In [2]:

```
import numpy as np
import pandas as pd
```

In [98]:

```
df = pd.read_csv("C:/harinumpy/Dataset/summer.csv")
df.head()
```

Out[98]:

	Year	City	Sport	Discipline	Athlete	Country	Gender	Event	Medal
0	1896	Athens	Aquatics	Swimming	HAJOS, Alfred	HUN	Men	100M Freestyle	Gold
1	1896	Athens	Aquatics	Swimming	HERSCHMANN, Otto	AUT	Men	100M Freestyle	Silver
2	1896	Athens	Aquatics	Swimming	DRIVAS, Dimitrios	GRE	Men	100M Freestyle For Sailors	Bronze
3	1896	Athens	Aquatics	Swimming	MALOKINIS, Ioannis	GRE	Men	100M Freestyle For Sailors	Gold
4	1896	Athens	Aquatics	Swimming	CHASAPIS, Spiridon	GRE	Men	100M Freestyle For Sailors	Silver

In [91]:

```
df.shape
```

Out[91]:

```
(31165, 9)
```

In [92]:

```
df.columns
```

Out[92]:

```
Index(['Year', 'City', 'Sport', 'Discipline', 'Athlete', 'Country', 'Gender',
       'Event', 'Medal'],
      dtype='object')
```

In [93]:

```
df.isnull().sum()
```

Out[93]:

```
Year          0
City          0
Sport         0
Discipline    0
Athlete       0
Country       4
Gender        0
Event         0
Medal         0
dtype: int64
```

In [95]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 31165 entries, 0 to 31164
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   Year        31165 non-null  int64 
 1   City        31165 non-null  object
 2   Sport       31165 non-null  object
```

```
 3   Discipline  31165 non-null  object
 4   Athlete     31165 non-null  object
 5   Country     31161 non-null  object
 6   Gender      31165 non-null  object
 7   Event       31165 non-null  object
 8   Medal       31165 non-null  object
dtypes: int64(1), object(8)
memory usage: 2.1+ MB
```

In [96]:

```
df.describe()
```

Out[96]:

	Year
count	31165.000000
mean	1970.482785
std	33.158416
min	1896.000000
25%	1948.000000
50%	1980.000000
75%	2000.000000
max	2012.000000

In how many cities Summer Olympics is held so far?

In [10]:

```
#In how many cities Summer Olympics is held so far?

Cities =  df["City"].unique()
Cities
# total cities
```

Out[10]:

```
array(['Athens', 'Paris', 'St Louis', 'London', 'Stockholm', 'Antwerp',
       'Amsterdam', 'Los Angeles', 'Berlin', 'Helsinki',
       'Melbourne / Stockholm', 'Rome', 'Tokyo', 'Mexico', 'Munich',
       'Montreal', 'Moscow', 'Seoul', 'Barcelona', 'Atlanta', 'Sydney',
       'Beijing'], dtype=object)
```

In [12]:

```
len(Cities)
#total number of cities
```

Out[12]:

```
22
```

Which sport is having most number of Gold Medals so far?

In [111]:

```
Max_Gold_medals = df[ (df["Medal"]=="Gold")].value_counts(subset = df["Sport"]).head()
Max_Gold_medals
```

Out[111]:

```
Sport
Aquatics      1421
Athletics     1215
Rowing         890
```

```
Gymnastics      820
Fencing         552
dtype: int64
```

Which sport is having most number of medals so far?

In [116]:

```
most_medals = []

for m in df['Sport'].unique():
    most_medals.append([m, len(df[df['Sport'] == m])])

most_medals = pd.DataFrame(most_medals, columns = ['Sport', 'Total Medals'])
most_medals = most_medals.sort_values(by = 'Total Medals', ascending = False).head()

most_medals
```

Out[116]:

	Sport	Total Medals
0	Aquatics	4170
1	Athletics	3638
17	Rowing	2667
4	Gymnastics	2307
3	Fencing	1613

Which player has won most number of medals?

In [120]:

```
most_medals = df[ (df["Medal"]==df["Medal"])].value_counts(subset = df["Athlete"]).head(
)
most_medals
```

Out[120]:

```
Athlete
PHELPS, Michael          22
LATYNINA, Larisa         18
ANDRIANOV, Nikolay       15
MANGIAROTTI, Edoardo     13
ONO, Takashi             13
dtype: int64
```

In which year India won first Gold Medal in Summer Olympics?

In [144]:

```
X = df[df["Medal"]=="Gold"]
Y = X.loc[X["Country"]=="IND"]

Y.iloc[0]
```

Out[144]:

```
Year                          1928
City                     Amsterdam
Sport                       Hockey
Discipline                  Hockey
Athlete       ALLEN, Richard James
Country                        IND
Gender                         Men
Event                       Hockey
Medal                         Gold
```

```
Name: 5512, dtype: object
```

Which event is most popular in terms on number of players?

In [147]:

```
E = []
for i in df['Event'].unique():
    E.append([i, len(df[df['Event'] == i])])
E = pd.DataFrame(E, columns = ['Event', 'Total Players'])
E = E.sort_values(by = 'Total Players', ascending = False).head()
E
```

Out[147]:

	Event	Total Players
80	Football	1497
176	Hockey	1422
138	Team Competition	1147
327	Basketball	1012
337	Handball	973

In [148]:

```
df
```

Out[148]:

	Year	City	Sport	Discipline	Athlete	Country	Gender	Event	Medal
0	1896	Athens	Aquatics	Swimming	HAJOS, Alfred	HUN	Men	100M Freestyle	Gold
1	1896	Athens	Aquatics	Swimming	HERSCHMANN, Otto	AUT	Men	100M Freestyle	Silver
2	1896	Athens	Aquatics	Swimming	DRIVAS, Dimitrios	GRE	Men	100M Freestyle For Sailors	Bronze
3	1896	Athens	Aquatics	Swimming	MALOKINIS, Ioannis	GRE	Men	100M Freestyle For Sailors	Gold
4	1896	Athens	Aquatics	Swimming	CHASAPIS, Spiridon	GRE	Men	100M Freestyle For Sailors	Silver
...	...	...	...	...	...	...	...	...	...
31160	2012	London	Wrestling	Wrestling Freestyle	JANIKOWSKI, Damian	POL	Men	Wg 84 KG	Bronze
31161	2012	London	Wrestling	Wrestling Freestyle	REZAEI, Ghasem Gholamreza	IRI	Men	Wg 96 KG	Gold
31162	2012	London	Wrestling	Wrestling Freestyle	TOTROV, Rustam	RUS	Men	Wg 96 KG	Silver
31163	2012	London	Wrestling	Wrestling Freestyle	ALEKSANYAN, Artur	ARM	Men	Wg 96 KG	Bronze
31164	2012	London	Wrestling	Wrestling Freestyle	LIDBERG, Jimmy	SWE	Men	Wg 96 KG	Bronze

31165 rows × 9 columns

In [207]:

```
x = df[df['Medal'] == 'Gold']
f = x[x['Gender'] == 'Women']
wgold = []
for i in f['Sport'].unique():
    wgold.append([i, len(f[f['Sport'] == i])])
```

```
wgold = pd.DataFrame(wgold, columns = ['Sport', 'Female Gold Medalists'])
wgold = wgold.sort_values(by = 'Female Gold Medalists', ascending = False).head()
wgold
```

Out[207]:

	Sport	Female Gold Medalists
4	Aquatics	589
7	Athletics	389
8	Gymnastics	268
14	Rowing	217
11	Volleyball	166

In [ ]:

```
```