REPORT OUT: Sport Statistics

Analysis in Olympic events through Time

SECTION 1: Questions To Answer

- **Q1: Is there any correlation between the performance of a country in winter Olympics and that in summer Olympics?**

> Answering this question could help identify whether countries that perform well in one type of Olympics (summer or winter) tend to perform well in the other as well. This could highlight countries with versatile athletic programs and training systems.

- **Q2: Does performance change with age?**

> Analyzing how the age of athletes correlates with their performance (measured by medals won) could reveal patterns in the age at which athletes tend to excel in various sports. This might indicate optimal age ranges for peak performance in different disciplines.

- **Q3: How has the male:female ratio evolved through time?**

> Studying the historical trends in the male-to-female ratio of athletes over different Olympic games could provide insights into gender equality and representation in sports. It might indicate changes in societal attitudes towards women in sports and the impact of initiatives aimed at promoting gender diversity in the Olympics.

Overall, these questions help us gain a deeper understanding of the dynamics of Olympic performance, the factors influencing it, and the broader social and cultural trends shaping the Olympics over time.

SECTION 2: Initial Hypothesis

- **Q1: Is there any correlation between the performance of a country in winter Olympics and that in summer Olympics?**

> Hypothesis: Countries with strong athletic programs and training systems tend to perform consistently well in both summer and winter Olympics, indicating a correlation between the two types of Games.

- **Q2: Does performance change with age?**

> Hypothesis: Athletes' performance follows an age-related curve, with younger athletes excelling in sports requiring agility and speed, while older athletes excel in sports demanding experience and strategy..

- **Q3: How has the male:female ratio evolved through time?**

> Hypothesis: Over time, there has been a gradual increase in the representation of female athletes in the Olympics, reflecting the global movement towards gender equality in sports.

SECTION 3: Data Analysis Approach

> We will look for correlations and trends. Visuallisations and predictive analytics for future trends. Statistical tests wil be done to validate or reject the initial assumptions, and p-values and confidence intervals to determine the significance of findings.

