Olympic Sports and Medals

[From 1896 – 2014]

By: Alec Mintz, Natasha Fidler, Bianca Santalla Terán

Intro to the data: More than 35,000 medals have been awarded at the olympics since 1896. Our dataset, "Olympic Sports and Medals, 1896 - 2014" dives into the history of the olympics, showing data for every olympic medal winner in the last 100+ years

Our dataset seeks to explore:

- Why has there been a recent increase in the number of athletes winning medals?
- Which sports are the most globally participated in?
- Who are the most successful olympic athletes?
- Which country are they from, and how has their success changed over time?
- Which events are the most popular?
- Do the countries with more medals overlap with higher GDP?
- Do men and women athletes earn comparable amount of medals?
- What countries have the most women medalists?
- What are ar

This data was provided by the IOC Research and Reference Service and published by the Guardian's Datablog (https://www.kaggle.com/datasets/the-guardian/olympic-games)

- Merge the 3 CSV files into 1 final datasource
  - Summer Olympics
  - Winter Olympics
  - Country Data
- Remove NaN values; originating from "country" dataset (GDP per capita)

Total Athletes	36,935
Total Countries	118
Total Cities	41
Total Events	733
Total Sports	48

Year	City	Sport	Discipline	Athlete	Code	Gender	Event	Medal	Season	Country	Population	GDP per Capita
1896	Athens	Aquatics	Swimming	HAJOS, Alfred	HUN	Men	100M Freestyle	Gold	Summer	Hungary	9844686	12363.54
1896	Athens	Aquatics	Swimming	HAJOS, Alfred	HUN	Men	1200M Freestyle	Gold	Summer	Hungary	9844686	12363.54
1896	Athens	Athletics	Athletics	SZOKOLYI, Alajos	HUN	Men	100M	Bronze	Summer	Hungary	9844686	12363.54
1896	Athens	Athletics	Athletics	DANI, Nandor	HUN	Men	800M	Silver	Summer	Hungary	9844686	12363.54
1896	Athens	Athletics	Athletics	KELLNER, Gyula	HUN	Men	Marathon	Bronze	Summer	Hungary	9844686	12363.54
...	...	...	...	...	...	...	...	...	...	...	...	...
2012	London	Athletics	Athletics	BARRONDO, Erick	GUA	Men	20KM Walk	Silver	Summer	Guatemala	16342897	3903.48
2012	London	Athletics	Athletics	JAMES, Kirani	GRN	Men	400M	Gold	Summer	Grenada	106825	9212.02
2012	London	Athletics	Athletics	AMOS, Nijel	BOT	Men	800M	Silver	Summer	Botswana	2262485	6360.14
2012	London	Sailing	Sailing	KONTIDES, Pavlos	CYP	Men	Laser	Silver	Summer	Cyprus	1165300	23242.84
2012	London	Taekwondo	Taekwondo	OBAME, Anthony	GAB	Men	+ 80 KG	Silver	Summer	Gabon	1725292	8266.45

Recent Olympic Boom

- The data shows a significant increase in the number of athletes receiving medals after the year 1990

