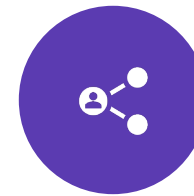# Olympia 120 Years

Alexander Gandji

Content

Review of Questions to Answer/Hypotheses Approach

Discussion of Technical Challenges

Detail: Entity Relationship Diagram (ERD)

Initial Findings

Deeper Analysis

Hypotheses Results

Which Data set was chosen and who will benefit from the analysis

- In the following analysis, I want to analyze the last 120 years of Olympic events

- Who will benefit from this analysis?
  - Sports magazines and sports broadcasting TV channel
    - They communicate the outcome of the analysis

Initial Questions and Hypotheses

Questions I wanted to answer

- Which was the most successful nation?
    - Measured by overall medal and gold medal count
- Who is the most successful athlete?
    - Measured by overall medal and gold medal count
- Do men and women have the same number of events?
- Did the profile of the athletes change over time?
    - Did the height and weight of summer game athletes change over the last 120 years?
- Is there a relationship between age and the number of medals an athlete won?

Hypotheses I want to proof

The US is the most successful nation

- All-time medal leader in overall medals and gold medals (count)

The most successful athlete is a male

- A man will hold the most gold medals (count)
- Most overall medals (count)

The weight and height have increased for summer game athletes

China was the runner-up in the last four summer games (2004-2016)

- China succeeds in less events than the most successful nation in this period
 - But these events are in majority different to the most successful nation

There is a negative correlation between age and medals won

In 2016, women and men had an equal number of participants

Technical Challenges

- All in all, I did not face significant challenges and issues
  - The majority of data had a good standard
    - The only issue I had to face regarding the data was that a few data points regarding height and weight were missing for the 1896 Olympics and then in the early 1900's. I have solved the issue with the ffill() function in Python.
      - For those that could not be fixed (too ambiguous), I had to remove them from the dataset
    - The data points which did not make any sense were not used for the analysis of height and weight
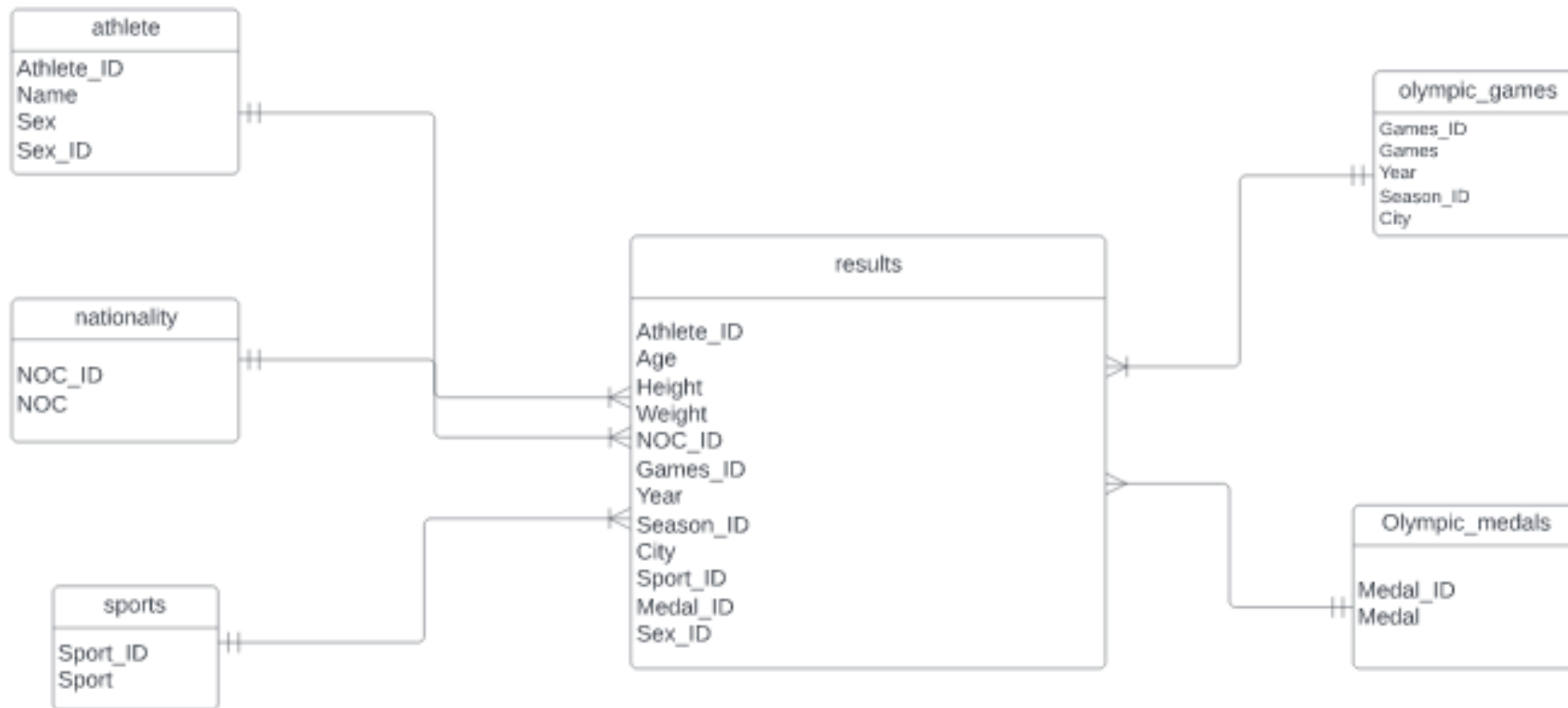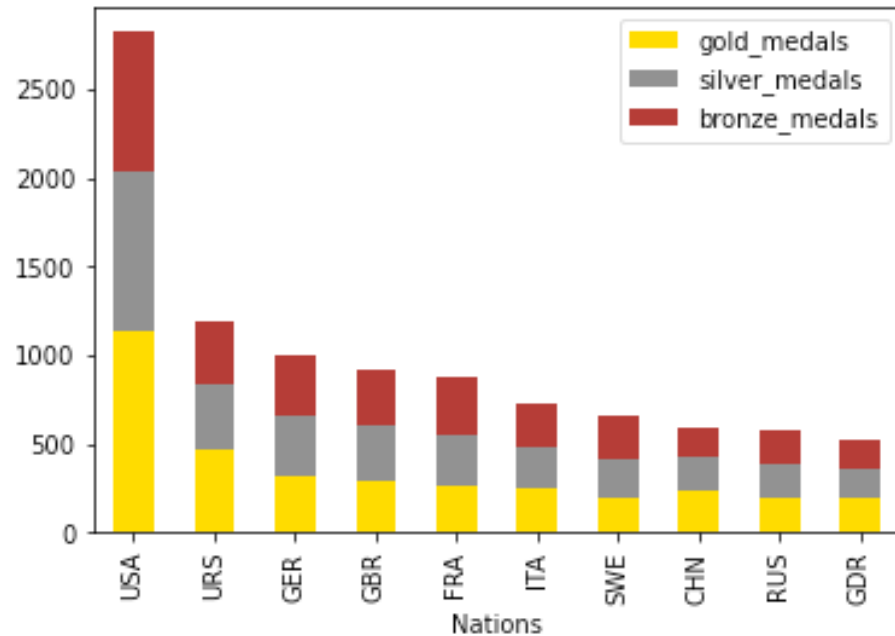
The ERD Model of the Analysis

