

Data Analysis Problem on Data Wrangling and Exploring

1. Import the necessary libraries

```
In [2]: import pandas as pd
```

2. Import the dataset from this address.

```
In [1]: import pandas as pd
olympic_bdata=pd.read_csv("https://raw.githubusercontent.com/svkarthik86/Advance-Python-Numpy/main/olympics_dirty.csv",skiprows=4)
olympic_bdata
```

```
Out[1]:
```

	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
...
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

3. Assign it to a variable called odata.

```
In [3]: odata=olympic_bdata
odata
```

```
Out[3]:
```

	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
...
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

4. show the first 5 entries

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

```
Out[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 [6]: odata.size
```

```
Out[6]: 292160
```

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

```
In [16]: odata.shape
```

```
Out[16]: (29216, 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 [18]: type(odata)
```

```
Out[18]: pandas.core.frame.DataFrame
```

```
In [19]: odata.index
```

```
Out[19]: 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 [39]: import pandas as pd
olympic_bdata=pd.read_csv("https://raw.githubusercontent.com/svkarthik86/Advance-Python-Numpy/main/olympics_dirty.csv",skiprows=4)
olympic_bdata
```

```
Out[39]:
```

	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
...
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

```
In [41]: odata=olympic_bdata
odata.rename(columns={' City':'Available_City',' Edition':'Edition_year',' Sport':'Sport_name'},inplace=True)
odata
```

```
Out[41]:
```

	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
...
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 [35]: odata[(odata.Medal=="Gold")]["Athlete"].value_counts()
```

```
Out[35]: PHELPS, Michael      14
LATYNINA, Larisa           9
SPITZ, Mark                 9
NURMI, Paavo                9
LEWIS, Carl                 9
...
LTSENKO-SHEVTSOVA, Lyudmila 1
THOMPSON, Donald James      1
HALBERG, Murray Gordon      1
YOUNG, Earl Verdelle        1
LOPEZ, Mijain               1
Name: Athlete, Length: 7840, dtype: int64
```

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

```
In [42]: od1=odata[(odata.Medal=="Gold")&(odata.Gender=="Men")&(odata.Event=="100m")&(odata.Discipline=="Athletics")]
od1
```

```
Out[42]:
```

	Available_City	Edition_year	Sport_name	Discipline	Athlete	NOC	Gender	Event	Event_gender	Medal
13	Athens	1896	Athletics	Athletics	BURKE, Thomas	USA	Men	100m	M	Gold
231	Paris	1900	Athletics	Athletics	JARVIS, Francis	USA	Men	100m	M	Gold
737	St Louis	1904	Athletics	Athletics	HAHN, Archie	USA	Men	100m	M	Gold
1191	London	1908	Athletics	Athletics	WALKER, Reginald	RSA	Men	100m	M	Gold
2022	Stockholm	1912	Athletics	Athletics	CRAIG, Ralph	USA	Men	100m	M	Gold
2996	Antwerp	1920	Athletics	Athletics	PADDOCK, Charles	USA	Men	100m	M	Gold
4236	Paris	1924	Athletics	Athletics	ABRAHAMS, Harold	GBR	Men	100m	M	Gold
5095	Amsterdam	1928	Athletics	Athletics	WILLIAMS, Percy	CAN	Men	100m	M	Gold
5806	Los Angeles	1932	Athletics	Athletics	TOLAN, Eddie	USA	Men	100m	M	Gold
6427	Berlin	1936	Athletics	Athletics	OWENS, Jesse	USA	Men	100m	M	Gold
7302	London	1948	Athletics	Athletics	DILLARD, Harrison	USA	Men	100m	M	Gold
8121	Helsinki	1952	Athletics	Athletics	REMIGINO, Lindy	USA	Men	100m	M	Gold
9009	Melbourne / Stockholm	1956	Athletics	Athletics	MORROW, Robert Joseph	USA	Men	100m	M	Gold
9924	Rome	1960	Athletics	Athletics	HARY, Armin	EUA	Men	100m	M	Gold
10823	Tokyo	1964	Athletics	Athletics	HAYES, Robert	USA	Men	100m	M	Gold
11865	Mexico	1968	Athletics	Athletics	HINES, James Ray	USA	Men	100m	M	Gold
12902	Munich	1972	Athletics	Athletics	BORZOV, Valery	URS	Men	100m	M	Gold
14069	Montreal	1976	Athletics	Athletics	CRAWFORD, Hasely	TRI	Men	100m	M	Gold
15374	Moscow	1980	Athletics	Athletics	WELLS, Allan	GBR	Men	100m	M	Gold
16794	Los Angeles	1984	Athletics	Athletics	LEWIS, Carl	USA	Men	100m	M	Gold
18284	Seoul	1988	Athletics	Athletics	LEWIS, Carl	USA	Men	100m	M	Gold
19859	Barcelona	1992	Athletics	Athletics	CHRISTIE, Linford	GBR	Men	100m	M	Gold
21598	Atlanta	1996	Athletics	Athletics	BAILEY, Donovan	CAN	Men	100m	M	Gold
23521	Sydney	2000	Athletics	Athletics	GREENE, Maurice	USA	Men	100m	M	Gold
25539	Athens	2004	Athletics	Athletics	GATLIN, Justin	USA	Men	100m	M	Gold
27552	Beijing	2008	Athletics	Athletics	BOLT, Usain	JAM	Men	100m	M	Gold

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