- This remains consistent for both summer and winter athletes

- Why might this be the case?

```python
unique_athletes_per_year = final_df.groupby(['Year', 'Season'])['Athlete'].nunique().reset_index()

summer_data = unique_athletes_per_year[unique_athletes_per_year['Season'] == 'Summer']
winter_data = unique_athletes_per_year[unique_athletes_per_year['Season'] == 'Winter']

plt.figure(figsize=(10, 6))
plt.scatter(summer_data['Year'], summer_data['Athlete'], color='orange', marker='o', label='Summer', s=100, alpha=1)
plt.scatter(winter_data['Year'], winter_data['Athlete'], color='blue', marker='o', label='Winter', s=100, alpha=1)

plt.xlabel('Year')
plt.ylabel('Number of Athletes')
plt.title('Number of Unique Athletes per Olympic Year')

plt.grid(True, linestyle='--', alpha=0.5)

plt.legend()

plt.show()
```
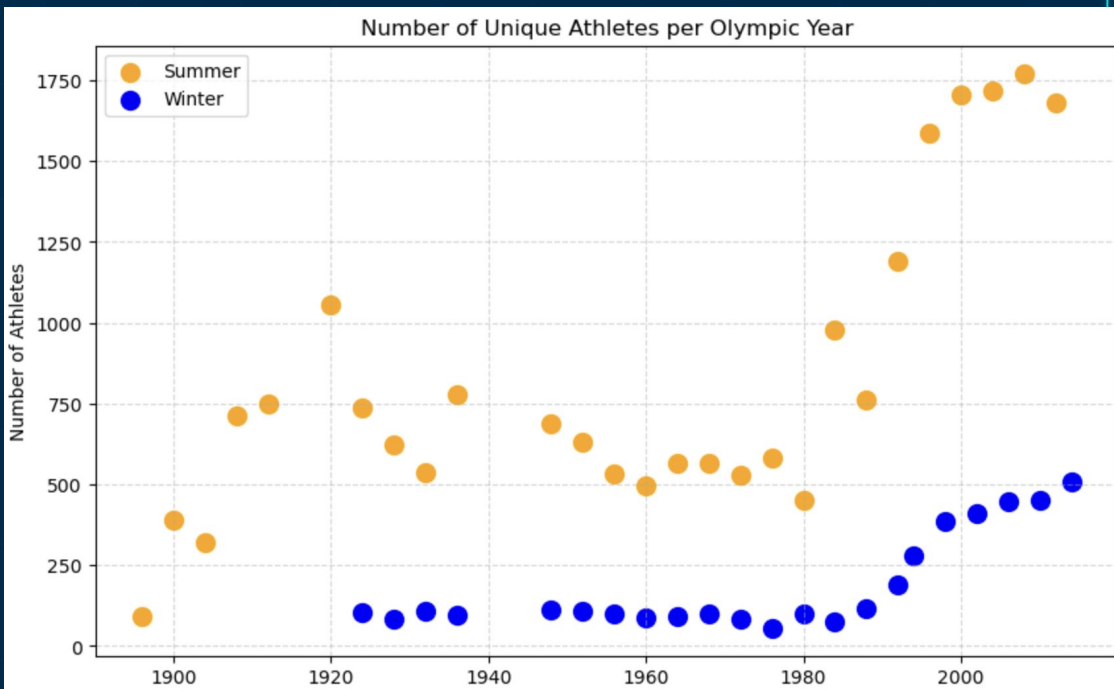

Number of Unique Athletes per Olympic Year

Recent Olympic Boom Explained

- From the years 1960 - 1980, there were 27 different sports, with 206 unique events being played at the Olympics

- From the years 1990 - 2010, there were 35 different sports, with 325 unique events being played at the Olympics

- The addition of 8 new sports and 119 new events has led to the dramatic increase in olympic medal winners

```python
In [50]: sports_1960_to_1980 = final_df[(final_df['Year'] >= 1960) & (final_df['Year'] <= 1980)]['Sport'].unique()

         print("Unique Sports from 1960 to 1980:")
         print(len(sports_1960_to_1980))

         Unique Sports from 1960 to 1980:
         27
```

```python
In [54]: events_1960_to_1980 = final_df[(final_df['Year'] >= 1960) & (final_df['Year'] <= 1980)]['Event'].unique()

         print("Unique Events from 1960 to 1980:")
         print(len(events_1960_to_1980))

         Unique Events from 1960 to 1980:
         206
```

```python
In [52]: sports_1990_to_2010 = final_df[(final_df['Year'] >= 1990) & (final_df['Year'] <= 2010)]['Sport'].unique()

         print("Unique Sports from 1990 to 2010:")
         print(len(sports_1990_to_2010))

         Unique Sports from 1990 to 2010:
         35
```

```python
In [53]: events_1990_to_2010 = final_df[(final_df['Year'] >= 1990) & (final_df['Year'] <= 2010)]['Event'].unique()

         print("Unique Events from 1990 to 2010:")
         print(len(events_1990_to_2010))

         Unique Events from 1990 to 2010:
         325
```

Most popular Olympic Sports by Participating Countries

- The sport with the highest number of countries winning medals is "Athletics"

- 79 different countries have won medals in the last 100+ years

- There are 56 different different events in the sports "Athletics"

