**120 Years of Olympic History**

INTRODUCTION

1. Meta Information:

The ‘modern Olympics’ comprises all the Games from Athens 1986 to Rio 2016. The Olympics is more than just a quadrennial multi-sport world championship. It is a lens through which to understand global history, including shifting geopolitical power dynamics, women’s empowerment, and the evolving values of society.

This is a historical dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016. The data is being used to study how many countries participated in each year, total medals won by each country, average age of players, number of teams sent by each country in all of the Olympics and their performance.

In [43]:

import pandas as pd

import numpy as np

import seaborn as sns

%pylab inline

Populating the interactive namespace from numpy and matplotlib

In [44]:

dataset = pd.read\_csv('C:\Users\Dheeraj\Desktop/athlete\_events.csv')

dataset.head()

Out[44]:

|  | **ID** | **Name** | **Sex** | **Age** | **Height** | **Weight** | **Team** | **NOC** | **Games** | **Year** | **Season** | **City** | **Sport** | **Event** | **Medal** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 1 | A Dijiang | M | 24.0 | 180.0 | 80.0 | China | CHN | 1992 Summer | 1992 | Summer | Barcelona | Basketball | Basketball Men's Basketball | NaN |
| **1** | 2 | A Lamusi | M | 23.0 | 170.0 | 60.0 | China | CHN | 2012 Summer | 2012 | Summer | London | Judo | Judo Men's Extra-Lightweight | NaN |
| **2** | 3 | Gunnar Nielsen Aaby | M | 24.0 | NaN | NaN | Denmark | DEN | 1920 Summer | 1920 | Summer | Antwerpen | Football | Football Men's Football | NaN |
| **3** | 4 | Edgar Lindenau Aabye | M | 34.0 | NaN | NaN | Denmark/Sweden | DEN | 1900 Summer | 1900 | Summer | Paris | Tug-Of-War | Tug-Of-War Men's Tug-Of-War | Gold |
| **4** | 5 | Christine Jacoba Aaftink | F | 21.0 | 185.0 | 82.0 | Netherlands | NED | 1988 Winter | 1988 | Winter | Calgary | Speed Skating | Speed Skating Women's 500 metres | NaN |

In [45]:

dataset.isnull().sum()

Out[45]:

ID 0

Name 0

Sex 0

Age 9474

Height 60171

Weight 62875

Team 0

NOC 0

Games 0

Year 0

Season 0

City 0

Sport 0

Event 0

Medal 231333

dtype: int64

In [46]:

dataset['Medal'].fillna('NP', inplace=True)

dataset.isnull().sum()

Out[46]:

ID 0

Name 0

Sex 0

Age 9474

Height 60171

Weight 62875

Team 0

NOC 0

Games 0

Year 0

Season 0

City 0

Sport 0

Event 0

Medal 0

dtype: int64

In [47]:

dataset['Age']=dataset['Age'].replace(np.NaN,dataset['Age'].mean())

dataset['Height']=dataset['Height'].replace(np.NaN,dataset['Height'].mean())

dataset['Weight']=dataset['Weight'].replace(np.NaN,dataset['Weight'].mean())

dataset.isnull().sum()

Out[47]:

ID 0

Name 0

Sex 0

Age 0

Height 0

Weight 0

Team 0

NOC 0

Games 0

Year 0

Season 0

City 0

Sport 0

Event 0

Medal 0

dtype: int64

In [48]:

dup = dataset[dataset.duplicated(subset = None, keep=False)]

dup

Out[48]:

|  | **ID** | **Name** | **Sex** | **Age** | **Height** | **Weight** | **Team** | **NOC** | **Games** | **Year** | **Season** | **City** | **Sport** | **Event** | **Medal** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1251** | 704 | Dsir Antoine Acket | M | 27.000000 | 175.33897 | 70.702393 | Belgium | BEL | 1932 Summer | 1932 | Summer | Los Angeles | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **1252** | 704 | Dsir Antoine Acket | M | 27.000000 | 175.33897 | 70.702393 | Belgium | BEL | 1932 Summer | 1932 | Summer | Los Angeles | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **4281** | 2449 | William Truman Aldrich | M | 48.000000 | 175.33897 | 70.702393 | United States | USA | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Drawings And ... | NP |
| **4282** | 2449 | William Truman Aldrich | M | 48.000000 | 175.33897 | 70.702393 | United States | USA | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Drawings And ... | NP |
| **4283** | 2449 | William Truman Aldrich | M | 48.000000 | 175.33897 | 70.702393 | United States | USA | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Drawings And ... | NP |
| **4861** | 2777 | Hermann Reinhard Alker | M | 43.000000 | 175.33897 | 70.702393 | Germany | GER | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Architecture, Designs F... | NP |
| **4862** | 2777 | Hermann Reinhard Alker | M | 43.000000 | 175.33897 | 70.702393 | Germany | GER | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Architecture, Designs F... | NP |
| **4863** | 2777 | Hermann Reinhard Alker | M | 43.000000 | 175.33897 | 70.702393 | Germany | GER | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Architecture, Architect... | NP |
| **4864** | 2777 | Hermann Reinhard Alker | M | 43.000000 | 175.33897 | 70.702393 | Germany | GER | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Architecture, Architect... | NP |
| **4866** | 2777 | Hermann Reinhard Alker | M | 51.000000 | 175.33897 | 70.702393 | Germany | GER | 1936 Summer | 1936 | Summer | Berlin | Art Competitions | Art Competitions Mixed Architecture, Unknown E... | NP |
| **4867** | 2777 | Hermann Reinhard Alker | M | 51.000000 | 175.33897 | 70.702393 | Germany | GER | 1936 Summer | 1936 | Summer | Berlin | Art Competitions | Art Competitions Mixed Architecture, Unknown E... | NP |
| **5104** | 2903 | Lucien Charles Edouard Alliot | M | 46.000000 | 175.33897 | 70.702393 | France | FRA | 1924 Summer | 1924 | Summer | Paris | Art Competitions | Art Competitions Mixed Sculpturing | NP |
| **5105** | 2903 | Lucien Charles Edouard Alliot | M | 46.000000 | 175.33897 | 70.702393 | France | FRA | 1924 Summer | 1924 | Summer | Paris | Art Competitions | Art Competitions Mixed Sculpturing | NP |
| **5106** | 2903 | Lucien Charles Edouard Alliot | M | 46.000000 | 175.33897 | 70.702393 | France | FRA | 1924 Summer | 1924 | Summer | Paris | Art Competitions | Art Competitions Mixed Sculpturing | NP |
| **7769** | 4319 | Ludwig Angerer | M | 41.000000 | 175.33897 | 70.702393 | Germany | GER | 1932 Summer | 1932 | Summer | Los Angeles | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **7770** | 4319 | Ludwig Angerer | M | 41.000000 | 175.33897 | 70.702393 | Germany | GER | 1932 Summer | 1932 | Summer | Los Angeles | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **7771** | 4319 | Ludwig Angerer | M | 41.000000 | 175.33897 | 70.702393 | Germany | GER | 1932 Summer | 1932 | Summer | Los Angeles | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **7772** | 4319 | Ludwig Angerer | M | 41.000000 | 175.33897 | 70.702393 | Germany | GER | 1932 Summer | 1932 | Summer | Los Angeles | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **7773** | 4319 | Ludwig Angerer | M | 45.000000 | 175.33897 | 70.702393 | Germany | GER | 1936 Summer | 1936 | Summer | Berlin | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **7774** | 4319 | Ludwig Angerer | M | 45.000000 | 175.33897 | 70.702393 | Germany | GER | 1936 Summer | 1936 | Summer | Berlin | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **9365** | 5146 | George Denholm Armour | M | 64.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **9366** | 5146 | George Denholm Armour | M | 64.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **9367** | 5146 | George Denholm Armour | M | 64.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **9368** | 5146 | George Denholm Armour | M | 64.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **9369** | 5146 | George Denholm Armour | M | 68.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1932 Summer | 1932 | Summer | Los Angeles | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **9370** | 5146 | George Denholm Armour | M | 68.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1932 Summer | 1932 | Summer | Los Angeles | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **10281** | 5623 | Konstantinos Aslanidis | M | 25.556898 | 175.33897 | 70.702393 | Greece | GRE | 1948 Summer | 1948 | Summer | London | Art Competitions | Art Competitions Mixed Architecture, Unknown E... | NP |