Initial Findings
Who is the most successful Nation?

- The US has the most gold and overall medals in the Olympic history (120 years)
  - Numbers include Summer and Winter Games

Overall Medal Count

Nations	all_time_medals	gold_medals	silver_medals	bronze_medals
USA	2823	1131	901	791
URS	1197	471	373	353
GER	1004	319	345	340
GBR	919	289	320	310
FRA	879	264	286	329
ITA	722	256	225	241
SWE	657	200	215	242

Who is the most successful Athlete in Olympic History?

- First, I wanted to see who are the top 5 most successful athletes in Olympic history based on the overall medal count

- Based on the overall medal count, we can see that Michael Phelps is the most successful athlete

Distribution of Medal Types of the most Successful Athletes
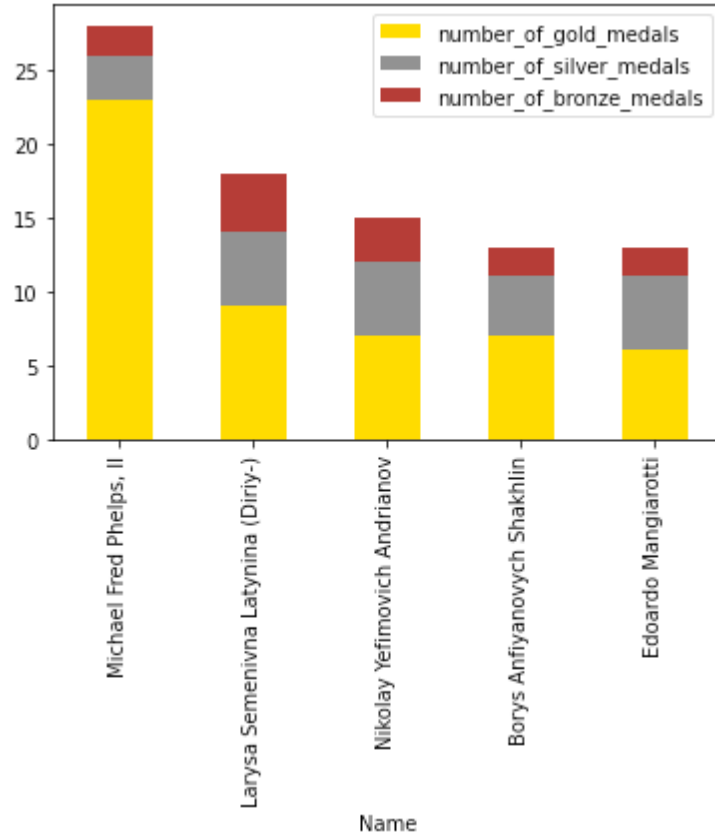
Name	NOC	Sex	number_of_medals	number_of_gold_medals	number_of_silver_medals	number_of_bronze_medals
Michael Fred Phelps, II	USA	M	28	23	3	2
Larysa Semenivna Latynina (Diriy-)	URS	F	18	9	5	4
Nikolay Yefimovich Andrianov	URS	M	15	7	5	3
Borys Anfiyanovych Shakhlin	URS	M	13	7	4	2
Edoardo Mangiarotti	ITA	M	13	6	5	2

Who is the most successful Athlete in Olympic History?

- When it comes to most gold medals, we see a familiar face, Michael Phelps is again the most successful athlete

Most Successful Athletes-Gold Medals
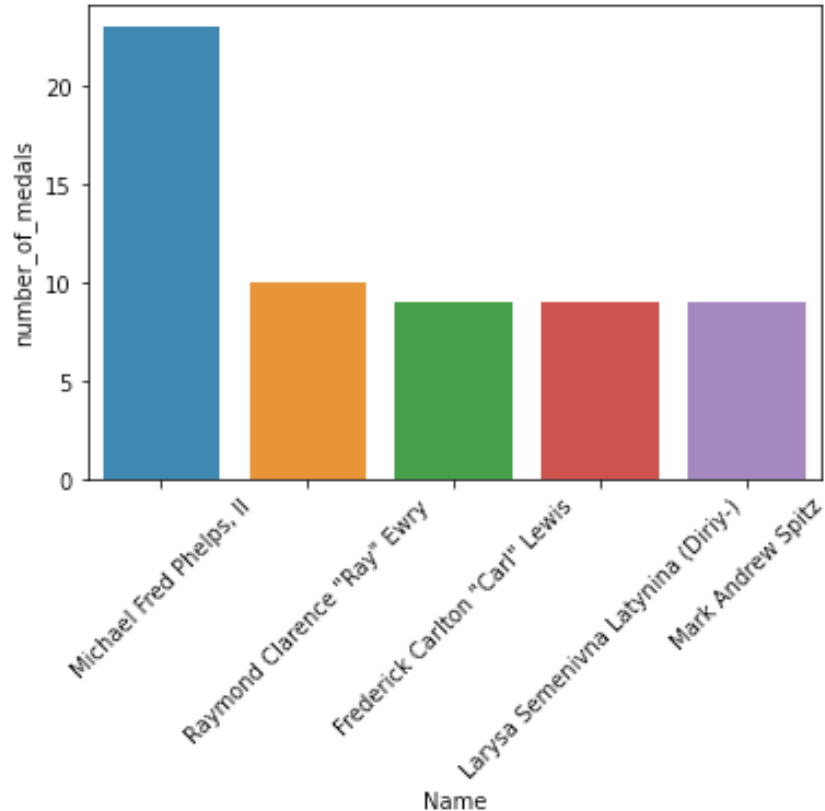
|	Name	NOC	Sex	number_of_gold_medals
0	Michael Fred Phelps, II	USA	M	23
1	Raymond Clarence "Ray" Ewry	USA	M	10
2	Frederick Carlton "Carl" Lewis	USA	M	9
3	Larysa Semenivna Latynina (Diriy-)	URS	F	9
4	Mark Andrew Spitz	USA	M	9

Who is the most successful Athlete in Olympic History?

