In [86]:
```
import pandas as pd
df = pd.DataFrame({
    'A': [1, 2, 3],
    'B': [4, 5, 6],
    'C': [7, 8, 9]
})
df
```

Out[86]:

	A	B	C
0	1	4	7
1	2	5	8
2	3	6	9

In [87]:
```
# axis=0 → سطر | axis=1 → ستون
# inplace=True → تغییر مستقیم روی دیتافریم اصلی
# inplace=False → یه کپی جدید میسازه (پیش‌فرض)  باید بریزی تو یک متغییر دیگه تا ببینی

# df.drop('B', axis=1, inplace=True)
# df

# ************************************************************************
# del df['ستون']  # مثل - مستقیماً حذف میکنه inplace=True
# فقط برای حذف تک ستون کاربرد داره
# سینتکس ساده‌تر
# مستقیماً تغییر میده

# del df['B']
# df
```

In [88]:
```
# حذف چند ستون با drop
# df.drop(['A','B'], axis=1, inplace=True)
# df
# coffee.drop(columns=['pvpvp'],inplace=True)


# حذف چند سطر با drop
```

```
# df.drop([0, 1], axis=0, inplace=True)
# df
```

In [89]:
```
df['D'] = [10, 11, 12]
df
```

Out[89]:

	A	B	C	D
0	1	4	7	10
1	2	5	8	11
2	3	6	9	12

In [134…
```
import pandas as pd
coffee=pd.read_csv('https://raw.githubusercontent.com/KeithGalli/complete-pandas-tutorial/refs/heads/master/warmup-da
coffee
```

Out[134...

	Day	Coffee Type	Units Sold
0	Monday	Espresso	25
1	Monday	Latte	15
2	Tuesday	Espresso	30
3	Tuesday	Latte	20
4	Wednesday	Espresso	35
5	Wednesday	Latte	25
6	Thursday	Espresso	40
7	Thursday	Latte	30
8	Friday	Espresso	45
9	Friday	Latte	35
10	Saturday	Espresso	45
11	Saturday	Latte	35
12	Sunday	Espresso	45
13	Sunday	Latte	35

In [135...

```
# **** شرط در بررسی داده با  Where ****
# np.where(condition , if it's true , else)

# import numpy as np

# coffee_price = 2
# coffee["Price"] = np.where(
#     coffee["Coffee Type"] == "Espresso", coffee_price * 1.5, "not set"
# )
# coffee

# **** بررسی چند شرط با nested where ****
# import numpy as np
```

```
# coffee_price = 2
# coffee["Price"] = np.where(
#     coffee["Coffee Type"] == "Espresso",
#     coffee_price * 1.5,
#     np.where(coffee["Coffee Type"] == "Latte", coffee_price * 3, "not set"),
# )
# coffee

# **** بررسی چند شرط و مقداردهی با np.select ****
import numpy as np

coffee_price = 2
conditions = [coffee["Coffee Type"] == "Espresso", coffee["Coffee Type"] == "Latte"]

choices = [coffee_price * 1.5, coffee_price * 3]

coffee["Price"] = np.select(conditions, choices, default="not set")
coffee
```

Out[135…

	Day	Coffee Type	Units Sold	Price
0	Monday	Espresso	25	3.0
1	Monday	Latte	15	6
2	Tuesday	Espresso	30	3.0
3	Tuesday	Latte	20	6
4	Wednesday	Espresso	35	3.0
5	Wednesday	Latte	25	6
6	Thursday	Espresso	40	3.0
7	Thursday	Latte	30	6
8	Friday	Espresso	45	3.0
9	Friday	Latte	35	6
10	Saturday	Espresso	45	3.0
11	Saturday	Latte	35	6
12	Sunday	Espresso	45	3.0
13	Sunday	Latte	35	6

In [92]:
```
#   ********تغییر نام ستون********

# coffee.rename(columns={'pvpvpv':'Price'}, inplace=True)
# coffee.rename(columns={coffee.columns[3]:'Price'}, inplace=True)
# coffee

#   ******* کپی دیتافریم    ********
# coffee_copy = coffee.copy()
# coffee_copy
```

In [136…
```
# ***** ساخت ستون جدید *****
# ***** تبدیل object به float *****
coffee['Revenue'] = coffee['Price'].astype(float) * coffee['Units Sold']
```

```
# coffee.dtypes          # بررسی نوع داده‌ها
coffee
```

Out[136...

