

```
In [2]: pip install import-ipynb

Requirement already satisfied: import-ipynb in /opt/anaconda3/lib/python3.8/site-packages (0.1.4)
Requirement already satisfied: IPython in /opt/anaconda3/lib/python3.8/site-packages (from import-ipynb) (7.22.0)
Requirement already satisfied: nbformat in /opt/anaconda3/lib/python3.8/site-packages (from import-ipynb) (5.1.3)
Requirement already satisfied: backcall in /opt/anaconda3/lib/python3.8/site-packages (from IPython->import-ipynb) (0.2.0)
Requirement already satisfied: traitlets>=4.2 in /opt/anaconda3/lib/python3.8/site-packages (from IPython->import-ipynb) (5.0.5)
Requirement already satisfied: appnope in /opt/anaconda3/lib/python3.8/site-packages (from IPython->import-ipynb) (0.1.2)
Requirement already satisfied: pickleshare in /opt/anaconda3/lib/python3.8/site-packages (from IPython->import-ipynb) (0.7.5)
Requirement already satisfied: jedi>=0.16 in /opt/anaconda3/lib/python3.8/site-packages (from IPython->import-ipynb) (0.17.2)
Requirement already satisfied: prompt-toolkit!=3.0.0,!3.0.1,<3.1.0,>=2.0.0 in /opt/anaconda3/lib/python3.8/site-packages (from IPython->import-ipynb) (3.0.17)
Requirement already satisfied: setuptools>=18.5 in /opt/anaconda3/lib/python3.8/site-packages (from IPython->import-ipynb) (65.6.3)
Requirement already satisfied: pexpect>4.3 in /opt/anaconda3/lib/python3.8/site-packages (from IPython->import-ipynb) (4.8.0)
Requirement already satisfied: pygments in /opt/anaconda3/lib/python3.8/site-packages (from IPython->import-ipynb) (2.8.1)
Requirement already satisfied: decorator in /opt/anaconda3/lib/python3.8/site-packages (from IPython->import-ipynb) (5.0.6)
Requirement already satisfied: jupyter-core in /opt/anaconda3/lib/python3.8/site-packages (from nbformat->import-ipynb) (4.7.1)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /opt/anaconda3/lib/python3.8/site-packages (from nbformat->import-ipynb) (3.2.0)
Requirement already satisfied: ipython-genutils in /opt/anaconda3/lib/python3.8/site-packages (from nbformat->import-ipynb) (0.2.0)
Requirement already satisfied: parso<0.8.0,>=0.7.0 in /opt/anaconda3/lib/python3.8/site-packages (from jedi=>0.16->IPython->import-ipynb) (0.7.0)
Requirement already satisfied: attrs>=17.4.0 in /opt/anaconda3/lib/python3.8/site-packages (from jsonschema!=2.5.0,>=2.4->nbformat->import-ipynb) (20.3.0)
Requirement already satisfied: six>=1.11.0 in /opt/anaconda3/lib/python3.8/site-packages (from jsonschema!=2.5.0,>=2.4->nbformat->import-ipynb) (1.16.0)
Requirement already satisfied: pyrsistent>=0.14.0 in /opt/anaconda3/lib/python3.8/site-packages (from jsonschema!=2.5.0,>=2.4->nbformat->import-ipynb) (0.17.3)
Requirement already satisfied: ptyprocess>=0.5 in /opt/anaconda3/lib/python3.8/site-packages (from pexpect>4.3->IPython->import-ipynb) (0.7.0)
Requirement already satisfied: wcwidth in /opt/anaconda3/lib/python3.8/site-packages (from prompt-toolkit!=3.0.0,!3.0.1,<3.1.0,>=2.0.0->IPython->import-ipynb) (0.2.5)

[notice] A new release of pip is available: 23.0 -> 23.1.2
[notice] To update, run: /opt/anaconda3/bin/python -m pip install --upgrade pip
Note: you may need to restart the kernel to use updated packages.
```

```
In [3]: pip install sqldf

Requirement already satisfied: sqldf in /opt/anaconda3/lib/python3.8/site-packages (0.4.2)

[notice] A new release of pip is available: 23.0 -> 23.1.2
[notice] To update, run: /opt/anaconda3/bin/python -m pip install --upgrade pip
Note: you may need to restart the kernel to use updated packages.
```

```
In [12]: import pandas as pd
import matplotlib.pyplot as plt
import sqldf
import import_ipynb
import numpy as np
```

```
In [5]: athlete_df = pd.read_csv("athlete_events.csv")
noc_df = pd.read_csv("noc_regions.csv")
```

```
In [7]: summer_events = sqldf.run('''
SELECT
    Year,
    COUNT(*) AS total_count,
    SUM(CASE
        WHEN Medal IS NOT NULL THEN 1 ELSE 0
    END) AS medal_count,
    SUM(CASE
        WHEN Medal = "Gold" THEN 1 ELSE 0
    END) AS gold_count,
    SUM(CASE
        WHEN Medal = "Silver" THEN 1 ELSE 0
    END) AS silver_count,
    SUM(CASE
        WHEN Medal = "Bronze" THEN 1 ELSE 0
    END) AS bronze_count
FROM
    summer_events
GROUP BY
    Year
''')

winter_events = sqldf.run('''
SELECT
    Year,
    COUNT(*) AS total_count,
    SUM(CASE
        WHEN Medal IS NOT NULL THEN 1 ELSE 0
    END) AS medal_count,
    SUM(CASE
        WHEN Medal = "Gold" THEN 1 ELSE 0
    END) AS gold_count,
    SUM(CASE
        WHEN Medal = "Silver" THEN 1 ELSE 0
    END) AS silver_count,
    SUM(CASE
        WHEN Medal = "Bronze" THEN 1 ELSE 0
    END) AS bronze_count
FROM
    winter_events
GROUP BY
    Year
''')
```