- When we look at the conversation rate for the overall medals won

- We see a different picture.

- Larysa Semenivna has the highest conversion rate

Conversion Rate from Events to Medals
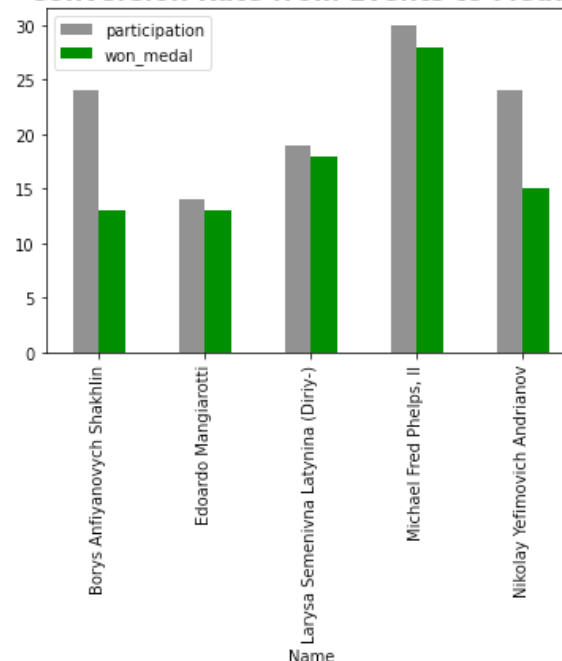
Conversion Rate on Overall Medal Level
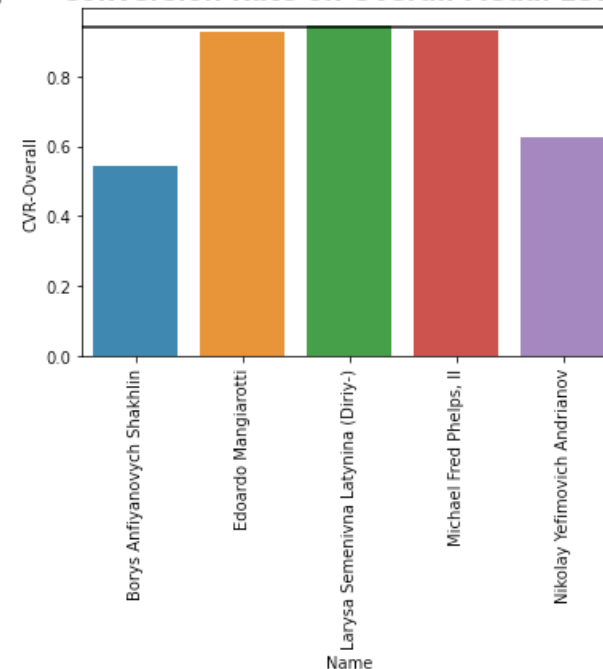
Name	NOC	Sex	won_medal	participation	CVR_Overall
Borys Anfiyanovych Shakhlin	URS	M	13	24	0.5417
Edoardo Mangiarotti	ITA	M	13	14	0.9286
Larysa Semenivna Latynina (Diriy-)	URS	F	18	19	0.9474
Michael Fred Phelps, II	USA	M	28	30	0.9333
Nikolay Yefimovich Andrianov	URS	M	15	24	0.6250

Who is the most successful Athlete in Olympic History?

- For the gold medal conversation rate of the most successful athletes based on the overall medal count

- We can see that Michael Phelps is swimming circles around his competition

Conversation Rate for Gold Medal of the most Successful Athletes based on the Overall Medal Count
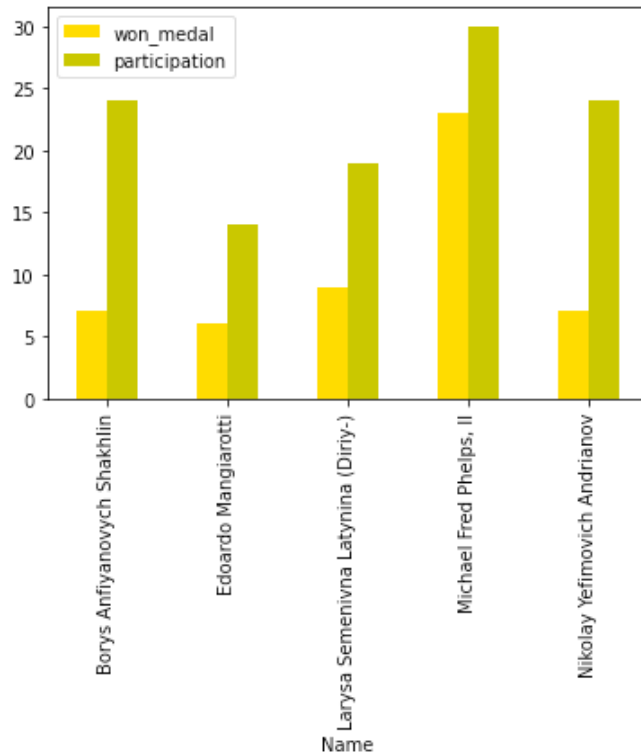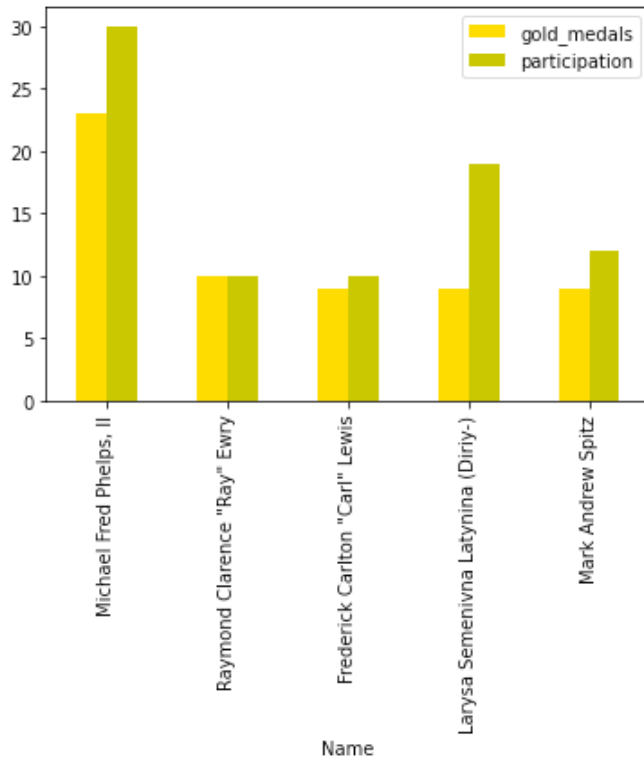