	Day	Coffee Type	Units Sold	Price	Revenue
0	Monday	Espresso	25	3.0	75.0
1	Monday	Latte	15	6	90.0
2	Tuesday	Espresso	30	3.0	90.0
3	Tuesday	Latte	20	6	120.0
4	Wednesday	Espresso	35	3.0	105.0
5	Wednesday	Latte	25	6	150.0
6	Thursday	Espresso	40	3.0	120.0
7	Thursday	Latte	30	6	180.0
8	Friday	Espresso	45	3.0	135.0
9	Friday	Latte	35	6	210.0
10	Saturday	Espresso	45	3.0	135.0
11	Saturday	Latte	35	6	210.0
12	Sunday	Espresso	45	3.0	135.0
13	Sunday	Latte	35	6	210.0

In [94]:
```
import random

a=random.randint(1, 1000)
a
```

Out[94]: 289

In [95]:
```
import pandas as pd

bios=pd.read_excel("C:/Users/LENOVO/Downloads/olympics-data.xlsx")
# bios.columns
```

```
biosn=bios.copy()
# ***** برای ساخت ستون نام و نام خانوادگی split جدا کردن نام به دو بخش با *****
# biosn['first_name'] = biosn['name'].str.split(' ').str[0]
# biosn['last_name'] = biosn['name'].str.split(' ').str[1]

# ***** همراه با ترتیب insert افزودن ستون‌های جدید با *****

biosn.insert(2,'first_name',biosn['name'].str.split(' ').str[0])
biosn.insert(3,'last_name',biosn['name'].str.split(' ').str[1])
biosn
```

Out[95]:

	athlete_id	name	first_name	last_name	born_date	born_city	born_region	born_country	NOC	height_cr
0	1	Jean-François Blanchy	Jean-François	Blanchy	1886-12-12	Bordeaux	Gironde	FRA	France	Na
1	2	Arnaud Boetsch	Arnaud	Boetsch	1969-04-01	Meulan	Yvelines	FRA	France	183.
2	3	Jean Borotra	Jean	Borotra	1898-08-13	Biarritz	Pyrénées-Atlantiques	FRA	France	183.
3	4	Jacques Brugnon	Jacques	Brugnon	1895-05-11	Paris VIIIe	Paris	FRA	France	168.
4	5	Albert Canet	Albert	Canet	1878-04-17	Wandsworth	England	GBR	France	Na
...	...	...	...	...	...	...	...	...	...	
145495	149222	Polina Luchnikova	Polina	Luchnikova	2002-01-30	Serov	Sverdlovsk	RUS	ROC	167.
145496	149223	Valeriya Merkusheva	Valeriya	Merkusheva	1999-09-20	Moskva (Moscow)	Moskva	RUS	ROC	168.
145497	149224	Yuliya Smirnova	Yuliya	Smirnova	1998-05-08	Kotlas	Arkhangelsk	RUS	ROC	163.
145498	149225	André Foussard	André	Foussard	1899-05-19	Niort	Deux-Sèvres	FRA	France	166.
145499	149814	Bill Phillips	Bill	Phillips	1913-07-15	Dulwich Hill	New South Wales	AUS	Australia	Na

145500 rows × 12 columns

In [96]:
```
# ***** جستوجو نام Pouria *****
biosn.query('first_name == "Pouria"')
```

Out[96]:

	athlete_id	name	first_name	last_name	born_date	born_city	born_region	born_country	NOC	height_cm	w
117007	118449	Pouria Saveh Shemshaki	Pouria	Saveh	1987-04-30	Tehran	Tehran	IRI	Islamic Republic of Iran	177.0	