| **10282** | 5623 | Konstantinos Aslanidis | M | 25.556898 | 175.33897 | 70.702393 | Greece | GRE | 1948 Summer | 1948 | Summer | London | Art Competitions | Art Competitions Mixed Architecture, Unknown E... | NP |
| **10294** | 5634 | Axel Edvin "Acke" slund | M | 50.000000 | 175.33897 | 70.702393 | Sweden | SWE | 1932 Summer | 1932 | Summer | Los Angeles | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **10295** | 5634 | Axel Edvin "Acke" slund | M | 50.000000 | 175.33897 | 70.702393 | Sweden | SWE | 1932 Summer | 1932 | Summer | Los Angeles | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **266300** | 133226 | Mahonri Mackintosh Young | M | 54.000000 | 175.33897 | 70.702393 | United States | USA | 1932 Summer | 1932 | Summer | Los Angeles | Art Competitions | Art Competitions Mixed Sculpturing, Unknown Event | NP |
| **267330** | 133749 | Oldich "Olda" k | M | 35.000000 | 175.33897 | 70.702393 | Czechoslovakia | TCH | 1936 Summer | 1936 | Summer | Berlin | Art Competitions | Art Competitions Mixed Sculpturing, Unknown Event | NP |
| **267331** | 133749 | Oldich "Olda" k | M | 35.000000 | 175.33897 | 70.702393 | Czechoslovakia | TCH | 1936 Summer | 1936 | Summer | Berlin | Art Competitions | Art Competitions Mixed Sculpturing, Unknown Event | NP |
| **267332** | 133749 | Oldich "Olda" k | M | 35.000000 | 175.33897 | 70.702393 | Czechoslovakia | TCH | 1936 Summer | 1936 | Summer | Berlin | Art Competitions | Art Competitions Mixed Sculpturing, Unknown Event | NP |
| **267928** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267929** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267930** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267931** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267932** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267933** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267934** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267935** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267936** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267937** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267938** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267939** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267940** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267941** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267942** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **267943** | 134046 | ngel Zrraga Argelles | M | 41.000000 | 175.33897 | 70.702393 | Mexico | MEX | 1928 Summer | 1928 | Summer | Amsterdam | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **269992** | 135072 | Anna Katrina Zinkeisen (-Heseltine) | F | 46.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1948 Summer | 1948 | Summer | London | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **269993** | 135072 | Anna Katrina Zinkeisen (-Heseltine) | F | 46.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1948 Summer | 1948 | Summer | London | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **269994** | 135072 | Anna Katrina Zinkeisen (-Heseltine) | F | 46.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1948 Summer | 1948 | Summer | London | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **269995** | 135072 | Anna Katrina Zinkeisen (-Heseltine) | F | 46.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1948 Summer | 1948 | Summer | London | Art Competitions | Art Competitions Mixed Painting, Paintings | NP |
| **269996** | 135072 | Anna Katrina Zinkeisen (-Heseltine) | F | 46.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1948 Summer | 1948 | Summer | London | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **269997** | 135072 | Anna Katrina Zinkeisen (-Heseltine) | F | 46.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1948 Summer | 1948 | Summer | London | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **269998** | 135073 | Doris Clare Zinkeisen (-Johnstone) | F | 49.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1948 Summer | 1948 | Summer | London | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **269999** | 135073 | Doris Clare Zinkeisen (-Johnstone) | F | 49.000000 | 175.33897 | 70.702393 | Great Britain | GBR | 1948 Summer | 1948 | Summer | London | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **270199** | 135173 | Henri Achille Zo | M | 58.000000 | 175.33897 | 70.702393 | France | FRA | 1932 Summer | 1932 | Summer | Los Angeles | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |
| **270200** | 135173 | Henri Achille Zo | M | 58.000000 | 175.33897 | 70.702393 | France | FRA | 1932 Summer | 1932 | Summer | Los Angeles | Art Competitions | Art Competitions Mixed Painting, Unknown Event | NP |