Name	NOC	Sex	won_medal	participation	CVR_Gold
Borys Anfiyanovych Shakhlin	URS	M	7	24	0.29
Edoardo Mangiarotti	ITA	M	6	14	0.43
Larysa Semenivna Latynina (Diriy-)	URS	F	9	19	0.47
Michael Fred Phelps, II	USA	M	23	30	0.77
Nikolay Yefimovich Andrianov	URS	M	7	24	0.29

Who is the most successful Athlete in Olympic History?

- When we look at the gold medal CVR of the most successful gold medal athletes

- We see that Michael Phelps has the most gold medals but the third best CVR

- Raymond Clarence "Ray" Ewry (100%) and Frederick Carlton "Carl" Lewis (90%)

- At this point, it depends on the interpretation
  - As already stated, we want to find out if a male is the most successful athlete based on the overall and gold medal count
  - Therefore, Michael Phelps (Male) is the most successful athlete in Olympic history

**Conversation Rate for Gold Medals of the most Successful Gold Winning Athletes**

Name	NOC	Sex	gold_medals	participation	CVR_Gold
Michael Fred Phelps, II	USA	M	23	30	0.77
Raymond Clarence "Ray" Ewry	USA	M	10	10	1.00
Frederick Carlton "Carl" Lewis	USA	M	9	10	0.90
Larysa Semenivna Latynina (Diriy-)	URS	F	9	19	0.47
Mark Andrew Spitz	USA	M	9	12	0.75

Do men or women have the same number of events?

- The graph is supposed to show the numbers of events for men and women and how they have developed

- As we can see, women tend to have fewer events than men

- But we can observe a trend, that the difference between men and women is decreasing and moving towards zero (green line)

- The up and downs in the line graph are explained through the different numbers of events between summer and winter games
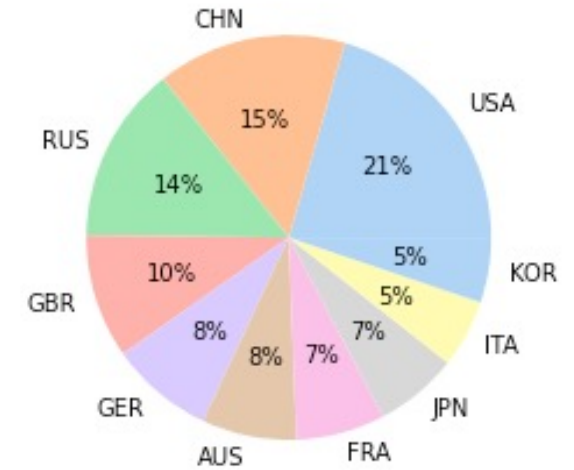 - Winter games do have fewer events than summer games

Differencen in number of events between Men and Women in 120 years
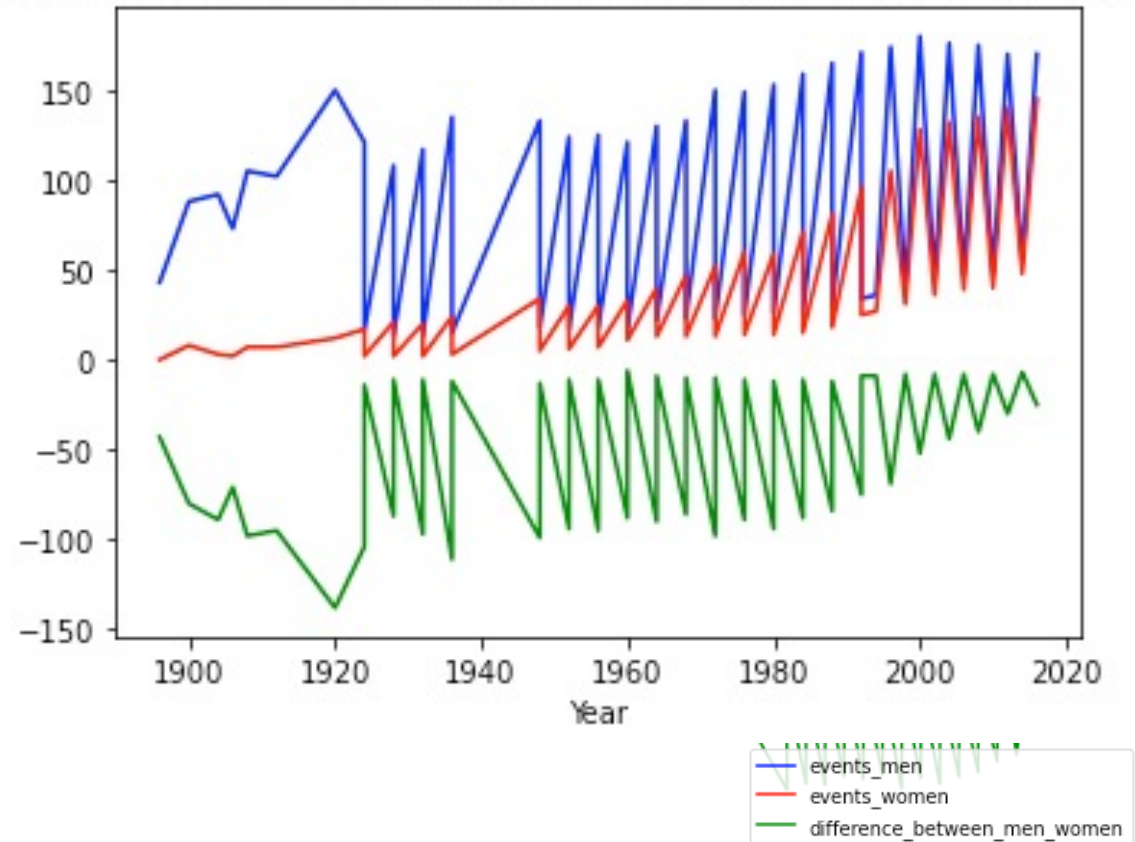
NOC	summer_medals
0 USA	435
1 CHN	321
2 RUS	300
3 GBR	210
4 GER	176
5 AUS	160
6 FRA	151
7 JPN	141