In [97]:
```
# ***** تبدیل نوع داده تاریخ تولد به datetime *****
biosn["born_date"] = pd.to_datetime(biosn["born_date"] , format='%Y-%m-%d')

biosn.head()
```

Out[97]:

	athlete_id	name	first_name	last_name	born_date	born_city	born_region	born_country	NOC	height_cm	weight_k
0	1	Jean-François Blanchy	Jean-François	Blanchy	1886-12-12	Bordeaux	Gironde	FRA	France	NaN	Na
1	2	Arnaud Boetsch	Arnaud	Boetsch	1969-04-01	Meulan	Yvelines	FRA	France	183.0	76.
2	3	Jean Borotra	Jean	Borotra	1898-08-13	Biarritz	Pyrénées-Atlantiques	FRA	France	183.0	76.
3	4	Jacques Brugnon	Jacques	Brugnon	1895-05-11	Paris VIIIe	Paris	FRA	France	168.0	64.
4	5	Albert Canet	Albert	Canet	1878-04-17	Wandsworth	England	GBR	France	NaN	Na

In [98]:
```
# ***** ذخیره‌سازی داده‌های تغییر یافته *****
biosn.to_csv("C:/Users/LENOVO/Downloads/olympics-data-transformed.csv", index=False)
```

In [99]:
```
nocs=pd.read_csv("C:/Users/LENOVO/Downloads/noc_regions.csv")
nocs
```

Out[99]:

	NOC	region	notes
0	AFG	Afghanistan	NaN
1	AHO	Curacao	Netherlands Antilles
2	ALB	Albania	NaN
3	ALG	Algeria	NaN
4	AND	Andorra	NaN
...	...	...	...
225	YEM	Yemen	NaN
226	YMD	Yemen	South Yemen
227	YUG	Serbia	Yugoslavia
228	ZAM	Zambia	NaN
229	ZIM	Zimbabwe	NaN

230 rows × 3 columns

In [100…

```
# Left Join: همه سطرهای bios رو نگه میداره و اطلاعات nocs رو بر اساس تطابق کشور اضافه میکنه
# جایگزین میشه NaN پیدا نشه، مقادیر nocs اگر کشوری در
bios_new=pd.merge(bios , nocs ,left_on='born_country', right_on='NOC', how='left')
# Right Join: همه سطرهای nocs رو نگه میداره و اطلاعات bios رو بر اساس تطابق کشور اضافه میکنه
# جایگزین میشه NaN پیدا نشه، مقادیر nocs اگر ورزشکاری برای کشوری در
# combined=pd.merge(bios, nocs, left_on='born_country', right_on='NOC', how='right')
bios_new


# Inner: فقط موارد مشترک
# Left: همه چپ + مشترکات
# Right: همه راست + مشترکات
# Outer: همه چیز از هر دو
```

Out[100…

	athlete_id	name	born_date	born_city	born_region	born_country	NOC_x	height_cm	weight_kg	died_date
0	1	Jean-François Blanchy	1886-12-12	Bordeaux	Gironde	FRA	France	NaN	NaN	1960-10-02
1	2	Arnaud Boetsch	1969-04-01	Meulan	Yvelines	FRA	France	183.0	76.0	NaN
2	3	Jean Borotra	1898-08-13	Biarritz	Pyrénées-Atlantiques	FRA	France	183.0	76.0	1994-07-17
3	4	Jacques Brugnon	1895-05-11	Paris VIIIe	Paris	FRA	France	168.0	64.0	1978-03-20
4	5	Albert Canet	1878-04-17	Wandsworth	England	GBR	France	NaN	NaN	1930-07-25
...	...	...	...	...	...	...	...	...	...	...
145495	149222	Polina Luchnikova	2002-01-30	Serov	Sverdlovsk	RUS	ROC	167.0	61.0	NaN
145496	149223	Valeriya Merkusheva	1999-09-20	Moskva (Moscow)	Moskva	RUS	ROC	168.0	65.0	NaN
145497	149224	Yuliya Smirnova	1998-05-08	Kotlas	Arkhangelsk	RUS	ROC	163.0	55.0	NaN
145498	149225	André Foussard	1899-05-19	Niort	Deux-Sèvres	FRA	France	166.0	NaN	1986-03-18
145499	149814	Bill Phillips	1913-07-15	Dulwich Hill	New South Wales	AUS	Australia	NaN	NaN	2003-10-20

