

# Tokyo 2021 Olympics Analysis

Dataset source: Kaggle.com

Nofar Azilik | Data Analyst

## Introduction:

The 2020 Summer Olympics officially was an international multi-sport event held from 23 July to 8 August 2021 in Tokyo, Japan, with some preliminary events that began on 21 July. It includes 11,483 athletes from 206 NOCs.

## Research Questions:

- Which countries had the greatest number of athletics in Tokyo 2020?
- Which countries had the greatest number of medals in Tokyo 2020?
- Is there a correlation between number of medals of team and team size?
- In which disciplines the number of women out of the total number of athletes was high?
- Is there any relation between average age of country's athletes and the success (number of total medals)?

## Findings:

### Research Question 1

Which countries had the greatest number of athletes in Tokyo 2020?

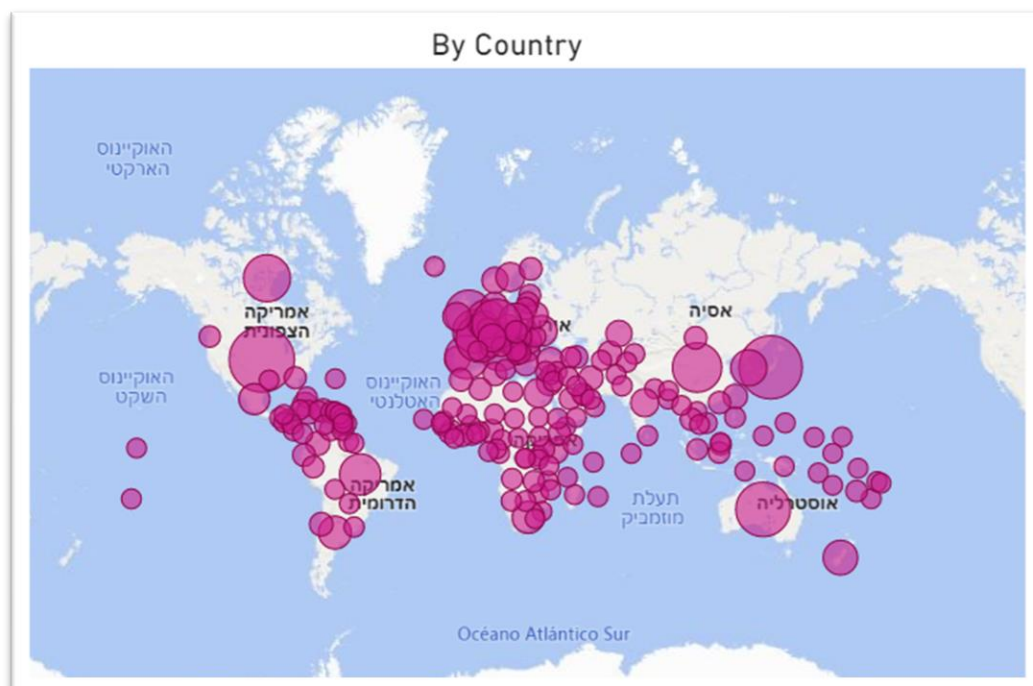
Input:

```
SELECT TOP 10 noc AS 'Country', COUNT(*) AS 'Athletes'  
FROM dbo.Athletes$  
GROUP BY noc  
ORDER BY COUNT(*) DESC
```

Output:

	Country	Athletes
1	United States of America	615
2	Japan	586
3	Australia	470
4	People's Republic of China	401
5	Germany	400
6	France	377
7	Canada	368
8	Great Britain	366
9	Italy	356
10	Spain	324

Figure 1:



## Research Question 2

Which countries (Top 10) had the greatest number of medals in Tokyo 2020?

