

1. Import the necessary libraries

```
In [5]: import pandas as pd
import numpy as np
```

2. Import the dataset from this address.

```
In [6]: olympic_data=pd.read_csv("https://raw.githubusercontent.com/svkarthik86/Advance-Python-Numpy/main/olympics_dirty.csv", skiprows=4)
```

3. Assign it to a variable called odata.

```
In [7]: odata=olympic_data
```

4. show the first 5 entries

```
In [8]: odata.head(5)
Out[8]:
```

	Available_City	Edition_year	Sport_name	Discipline	Athlete	NOC	Gender	Event	Event_gender	Medal
0	Athens	1896	Aquatics	Swimming	HAJOS, Alfred	HUN	Men	100m freestyle	M	Gold
1	Athens	1896	Aquatics	Swimming	HERSCHMANN, Otto	AUT	Men	100m freestyle	M	Silver
2	Athens	1896	Aquatics	Swimming	DRIVAS, Dimitrios	GRE	Men	100m freestyle for sailors	M	Bronze
3	Athens	1896	Aquatics	Swimming	MALOKINIS, Ioannis	GRE	Men	100m freestyle for sailors	M	NaN
4	Athens	1896	Aquatics	Swimming	CHASAPIS, Spiridon	NaN	Men	100m freestyle for sailors	M	Silver

5. What is the number of observations in the dataset?

```
In [9]: len(odata)
Out[9]: 29216
```

6. What is the number of columns in the dataset?

```
In [10]: odata.shape[1]
Out[10]: 10
```

7. Print the name of all the columns.

```
In [12]: for i in odata.columns:
print(i,end="\t")
Available_City  Edition_year  Sport_name  Discipline  Athlete NOC  Gender  Event  Event_gender  Medal
```

8. How is the dataset indexed?

```
In [13]: odata.index
Out[13]: RangeIndex(start=0, stop=29216, step=1)
```

9. Rename the column as Available\_City to City, Edition\_year to Edition , Sport\_name to Sport

```
In [14]: odata.rename(columns={"Available_City":"City","Edition_year":"Edition","Sport_name":"Sport"},inplace=True)
```

10. How many Athletes own the Gold medal?

```
In [15]: len(odata.query("Medal=='Gold'").Athlete.drop_duplicates())
Out[15]: 7841
```

11. Display the male gold medal winners for the 100m Track?

```
In [16]: odata.query("(Medal=='Gold')and(Event=='100m')and(Gender=='Men')").Athlete.value_counts()
Out[16]:
```

LEWIS, Carl	2
BURKE, Thomas	1
HARY, Armin	1
GATLIN, Justin	1
GREENE, Maurice	1
BAILEY, Donovan	1
CHRISTIE, Linford	1
WELLS, Allan	1
CRAWFORD, Hasely	1
BORZOV, Valery	1
HINES, James Ray	1
HAYES, Robert	1
MORROW, Robert Joseph	1
JARVIS, Francis	1
REMINGNO, Lindy	1
DILLARD, Harrison	1
OWENS, Jesse	1
TOLAN, Eddie	1
WILLIAMS, Percy	1
ABRAHAMMS, Harold	1
PADDOCK, Charles	1
CRAIG, Ralph	1
WALKER, Reginald	1
HAHN, Archie	1
BOLT, Usain	1

Name: Athlete, dtype: int64

12. Which three countries have won the most medals in recent years (from 1984 to 2008)?

```
In [17]: odata[odata.Edition>=1984].NOC.value_counts()[:3]
Out[17]:
```

USA	1837
AUS	762
GER	691

Name: NOC, dtype: int64

13. Which country has won the most men's gold medals in singles badminton over the years? Sort the results alphabetically by the player's names.

```
In [18]: print(odata.query("(Medal=='Gold')and(Gender=='Men')and(Event=='singles')and(Discipline=='Badminton')").NOC.value_counts()[:2])
odata.query("(Medal=='Gold')and(Gender=='Men')and(Event=='singles')and(Discipline=='Badminton')").sort_values("Athlete")
INA 2
CHN 2
Name: NOC, dtype: int64
Out[18]:
```