145500 rows × 13 columns

In [101…

```
bios_new.rename(columns={'region': 'born_country_full'}, inplace=True)
bios_new
```

Out[101…

	athlete_id	name	born_date	born_city	born_region	born_country	NOC_x	height_cm	weight_kg	died_date
0	1	Jean-François Blanchy	1886-12-12	Bordeaux	Gironde	FRA	France	NaN	NaN	1960-10-02
1	2	Arnaud Boetsch	1969-04-01	Meulan	Yvelines	FRA	France	183.0	76.0	NaN
2	3	Jean Borotra	1898-08-13	Biarritz	Pyrénées-Atlantiques	FRA	France	183.0	76.0	1994-07-17
3	4	Jacques Brugnon	1895-05-11	Paris VIIIe	Paris	FRA	France	168.0	64.0	1978-03-20
4	5	Albert Canet	1878-04-17	Wandsworth	England	GBR	France	NaN	NaN	1930-07-25
…	…	…	…	…	…	…	…	…	…	…
145495	149222	Polina Luchnikova	2002-01-30	Serov	Sverdlovsk	RUS	ROC	167.0	61.0	NaN
145496	149223	Valeriya Merkusheva	1999-09-20	Moskva (Moscow)	Moskva	RUS	ROC	168.0	65.0	NaN
145497	149224	Yuliya Smirnova	1998-05-08	Kotlas	Arkhangelsk	RUS	ROC	163.0	55.0	NaN
145498	149225	André Foussard	1899-05-19	Niort	Deux-Sèvres	FRA	France	166.0	NaN	1986-03-18
145499	149814	Bill Phillips	1913-07-15	Dulwich Hill	New South Wales	AUS	Australia	NaN	NaN	2003-10-20

145500 rows × 13 columns

In [102…

```
# با نام کامل کشور محل تولدشان تطابق ندارد (NOC) پیدا کردن ورزشکارانی که کد کشور
# این برای شناسایی خطاهای داده‌ای یا تفاوت‌های نامگذاری مفید است

# bios_new[bios_new['NOC_x'] !=bios_new['born_country_full']][['name','NOC_x','born_country_full']]
```

```
bios_new[

    # با نام کامل کشور (NOC) شرط: مقایسه ستون کد کشور
    # (یعنی مشکل دارن) = True اگر متفاوت باشند
    bios_new['NOC_x'] != bios_new['born_country_full']

][
    # انتخاب فقط این سه ستون برای نمایش نتیجه
    ['name', 'NOC_x', 'born_country_full']

    # name: نام ورزشکار
    # NOC_x: کد سه حرفی کشور (مثلاً IRI برای ایران)
    # born_country_full: نام کامل کشور (مثلاً Iran)
]
```

Out[102…

	name	NOC_x	born_country_full
4	Albert Canet	France	UK
12	J. Defert	France	NaN
13	Étienne Durand	France	NaN
16	Guy Forget	France	Morocco
27	Guy, Baron Lejeune	France	NaN
…	…	…	…
145491	Matthew Wepke	Jamaica	NaN
145493	Landysh Falyakhova	ROC	Russia
145495	Polina Luchnikova	ROC	Russia
145496	Valeriya Merkusheva	ROC	Russia
145497	Yuliya Smirnova	ROC	Russia