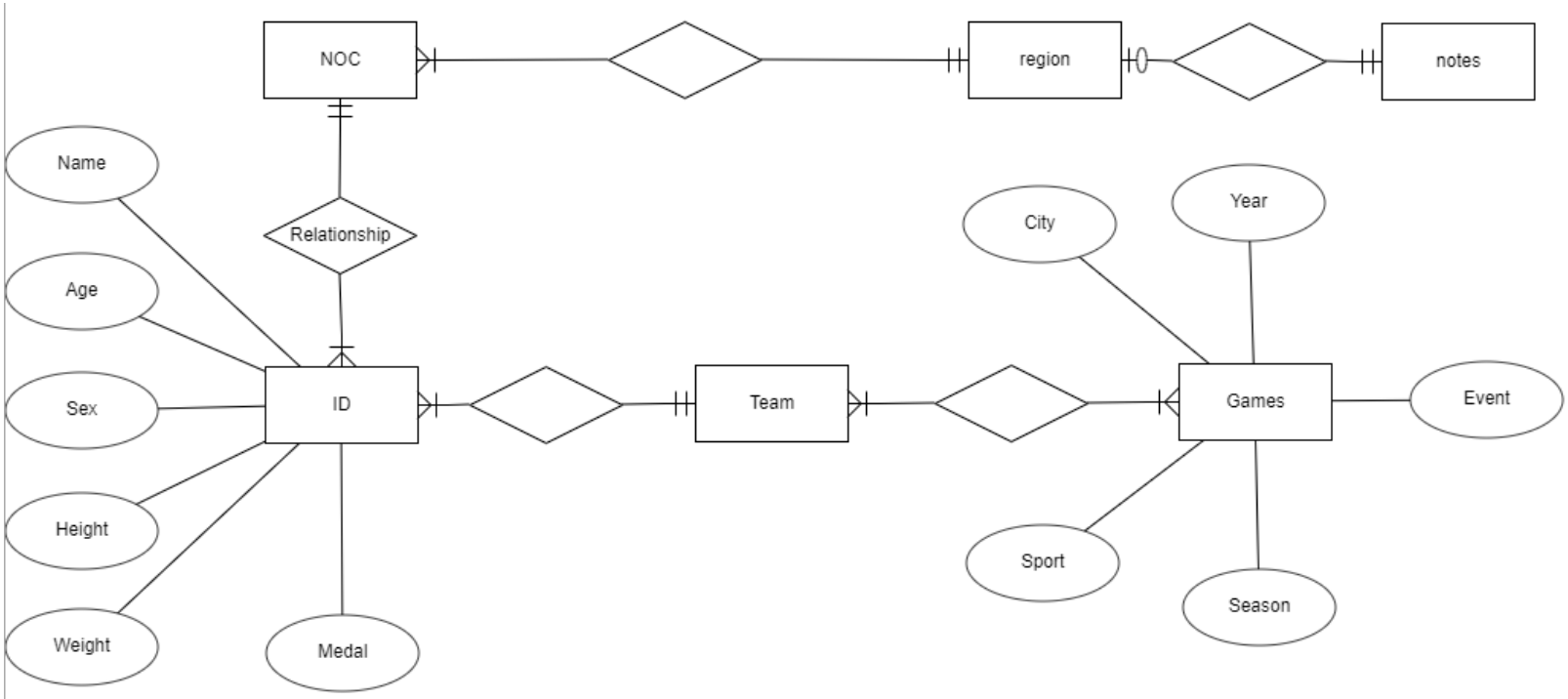
Technical Challenges

> Missing data from Countries hosted the Olympics. We needed to do a web search and add new data to our analysis.

Entry Relationship Diagram

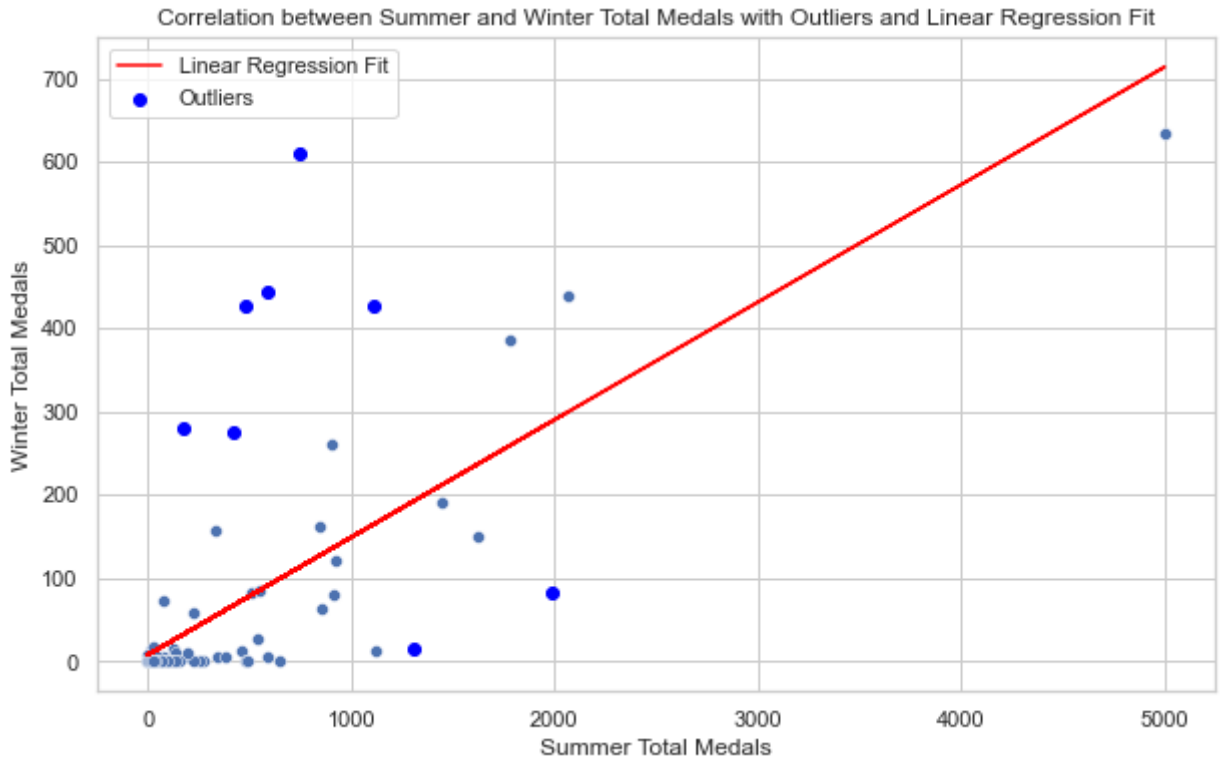
In [15]:
```
from IPython.display import Image
Image(filename = r"C:\Users\amanta\Documents\Coursera\ERD.png", width = 800, height = 400)
```