- The United States has won the most medals in athletics with 994 medals

```python
medals_by_country = athletics_df.groupby('Country')['Medal'].count()

most_medals_country = medals_by_country.idxmax()
num_medals_most_country = medals_by_country.max()

print(f"Country with the most medals in Athletics: {most_medals_country}")
print(f"Number of medals won: {num_medals_most_country}")
```
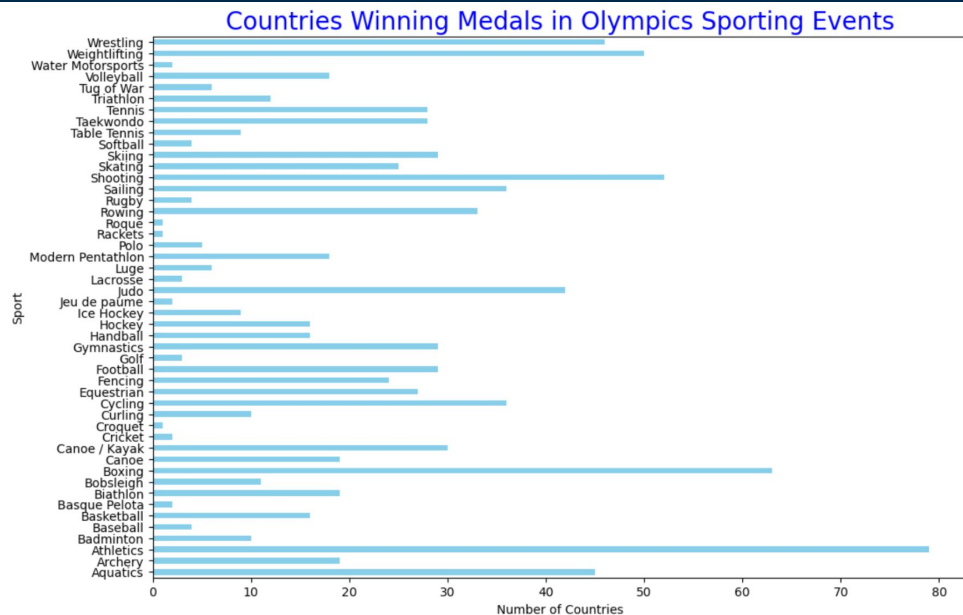```
Country with the most medals in Athletics: United States
Number of medals won: 994
```

```python
sport_countries = final_df.groupby('Sport')['Country'].nunique()

plt.figure(figsize=(12, 8))
sport_countries.plot(kind='barh', color='skyblue')

plt.xlabel('Number of Countries')
plt.ylabel('Sport')
plt.title('Countries Winning Medals in Olympics Sporting Events', color='blue', fontsize=20)