73210 rows × 3 columns

In [103…

```
usa=bios[bios['born_country'] == 'USA']
gbr=bios[bios['born_country'] == 'GBR']
#**** چینش کد ها زیر هم ****

new_df=pd.concat([usa, gbr])
new_df

#**** قوانین و شرایط ****
# اگر ستون‌ها متفاوت باشن:
# usa: ['name', 'age'] + gbr: ['name', 'height'] = new_df: ['name', 'age', 'height']
# مقادیر ستون‌های missing میشه NaN
# new_df = pd.concat([usa, gbr],
#     ignore_index=True,    # ایندکس جدید از ۰ میسازه
#     sort=False,           # ستون‌ها رو مرتب نکنه
#     axis=0                # (پیش‌فرض) عمودی ادغام کن
# )
```

Out[103…

	athlete_id	name	born_date	born_city	born_region	born_country	NOC	height_cm	weight_kg	died_date
54	55	Monique Javer	1967-07-22	Burlingame	California	USA	Great Britain	177.0	64.0	NaN
960	964	Xóchitl Escobedo	1968-09-17	West Covina	California	USA	Mexico	170.0	60.0	NaN
961	965	Angélica Gavaldón	1973-10-03	El Centro	California	USA	Mexico	160.0	54.0	NaN
1231	1238	Bert Schneider	1897-07-01	Cleveland	Ohio	USA	Canada	NaN	NaN	1986-02-20
1345	1352	Laura Berg	1975-01-06	Santa Fe Springs	California	USA	United States	168.0	61.0	NaN
…	…	…	…	…	…	…	…	…	…	…
144811	148512	Benjamin Alexander	1983-05-08	London	England	GBR	Jamaica	NaN	NaN	NaN
144815	148517	Ashley Watson	1993-10-28	Peterborough	England	GBR	Jamaica	NaN	NaN	NaN
145005	148716	Peder Kongshaug	2001-08-13	Wimbledon	England	GBR	Norway	184.0	86.0	NaN
145319	149041	Axel Brown	1992-04-02	Harrogate	England	GBR	Trinidad and Tobago	NaN	NaN	NaN
145388	149111	Jean-Luc Baker	1993-10-07	Burnley	England	GBR	United States	NaN	NaN	NaN

15433 rows × 10 columns

In [104…

```
results=pd.read_parquet("C:/Users/LENOVO/Downloads/results.parquet")
results
```

Out[104...

	year	type	discipline	event	as	athlete_id	noc	team	place	tied	medal
0	1912.0	Summer	Tennis	Singles, Men (Olympic)	Jean-François Blanchy	1	FRA	None	17.0	True	None
1	1912.0	Summer	Tennis	Doubles, Men (Olympic)	Jean-François Blanchy	1	FRA	Jean Montariol	NaN	False	None
2	1920.0	Summer	Tennis	Singles, Men (Olympic)	Jean-François Blanchy	1	FRA	None	32.0	True	None
3	1920.0	Summer	Tennis	Doubles, Mixed (Olympic)	Jean-François Blanchy	1	FRA	Jeanne Vaussard	8.0	True	None
4	1920.0	Summer	Tennis	Doubles, Men (Olympic)	Jean-François Blanchy	1	FRA	Jacques Brugnon	4.0	False	None
...	...	...	...	...	...	...	...	...	...	...	...
308403	2022.0	Winter	Luge	Singles, Men (Olympic)	Marián Skupek	148983	SVK	None	26.0	False	None
308404	2022.0	Winter	Alpine Skiing (Skiing)	Slalom, Women (Olympic)	Elsa Fermbäck	148984	SWE	None	28.0	False	None
308405	2022.0	Winter	Alpine Skiing (Skiing)	Team, Mixed (Olympic)	Hilma Lövblom	148985	SWE	Sweden	13.0	False	None
308406	2022.0	Winter	Alpine Skiing (Skiing)	Giant Slalom, Women (Olympic)	Hilma Lövblom	148985	SWE	None	NaN	False	None
308407	2022.0	Winter	None	Slalom, Women (Olympic)	Charlotta Säfvenberg	148986	None	None	24.0	False	None