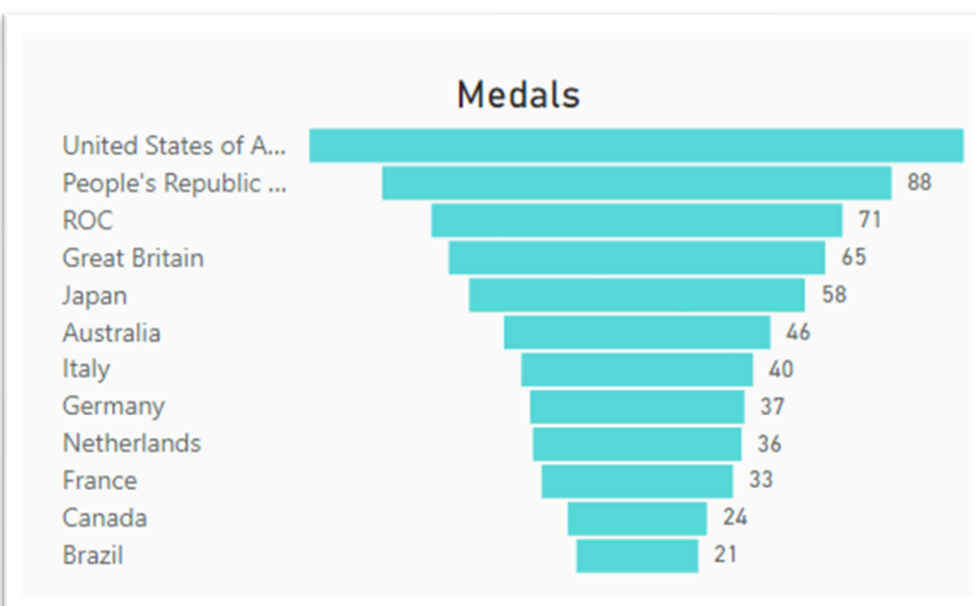
Input:

```
SELECT TOP 10 [Team/NOC], total FROM dbo.Medals$  
ORDER BY (CAST(total AS INTEGER)) DESC
```

Output:

	Team/NOC	total
1	United States of America	113
2	People's Republic of China	88
3	ROC	71
4	Great Britain	65
5	Japan	58
6	Australia	46
7	Italy	40
8	Germany	37
9	Netherlands	36
10	France	33

Figuer 2:



### Research Question 3

Is there a correlation between number of medals of team and team size?

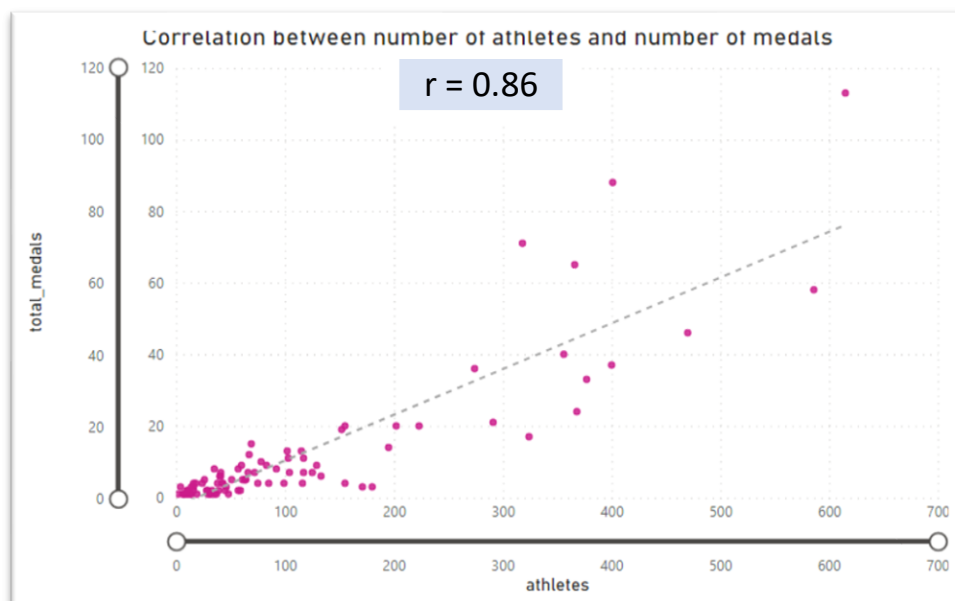
input:

```
SELECT COUNT(dbo.Athletes$.noc) AS 'athletes',  
       dbo.Athletes$.noc AS 'team',  
       dbo.Medals$.Total AS 'total_medals'  
FROM   dbo.Athletes$ INNER JOIN dbo.Medals$  
ON     dbo.Athletes$.noc = dbo.Medals$.Team  
GROUP BY dbo.Athletes$.noc, dbo.Medals$.Total  
ORDER BY CAST(dbo.Medals$.Total AS INTEGER) DESC
```

output:

	athletes	team	total_medals
1	615	United States of America	113
2	401	People's Republic of China	88
3	318	ROC	71
4	366	Great Britain	65
5	586	Japan	58
6	470	Australia	46
7	356	Italy	40
8	400	Germany	37
9	274	Netherlands	36
10	377	France	33
11	368	Canada	24

figure 3:



This result shows that there is a strong positive correlation between average age of team and the number of medals (Pearson  $r = 0.86$ ).