plt.show()
```
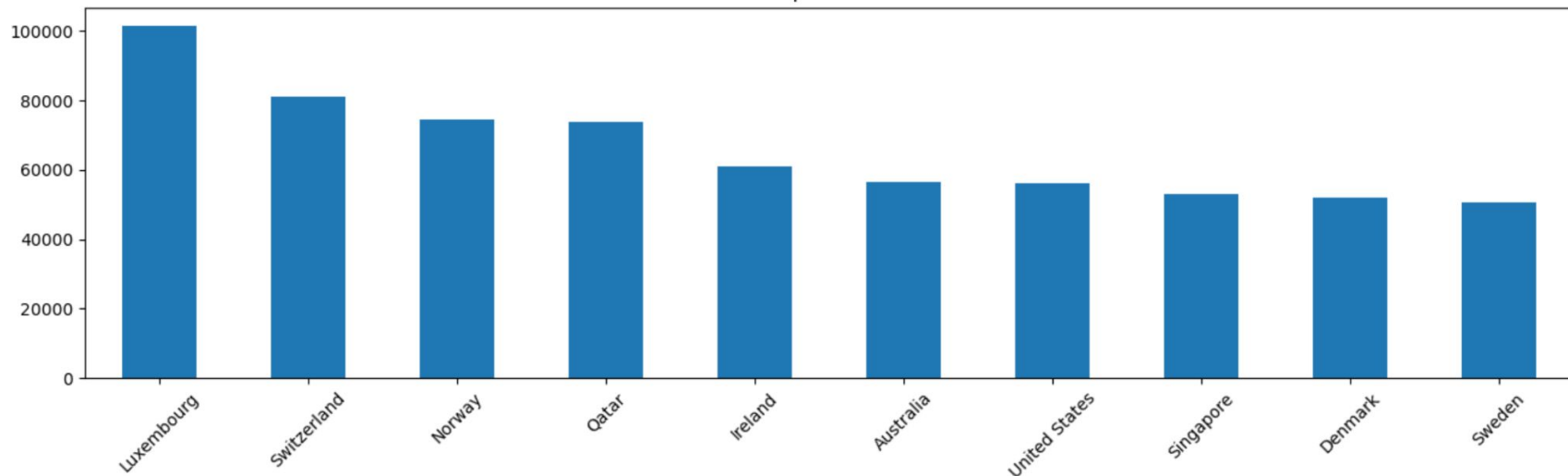
Top 10 Highest GDP Countries

- Top 10 Countries includes 6 European Countries
- Luxembourg has the Highest GDP, but does not have the most medals
- USA is 7th Highest GDP, despite having the most medals
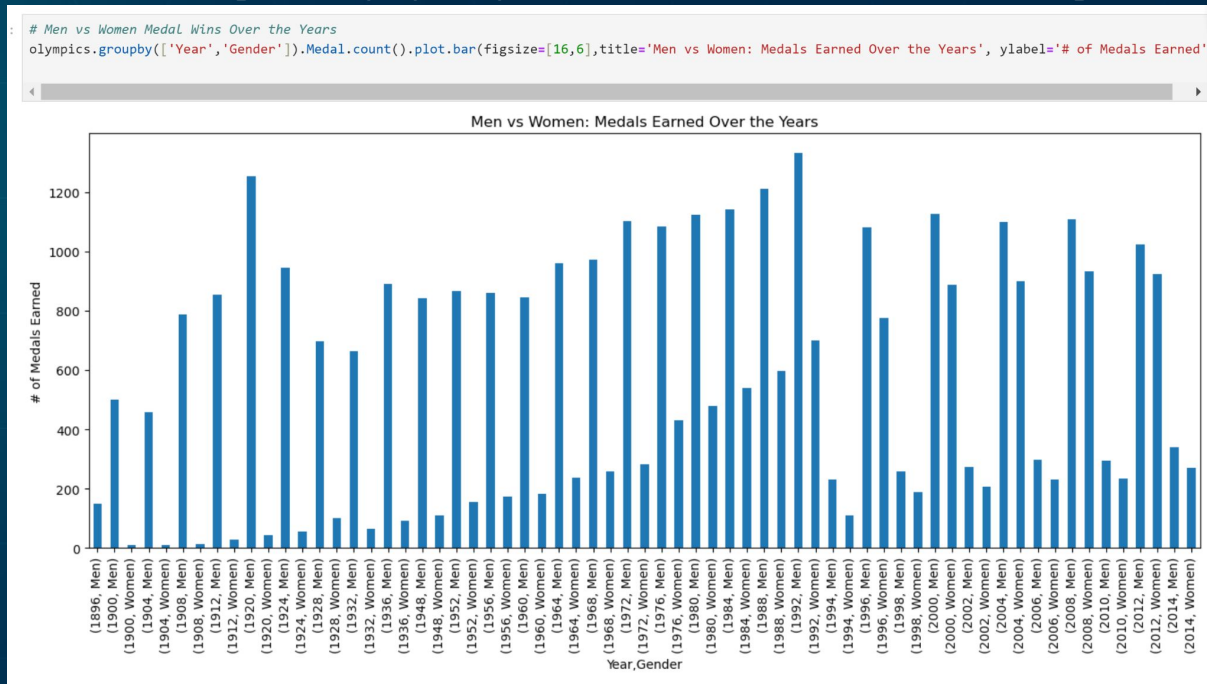
```
#Top10GDP Countries Calculated
clean.groupby('Country')['GDP per Capita'].mean().sort_values(ascending=False)[:10].plot.bar(figsize=[16,4],rot=45,title='Top 10 GDP');
```

Men vs Women: Medals Earned Over the Years

- Men have consistently earned more medals than women in olympic sports
    - Likely, most sports were only available for men to start
- Women have increased their medal count over the years
    - Likely, more sports have become available for women to participate in
- Note: 1994 started a change in Olympic Sports. Winter/Summer were no longer held in the same year.
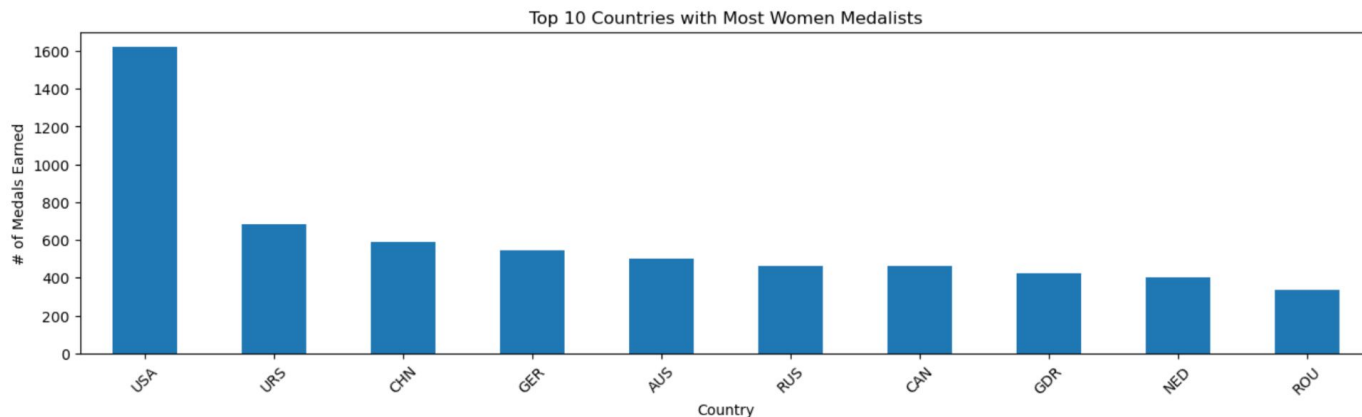
Countries with Most Women Medalists

- Australia and USA the only countries also with Top 10 GDP
- High GDP does not correlate to best women medalists
- High GDP does not have direct causation to accessibility for women sports
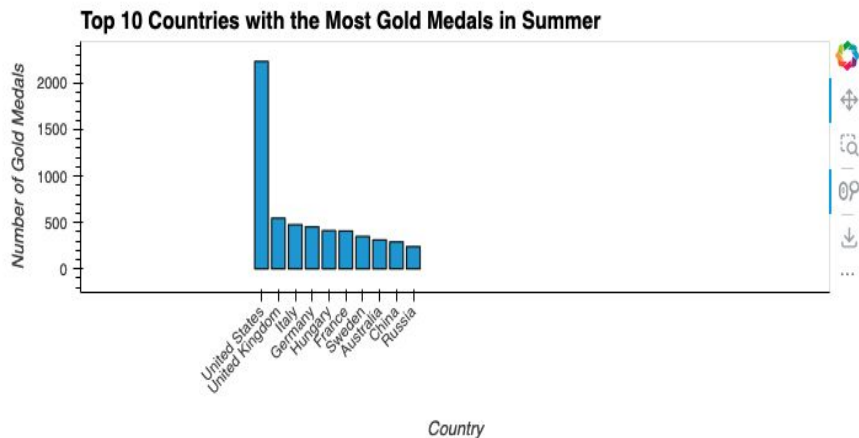
Top 10 Countries with Gold Medals: Summer

