

# Python Code Answers to Summer Olympics Data Analysis Questions

Dr. Roi Yozevitch

May 2024

## Introduction

Below are the Python (Pandas) code snippets designed to answer specific data analysis questions using the Summer Olympics dataset. These commands facilitate the exploration of various aspects of Olympic data through DataFrame manipulation.

## Questions and Answers

1. Which country won the largest number of Bronze medals ever?

```
df[df['Medal'] == 'Bronze'].groupby('Country').size()  
().idxmax()
```

2. How many athletes participated in the 2012 Olympics?

```
df[df['Year'] == 2012]['Athlete'].nunique()
```

3. List all the sports in which the United States has won at least one gold medal.

```
df[(df['Country'] == 'USA') & (df['Medal'] == 'Gold')  
]['Sport'].unique()
```

4. Which male athlete has won the most medals in a single edition of the Olympics?

```
df[df['Gender'] == 'Men'].groupby(['Year', 'Athlete']  
).size().idxmax()
```

5. How many different countries have participated in the Olympics since 2000?

```
df[df['Year'] >= 2000]['Country'].nunique()
```

6. In which year did Russia win the most total medals, and how many did they win?

```
df[df['Country'] == 'RUS'].groupby('Year').size().idxmax()
```

7. Identify the top 5 athletes with the most silver medals across all editions of the Olympics.

```
df[df['Medal'] == 'Silver'].groupby('Athlete').size().nlargest(5)
```

8. Which sport has the smallest number of different events?

```
df.groupby('Sport')['Event'].nunique().idxmin()
```

9. For each Olympics, how many female versus male athletes participated?

```
df.groupby(['Year', 'Gender']).size()
```

10. What is the total number of medals won by China in swimming?

```
df[(df['Country'] == 'CHN') & (df['Sport'] == 'Swimming')].shape[0]
```

11. Rank countries by the number of gold medals won in athletics historically.

```
df[(df['Sport'] == 'Athletics') & (df['Medal'] == 'Gold')].groupby('Country').size().sort_values(ascending=False)
```

12. Which discipline had athletes from the most number of different countries in the latest Olympics?

```
latest_year = df['Year'].max()
df[df['Year'] == latest_year].groupby('Discipline')['Country'].nunique().idxmax()
```

13. Find the athlete who improved the most in terms of medal color (e.g., from Bronze in one Olympics to Gold in the next) across successive Olympics.

```
# This is a complex query requiring custom logic to  
track medal progression
```

14. What is the average number of participating countries in each Olympic sport?

```
df.groupby('Sport')['Country'].nunique().mean()
```

15. Compare the total number of medals won by Germany before and after the reunification in 1990.

```
before_1990 = df[(df['Country'] == 'GER') & (df['  
Year'] < 1990)].shape[0]  
after_1990 = df[(df['Country'] == 'GER') & (df['Year'  
' ] >= 1990)].shape[0]
```