NOC	summer_gold_medals
0 USA	164
1 CHN	147
2 RUS	94
3 GBR	84
4 GER	57
5 AUS	46
6 KOR	44
7 JPN	44
8 FRA	39

Medals won in the Summer Games 2004-2016

China was a runner-up in the summer last four summer games (2004-2016)

- The table confirms one part of my hypothesis, that China was the runner-up in the last four summer games

- I will compare China with the US, because the the US was the most successful Nation in this time period as well

China was a runner-up in the summer last four summer games

	Sport	total_medals	number_of_gold_medals	number_of_silver_medals	number_of_bronze_medals
0	Table Tennis Men's Singles	9	3	4	2
1	Table Tennis Women's Singles	8	4	3	1
2	Diving Women's Springboard	8	4	3	1
3	Trampolining Men's Individual	6	2	1	3
4	Diving Women's Platform	6	3	2	1
5	Diving Men's Springboard	6	3	1	2
6	Badminton Women's Singles	6	3	2	1
7	Trampolining Women's Individual	5	1	1	3
8	Shooting Women's Air Rifle, 10 metres	5	2	1	2
9	Diving Men's Platform	5	2	2	1

	Sport	total_medals	number_of_gold_medals	number_of_silver_medals	number_of_bronze_medals
0	Athletics Women's 100 metres Hurdles	8	3	2	3
1	Swimming Men's 200 metres Individual Medley	7	4	2	1
2	Swimming Men's 100 metres Backstroke	7	4	2	1
3	Athletics Men's 400 metres	7	2	2	3
4	Swimming Men's 400 metres Individual Medley	6	3	2	1
5	Swimming Men's 200 metres Backstroke	6	4	1	1
6	Gymnastics Women's Individual All-Around	6	4	2	0
7	Gymnastics Women's Balance Beam	6	1	3	2
8	Beach Volleyball Women's Beach Volleyball	6	3	1	2
9	Swimming Women's 400 metres Freestyle	5	1	2	2

10 of 143   **Succesful events of China**

10 of 179   **Successful events for USA**

These tables show what the most successful events for the USA and China were
The chinese table confirms one aspect of my hypothesis, China had less events where they won medals, than the most successful nation

China was a runner-up in the last four summer games

- When we compare both tables (events from China and the US), we will see that they have 73 events in common
- This table alone would disprove the hypothesis, but we should check if China performs better or worse than the US in these events
- We will measure that by the different medal types
  - Number of events where one nation had more medals than the other

|	Sport
0	Swimming Men's 200 metres Individual Medley
1	Swimming Men's 100 metres Backstroke
2	Gymnastics Women's Individual All-Around
3	Gymnastics Women's Balance Beam
4	Beach Volleyball Women's Beach Volleyball
...	...
68	Athletics Women's Shot Put
69	Athletics Women's Marathon
70	Athletics Women's Discus Throw
71	Athletics Women's 10,000 metres
72	Archery Men's Individual