```python
gold_df = final_df[(final_df['Medal'] == 'Gold') & (final_df['Seasons'] == 'summer')]

# Group by country, sum the gold medals
gold_grouped = gold_df.groupby('Country')['Medal'].count().reset_index()

# Sort by the number of gold medals in descending order
gold_top10_summer = gold_grouped.sort_values(by='Medal', ascending=False).head(10)

# Plot the bar graph using hvplot and directly chain the opts
gold_top10_summer.hvplot.bar(x='Country', y='Medal', rot=45, xlabel='Country', ylabel='Number of Gold Medals', title='Top 10 Countries with th
```


Top 10 Countries with the Most Gold Medals in Summer

What the data shows:
- ➔ United States has collected the most Gold Medals in the Summer Olympics over time.
- ➔ The US have almost quadrupled the amount as other countries.
- ➔ The United States and Sweden are also both on the top 10 GDP list.

Top 10 Countries with Gold Medals: Winter

```python
gold_df = final_df[(final_df['Medal'] == 'Gold') & (final_df['Seasons'] == 'winter')]

# Group by country, sum the gold medals
gold_grouped = gold_df.groupby('Country')['Medal'].count().reset_index()

# Sort by the number of gold medals in descending order
gold_top10_winter = gold_grouped.sort_values(by='Medal', ascending=False).head(10)