```
In [44]: odata[(odata.Edition_year>=1984)&(odata.Edition_year<=2008)][["NOC"]].value_counts().head(3)
```

```
Out[44]: 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 [45]: odata[(odata.Medal=="Gold")&(odata.Gender=="Men")&(odata.Event=="singles")&(odata.Discipline=="Badminton")].sort_values(['Athlete'])
```

```
Out[45]:
```

	Available_City	Edition_year	Sport_name	Discipline	Athlete	NOC	Gender	Event	Event_gender	Medal
20045	Barcelona	1992	Badminton	Badminton	BUDI KUSUMA, Alan	INA	Men	singles	M	Gold
27534	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 [46]: od1=odata[(odata.Medal=="Gold")&(odata.Gender=="Women")&(odata.Event=="100m")&(odata.Discipline=="Athletics")]
od1
```

```
Out[46]:
```

	Available_City	Edition_year	Sport_name	Discipline	Athlete	NOC	Gender	Event	Event_gender	Medal
5098	Amsterdam	1928	Athletics	Athletics	ROBINSON, Elizabeth	USA	Women	100m	W	Gold
5809	Los Angeles	1932	Athletics	Athletics	WALASIEWICZ, Stanislaw	POL	Women	100m	W	Gold
6430	Berlin	1936	Athletics	Athletics	STEPHENS, Helen Herring	USA	Women	100m	W	Gold
7305	London	1948	Athletics	Athletics	BLANKERS-KOEN, Fanny	NED	Women	100m	W	Gold
8124	Helsinki	1952	Athletics	Athletics	JACKSON, Marjorie	AUS	Women	100m	W	Gold
9012	Melbourne / Stockholm	1956	Athletics	Athletics	CUTHBERT, Betty	AUS	Women	100m	W	Gold
9927	Rome	1960	Athletics	Athletics	RUDOLPH, Wilma	USA	Women	100m	W	Gold
10826	Tokyo	1964	Athletics	Athletics	TYUS, Wyomia	USA	Women	100m	W	Gold
11868	Mexico	1968	Athletics	Athletics	TYUS, Wyomia	USA	Women	100m	W	Gold
12905	Munich	1972	Athletics	Athletics	STECHER, Renate	GDR	Women	100m	W	Gold
14072	Montreal	1976	Athletics	Athletics	RICHTER-IRRGANG, Annegret	FRG	Women	100m	W	Gold
15377	Moscow	1980	Athletics	Athletics	KONDRATIEVA, Liudmila	URS	Women	100m	W	Gold
16797	Los Angeles	1984	Athletics	Athletics	ASHFORD, Evelyn	USA	Women	100m	W	Gold
18287	Seoul	1988	Athletics	Athletics	GRIFFITH-JOYNER, Florence	USA	Women	100m	W	Gold
19862	Barcelona	1992	Athletics	Athletics	DEVERS, Gail	USA	Women	100m	W	Gold
21601	Atlanta	1996	Athletics	Athletics	DEVERS, Gail	USA	Women	100m	W	Gold
25542	Athens	2004	Athletics	Athletics	NESTERENKO, Yuliya	BLR	Women	100m	W	Gold
27554	Beijing	2008	Athletics	Athletics	FRASER, Shelly-ann	JAM	Women	100m	W	Gold

15.list out silver medal winners at 2008

```
In [48]: odata.dropna()
```

```
Out[48]:
```

	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
7	Athens	1896	Aquatics	Swimming	ANDREOU,					