	City	Edition	Sport	Discipline	Athlete	NOC	Gender	Event	Event_gender	Medal
20045	Barcelona	1992	Badminton	Badminton	BUDI KUSUMA, Alan	INA	Men	singles	M	Gold
25734	Athens	2004	Badminton	Badminton	HIDAYAT, Taufik	INA	Men	singles	M	Gold
21787	Atlanta	1996	Badminton	Badminton	HOYER-LARSEN, Poul Erik	DEN	Men	singles	M	Gold
23717	Sydney	2000	Badminton	Badminton	Ji, Xinpeng	CHN	Men	singles	M	Gold
27741	Beijing	2008	Badminton	Badminton	LIN, Dan	CHN	Men	singles	M	Gold

14. Display the women gold medal winners for the 100m Track?

```
In [19]: odata.query("(Medal=='Gold')and(Event=='100m')and(Gender=='Women')").Athlete.value_counts()
Out[19]:
```

TYUS, Wyomia	2
DEVERS, Gail	2
ROBINSON, Elizabeth	1
WALASIEWICZ, Stanislaw	1
STEPHENS, Helen Herring	1
BLANKERS-KOEN, Fanny	1
JACKSON, Marjorie	1
CUTHBERT, Betty	1
RUDOLPH, Wilma	1
STECHEr, Renate	1
RICHTER-IRRGANG, Annegret	1
KONDRATIEVA, Liudmila	1
ASHFORD, Evelyn	1
GRIFFITH-JOYNER, Florence	1
NESTERENKO, Yuliya	1
FRASER, Shelly-ann	1

Name: Athlete, dtype: int64

15. list out the Sliver medal winner @ year 2008?

```
In [20]: odata.query("(Edition==2008)and(Medal=='Silver')")
Out[20]:
```

	City	Edition	Sport	Discipline	Athlete	NOC	Gender	Event	Event_gender	Medal
27176	Beijing	2008	Aquatics	Diving	ZHOu, Luxin	CHN	Men	10m platform	M	Silver
27179	Beijing	2008	Aquatics	Diving	HEYMANS, Emilie	CAN	Women	10m platform	W	Silver
27182	Beijing	2008	Aquatics	Diving	DESPATIE, Alexandre	CAN	Men	3m springboard	M	Silver
27185	Beijing	2008	Aquatics	Diving	PAKHALINA, Julia	RUS	Women	3m springboard	W	Silver
27190	Beijing	2008	Aquatics	Diving	HAUSDING, Patrick	GER	Men	synchronized diving 10m platform	M	Silver
...	...	...	...	...	...	...	...	...	...	...
29200	Beijing	2008	Wrestling	Wrestling Gre-R	BEGALIEV, Kanatbek	KGZ	Men	60 - 66kg	M	Silver
29204	Beijing	2008	Wrestling	Wrestling Gre-R	CHANG, Yongxiang	CHN	Men	66 - 74kg	M	Silver
29207	Beijing	2008	Wrestling	Wrestling Gre-R	FODOR, Zoltan	HUN	Men	74 - 84kg	M	Silver
29211	Beijing	2008	Wrestling	Wrestling Gre-R	ENGLICH, Mirko	GER	Men	84 - 96kg	M	Silver
29215	Beijing	2008	Wrestling	Wrestling Gre-R	BAROEV, Khasan	RUS	Men	96 - 120kg	M	Silver

663 rows × 10 columns

16.Group by Countries and apply aggregate function min,max,count on over the year

```
In [21]: odata.groupby("NOC")["Edition"].agg([min, max, len])
Out[21]:
```

	min	max	len
NOC			
AFG	2008	2008	1
AHO	1988	1988	1
ALG	1984	2008	14
ANZ	1908	1912	29
ARG	1924	2008	239
...	...	...	...
VIE	2000	2008	2
YUG	1924	2000	435
ZAM	1984	1996	2
ZIM	1980	2008	23
ZZX	1896	1904	48

138 rows × 3 columns