#### Research Question 4

In which disciplines the number of women out of the total number of athletes was high?

input:

```
SELECT * FROM dbo.Gender$  
ORDER BY (CAST(Female AS INTEGER) / CAST(Total AS INTEGER)) DESC
```

output:

	Discipline	Female	Male	Total
1	Artistic Swimming	105	0	105
2	Rhythmic Gymnastics	96	0	96
3	Rowing	257	265	522
4	Rugby Sevens	146	151	297
5	Sailing	175	175	350
6	Shooting	178	178	356
7	Skateboarding	40	40	80
8	Sport Climbing	20	20	40
9	Surfing	20	20	40
10	Swimming	361	418	779
11	Table Tennis	86	86	172
12	Taekwondo	65	65	130
13	Tennis	94	97	191
14	Trampoline Gymnastics	16	16	32
15	Triathlon	55	55	110
16	Volleyball	144	144	288

Figure 4:

By Gender	
Discipline	percents of females ^
Artistic Swimming	100.00%
Rhythmic Gymnastics	100.00%
Cycling BMX Freestyle	52.63%
Diving	50.35%
3x3 Basketball	50.00%
Archery	50.00%
Artistic Gymnastics	50.00%
Basketball	50.00%
Beach Volleyball	50.00%
Canoe Slalom	50.00%
Cycling BMX Racing	50.00%
Cycling Mountain Bike	50.00%
Golf	50.00%
Handball	50.00%

### Research Question 5

Is there any relation between average age of team and success (measured by number of total medals)?

The database include a partial data of athletes ages. Therefore, this result presents a representative sample of 472 Athletes in Tokyo Olympics 2021.

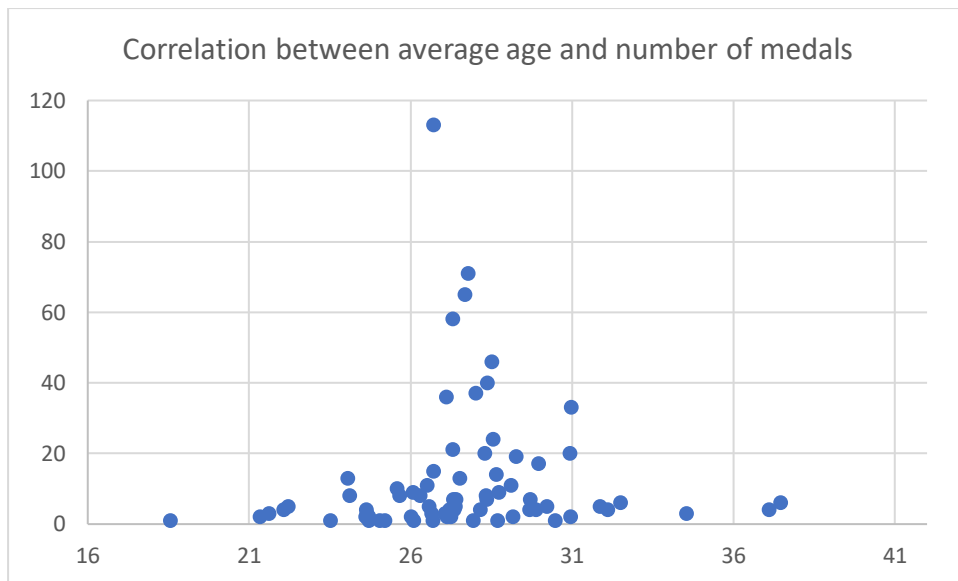
Input:

```
SELECT
  dbo.Athletes$.NOC AS 'Country',
  AVG(FLOOR((CAST (GetDate()) AS INTEGER) - CAST(Date_of_birth AS INTEGER))) / 365.25 AS 'Average_Age',
  AVG(CAST(dbo.Medals$.Total AS INTEGER)) AS 'Total medals'
FROM dbo.Athletes$ INNER JOIN dbo.Athletes_addition$
ON SUBSTRING(dbo.Athletes$.full_name, 2, 8) = SUBSTRING(dbo.Athletes_addition$.full_name, 2, 8)
INNER JOIN dbo.Medals$
ON dbo.Medals$.[Team/NOC] = dbo.Athletes$.NOC
GROUP BY dbo.Athletes$.NOC
ORDER BY AVG(CAST(dbo.Medals$.Total AS INTEGER)) DESC
```

Output:

	Country	Average_Age	Total medals
1	United States of America	26.711196	113
2	ROC	27.785352	71
3	Great Britain	27.681268	65
4	Japan	27.316563	58
5	Australia	28.513346	46
6	Italy	28.388774	40
7	Germany	28.022546	37
8	Netherlands	27.106502	36
9	France	30.985625	33
10	Canada	28.562236	24
11	Brazil	27.314355	21
12	Hungary	30.929499	20
13	New Zealand	28.305543	20
14	Ukraine	29.281085	19
15	Spain	29.957562	17
16	Cuba	26.718685	15
17	Poland	28.654756	14
18	Switzerland	24.053843	13

Figuer 4:



It was found that there is a no relationship between average age of team and the number of medals.