# Plot the bar graph using hvplot and set the color to dark blue
gold_top10_winter.hvplot.bar(x='Country', y='Medal', rot=45, xlabel='Country', ylabel='Number of Gold Medals', title='Top 10 Countries with th
```
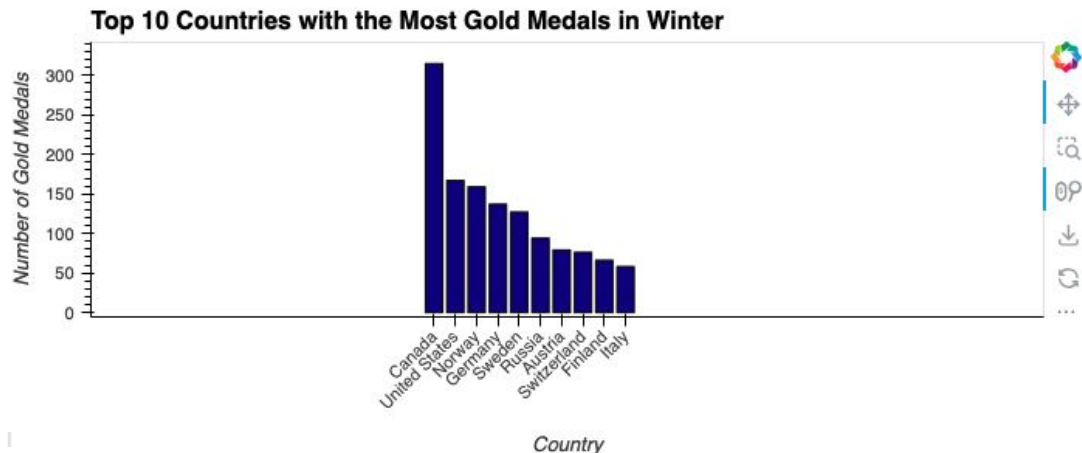

Top 10 Countries with the Most Gold Medals in Winter

What the data shows:
➔   The United States and Sweden are also both on the top 10 GDP list same as Summer Top 10 Gold Medals.
➔   Canada is the Country with the most Gold medals almost doubling the United States.
➔   4 of the same countries are on both the Summer and Winter list.
➔   These larger countries with higher GDPs are on the lists of top 10 Countries in the world with the most amount of Gold Medals.
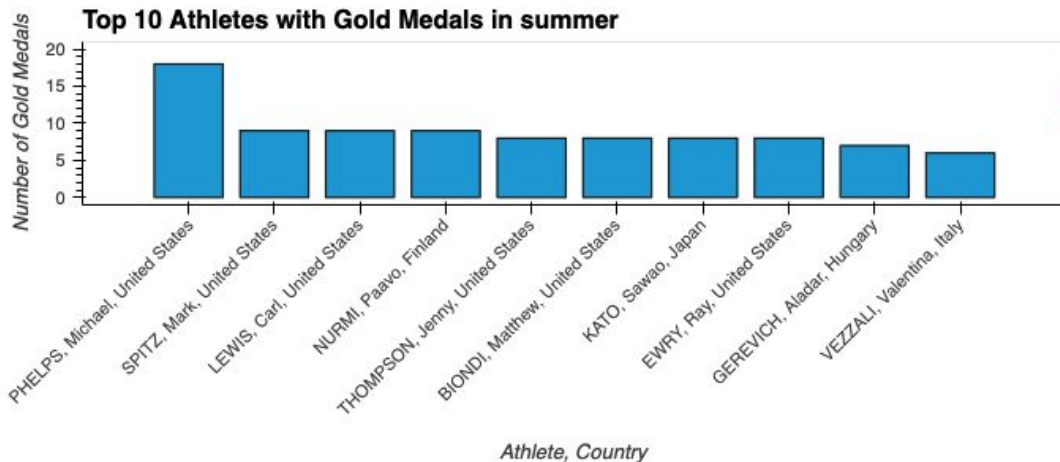
Top 10 Athletes with Gold Medals: Summer

```python
# Top 10 Athletes with gold wins
Athlete_gold_df = final_df[(final_df['Medal'] == 'Gold')& (final_df['Seasons'] == 'summer')]

# Group by athlete and country, sum the gold medals
Athlete_gold_grouped = Athlete_gold_df.groupby(['Athlete', 'Country'])['Medal'].count().reset_index()

# Combine 'Athlete' and 'Country' into a new column
Athlete_gold_grouped['Athlete_Country'] = Athlete_gold_grouped['Athlete'] + ', ' + Athlete_gold_grouped['Country']

# Sort by the number of gold medals in descending order
Athlete_gold_top10 = Athlete_gold_grouped.sort_values(by='Medal', ascending=False).head(10)