Out[15]:



Entry Deeper Analysis (Performance Winter vs Summer Events)

In [18]:

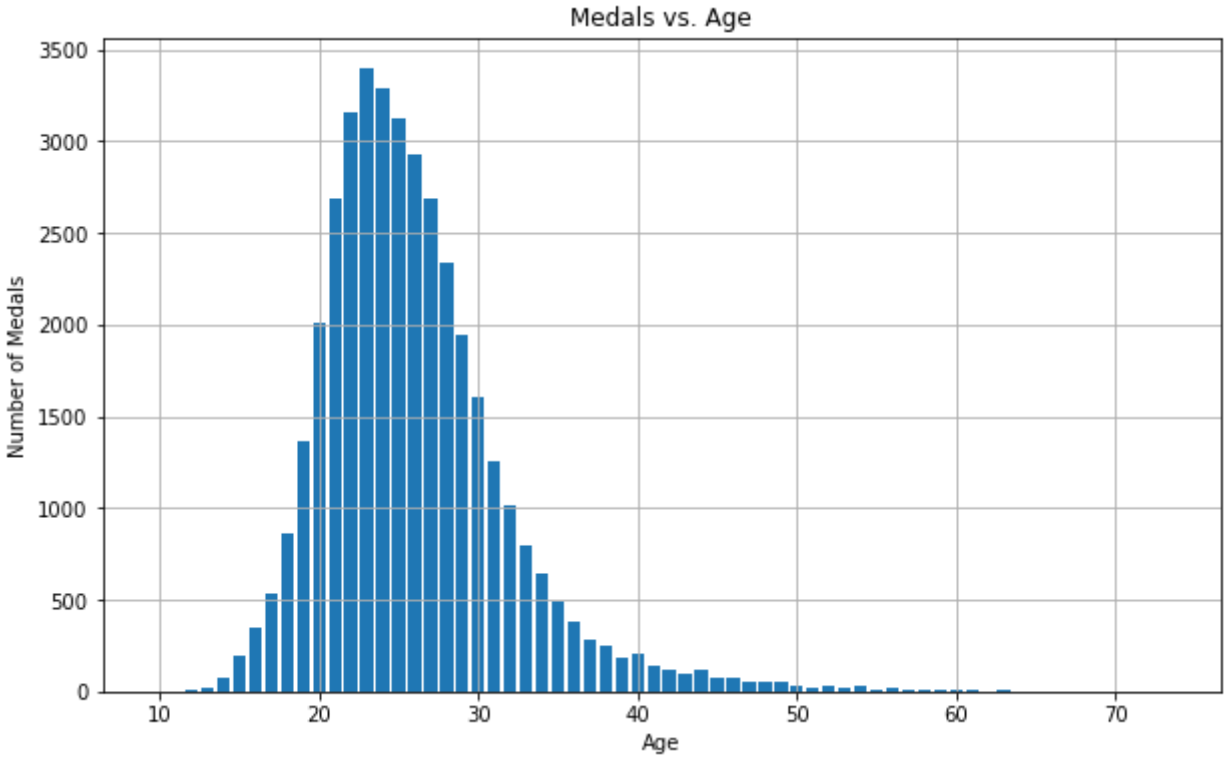



```
Correlation Coefficient: 0.7005179556904234
```

Country with Around 5000 Summer Medals: Country: USA Summer Medals: 5002.0 Winter Medals: 635.0

Deeper Analysis (Performance vs Age)