China is a runner-up in the last four summer games

- As we can see, the US performed in all three medal types better than China
  - It got very close in the gold medals segment
- The majority of China's events are similar to the US's ones
  - 73 of 143 => 51%
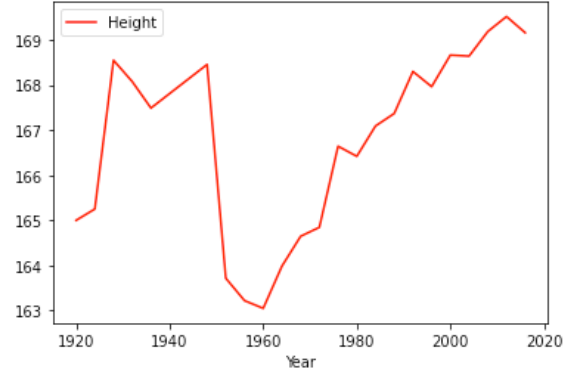- In these similar events, China always came up short of the US
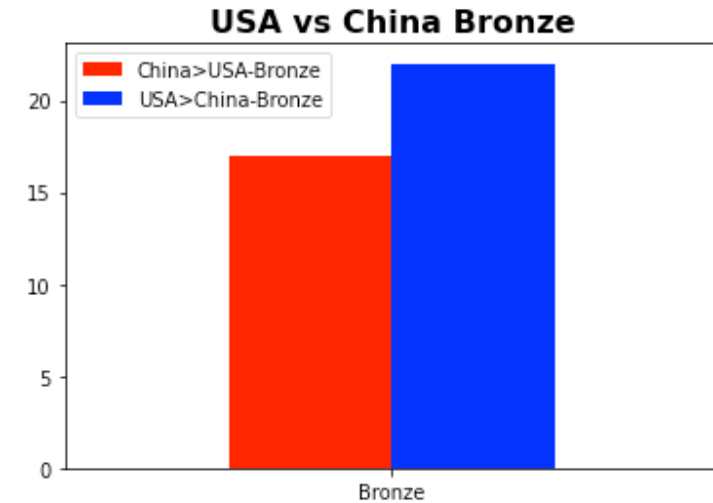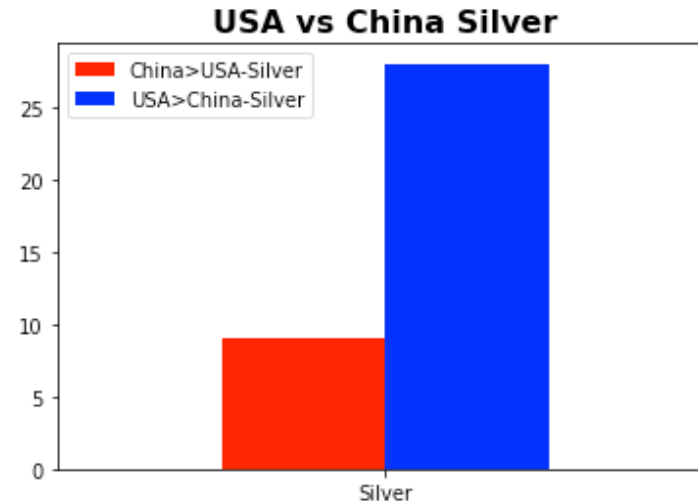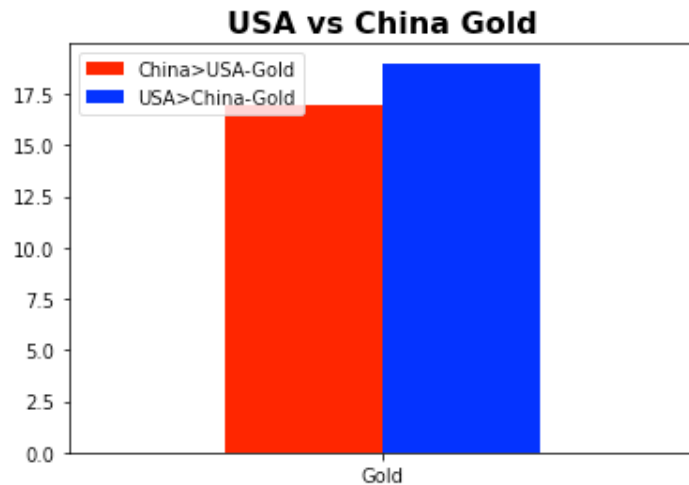
Did the Height and Weight of summer game athletes change over the last 120 years?
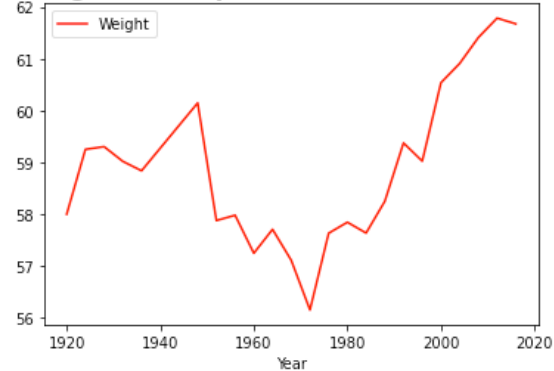

Height Development of Male Athletes
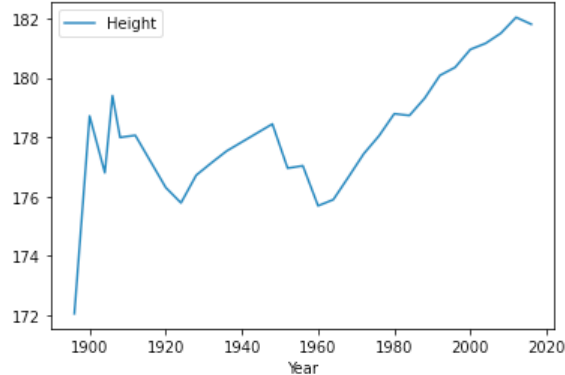
Height Development of Female Athletes

Weight Development of Male Athletes

Weight Development of Female Athletes

- We can observe that the height and weight have increased. (I have used the average weight and height of summer game athletes (grouped by 120 years of Olympic games)
  - It is essential to keep in mind that I have used ffill() to get ride of the na-values
  - We need to get rid of those values which do not make sense
  - We got ride of the values 176 cm (height) and 67 Kg (weight)
    - See the appendix for the additional table, which supports the decision

- The analysis has shown that the height and weight of both genders have increased
  - But it is important to mention that the male athletes have recorded the more significant increase

Deeper Analysis

After analysing the descriptive aspects of the data, we want to get a better/deeper understanding of the data

We will analyse the relationship between age, and medals won

Deeper Analysis

- The data indicates that most athletes were in the age group 20-29 when they won a medal
  - The histogram supports that because the distribution is right skewed
- 20-29 is the most substantial age group for men and women, but women have a higher number of medal-winning athletes in the age group 10-19 compared to men
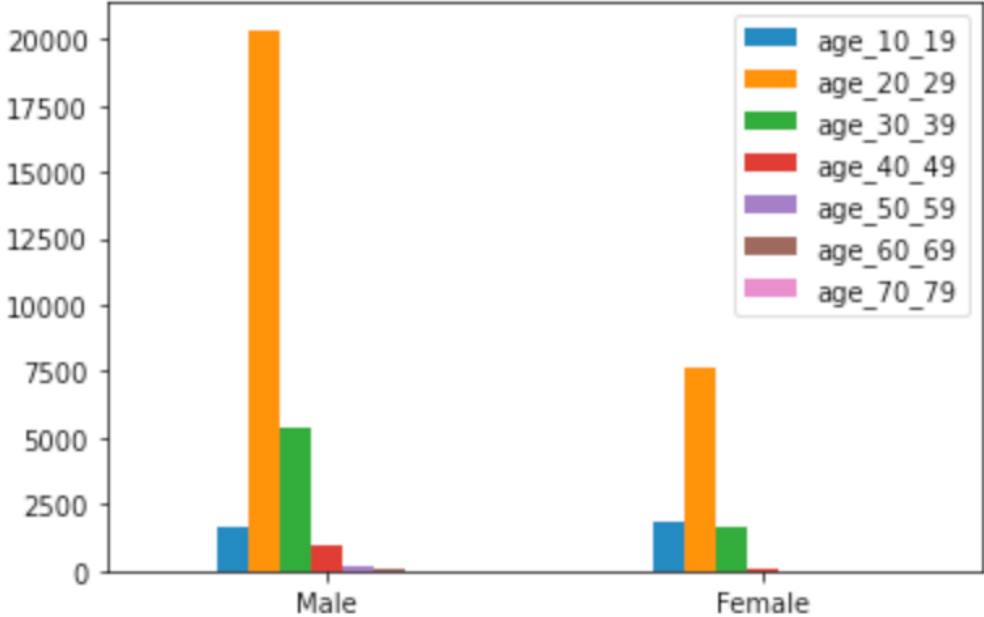- On the other side, men are drastically more successful than women in the age group 30-39

