



Data Analysis Problem on Data Wrangling and Exploring

1. Import the necessary libraries

```
In [1]: import pandas as pd
```

2. Import the dataset from this address.

```
In [2]: address=("https://raw.githubusercontent.com/svkarthik86/Advance-Python-Numpy/main/olympics_dirty.csv")
```

3. Assign it to a variable called odata

```
In [3]: odata=pd.read_csv(address,skiprows=4)
```

4. show the first 5 entries

```
In [4]: odata.head(5)
```

	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 [5]: len(odata)
```

```
Out[5]: 29216
```

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

```
In [6]: len(odata.columns)
```

```
Out[6]: 10
```

7. Print the name of all the columns.

```
In [7]: odata.columns
```

```
Out[7]: Index(['Available_City', 'Edition_year', 'Sport_name', 'Discipline', 'Athlete', 'NOC', 'Gender', 'Event', 'Event_gender', 'Medal'], dtype='object')
```

8. How is the dataset indexed?

```
In [9]: odata.index
```

```
Out[9]: 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 [10]: odata.rename(columns={'Available_City':'City','Edition_year':'Edition','Sport_name':'Sport'},inplace=True)
```

	City	Edition	Sport	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
...	...	...	...	...	...	...	...	...	...	...
29211	Beijing	2008	Wrestling	Wrestling Gre-R	ENGLISH, Mirko	GER	Men	84 - 96kg	M	Silver
29212	Beijing	2008	Wrestling	Wrestling Gre-R	MIZGAITIS, Mindaugas	LTU	Men	96 - 120kg	M	Bronze
29213	Beijing	2008	Wrestling	Wrestling Gre-R	PATRIKEEV, Yuri	ARM	Men	96 - 120kg	M	Bronze
29214	Beijing	2008	Wrestling	Wrestling Gre-R	LOPEZ, Mijain	CUB	Men	96 - 120kg	M	Gold
29215	Beijing	2008	Wrestling	Wrestling Gre-R	BAROEV, Khasan	RUS	Men	96 - 120kg	M	Silver

29216 rows × 10 columns

10. How many Athletes own the Gold medal?

```
In [11]: len(odata[odata.Medal=='Gold'])
```

```
Out[11]: 9849
```

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

```
In [12]: male_gold_medal_winners=odata[(odata.Gender=='Men')&(odata.Event=='100m')&(odata.Medal=='Gold')][['Athlete','Event','Medal']]
```

	Athlete	Event	Medal
13	BURKE, Thomas	100m	Gold
231	JARVIS, Francis	100m	Gold
737	HAHN, Archie	100m	Gold
1191	WALKER, Reginald	100m	Gold
2022	CRAIG, Ralph	100m	Gold
2996	PADDOCK, Charles	100m	Gold
4236	ABRAHAM, Harold	100m	Gold
5095	WILLIAMS, Percy	100m	Gold
5806	TOLAN, Eddie	100m	Gold
6427	OWENS, Jesse	100m	Gold
7302	DILLARD, Harrison	100m	Gold
8121	REMIGINO, Lindy	100m	Gold
9009	MORROW, Robert Joseph	100m	Gold
9924	HARY, Armin	100m	Gold
10823	HAYES, Robert	100m	Gold
11865	HINES, James Ray	100m	Gold
12902	BORZOV, Valery	100m	Gold
14069	CRAWFORD, Hasely	100m	Gold
15374	WELLS, Allan	100m	Gold
16794	LEWIS, Carl	100m	Gold
18284	LEWIS, Carl	100m	Gold
19859	CHRISTIE, Linford	100m	Gold
21598	BAILEY, Donovan	100m	Gold
23521	GREENE, Maurice	100m	Gold
25539	GATLIN, Justin	100m	Gold
27552	BOLT, Usain	100m	Gold

12.Which three countries have won the most medals in recent years

(from 1984 to 2008)?

```
In [13]: odata[(odata.Edition>=1984)&(odata.Edition<=2008)][['NOC']].value_counts().head(3)
```

```
Out[13]: 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 [14]: odata[(odata.Medal=='Gold')&(odata.Gender=='Men')&(odata.Event=='singles') & (odata.Discipline=='Badminton')].sort_values(['Athlete'])
```

	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 [16]: women_gold_medal_winners=odata[(odata.Gender=='Women')&(odata.Medal=='Gold')&(odata.Event=='100m')][['Athlete','Gender','Event','Medal']]
```

	Athlete	Gender	Event	Medal
5098	ROBINSON, Elizabeth	Women	100m	Gold
5809	WALASIEWICZ, Stanisława	Women	100m	Gold
6430	STEPHENS, Helen Herring	Women	100m	Gold
7305	BLANKERS-KOEN, Fanny	Women	100m	Gold
8124	JACKSON, Marjorie	Women	100m	Gold
9012	CUTHBERT, Betty	Women	100m	Gold
9927	RUDOLPH, Wilma	Women	100m	Gold
10826	TYUS, Wyomia	Women	100m	Gold
11868	TYUS, Wyomia	Women	100m	Gold
12905	STECHE, Renate	Women	100m	Gold
14072	RICHTER-IRRGANG, Annegret	Women	100m	Gold
15377	KONDRATIEVA, Liudmila	Women	100m	Gold
16797	ASHFORD, Evelyn	Women	100m	Gold
18287	GRIFFITH-JOYNER, Florence	Women	100m	Gold
19862	DEVERS, Gail	Women	100m	Gold
21601	DEVERS, Gail	Women	100m	Gold
25542	NESTERENKO, Yuliya	Women	100m	Gold
27554	FRASER, Shelly-ann	Women	100m	Gold

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

```
In [17]: silver_2008_winner=odata[(odata.Medal=='Silver')&(odata.Edition==2008)][['Athlete','Edition','Medal']]
```

	Athlete	Edition	Medal
27176	ZHOU, Luxin	2008	Silver
27179	HEYMANS, Emilie	2008	Silver
27182	DESPATIE, Alexandre	2008	Silver
27185	PAKHALINA, Julia	2008	Silver
27190	HAUSDING, Patrick	2008	Silver
...	...	...	...
29200	BEGALIEV, Kanatbek	2008	Silver
29204	CHANG, Yongxiang	2008	Silver
29207	FODOR, Zoltan	2008	Silver
29211	ENGLISH, Mirko	2008	Silver
29215	BAROEV, Khasan	2008	Silver

663 rows × 3 columns

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

```
In [18]: odata.groupby('NOC')['Edition'].agg([min,max,'count'])
```

	min	max	count
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