In [116...




```
EventGroup  Correlation
```

0 Speed Skating 0.002114 1 Cross Country -0.018487 2 Ice Hockey 0.013334 3 Art Competitions -0.049586 4 Alpine Skiing -0.012798 5 Water Polo 0.084946 6 Sailing Mixed -0.034536 7 Equestrianism Mixed -0.053136 8 Shooting Mixed 0.010006 9 Modern Pentathlon 0.023183 10 Figure Skating 0.016688 11 Synchronized Swimming 0.142025 12 Table Tennis -0.131537 13 Nordic Combined 0.151451 14 Rhythmic Gymnastics -0.131599 15 Freestyle Skiing -0.029311 16 Rugby Sevens -0.032574 17 Biathlon Mixed 0.511759 18 Beach Volleyball 0.053426 19 Ski Jumping -0.042966 20 Badminton Mixed -0.325163 21 Tennis Mixed 0.074267 22 Short Track -0.040403 23 Luge Mixed -0.292947 24 Motorboating Mixed NaN 25 Military Ski 0.181735 26 Croquet Mixed -0.436288 27 Jeu De -0.500000 28 Alpinism Mixed NaN 29 Basque Pelota NaN 30 Aeronautics Mixed NaN

- Positive Correlation: Event groups with positive correlation values (e.g., Speed Skating, Ice Hockey, Water Polo, Modern Pentathlon, etc.) indicate a slight positive relationship between the age of athletes and their chances of winning medals. This suggests that older athletes in these

event groups may have a slightly higher likelihood of winning medals.

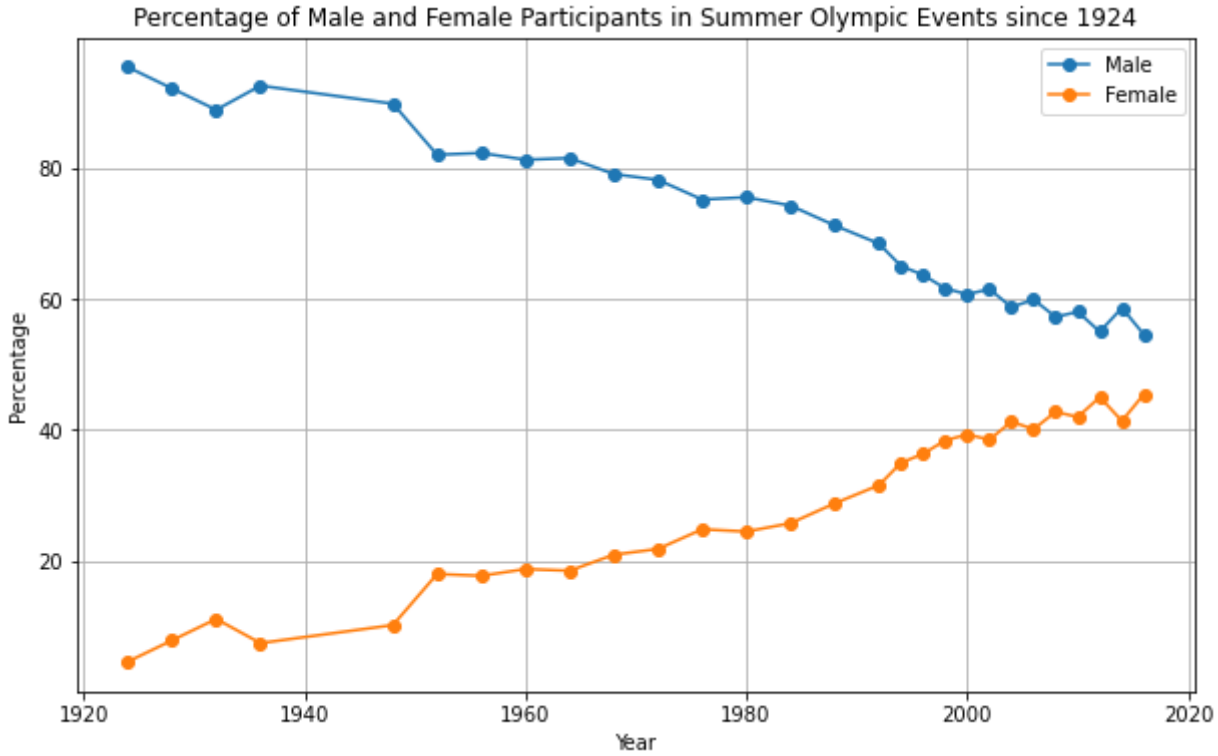
- Negative Correlation: Event groups with negative correlation values (e.g., Cross Country, Art Competitions, Alpine Skiing, Sailing Mixed, Equestrianism Mixed, Table Tennis, Rhythmic Gymnastics, etc.) suggest a negative relationship between age and medal performance. This implies that younger athletes in these event groups might have a better chance of winning medals.