308408 rows × 11 columns

In [105...

```
# ***** در  merge کردن *****
# on='ستون' - وقتی اسم ستون مشترکه
# left_on و right_on - وقتی اسم ستون‌ها متفاوته:

# ***** با suffixes: *****
# pd.merge(df1, df2, on='id', suffixes=('', '_bio'))  # ستون‌ها: name, name_bio
```

```
combined_df=pd.merge(results,bios,on='athlete_id',how='left',suffixes=('', '_bio'))
pd.set_option('display.max_columns', None)
combined_df
```

Out[105...

	year	type	discipline	event	as	athlete_id	noc	team	place	tied	medal	name	born_da
0	1912.0	Summer	Tennis	Singles, Men (Olympic)	Jean-François Blanchy	1	FRA	None	17.0	True	None	Jean-François Blanchy	1886-1
1	1912.0	Summer	Tennis	Doubles, Men (Olympic)	Jean-François Blanchy	1	FRA	Jean Montariol	NaN	False	None	Jean-François Blanchy	1886-1
2	1920.0	Summer	Tennis	Singles, Men (Olympic)	Jean-François Blanchy	1	FRA	None	32.0	True	None	Jean-François Blanchy	1886-1
3	1920.0	Summer	Tennis	Doubles, Mixed (Olympic)	Jean-François Blanchy	1	FRA	Jeanne Vaussard	8.0	True	None	Jean-François Blanchy	1886-1
4	1920.0	Summer	Tennis	Doubles, Men (Olympic)	Jean-François Blanchy	1	FRA	Jacques Brugnon	4.0	False	None	Jean-François Blanchy	1886-1
...	...	...	...	...	...	...	...	...	...	...	...	...	
308403	2022.0	Winter	Luge	Singles, Men (Olympic)	Marián Skupek	148983	SVK	None	26.0	False	None	Marián Skupek	2001-0
308404	2022.0	Winter	Alpine Skiing (Skiing)	Slalom, Women (Olympic)	Elsa Fermbäck	148984	SWE	None	28.0	False	None	Elsa Fermbäck	1998-0
308405	2022.0	Winter	Alpine Skiing (Skiing)	Team, Mixed (Olympic)	Hilma Lövblom	148985	SWE	Sweden	13.0	False	None	Hilma Lövblom	2000-0
308406	2022.0	Winter	Alpine Skiing (Skiing)	Giant Slalom, Women (Olympic)	Hilma Lövblom	148985	SWE	None	NaN	False	None	Hilma Lövblom	2000-0
308407	2022.0	Winter	None	Slalom, Women (Olympic)	Charlotta Säfvenberg	148986	None	None	24.0	False	None	Charlotta Säfvenberg	1994-1

308408 rows × 20 columns

In [ ]:
```
# coffee.loc[[8],'Units Sold']=np.nan
# coffee.loc[[0,1],'Units Sold']=[15,42]

# **** mean میانگین ****
# coffee.fillna(coffee['Units Sold'].mean(),inplace=True)

# **** interpolate به همسایه های اون نگاه میکنه و طبق اونها پر میکنه ****
# coffee['Units Sold'] = coffee['Units Sold'].interpolate()

# **** dropna() خالی ها رو پاک میکنه ****
# coffee.dropna(inplace=True) #→ سطرهایی که هر ستونی NaN داشته باشه
# coffee.dropna(subset=['Units Sold'],inplace=True) #→ فقط سطرهایی که ستون‌های مشخص شده NaN داشته باشن

# **** فیلتر سطرهایی که مقدار ندارند ****
# نیاز به فراخوانی coffee ندارد
# coffee[coffee['Units Sold'].isna()]

# **** فیلتر سطرهایی که مقدار دارند ****
# نیاز به فراخوانی coffee ندارد
# coffee[coffee['Units Sold'].notna()]


# coffee

# **** Not Working ****
# ****نکته خیلی مهمممممممم****
# ****interpolate عملا خروجیش عدد نیست سری هست و نیمتونم سری رو بدیم برای پر کردن سری دیگه ****
# coffee.fillna(coffee['Units Sold'].interpolate(),inplace=True)
```

Out[ ]:

	Day	Coffee Type	Units Sold	Price	Revenue
8	Friday	Espresso	NaN	3.0	135.0

In [ ]: