows 3.62 seconds runtime			

[Back](#)

Question 1 Source Query 2

```
select
  count(case
    when a.medal = 'Gold' then 'Gold'
  end) as Gold_medal,
  count(case
    when a.medal = 'Silver' then 'Silver'
  end) as Silver_Medal,
  count(case
    when a.medal = 'Bronze' then 'Bronze'
  end) as Bronze_Medal,
  count(case
    when a.medal = 'Gold' then 'Gold'
    when a.medal = 'Silver' then 'Silver'
    when a.medal = 'Bronze' then 'Bronze'
    when a.medal = 'NA' then null
  end) as Total_medals,
  n.region
from
  athlete_events_1_csv as a
inner join
  noc_regions_3_csv as n
on
  a.NOC = n.NOC
group by
  n.region
order by
  count(case
    when a.medal = 'Gold' then 'Gold'
    when a.medal = 'Silver' then 'Silver'
    when a.medal = 'Bronze' then 'Bronze'
    when a.medal = 'NA' then null
  end) desc
limit 20
```

	Gold_medal ▲	Silver_Medal ▲	Bronze_Medal ▲	Total_medals ▲	region ▲
1	2638	1641	1358	5637	USA
2	1599	1170	1178	3947	Russia
3	1301	1195	1260	3756	Germany
4	678	739	651	2068	UK
5	501	610	666	1777	France
6	575	531	531	1637	Italy
7	479	522	535	1536	Sweden
8	463	438	451	1352	Canada
9	368	459	522	1349	Australia
10	432	332	371	1135	Hungary
11	287	340	413	1040	Netherlands
12	378	361	294	1033	Norway
13	351	349	293	993	China
14	247	309	357	913	Japan
15	198	270	432	900	Finland
16	175	248	268	691	Switzerland
17	161	200	292	653	Romania
18	123	262	259	644	Czech Republic
19	221	232	185	638	South Korea
20	179	241	177	597	Denmark

[Back](#)

Question 1 Source Query 3

```
select
  region,
  Total_medals,
  Gold_medal,
  cast(Gold_medal / Total_medals as decimal(9,2)) as Gold_Percent,
  Silver_medal,
  cast(Silver_medal / Total_medals as decimal(9,2)) as Silver_Percent,
  Bronze_medal,
  cast(Bronze_medal / Total_medals as decimal(9,2)) as Bronze_Percent
from
(
  select
    count(case
      when a.medal = 'Gold' then 'Gold'
    end) as Gold_medal,
    count(case
      when a.medal = 'Silver' then 'Silver'
    end) as Silver_Medal,
    count(case
      when a.medal = 'Bronze' then 'Bronze'
    end) as Bronze_Medal,
    count(case
      when a.medal = 'Gold' then 'Gold'
      when a.medal = 'Silver' then 'Silver'
      when a.medal = 'Bronze' then 'Bronze'
      when a.medal = 'NA' then null
    end) as Total_medals,
    n.region
  from
    athlete_events as a
  inner join
    noc_regions as n
  on
    a.NOC = n.NOC
  group by
    n.region
  order by
    count(case
      when a.medal = 'Gold' then 'Gold'
      when a.medal = 'Silver' then 'Silver'
      when a.medal = 'Bronze' then 'Bronze'
      when a.medal = 'NA' then null
    end) desc)
limit 20
```

	region	Total_medals	Gold_medal	Gold_Percent	Silver_medal	Silver_Percent	Bronze_medal
1	USA	5637	2638	0.47	1641	0.29	1358
2	Russia	3947	1599	0.41	1170	0.3	1178
3	Germany	3756	1301	0.35	1195	0.32	1260
4	UK	2068	678	0.33	739	0.36	651
5	France	1777	501	0.28	610	0.34	666
6	Italy	1637	575	0.35	531	0.32	531
7	Sweden	1536	479	0.31	522	0.34	535
8	Canada	1352	463	0.34	438	0.32	451
9	Australia	1349	368	0.27	459	0.34	522
10	Hungary	1135	432	0.38	332	0.29	371
11	Netherlands	1040	287	0.28	340	0.33	413
12	Norway	1033	378	0.37	361	0.35	294
13	China	993	351	0.35	349	0.35	293
14	Japan	913	247	0.27	309	0.34	357
15	Finland	900	198	0.22	270	0.3	432
16	Switzerland	691	175	0.25	248	0.36	268
17	Romania	653	161	0.25	200	0.31	292
18	Czech Republic	644	123	0.19	262	0.41	259
19	South Korea	638	221	0.35	232	0.36	185
20	Denmark	597	179	0.3	241	0.4	177