# Plot the bar graph using hvplot and directly chain the opts
Athlete_gold_top10.hvplot.bar(x='Athlete_Country', y='Medal', rot=45, xlabel='Athlete, Country', ylabel='Number of Gold Medals', title='To
```


Top 10 Athletes with Gold Medals in summer

What the data shows:
- ➔ Within the data of the Top 10 Athletes with Gold Medals 6 out 10 are from the USA for Summer.
- ➔ This leaves Finland, Japan, Hungary, and Italy as the last four.
- ➔ After Michael Phelps all the other Athletes have very similar Medal numbers.
- ➔ USA, Hungary, and Italy are all also on the Top 10 Countries with Gold Medals List.
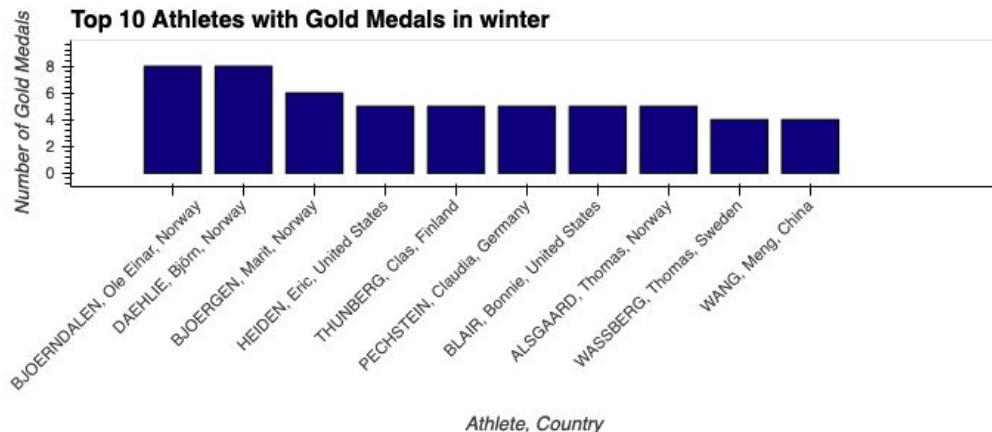
Top 10 Athletes with Gold Medals: Winter

```python
# Top 10 Athletes with gold wins
Athlete_gold_df = final_df[(final_df['Medal'] == 'Gold')& (final_df['Seasons'] == 'winter')]

# Group by athlete and country, sum the gold medals
Athlete_gold_grouped = Athlete_gold_df.groupby(['Athlete', 'Country'])['Medal'].count().reset_index()

# Combine 'Athlete' and 'Country' into a new column
Athlete_gold_grouped['Athlete_Country'] = Athlete_gold_grouped['Athlete'] + ', ' + Athlete_gold_grouped['Country']

# Sort by the number of gold medals in descending order
Athlete_gold_top10 = Athlete_gold_grouped.sort_values(by='Medal', ascending=False).head(10)

# Plot the bar graph using hvplot and directly chain the opts
Athlete_gold_top10.hvplot.bar(x='Athlete_Country', y='Medal', rot=45, xlabel='Athlete, Country', ylabel='Number of Gold Medals', title='Top 1
```


Top 10 Athletes with Gold Medals in winter

What the data shows:
➔ Within the data of the Top 10 Athletes with Gold Medals 4 out 10 are from the Norway for winter, which is the majority Country.
➔ This leaves US with 2, Finland with 1, Germany with 1, Sweden with 1, and China with 1.
➔ USA, Germany, Finland, Norway, and Sweden are all also on the Top 10 Countries with Gold Medals List.

Summary and Next Steps:

- The recent boom in the number of athletes winning medals can be explained by the addition 8 new sports and 119 new events following the year 1990

- Athletics is the most globally popular sport, with the United States winning the most amount of medals

- Looking through the data, the athletes that were in the top 10 for gold medals for each season often are also from countries that are in the top 10 for gold medals. This happens for both the Summer and Winter data.

- Women have continued to trend to get more and more medals each year in sports as more olympic sports become available for women.

- Further EDA of our data could explore the relationship of countries winning gold medals, to their GDP, to see if there's a trend in more "developed" countries winning gold medals
    - This analysis could also identify which sports the countries with higher GDP tend to win

- Further analysis can also focus on the most recent 10-20 years GDP rates to accurately compare current data in relation to olympic sports.