1997 rows × 15 columns

In [49]:

dataset.drop\_duplicates(subset="Name", keep=False, inplace=True)

dup = dataset[dataset.duplicated(subset = None, keep=False)]

dup

Out[49]:

|  | **ID** | **Name** | **Sex** | **Age** | **Height** | **Weight** | **Team** | **NOC** | **Games** | **Year** | **Season** | **City** | **Sport** | **Event** | **Medal** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

In [50]:

most\_gold = dataset[dataset['Medal'] == 'Gold']['Team'].value\_counts().head()

gold = pd.DataFrame(most\_gold)

gold.reset\_index(inplace=True)

gold

Out[50]:

|  | **index** | **Team** |
| --- | --- | --- |
| **0** | United States | 770 |
| **1** | Soviet Union | 288 |
| **2** | Canada | 174 |
| **3** | Great Britain | 162 |
| **4** | Germany | 159 |

In [51]:

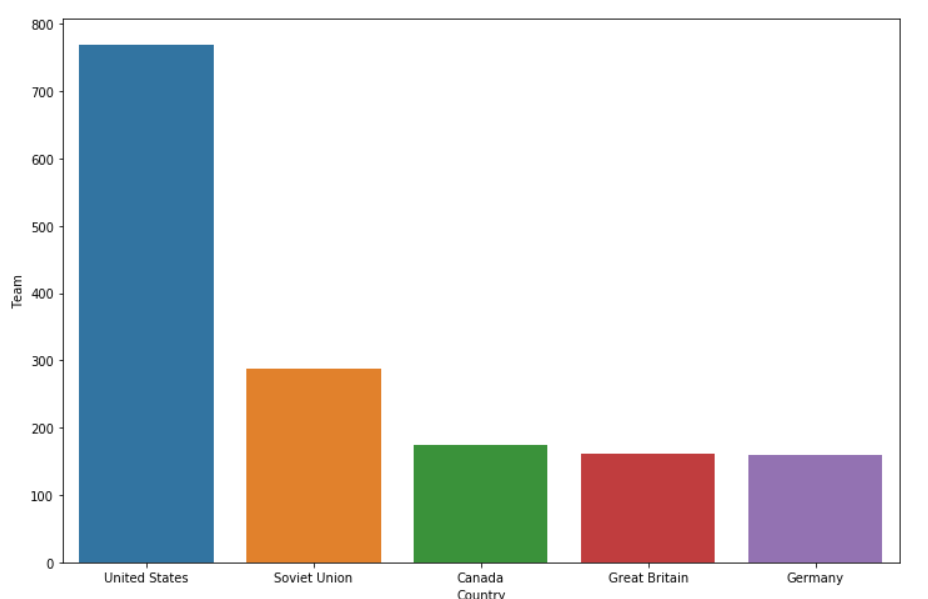
gold.rename(columns={'index':'Country'}, inplace=True)

plt.figure(figsize=(12,8))

sns.barplot(x='Country', y='Team', data=gold)

Out[51]:

<matplotlib.axes.\_subplots.AxesSubplot at 0xf3f4358>



In [55]:

dataset.describe(include=[np.number])

Out[55]:

|  | **ID** | **Age** | **Height** | **Weight** | **Year** |
| --- | --- | --- | --- | --- | --- |
| **count** | 77009.000000 | 77009.000000 | 77009.000000 | 77009.000000 | 77009.000000 |
| **mean** | 67266.393590 | 25.092461 | 176.275986 | 71.892533 | 1976.867379 |
| **std** | 39188.417047 | 5.518379 | 8.765076 | 12.248978 | 31.156368 |
| **min** | 1.000000 | 10.000000 | 127.000000 | 28.000000 | 1896.000000 |
| **25%** | 33193.000000 | 22.000000 | 172.000000 | 66.000000 | 1956.000000 |
| **50%** | 67311.000000 | 24.000000 | 175.338970 | 70.702393 | 1984.000000 |
| **75%** | 101016.000000 | 27.000000 | 180.000000 | 76.000000 | 2004.000000 |
| **max** | 135569.000000 | 97.000000 | 223.000000 | 198.000000 | 2016.000000 |

In [57]:

dataset.describe(exclude=[np.number])

Out[57]:

|  | **Name** | **Sex** | **Team** | **NOC** | **Games** | **Season** | **City** | **Sport** | **Event** | **Medal** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **count** | 77009 | 77009 | 77009 | 77009 | 77009 | 77009 | 77009 | 77009 | 77009 | 77009 |
| **unique** | 77009 | 2 | 892 | 229 | 51 | 2 | 42 | 66 | 661 | 4 |
| **top** | Forrest Grady Towns | M | United States | USA | 2016 Summer | Summer | London | Athletics | Football Men's Football | NP |
| **freq** | 1 | 59529 | 5307 | 5721 | 6288 | 68883 | 7334 | 12300 | 5094 | 65517 |

In [59]:

year = dataset.groupby('Year')['Sport'].nunique()

year = pd.DataFrame(year)

year.reset\_index(inplace=True)

year.head()

Out[59]:

|  | **Year** | **Sport** |
| --- | --- | --- |
| **0** | 1896 | 8 |
| **1** | 1900 | 20 |
| **2** | 1904 | 18 |
| **3** | 1906 | 13 |
| **4** | 1908 | 24 |