Distribution of Age for Medal Winning Parcitpants
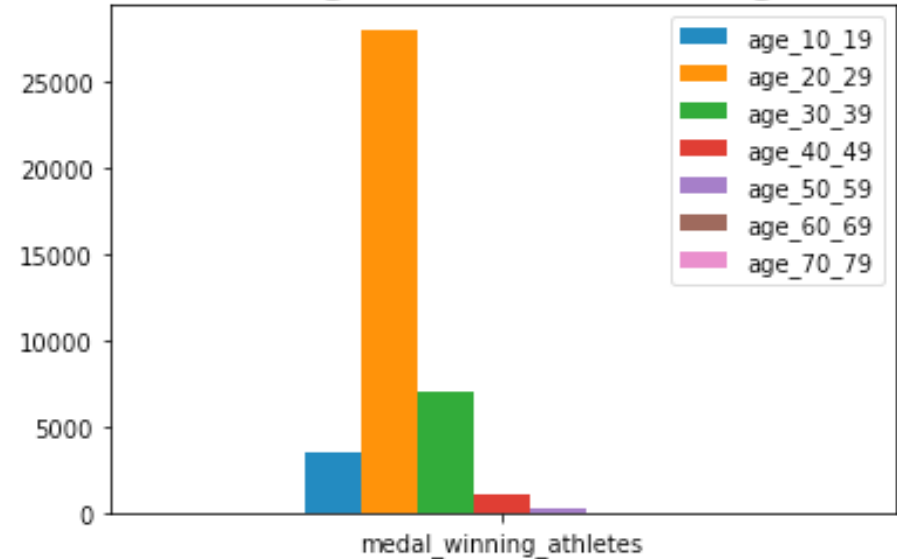
	age_10_19	age_20_29	age_30_39	age_40_49	age_50_59	age_60_69	age_70_79
medal_winning_athletes	3481	28002	7003	1039	209	44	5

Age Distribution Males vs Female

	age_10_19	age_20_29	age_30_39	age_40_49	age_50_59	age_60_69	age_70_79
Male	1614	20353	5383	937	198	40	5
Female	1867	7649	1620	102	11	4	0

Correlation between Age and Medals won

Correlation between Age and Medals won

- As I have previously mentioned, the strongest age group is 20-29

- To get a fair compression/analysis, I have sampled 50 data points from every age
  - The cutoff point was at age 46
  - We got a sample size of 1650

- As we can see, we are dealing with a non-normal distribution. Therefore I have used the Kendall Rank correlation

- The Outcome shows that the correlation is very weak and not statically significant

**Scatterplot Age and Medals Won**
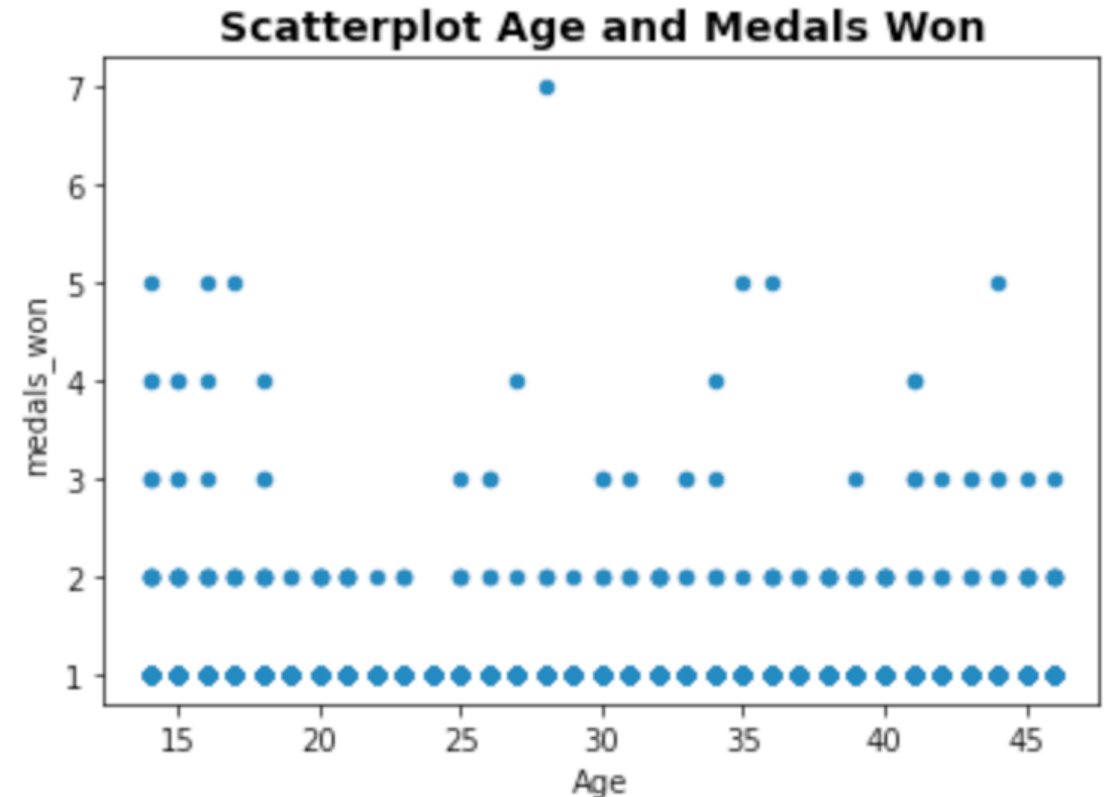

```
# continous=> Age and number of medals
# monotonic=> we cannot observe a monotonic development
corr_kdt=kendalltau(age_meds.Age, age_meds.medals_won)
corr_kdt
```

KendalltauResult(correlation=−0.02305034225633033, pvalue=0.2534944857758197)