[Back](#)

Question 1 Source Query 4

Table Visualization 1 +

	region	season	Total_medals	Gold_medal	Gold_Percent	Silver_medal	Silver_Percent	Bro
1	USA	Summer	5002	2472	0.49	1333	0.27	119
2	Russia	Summer	3188	1220	0.38	974	0.31	994
3	Germany	Summer	3126	1075	0.34	987	0.32	106
4	UK	Summer	1985	636	0.32	729	0.37	620
5	France	Summer	1627	465	0.29	575	0.35	587
6	Italy	Summer	1446	518	0.36	474	0.33	454
7	Australia	Summer	1333	362	0.27	456	0.34	515
8	Hungary	Summer	1123	432	0.38	328	0.29	363
9	Sweden	Summer	1108	354	0.32	396	0.36	358
10	Netherlands	Summer	918	245	0.27	302	0.33	371
11	China	Summer	913	335	0.37	319	0.35	259
12	Japan	Summer	850	230	0.27	287	0.34	333
13	Canada	Summer	741	158	0.21	239	0.32	344
14	Romania	Summer	651	161	0.25	200	0.31	290
15	Denmark	Summer	592	179	0.3	236	0.4	177
16	Norway	Summer	590	227	0.38	196	0.33	167
17	South Korea	Summer	552	171	0.31	206	0.37	175
18	Poland	Summer	538	111	0.21	185	0.34	242
19	Serbia	Summer	532	157	0.3	216	0.41	159
20	Spain	Summer	487	109	0.22	243	0.5	135

```
select
  region,
  season,
  Total_medals,
  Gold_medal,
  cast(Gold_medal / Total_medals as decimal(9,2)) as Gold_Percent,
  Silver_medal,
  cast(Silver_medal / Total_medals as decimal(9,2)) as Silver_Percent,
  Bronze_medal,
  cast(Bronze_medal / Total_medals as decimal(9,2)) as Bronze_Percent
from
(
  select
    a.season,
    count(case
      when a.medal = 'Gold' then 'Gold'
    end) as Gold_medal,
    count(case
      when a.medal = 'Silver' then 'Silver'
    end) as Silver_Medal,
    count(case
      when a.medal = 'Bronze' then 'Bronze'
    end) as Bronze_Medal,
    count(case
      when a.medal = 'Gold' then 'Gold'
      when a.medal = 'Silver' then 'Silver'
      when a.medal = 'Bronze' then 'Bronze'
      when a.medal = 'NA' then null
    end) as Total_medals,
    n.region
  from
    athlete_events as a
  inner join
    noc_regions as n
  on
    a.NOC = n.NOC
  group by
    a.season,
    n.region
  order by
    count(case
      when a.medal = 'Gold' then 'Gold'
      when a.medal = 'Silver' then 'Silver'
      when a.medal = 'Bronze' then 'Bronze'
      when a.medal = 'NA' then null
    end) desc)
where
  season = 'Summer'
limit 20
```

[Back](#)

Question 1 Source Query 5

	region	season	Total_medals	Gold_medal	Gold_Percent	Silver_medal	Silver_Percent
1	Russia	Winter	759	379	0.5	196	0.26
2	USA	Winter	635	166	0.26	308	0.49
3	Germany	Winter	630	226	0.36	208	0.33
4	Canada	Winter	611	305	0.5	199	0.33
5	Norway	Winter	443	151	0.34	165	0.37
6	Sweden	Winter	428	125	0.29	126	0.29
7	Finland	Winter	426	66	0.15	145	0.34
8	Austria	Winter	280	79	0.28	98	0.35
9	Switzerland	Winter	275	76	0.28	70	0.25
10	Czech Republic	Winter	231	29	0.13	93	0.4
11	Italy	Winter	191	57	0.3	57	0.3
12	France	Winter	150	36	0.24	35	0.23
13	Netherlands	Winter	122	42	0.34	38	0.31
14	South Korea	Winter	86	50	0.58	26	0.3
15	UK	Winter	83	42	0.51	10	0.12
16	China	Winter	80	16	0.2	30	0.38
17	Japan	Winter	63	17	0.27	22	0.35
18	Poland	Winter	27	6	0.22	10	0.37
19	Slovenia	Winter	18	2	0.11	4	0.22
20	Australia	Winter	16	6	0.38	3	0.19