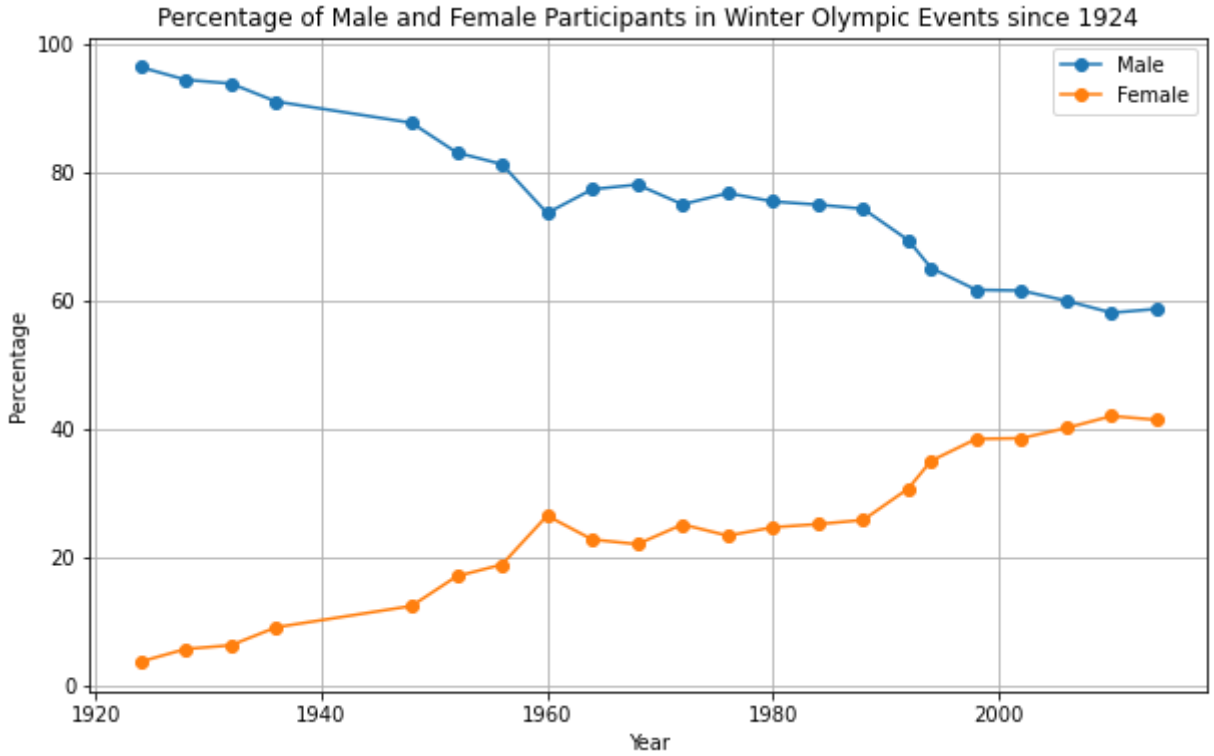
- Strong Correlation: Some event groups like Nordic Combined and Biathlon Mixed show relatively strong positive correlations. This could indicate a stronger relationship between age and medal performance in these groups.

- No Correlation: In some cases where the correlation value is close to 0 or NaN (like Alpinism Mixed, Basque Pelota, Aeronautics Mixed, etc.), it suggests that there is no significant linear relationship between age and medal performance in those event groups.

