Does hosting the Olympics improve performance on the medals table?

Countries who hosted atleast 2 times

In [2]:
```
host_data = []

for city in olympics['City'].unique():
    host_count = 0
    for year in olympics['Year'].unique():
        if len(olympics[(olympics['City'] == city) & (olympics['Year'] == y
            host_count += 1
    host_data.append([city, host_count])

host_data = pd.DataFrame(host_data,columns = ['City','No. of times hosted']

host_data.reset_index(inplace = True, drop = True)

host_data.head(10)
```

Out[2]:

	City	No. of times hosted
0	London	3
1	Athina	3
2	Innsbruck	2
3	Sankt Moritz	2
4	Paris	2
5	Los Angeles	2
6	Lake Placid	2
7	Stockholm	2
8	Amsterdam	1
9	Berlin	1

So ["UK", "Greece", "Austria", "Switzerland", "France", "USA", "Sweden"], these are the countries associated with the cities which have hosted the olympics 2 or more times. Let's map these countries with the cities.

In [3]:
```
countries_cities_map = {"USA": ["Lake Placid", "Los Angeles"], "UK": ["Lond

countries = ["USA", "UK", "Greece", "Austria", "Switzerland", "France", "Sw
no_of_times = [4, 3, 3, 2, 2, 2, 2]

countries_hosted = pd.DataFrame(data=zip(countries,no_of_times), columns =

countries_hosted
```

Out[3]:

	Country	Hosted
0	USA	4
1	UK	3
2	Greece	3
3	Austria	2
4	Switzerland	2
5	France	2
6	Sweden	2

Total medal tally of these countries

In [6]:
```
olympics.rename(columns={'region':'Country'}, inplace=True)

selected_countries_medals = olympics.loc[(olympics["Country"].isin(countrie

selected_countries_medals.head()
```

Out[6]:

	ID	Name	Sex	Age	Team	NOC	Games	Year	Season	City	Sport	
150	56	Ren Abadie	M	21.0	France	FRA	1956 Summer	1956	Summer	Melbourne	Cycling	Cy
173	73	Luc Abalo	M	23.0	France	FRA	2008 Summer	2008	Summer	Beijing	Handball	
174	73	Luc Abalo	M	27.0	France	FRA	2012 Summer	2012	Summer	London	Handball	
175	73	Luc Abalo	M	31.0	France	FRA	2016 Summer	2016	Summer	Rio de Janeiro	Handball	
186	84	Stephen Anthony Abas	M	26.0	United States	USA	2004 Summer	2004	Summer	Athina	Wrestling	Fea

Let's analyse the medal tally oof USA

In [7]:
```
usa_hosted_medals = selected_countries_medals[selected_countries_medals["Ci
usa_hosted_medals = usa_hosted_medals[usa_hosted_medals["Country"] == "USA"
```

In [8]:
```
usa_hosted = usa_hosted_medals.groupby("Year")["Medal"].count()
```

In [9]:
```
usa_hosted
```

Out[9]:
```
Year
1932    223
1980     30
1984    352
Name: Medal, dtype: int64
```

In [10]:
```
usa_not_hosted_medals = selected_countries_medals[~selected_countries_medal
usa_not_hosted_medals = usa_not_hosted_medals[usa_not_hosted_medals["Countr
```

In [11]:
```
usa_not_hosted = usa_not_hosted_medals.groupby("Year")["Medal"].count()
usa_not_hosted.sort_values()
```

Out[11]:
```
Year
1984      9
1994     19
1896     20
1906     24
1998     34
2006     52
1900     63
2014     64
1908     65
2002     84
2010     97
1928    102
1912    107
1936    112
1956    149
1960    152
1952    164
1948    168
1968    173
1976    175
1964    177
1924    194
1920    194
1972    195
1988    214
1992    238
2000    242
2012    248
1996    259
2004    263
2016    264
2008    317
1904    394
Name: Medal, dtype: int64
```

Let's anaylse the medal tally of all the countries together

Medal tally of these countries when hosting and non hosting

In [12]:
```
hosted_medals = pd.DataFrame()
non_hosted_medals = pd.DataFrame()

for countries_cities in countries_cities_map:
  selected_medals = selected_countries_medals[selected_countries_medals["Ci
  hosted_medals = pd.concat([hosted_medals, selected_medals[selected_medals

  non_selected_medals = selected_countries_medals[~selected_countries_medal
  non_hosted_medals = pd.concat([non_hosted_medals, non_selected_medals[non
```

In [13]:
```
host_medals = hosted_medals.groupby(["Country", "Year"])["Medal"].count().v
```

In [14]:
```
host_medals
```

Out[14]:
```
array([ 17,   7, 235, 110,  48, 102,  31, 190,   5,  12,  28, 368,  61,
       126, 223,  30, 352], dtype=int64)
```

In [15]:
```
non_host_medals = non_hosted_medals.groupby(["Country", "Year"])["Medal"].c
```

In [16]:
```
len(host_medals)
```

Out[16]:
```
17
```

Statistical Testing

Null Hypothesis(H0) : There is no effect of hosting to medal tally.

Alternative Hypothesis(H1) : There is an effect of hosting to medal tally.

In [17]:
```
from scipy import stats

ttest,pval = stats.ttest_rel(host_medals, non_host_medals[:17])
print(pval)
if pval<0.05:
    print("reject null hypothesis")
else:
    print("accept null hypothesis")
```

```
0.0032795760579561457
reject null hypothesis
```

So, after doing statistical testing it appears that hosting olympics increases the chances of winning the medals.

Based on the tests above, it appears that we can reject the null hypothesis that hosting has no impact on performance.

In [ ]:
```
```