Correlation between Age and Medals won

Scatterplot Age and No Medals Won

- A scatter plot was created for the athletes who have not won a medal to compare it against the medal-winning scatter plot
    - If the scatter plot of the none winning athletes looks completely random, then there could be the slightest chance for a trend regarding age and medals

- Looking at the winning and non-winning, they look rather familiar
    - Nonetheless, the winning scatter plot did not indicate any trend that age and medals won have any correlation

Scatterplot Age and Medals Won
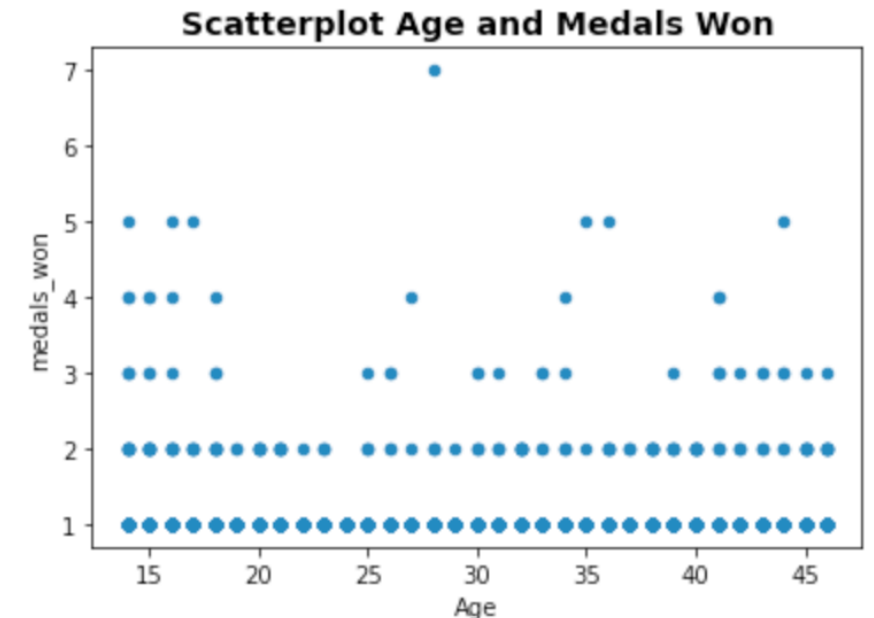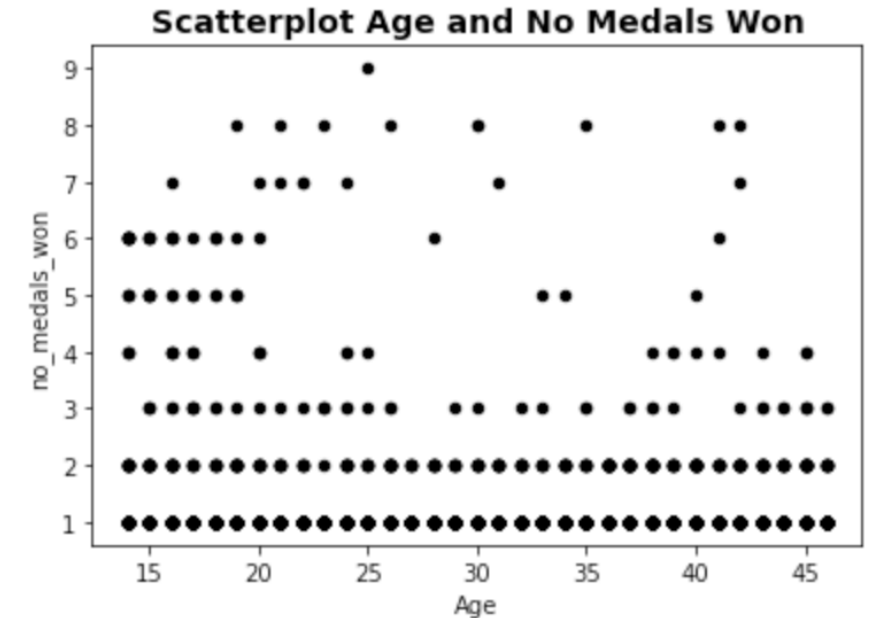
Hypotheses Results

- **The most successful athlete is a male**
  - A man will hold the most gold medals (count)=> A male athlete has won the most gold medals in Olympic history=> Michael Phelps
  - Most overall medals (count)=> A male athlete has won the most medals in Olympic history=> Michael Phelps

Hypothesis=> 👍

- **The US is the most successful nation**
  - All-time medal leader in overall medal and gold medal counts=> The US is the most successful nation on the overall and gold medal level

Hypothesis=> 👍

- **The weight and height have increased for summer game athletes**
  - => The height and weight of both genders (male and female) have increased

Hypothesis=> 👍

Hypotheses Results

- **China was a runner-up in the last four summer games**
  - China was a runner-up in the last four summer games; the US is the most successful nation in this period as well
  - China has less events than the most successful nation
  - But China had more events with the US in common than unique events
    - Regarding the events that the US and China had in common, the US was at every medal type more successful than China (had more medals)

  Hypothesis=> 👎

- **There is a negative correlation between age and medals won**
  - => We can see that the correlation was negative. But the correlation was not strong and not statistically significant

Hypothesis=> 👎

- **In 2016, women and men had an equal number of participants**
  - => The hypothesis is wrong; women did not have the same number of events as men in 2016

- Hypothesis=> 👎

Actions

- The media can communicate the outcome of the study
  - It will be interesting for people to understand which nation is the most successful one
  - Who is the most successful athlete
  - How the role of women has developed over 120 years of Olympic games
  - They can communicate that in the past, Russia/UDSSR was the biggest challenger of the US, but now it is China
  - People can be informed that there is no correlation between age and medals won

Sources

The Correlation Coefficient (r) - Anon, 2021

URL: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module9-Correlation-Regression/PH717-Module9-Correlation-Regression4.html.

Appendix

	Height	Weight	count(Height)	count(Weight)
0	176	67	38816	38816
1	170	67	346	346
2	173	67	274	274
3	178	67	246	246
4	168	67	244	244
5	167	67	219	219

- As we can see, the combination of 176 cm and 67 Kg, which were used in the ffill function, make a vast amount of data and seem very unrealistic

- Therefore, they were not included in the analysis