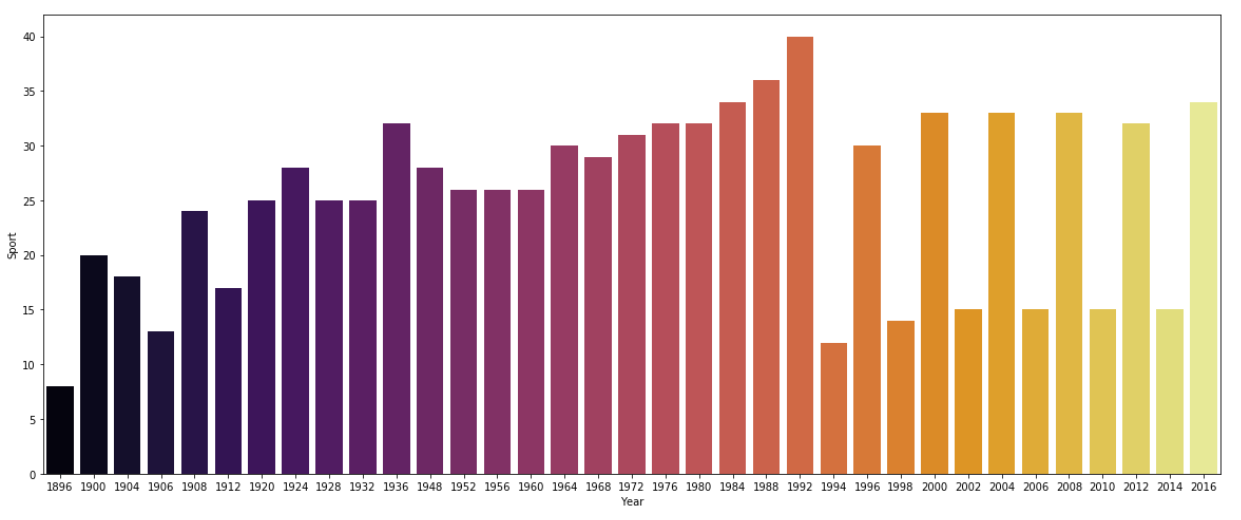
In [60]:

plt.figure(figsize=(20,8))

sns.barplot(x='Year', y='Sport', data=year, palette='inferno')

Out[60]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x10f93d68>



In [61]:

dataset.groupby('NOC')['Medal'].count().head()

medal = dataset.groupby('NOC')['Medal'].count()

medal = pd.DataFrame(medal)

medal.reset\_index(inplace=True)

medal.head()

Out[61]:

|  | **NOC** | **Medal** |
| --- | --- | --- |
| **0** | AFG | 82 |
| **1** | AHO | 37 |
| **2** | ALB | 26 |
| **3** | ALG | 251 |
| **4** | AND | 24 |

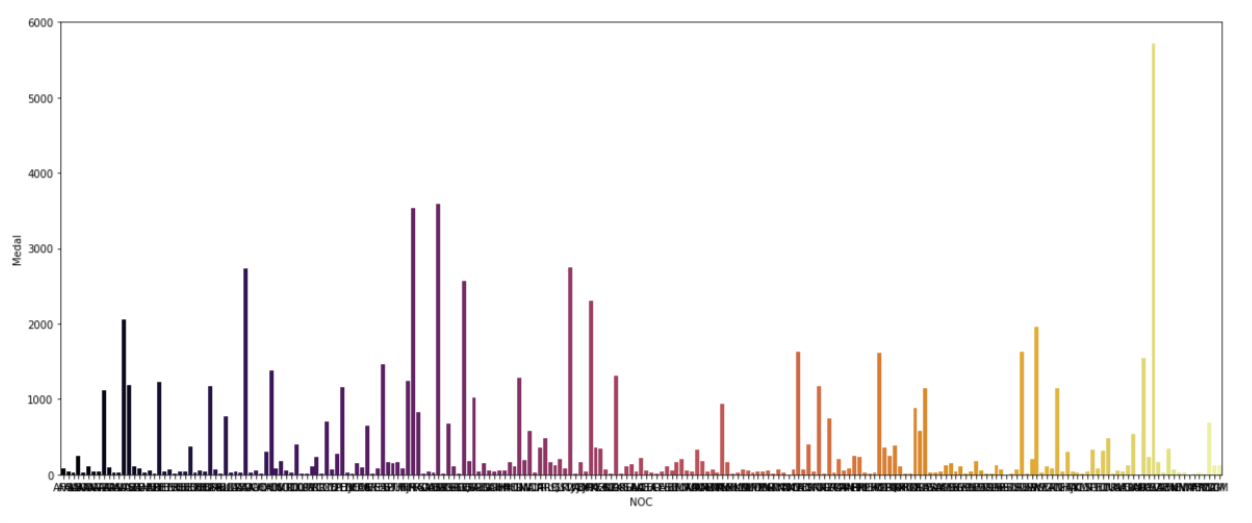
In [62]:

plt.figure(figsize=(20,8))

sns.barplot(x='NOC', y='Medal', data=medal, palette='inferno')

Out[62]:

<matplotlib.axes.\_subplots.AxesSubplot at 0xf592cc0>



In [63]:

dataset.corr()

Out[63]:

|  | **ID** | **Age** | **Height** | **Weight** | **Year** |
| --- | --- | --- | --- | --- | --- |
| **ID** | 1.000000 | 0.002684 | -0.000891 | -0.001213 | 0.015169 |
| **Age** | 0.002684 | 1.000000 | 0.044235 | 0.111739 | -0.112144 |
| **Height** | -0.000891 | 0.044235 | 1.000000 | 0.751670 | 0.049869 |
| **Weight** | -0.001213 | 0.111739 | 0.751670 | 1.000000 | 0.037631 |
| **Year** | 0.015169 | -0.112144 | 0.049869 | 0.037631 | 1.000000 |

In [72]:

categorical = list(dataset.select\_dtypes(include=['object']).columns.values)

categorical

Out[72]:

['Name',

'Sex',

'Team',

'NOC',

'Games',

'Season',

'City',

'Sport',

'Event',

'Medal']

In [74]:

from sklearn.preprocessing import LabelEncoder

le=LabelEncoder()

for i in range(0, len(categorical)):

dataset[categorical[i]] = le.fit\_transform(dataset[categorical[i]])

dataset.head()

Out[74]:

|  | **ID** | **Name** | **Sex** | **Age** | **Height** | **Weight** | **Team** | **NOC** | **Games** | **Year** | **Season** | **City** | **Sport** | **Event** | **Medal** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 1 | 4 | 1 | 24.0 | 180.00000 | 80.000000 | 154 | 41 | 37 | 1992 | 0 | 5 | 8 | 133 | 2 |
| **1** | 2 | 5 | 1 | 23.0 | 170.00000 | 60.000000 | 154 | 41 | 48 | 2012 | 0 | 17 | 32 | 337 | 2 |
| **2** | 3 | 25337 | 1 | 24.0 | 175.33897 | 70.702393 | 207 | 55 | 6 | 1920 | 0 | 2 | 24 | 297 | 2 |
| **3** | 4 | 16828 | 1 | 34.0 | 175.33897 | 70.702393 | 210 | 55 | 1 | 1900 | 0 | 26 | 61 | 607 | 1 |
| **28** | 9 | 6381 | 1 | 26.0 | 186.00000 | 96.000000 | 264 | 68 | 43 | 2002 | 1 | 29 | 30 | 334 | 2 |

In [81]:

from sklearn.model\_selection import train\_test\_split

x = dataset[['Age', 'Year', 'NOC', 'Event']]

y = dataset['Medal']

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x,y, random\_state=101, test\_size=0.30)

In [82]:

from sklearn.ensemble import RandomForestClassifier

In [84]:

rcf = RandomForestClassifier()

rcf.fit(x\_train, y\_train)

Out[84]:

RandomForestClassifier(bootstrap=True, class\_weight=None, criterion='gini',

max\_depth=None, max\_features='auto', max\_leaf\_nodes=None,

min\_impurity\_decrease=0.0, min\_impurity\_split=None,

min\_samples\_leaf=1, min\_samples\_split=2,

min\_weight\_fraction\_leaf=0.0, n\_estimators=10, n\_jobs=1,

oob\_score=False, random\_state=None, verbose=0,

warm\_start=False)

In [85]:

pred = rcf.predict(x\_test)

In [86]:

from sklearn.metrics import classification\_report

print(classification\_report(y\_test, pred))

precision recall f1-score support

0 0.68 0.47 0.56 1253

1 0.75 0.62 0.68 1132

2 0.93 0.98 0.95 19556

3 0.76 0.48 0.59 1162

avg / total 0.90 0.91 0.90 23103