The length of the number of Winter Olympics and Summer Olympics are different as Winter Olympics started in 1924, while Summer Olympics started in 1896. Hence, I've cut the number of Summer Olympics to include up till 1924, so as to match the number of Winter Olympics.

```
In [8]: summer_events = summer_events[7:]
```

```
In [9]: summer_events
```

	Year	total_count	medal_count	gold_count	silver_count	bronze_count
7	1924	5233	832	277	281	274
8	1928	4992	734	245	239	250
9	1932	2969	647	229	214	204
10	1936	6506	917	312	310	295
11	1948	6405	852	289	284	279
12	1952	8270	897	306	291	300
13	1956	5127	893	302	293	298
14	1960	8119	911	309	294	308
15	1964	7702	1029	347	339	343
16	1968	8588	1057	359	340	358
17	1972	10304	1215	404	392	419
18	1976	8641	1320	438	434	448
19	1980	7191	1384	457	458	469
20	1984	9454	1476	497	477	502
21	1988	12037	1582	520	513	549
22	1992	12977	1712	559	549	604
23	1996	13780	1842	608	605	629
24	2000	13821	2004	663	661	680
25	2004	13443	2001	664	660	677
26	2008	13602	2048	671	667	710
27	2012	12920	1941	632	630	679
28	2016	13688	2023	665	655	703

```
In [10]: winter_events
```

	Year	total_count	medal_count	gold_count	silver_count	bronze_count
0	1924	460	130	55	38	37
1	1928	582	89	30	28	31
2	1932	352	92	32	32	28
3	1936	895	108	36	37	35
4	1948	1075	135	41	48	46
5	1952	1088	136	45	44	47
6	1956	1307	150	51	49	50
7	1960	1116	147	50	48	49
8	1964	1778	186	61	67	58
9	1968	1891	199	66	70	63
10	1972	1655	199	70	63	66
11	1976	1861	211	70	71	70
12	1980	1746	218	72	73	73
13	1984	2134	222	74	74	74
14	1988	2639	263	87	88	88
15	1992	3436	318	104	108	106
16	1994	3160	331	110	109	112
17	1998	3605	440	145	145	150
18	2002	4109	478	162	157	159
19	2006	4382	526	176	175	175
20	2010	4402	520	174	175	171
21	2014	4891	597	202	197	198

```
In [15]: x_level = winter_events.medal_count
y_level = summer_events.medal_count
correlation = np.corrcoef(x_level, y_level)

print(correlation)

[[1.          0.94141801]
 [0.94141801  1.        ]]
```

In order to find the correlation between the 2 events, i have used the Pearson correlation coefficient between the total number of mdeals won between the 2 events. Since it is 0.94, this shows that the performance of a country in the Winter Olympics is highly correlated to that in the Summer Olympics.

```
In [16]: std_medal_winter = np.std(x_level)
std_medal_summer = np.std(y_level)

print(std_medal_winter)
print(std_medal_summer)

152.56899942903493
475.323015441357
```

Next, I have calculated the standard deviation in country performance for each event across the years. Since the standard deviation in the Summer Olympics is around 3 times that of the Winter Olympics, the country performance by year changed more in the Summer Olympics.