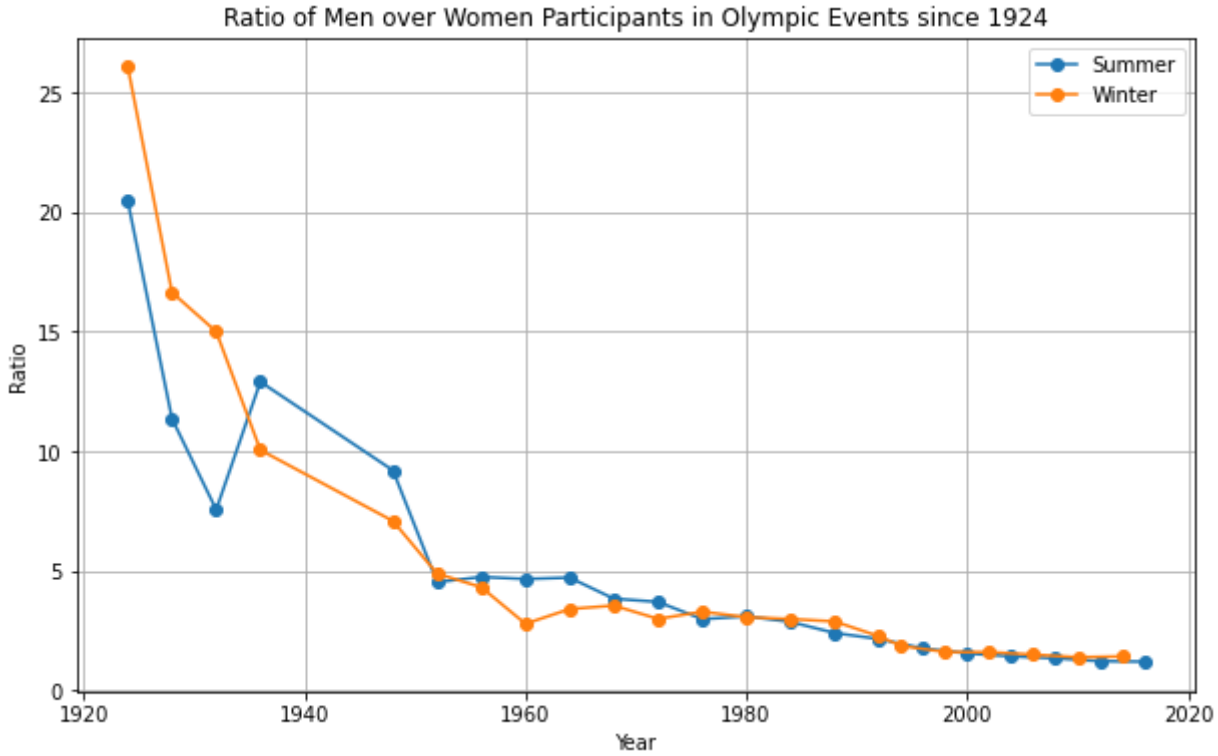
Deeper Analysis (Diversity)