[Back](#)

```
select
  region,
  season,
  Total_medals,
  Gold_medal,
  cast(Gold_medal / Total_medals as decimal(9,2)) as Gold_Percent,
  Silver_medal,
  cast(Silver_medal / Total_medals as decimal(9,2)) as Silver_Percent,
  Bronze_medal,
  cast(Bronze_medal / Total_medals as decimal(9,2)) as Bronze_Percent
from
(
  select
    a.season,
    count(case
      when a.medal = 'Gold' then 'Gold'
    end) as Gold_medal,
    count(case
      when a.medal = 'Silver' then 'Silver'
    end) as Silver_Medal,
    count(case
      when a.medal = 'Bronze' then 'Bronze'
    end) as Bronze_Medal,
    count(case
      when a.medal = 'Gold' then 'Gold'
      when a.medal = 'Silver' then 'Silver'
      when a.medal = 'Bronze' then 'Bronze'
      when a.medal = 'NA' then null
    end) as Total_medals,
    n.region
  from
    athlete_events_1_csv as a
  inner join
    noc_regions_3_csv as n
  on
    a.NOC = n.NOC
  group by
    a.season,
    n.region
  order by
    count(case
      when a.medal = 'Gold' then 'Gold'
      when a.medal = 'Silver' then 'Silver'
      when a.medal = 'Bronze' then 'Bronze'
      when a.medal = 'NA' then null
    end) desc)
where
  season = 'Winter'
limit 20
```

Question 2 Source Query 1

```
select
  count(distinct name) as total_athletes,
  region
from
  (select a.*, n.region, n.notes
   from athlete_events as a
   inner join
     noc_regions as n
   on
     a.NOC = n.NOC)
group by
  region
order by
  count(distinct name) desc
limit 20
```

[Back](#)

	total_athletes	region
1	9652	USA
2	7541	Germany
3	6273	UK
4	6161	France
5	5597	Russia
6	4921	Italy
7	4810	Canada
8	4036	Japan
9	3868	Australia
10	3782	Sweden
11	2964	Poland
12	2937	Netherlands
13	2868	Switzerland
14	2815	China
15	2741	Czech Republic
16	2681	Hungary
17	2637	Spain
18	2346	Finland
19	2334	Austria
20	2289	South Korea

Question 2 Source Query 2

```
select
  name,
  sport,
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
  end) as Total_medals,
  region
from
  (select a.*, n.region, n.notes
   from athlete_events as a
   inner join
     noc_regions as n
   on
     a.NOC = n.NOC)
where region in
  ('USA', 'Germany', 'France', 'UK', 'Russia', 'Italy', 'Canada', 'Japan', 'Sweden', 'Australia',
   'Hungary', 'Czech Republic', 'Poland', 'Switzerland', 'Netherlands', 'China', 'Finland', 'Spain',
   'Austria', 'Norway')
group by
  name,
  sport,
  region
order by
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
  end) desc
```

	name	sport	Total_medals	region
1	Michael Fred Phelps, II	Swimming	28	USA
2	Larysa Semenivna Latynina (Diriy-)	Gymnastics	18	Russia
3	Nikolay Yefimovich Andrianov	Gymnastics	15	Russia
4	Takashi Ono	Gymnastics	13	Japan
5	Ole Einar Bjørndalen	Biathlon	13	Norway
6	Edoardo Mangiarotti	Fencing	13	Italy
7	Borys Anfiyanovych Shakhlin	Gymnastics	13	Russia
8	Dara Grace Torres (-Hoffman, -Minas)	Swimming	12	USA
9	Natalie Anne Coughlin (-Hall)	Swimming	12	USA
10	Birgit Fischer-Schmidt	Canoeing	12	Germany
11	Aleksey Yuryevich Nemov	Gymnastics	12	Russia
12	Paavo Johannes Nurmi	Athletics	12	Finland
13	Jennifer Elisabeth "Jenny" Thompson (-Cumpelik)	Swimming	12	USA
14	Sawao Kato	Gymnastics	12	Japan
15	Ryan Steven Lochte	Swimming	12	USA
16	Matthew Nicholas "Matt" Biondi	Swimming	11	USA
17	Carl Townsend Osburn	Shooting	11	USA
18	Viktor Ivanovych Chukarin	Gymnastics	11	Russia
19	Vra slavsk (-Odloilov)	Gymnastics	11	Czech Re
20	Mark Andrew Spitz	Swimming	11	USA
21	Polina Hryhorivna Astakhova	Gymnastics	10	Russia
22	Raisa Petrovna Smetanina	Cross Country Skiing	10	Russia

[Back](#)

Question 2 Source Query 3

	name	region	sport	min(year)	max(year)	Total_me
1	Michael Fred Phelps, II	USA	Swimming	2000	2016	28
2	Larysa Semenivna Latynina (Diriy-)	Russia	Gymnastics	1956	1964	18
3	Nikolay Yefimovich Andrianov	Russia	Gymnastics	1972	1980	15
4	Takashi Ono	Japan	Gymnastics	1952	1964	13
5	Ole Einar Bjørndalen	Norway	Biathlon	1994	2014	13
6	Borys Anfiyanovich Shakhlin	Russia	Gymnastics	1956	1964	13
7	Edoardo Mangiarotti	Italy	Fencing	1936	1960	13
8	Ryan Steven Lochte	USA	Swimming	2004	2016	12
9	Paavo Johannes Nurmi	Finland	Athletics	1920	1928	12
10	Birgit Fischer-Schmidt	Germany	Canoeing	1980	2004	12
11	Aleksey Yuryevich Nemov	Russia	Gymnastics	1996	2004	12
12	Sawao Kato	Japan	Gymnastics	1968	1976	12
13	Natalie Anne Coughlin (-Hall)	USA	Swimming	2004	2012	12
14	Jennifer Elisabeth "Jenny" Thompson (-Cumpelik)	USA	Swimming	1992	2004	12
15	Dara Grace Torres (-Hoffman, -Minas)	USA	Swimming	1984	2008	12
16	Viktor Ivanovich Chukarin	Russia	Gymnastics	1952	1956	11
17	Matthew Nicholas "Matt" Biondi	USA	Swimming	1984	1992	11
18	Vra slavsk (-Odolilov)	Czech Republic	Gymnastics	1960	1968	11
19	Carl Townsend Osburn	USA	Shooting	1912	1924	11
20	Mark Andrew Spitz	USA	Swimming	1968	1972	11

```
select
  name,
  region,
  sport,
  min(year),
  max(year),
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
    end) as Total_medals,
  count(case
    when medal = 'Gold' then 'Gold'
    end) as Gold_medal,
  count(case
    when medal = 'Silver' then 'Silver'
    end) as Silver_Medal,
  count(case
    when medal = 'Bronze' then 'Bronze'
    end) as Bronze_Medal
from
  (select a.*, n.region, n.notes
   from athlete_events as a
   inner join
     noc_regions as n
   on
     a.NOC = n.NOC)
 where region in
   ('USA', 'Germany', 'France', 'UK', 'Russia', 'Italy', 'Canada', 'Japan', 'Sweden', 'Australia',
    'Hungary', 'Czech Republic', 'Poland', 'Switzerland', 'Netherlands', 'China', 'Finland', 'Spain',
    'Austria', 'Norway')
group by
  name,
  sport,
  region
order by
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
    end) desc
limit 20
```

[Back](#)

Question 3 Source Query 1.1

```
select
  event,
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
  end) as Total_medals
from
  athlete_events
where
  event not like '%Women%'
group by
  event
order by
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
  end) desc
limit 10
```

	event	Total_medals
1	Football Men's Football	1269
2	Ice Hockey Men's Ice Hockey	1230
3	Hockey Men's Hockey	1050
4	Water Polo Men's Water Polo	866
5	Rowing Men's Coxed Eights	730
6	Gymnastics Men's Team All-Around	713
7	Basketball Men's Basketball	687
8	Handball Men's Handball	588
9	Volleyball Men's Volleyball	495
10	Swimming Men's 4 x 200 metres Freestyle Relay	395

[Back](#)

Question 3 Source Query 1.2

```
select
  event,
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
    end) as Total_medals
from
  athlete_events
where
  event like '%Women%'
group by
  event
order by
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
    end) desc
limit 10
```

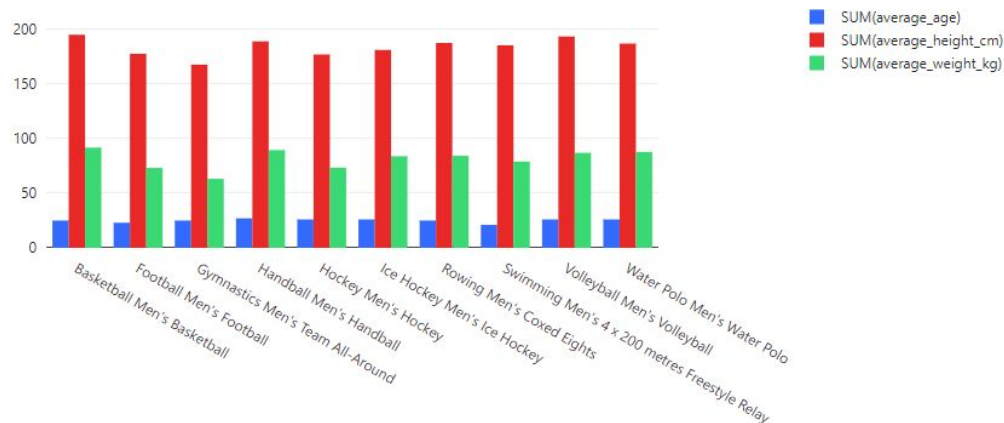
	event	Total_medals
1	Hockey Women's Hockey	478
2	Volleyball Women's Volleyball	474
3	Handball Women's Handball	472
4	Basketball Women's Basketball	393
5	Gymnastics Women's Team All-Around	390
6	Swimming Women's 4 x 100 metres Freestyle Relay	362
7	Football Women's Football	302
8	Ice Hockey Women's Ice Hockey	300
9	Rowing Women's Coxed Eights	299
10	Swimming Women's 4 x 100 metres Medley Relay	272

[Back](#)

Question 3 Source Query 2

```
select
  event,
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
  end) as Total_medals,
  cast(avg(age) as decimal(9,0)) as average_age,
  cast(avg(height) as decimal(9,2)) || ' cm' as average_height,
  cast(avg(weight) as decimal(9,2)) || ' kg' as average_weight
from
  athlete_events
where
  event not like '%Women%'
group by
  event
order by
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
  end) desc
limit 10
```

	event	Total_medals	average_age	average_height	average_weight
1	Football Men's Football	1269	23	177.48 cm	73.09 kg
2	Ice Hockey Men's Ice Hockey	1230	26	180.99 cm	83.78 kg
3	Hockey Men's Hockey	1050	26	176.85 cm	73.34 kg
4	Water Polo Men's Water Polo	866	26	186.80 cm	87.71 kg
5	Rowing Men's Coxed Eights	730	25	187.38 cm	84.18 kg
6	Gymnastics Men's Team All-Around	713	25	167.54 cm	63.13 kg
7	Basketball Men's Basketball	687	25	194.87 cm	91.68 kg
8	Handball Men's Handball	588	27	188.78 cm	89.39 kg
9	Volleyball Men's Volleyball	495	26	193.27 cm	86.93 kg
10	Swimming Men's 4 x 200 metres Freestyle Relay	395	21	185.22 cm	78.87 kg

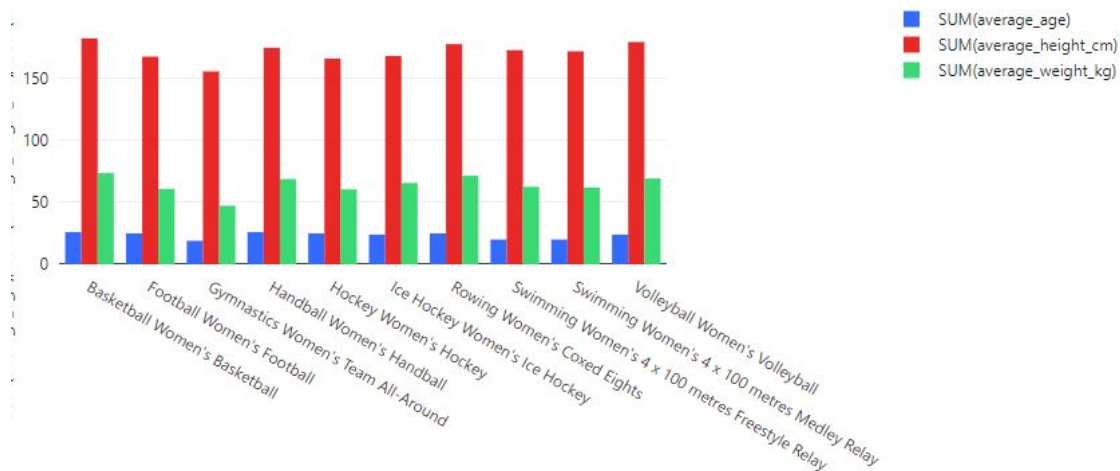


[Back](#)

Question 3 Source Query 3

```
select
  event,
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
    end) as Total_medals,
  cast(avg(age) as decimal(9,0)) as average_age,
  cast(avg(height) as decimal(9,2)) || ' cm' as average_height,
  cast(avg(weight) as decimal(9,2)) || ' kg' as average_weight
from
  athlete_events
where
  event like '%Women%'
group by
  event
order by
  count(case
    when medal = 'Gold' then 'Gold'
    when medal = 'Silver' then 'Silver'
    when medal = 'Bronze' then 'Bronze'
    when medal = 'NA' then null
    end) desc
limit 10
```

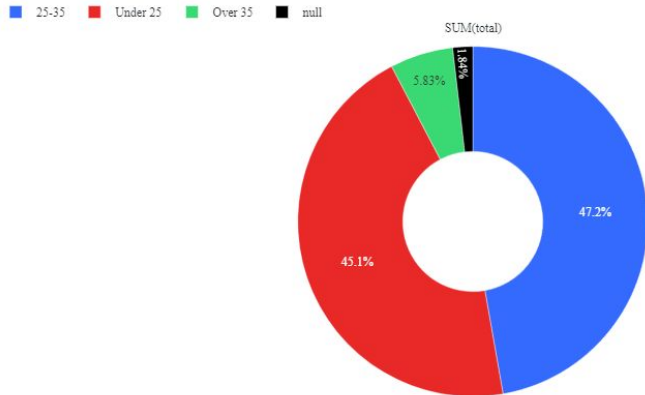
	event	Total_medals	average_age	average_height	average_weight
1	Hockey Women's Hockey	478	25	166.13 cm	60.53 kg
2	Volleyball Women's Volleyball	474	24	179.49 cm	69.33 kg
3	Handball Women's Handball	472	26	174.84 cm	68.88 kg
4	Basketball Women's Basketball	393	26	182.45 cm	73.69 kg
5	Gymnastics Women's Team All-Around	390	19	155.72 cm	47.32 kg
6	Swimming Women's 4 x 100 metres Freestyle Relay	362	20	172.89 cm	62.60 kg
7	Football Women's Football	302	25	167.68 cm	60.93 kg
8	Ice Hockey Women's Ice Hockey	300	24	168.21 cm	65.71 kg
9	Rowing Women's Coxed Eights	299	25	177.77 cm	71.50 kg
10	Swimming Women's 4 x 100 metres Medley Relay	272	20	172.01 cm	61.95 kg



[Back](#)

Metric Tracking Query

```
select
  age,
  count(medals) as total,
  count(case
    when medals = 'Gold' then 'Gold'
  end) as Gold_medal,
  count(case
    when medals = 'Silver' then 'Silver'
  end) as Silver_Medal,
  count(case
    when medals = 'Bronze' then 'Bronze'
  end) as Bronze_Medal
from
  (Select
    case
      when age < 25 then 'Under 25'
      when age between 25 and 35 then '25-35'
      when age > 35 then 'Over 35'
      when age = 'NA' then null
    end as age,
    case
      when medal = 'Gold' then 'Gold'
      when medal = 'Silver' then 'Silver'
      when medal = 'Bronze' then 'Bronze'
      when medal = 'NA' then null
    end as medals
  from
    athlete_events)
group by
  age
order by
  count(medals) desc,
```



	age	total	Gold_medal	Silver_Medal	Bronze_Medal
1	25-35	18791	6317	6227	6247
2	Under 25	17939	6129	5806	6004
3	Over 35	2321	778	788	755
4	null	732	148	295	289

[Back](#)