In [104…

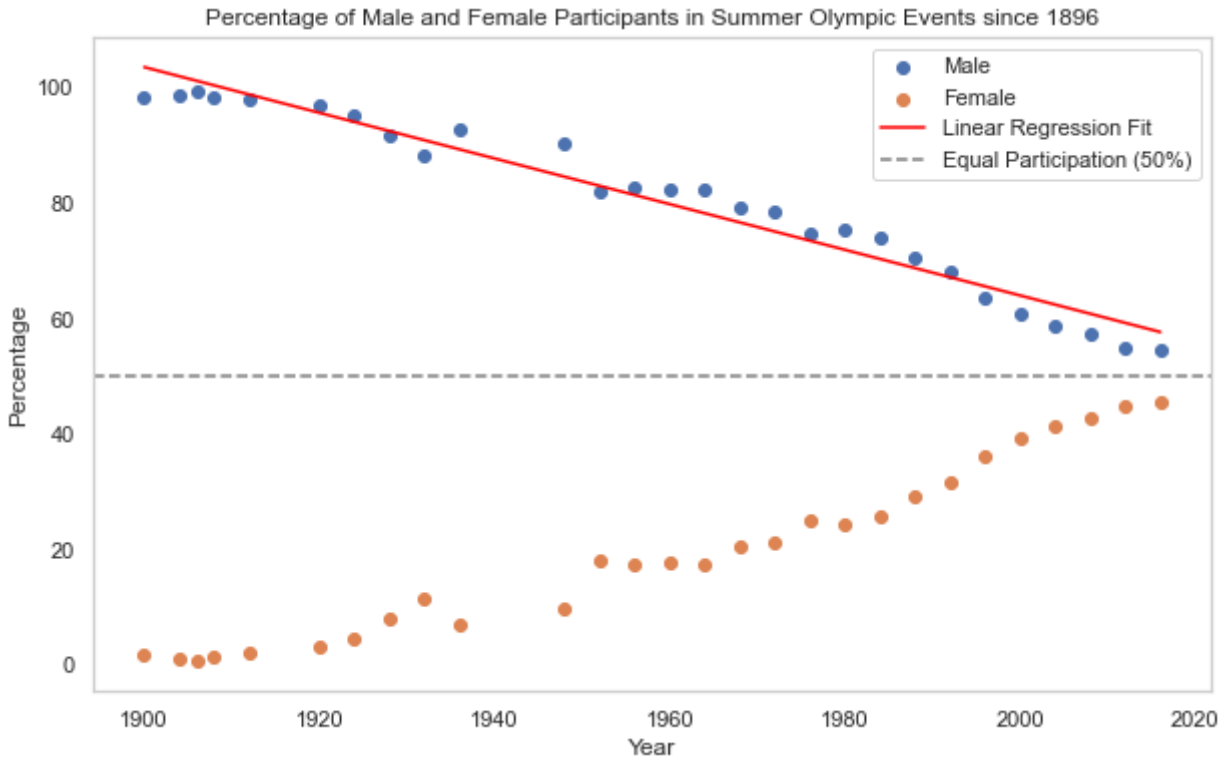

Percentage of Male and Female Participants in Summer Olympic Events since 1924
Male
Female
Percentage
Year
20
40
60
80
1920
1940
1960
1980
2000
2020


Percentage of Male and Female Participants in Winter Olympic Events since 1924
Male
Female
Percentage
Year
100
80
60
40
20
0
1920
1940
1960
1980
2000

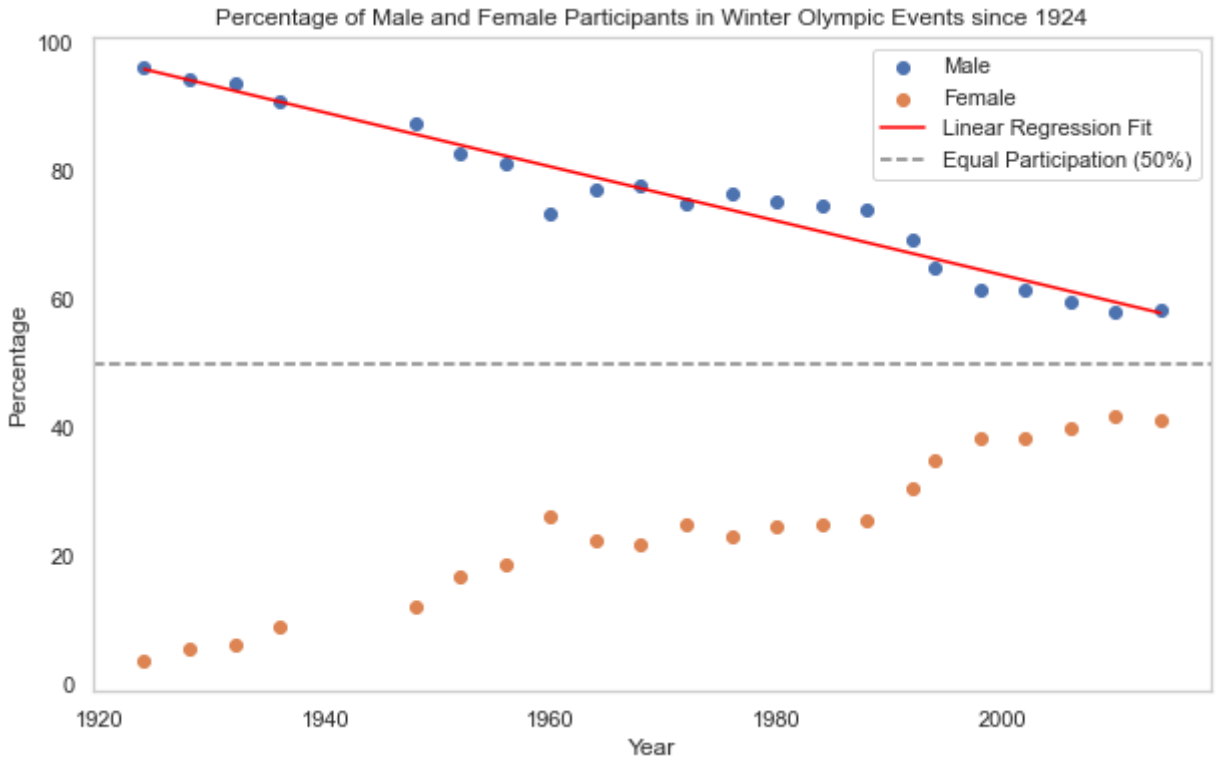
In [28]:


Ratio of Men over Women Participants in Olympic Events since 1924
Summer
Winter
Ratio
Year
0
5
10
15
20
25
1920
1940
1960
1980
2000
2020

In [161…



R-squared value for Summer events: 0.9578479113235537
Predicted Year when Male and Female Participation will be Equal for Summer events: 3931




```
R-squared value for Winter events: 0.9498287782760846
Predicted Year when Male and Female Participation will be Equal for Winter events: 2032
```

In [176…

```
Top 10 Countries with the Biggest Gap in Male and Female Participation (Last 50 Years, Summer Events):
                  region  Percentage_M  Percentage_F  Percentage_Difference
NOC
KUW               Kuwait     98.561151      1.438849              97.122302
KSA         Saudi Arabia     97.685185      2.314815              95.370370
QAT                Qatar     96.354167      3.645833              92.708333
IRQ                 Iraq     95.833333      4.166667              91.666667
PAK             Pakistan     95.510204      4.489796              91.020408
UAE  United Arab Emirates     95.138889      4.861111              90.277778
OMA                 Oman     93.939394      6.060606              87.878788
BOT             Botswana     92.222222      7.777778              84.444444
IRI                 Iran     92.081448      7.918552              84.162896
ZAM               Zambia     90.604027      9.395973              81.208054
Top 10 Countries with the Biggest Gap in Male and Female Participation (Last 50 Years, Winter Events):
           region  Percentage_M  Percentage_F  Percentage_Difference
NOC
SMR    San Marino     97.297297      2.702703              94.594595
PUR   Puerto Rico     94.594595      5.405405              89.189189
TPE        Taiwan     94.318182      5.681818              88.636364
LUX    Luxembourg     93.750000      6.250000              87.500000
HKG         China      8.333333     91.666667              83.333333
MGL      Mongolia     87.719298     12.280702              75.438596
MEX        Mexico     87.692308     12.307692              75.384615
IRI          Iran     87.500000     12.500000              75.000000
RSA  South Africa     85.714286     14.285714              71.428571
POR      Portugal     85.000000     15.000000              70.000000
```

Final Findings

- **Q1: Is there any correlation between the performance of a country in winter Olympics and that in summer Olympics?**

  > It seems that there is a correlation between the performance of a country in winter and summer events. This might indicate that countries with strong sporting infrastructures, training facilities, and resources tend to perform well in both Winter and Summer Olympics. This consistency could be attributed to a comprehensive approach to athlete development.

- **Q2: Does performance change with age?**

> Our analysis indicated that athletes in their late 20s to early 30s tend to win the most medals. This suggests that athletes in this age range might be at their peak physical condition and performance capabilities. On on average, the age of medalists has increased over the years. This suggests that athletes are competing and succeeding at older ages in recent Olympics. The findings also showed that in some events age is not strongly correlated.

- **Q3: How has the male:female ratio evolved through time?**

> SOur analysis revealed that the number of female athletes has been steadily increasing over the years, reflecting a greater gender inclusion and participation in the Olympics. We observed that the early Olympic Games had a significantly imbalanced gender ratio, with a majority of male athletes. However, this ratio has been progressively moving towards a more balanced distribution, indicating the Olympics' efforts towards gender equality. The analysis indicates a positive trend towards a more balanced male:female ratio in the Olympics over the years. This reflects the broader societal progress towards gender equality and highlights the Olympics' role in promoting inclusivity and diversity in sports.

In [ ]: