

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

```
In [2]: pd.__version__
```

```
Out[2]: '2.2.2'
```

```
In [5]: df = pd.read_csv(r'D:\sir gen Ai\pandas\data.csv')
df
```

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income
1	Afghanistan	AFG	35.253	5.9	Low income
2	Angola	AGO	45.985	19.1	Upper middle income
3	Albania	ALB	12.877	57.2	Upper middle income
4	United Arab Emirates	ARE	11.044	88.0	High income
...
190	Yemen, Rep.	YEM	32.947	20.0	Lower middle income
191	South Africa	ZAF	20.850	46.5	Upper middle income
192	Congo, Dem. Rep.	COD	42.394	2.2	Low income
193	Zambia	ZMB	40.471	15.4	Lower middle income
194	Zimbabwe	ZWE	35.715	18.5	Low income

195 rows × 5 columns

```
In [6]: id(df)
```

```
Out[6]: 2118218292624
```

```
In [7]: len(df)
```

```
Out[7]: 195
```

```
In [8]: df.columns
```

```
Out[8]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
       dtype='object')
```

```
In [9]: len(df.columns)
```

```
Out[9]: 5
```

In [10]: `df.isnull() #checking null value in data frames (df.isna() also does same job)`

Out[10]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
...
190	False	False	False	False	False
191	False	False	False	False	False
192	False	False	False	False	False
193	False	False	False	False	False
194	False	False	False	False	False

195 rows × 5 columns

In [11]: `df.isnull().sum() # count of the null values`

Out[11]:

CountryName	0
CountryCode	0
BirthRate	0
InternetUsers	0
IncomeGroup	0
dtype: int64	

In [12]: `df.head() # default top 5 rows`

Out[12]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income
1	Afghanistan	AFG	35.253	5.9	Low income
2	Angola	AGO	45.985	19.1	Upper middle income
3	Albania	ALB	12.877	57.2	Upper middle income
4	United Arab Emirates	ARE	11.044	88.0	High income

In [13]: `df.tail() #default bottom 5 rows`

Out[13]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
190	Yemen, Rep.	YEM	32.947	20.0	Lower middle income
191	South Africa	ZAF	20.850	46.5	Upper middle income
192	Congo, Dem. Rep.	COD	42.394	2.2	Low income
193	Zambia	ZMB	40.471	15.4	Lower middle income
194	Zimbabwe	ZWE	35.715	18.5	Low income

In [14]:

`df.info() #gives whole info of excel sheet`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 195 entries, 0 to 194
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
--- 
 0   CountryName    195 non-null    object  
 1   CountryCode    195 non-null    object  
 2   BirthRate      195 non-null    float64 
 3   InternetUsers 195 non-null    float64 
 4   IncomeGroup    195 non-null    object  
dtypes: float64(2), object(3)
memory usage: 7.7+ KB
```

In [15]:

`df[:]`

Out[15]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income
1	Afghanistan	AFG	35.253	5.9	Low income
2	Angola	AGO	45.985	19.1	Upper middle income
3	Albania	ALB	12.877	57.2	Upper middle income
4	United Arab Emirates	ARE	11.044	88.0	High income
...
190	Yemen, Rep.	YEM	32.947	20.0	Lower middle income
191	South Africa	ZAF	20.850	46.5	Upper middle income
192	Congo, Dem. Rep.	COD	42.394	2.2	Low income
193	Zambia	ZMB	40.471	15.4	Lower middle income
194	Zimbabwe	ZWE	35.715	18.5	Low income

195 rows × 5 columns

In [16]: df[1:]

Out[16]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
1	Afghanistan	AFG	35.253	5.9	Low income
2	Angola	AGO	45.985	19.1	Upper middle income
3	Albania	ALB	12.877	57.2	Upper middle income
4	United Arab Emirates	ARE	11.044	88.0	High income
5	Argentina	ARG	17.716	59.9	High income
...
190	Yemen, Rep.	YEM	32.947	20.0	Lower middle income
191	South Africa	ZAF	20.850	46.5	Upper middle income
192	Congo, Dem. Rep.	COD	42.394	2.2	Low income
193	Zambia	ZMB	40.471	15.4	Lower middle income
194	Zimbabwe	ZWE	35.715	18.5	Low income

194 rows × 5 columns

In [17]: df[1:100:10]

Out[17]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
1	Afghanistan	AFG	35.253	5.9000	Low income
11	Burundi	BDI	44.151	1.3000	Low income
21	Belize	BLZ	23.092	33.6000	Upper middle income
31	Switzerland	CHE	10.200	86.3400	High income
41	Cuba	CUB	10.400	27.9300	Upper middle income
51	Egypt, Arab Rep.	EGY	28.032	29.4000	Lower middle income
61	United Kingdom	GBR	12.200	89.8441	High income
71	Guatemala	GTM	27.465	19.7000	Lower middle income
81	Ireland	IRL	15.000	78.2477	High income
91	Kenya	KEN	35.194	39.0000	Lower middle income

In [18]: df[:::-1]

Out[18]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
194	Zimbabwe	ZWE	35.715	18.5	Low income
193	Zambia	ZMB	40.471	15.4	Lower middle income
192	Congo, Dem. Rep.	COD	42.394	2.2	Low income
191	South Africa	ZAF	20.850	46.5	Upper middle income
190	Yemen, Rep.	YEM	32.947	20.0	Lower middle income
...
4	United Arab Emirates	ARE	11.044	88.0	High income
3	Albania	ALB	12.877	57.2	Upper middle income
2	Angola	AGO	45.985	19.1	Upper middle income
1	Afghanistan	AFG	35.253	5.9	Low income
0	Aruba	ABW	10.244	78.9	High income

195 rows × 5 columns

In [19]: df[10:21]

Out[19]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
10	Azerbaijan	AZE	18.300	58.70000	Upper middle income
11	Burundi	BDI	44.151	1.30000	Low income
12	Belgium	BEL	11.200	82.17020	High income
13	Benin	BEN	36.440	4.90000	Low income
14	Burkina Faso	BFA	40.551	9.10000	Low income
15	Bangladesh	BGD	20.142	6.63000	Lower middle income
16	Bulgaria	BGR	9.200	53.06150	Upper middle income
17	Bahrain	BHR	15.040	90.00004	High income
18	Bahamas, The	BHS	15.339	72.00000	High income
19	Bosnia and Herzegovina	BIH	9.062	57.79000	Upper middle income
20	Belarus	BLR	12.500	54.17000	Upper middle income

In [20]: `df.describe() # gives only gives always numerical records`

Out[20]:

	BirthRate	InternetUsers
count	195.000000	195.000000
mean	21.469928	42.076471
std	10.605467	29.030788
min	7.900000	0.900000
25%	12.120500	14.520000
50%	19.680000	41.000000
75%	29.759500	66.225000
max	49.661000	96.546800

In [21]: `df.head(1)`

Out[21]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income

In [23]: `df['CountryName']`

Out[23]:

0	Aruba
1	Afghanistan
2	Angola
3	Albania
4	United Arab Emirates
	...
190	Yemen, Rep.
191	South Africa
192	Congo, Dem. Rep.
193	Zambia
194	Zimbabwe

Name: CountryName, Length: 195, dtype: object

In [25]: `df[['CountryName', 'CountryCode']] # subset`

Out[25]:

	CountryName	CountryCode
0	Aruba	ABW
1	Afghanistan	AFG
2	Angola	AGO
3	Albania	ALB
4	United Arab Emirates	ARE
...
190	Yemen, Rep.	YEM
191	South Africa	ZAF
192	Congo, Dem. Rep.	COD
193	Zambia	ZMB
194	Zimbabwe	ZWE

195 rows × 2 columns

In [26]:

`df[['CountryName', 'CountryCode', 'IncomeGroup']]`

Out[26]:

	CountryName	CountryCode	IncomeGroup
0	Aruba	ABW	High income
1	Afghanistan	AFG	Low income
2	Angola	AGO	Upper middle income
3	Albania	ALB	Upper middle income
4	United Arab Emirates	ARE	High income
...
190	Yemen, Rep.	YEM	Lower middle income
191	South Africa	ZAF	Upper middle income
192	Congo, Dem. Rep.	COD	Low income
193	Zambia	ZMB	Lower middle income
194	Zimbabwe	ZWE	Low income

195 rows × 3 columns

In [27]:

`df_cat = df[['CountryName', 'CountryCode', 'IncomeGroup']]`

In [28]:

`df_cat`

Out[28]:

	CountryName	CountryCode	IncomeGroup
0	Aruba	ABW	High income
1	Afghanistan	AFG	Low income
2	Angola	AGO	Upper middle income
3	Albania	ALB	Upper middle income
4	United Arab Emirates	ARE	High income
...
190	Yemen, Rep.	YEM	Lower middle income
191	South Africa	ZAF	Upper middle income
192	Congo, Dem. Rep.	COD	Low income
193	Zambia	ZMB	Lower middle income
194	Zimbabwe	ZWE	Low income

195 rows × 3 columns

In [31]:

```
print(df.columns)
print(df_cat.columns)
```

Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
 'IncomeGroup'],
 dtype='object')
Index(['CountryName', 'CountryCode', 'IncomeGroup'], dtype='object')

In [32]:

```
print(df_cat.columns)
```

Index(['CountryName', 'CountryCode', 'IncomeGroup'], dtype='object')

In [33]:

```
df_cat.describe()
```

Out[33]:

	CountryName	CountryCode	IncomeGroup
count	195	195	195
unique	195	195	4
top	Aruba	ABW	High income
freq	1	1	67

In [34]:

```
df_num = df[['BirthRate', 'InternetUsers']]
df_num
```

Out[34]:

	BirthRate	InternetUsers
0	10.244	78.9
1	35.253	5.9
2	45.985	19.1
3	12.877	57.2
4	11.044	88.0
...
190	32.947	20.0
191	20.850	46.5
192	42.394	2.2
193	40.471	15.4
194	35.715	18.5

195 rows × 2 columns

In [35]:

`df.describe().transpose() # rows are converted into columns`

Out[35]:

	count	mean	std	min	25%	50%	75%	max
BirthRate	195.0	21.469928	10.605467	7.9	12.1205	19.68	29.7595	49.6610
InternetUsers	195.0	42.076471	29.030788	0.9	14.5200	41.00	66.2250	96.5468

In [36]:

`df.describe().T`

Out[36]:

	count	mean	std	min	25%	50%	75%	max
BirthRate	195.0	21.469928	10.605467	7.9	12.1205	19.68	29.7595	49.6610
InternetUsers	195.0	42.076471	29.030788	0.9	14.5200	41.00	66.2250	96.5468

In [37]:

`df.columns = ['a', 'b', 'c', 'd', 'e']`

In [38]:

`df.columns`

Out[38]:

`Index(['a', 'b', 'c', 'd', 'e'], dtype='object')`

In [39]:

`df.head()`

Out[39]:

	a	b	c	d	e
0	Aruba	ABW	10.244	78.9	High income
1	Afghanistan	AFG	35.253	5.9	Low income
2	Angola	AGO	45.985	19.1	Upper middle income
3	Albania	ALB	12.877	57.2	Upper middle income
4	United Arab Emirates	ARE	11.044	88.0	High income

In [40]:

```
df.columns = ['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
              'IncomeGroup']
```

In [42]:

```
df.head(1)
```

Out[42]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income

In [43]:

```
df[['CountryName', 'CountryCode', 'IncomeGroup']][4:8]
```

Out[43]:

	CountryName	CountryCode	IncomeGroup
4	United Arab Emirates	ARE	High income
5	Argentina	ARG	High income
6	Armenia	ARM	Lower middle income
7	Antigua and Barbuda	ATG	High income

In [44]:

```
df[4:8][['CountryName', 'CountryCode', 'IncomeGroup']] # subset
```

Out[44]:

	CountryName	CountryCode	IncomeGroup
4	United Arab Emirates	ARE	High income
5	Argentina	ARG	High income
6	Armenia	ARM	Lower middle income
7	Antigua and Barbuda	ATG	High income

In [45]:

```
df.columns
```

Out[45]:

```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

In [46]:

```
df.BirthRate * df.InternetUsers
```

```
Out[46]: 0    808.2516
         1    207.9927
         2    878.3135
         3    736.5644
         4    971.8720
         ...
        190   658.9400
        191   969.5250
        192   93.2668
        193   623.2534
        194   660.7275
Length: 195, dtype: float64
```

```
In [47]: df.head()
```

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income
1	Afghanistan	AFG	35.253	5.9	Low income
2	Angola	AGO	45.985	19.1	Upper middle income
3	Albania	ALB	12.877	57.2	Upper middle income
4	United Arab Emirates	ARE	11.044	88.0	High income

```
In [48]: df['newcol']= df.BirthRate * df.InternetUsers
```

```
In [49]: df.head()
```

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	newcol
0	Aruba	ABW	10.244	78.9	High income	808.2516
1	Afghanistan	AFG	35.253	5.9	Low income	207.9927
2	Angola	AGO	45.985	19.1	Upper middle income	878.3135
3	Albania	ALB	12.877	57.2	Upper middle income	736.5644
4	United Arab Emirates	ARE	11.044	88.0	High income	971.8720

```
In [50]: df = df.drop('newcol',axis= 1)
```

```
In [51]: df
```

Out[51]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income
1	Afghanistan	AFG	35.253	5.9	Low income
2	Angola	AGO	45.985	19.1	Upper middle income
3	Albania	ALB	12.877	57.2	Upper middle income
4	United Arab Emirates	ARE	11.044	88.0	High income
...
190	Yemen, Rep.	YEM	32.947	20.0	Lower middle income
191	South Africa	ZAF	20.850	46.5	Upper middle income
192	Congo, Dem. Rep.	COD	42.394	2.2	Low income
193	Zambia	ZMB	40.471	15.4	Lower middle income
194	Zimbabwe	ZWE	35.715	18.5	Low income

195 rows × 5 columns

In [52]: df = df.drop('newcol', axis=1) # deleting a column

```

-----
KeyError                                         Traceback (most recent call last)
Cell In[52], line 1
----> 1 df = df.drop('newcol', axis= 1)

File ~\anaconda3\Lib\site-packages\pandas\core\frame.py:5581, in DataFrame.drop(self, labels, axis, index, columns, level, inplace, errors)
    5433 def drop(
    5434     self,
    5435     labels: IndexLabel | None = None,
    5436     ...
    5442     errors: IgnoreRaise = "raise",
    5443 ) -> DataFrame | None:
    5444     """
    5445     Drop specified labels from rows or columns.
    5446
    5447     ...
    5579         weight 1.0      0.8
    5580     """
-> 5581     return super().drop(
    5582         labels=labels,
    5583         axis=axis,
    5584         index=index,
    5585         columns=columns,
    5586         level=level,
    5587         inplace=inplace,
    5588         errors=errors,
    5589     )

File ~\anaconda3\Lib\site-packages\pandas\core\generic.py:4788, in NDFrame.drop(self, labels, axis, index, columns, level, inplace, errors)
    4786 for axis, labels in axes.items():
    4787     if labels is not None:
-> 4788         obj = obj._drop_axis(labels, axis, level=level, errors=errors)
    4790 if inplace:
    4791     self._update_inplace(obj)

File ~\anaconda3\Lib\site-packages\pandas\core\generic.py:4830, in NDFrame._drop_axis(self, labels, axis, level, errors, only_slice)
    4828     new_axis = axis.drop(labels, level=level, errors=errors)
    4829 else:
-> 4830     new_axis = axis.drop(labels, errors=errors)
    4831 indexer = axis.get_indexer(new_axis)
    4833 # Case for non-unique axis
    4834 else:

File ~\anaconda3\Lib\site-packages\pandas\core\indexes\base.py:7070, in Index.drop(self, labels, errors)
    7068 if mask.any():
    7069     if errors != "ignore":
-> 7070         raise KeyError(f"{labels[mask].tolist()} not found in axis")
    7071 indexer = indexer[~mask]
    7072 return self.delete(indexer)

KeyError: "['newcol'] not found in axis"

```

In [53]: df['newcol'] = df.BirthRate * df.InternetUsers # adding a column

In [54]: df.head()

Out[54]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	newcol
0	Aruba	ABW	10.244	78.9	High income	808.2516
1	Afghanistan	AFG	35.253	5.9	Low income	207.9927
2	Angola	AGO	45.985	19.1	Upper middle income	878.3135
3	Albania	ALB	12.877	57.2	Upper middle income	736.5644
4	United Arab Emirates	ARE	11.044	88.0	High income	971.8720

In [55]: `len(df.columns)`

Out[55]: 6

In [56]: `df`

Out[56]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	newcol
0	Aruba	ABW	10.244	78.9	High income	808.2516
1	Afghanistan	AFG	35.253	5.9	Low income	207.9927
2	Angola	AGO	45.985	19.1	Upper middle income	878.3135
3	Albania	ALB	12.877	57.2	Upper middle income	736.5644
4	United Arab Emirates	ARE	11.044	88.0	High income	971.8720
...
190	Yemen, Rep.	YEM	32.947	20.0	Lower middle income	658.9400
191	South Africa	ZAF	20.850	46.5	Upper middle income	969.5250
192	Congo, Dem. Rep.	COD	42.394	2.2	Low income	93.2668
193	Zambia	ZMB	40.471	15.4	Lower middle income	623.2534
194	Zimbabwe	ZWE	35.715	18.5	Low income	660.7275

195 rows × 6 columns

In [57]: `df.InternetUsers < 2`

```
Out[57]: 0    False
         1    False
         2    False
         3    False
         4    False
         ...
        190   False
        191   False
        192   False
        193   False
        194   False
Name: InternetUsers, Length: 195, dtype: bool
```

```
In [59]: df[df.InternetUsers < 2]
```

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	newcol
11	Burundi	BDI	44.151	1.3	Low income	57.3963
52	Eritrea	ERI	34.800	0.9	Low income	31.3200
55	Ethiopia	ETH	32.925	1.9	Low income	62.5575
64	Guinea	GIN	37.337	1.6	Low income	59.7392
117	Myanmar	MMR	18.119	1.6	Lower middle income	28.9904
127	Niger	NER	49.661	1.7	Low income	84.4237
154	Sierra Leone	SLE	36.729	1.7	Low income	62.4393
156	Somalia	SOM	43.891	1.5	Low income	65.8365
172	Timor-Leste	TLS	35.755	1.1	Lower middle income	39.3305

```
In [61]: len(df[df.InternetUsers < 2])
```

```
Out[61]: 9
```

```
In [62]: df.BirthRate > 40
```

```
Out[62]: 0    False
         1    False
         2    True
         3    False
         4    False
         ...
        190   False
        191   False
        192   True
        193   True
        194   False
Name: BirthRate, Length: 195, dtype: bool
```

```
In [63]: df[df.BirthRate > 40]
```

Out[63]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	newcol
2	Angola	AGO	45.985	19.1	Upper middle income	878.3135
11	Burundi	BDI	44.151	1.3	Low income	57.3963
14	Burkina Faso	BFA	40.551	9.1	Low income	369.0141
65	Gambia, The	GMB	42.525	14.0	Low income	595.3500
115	Mali	MLI	44.138	3.5	Low income	154.4830
127	Niger	NER	49.661	1.7	Low income	84.4237
128	Nigeria	NGA	40.045	38.0	Lower middle income	1521.7100
156	Somalia	SOM	43.891	1.5	Low income	65.8365
167	Chad	TCD	45.745	2.3	Low income	105.2135
178	Uganda	UGA	43.474	16.2	Low income	704.2788
192	Congo, Dem. Rep.	COD	42.394	2.2	Low income	93.2668
193	Zambia	ZMB	40.471	15.4	Lower middle income	623.2534

In [64]: `len(df[df.BirthRate > 40])`

Out[64]: 12

In [65]: `low_educate = df[df.InternetUsers < 2]`
`low_educate`

Out[65]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	newcol
11	Burundi	BDI	44.151	1.3	Low income	57.3963
52	Eritrea	ERI	34.800	0.9	Low income	31.3200
55	Ethiopia	ETH	32.925	1.9	Low income	62.5575
64	Guinea	GIN	37.337	1.6	Low income	59.7392
117	Myanmar	MMR	18.119	1.6	Lower middle income	28.9904
127	Niger	NER	49.661	1.7	Low income	84.4237
154	Sierra Leone	SLE	36.729	1.7	Low income	62.4393
156	Somalia	SOM	43.891	1.5	Low income	65.8365
172	Timor-Leste	TLS	35.755	1.1	Lower middle income	39.3305

In [67]: `high_birthrate = df[df.BirthRate > 40]`
`high_birthrate`

Out[67]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	newcol
2	Angola	AGO	45.985	19.1	Upper middle income	878.3135
11	Burundi	BDI	44.151	1.3	Low income	57.3963
14	Burkina Faso	BFA	40.551	9.1	Low income	369.0141
65	Gambia, The	GMB	42.525	14.0	Low income	595.3500
115	Mali	MLI	44.138	3.5	Low income	154.4830
127	Niger	NER	49.661	1.7	Low income	84.4237
128	Nigeria	NGA	40.045	38.0	Lower middle income	1521.7100
156	Somalia	SOM	43.891	1.5	Low income	65.8365
167	Chad	TCD	45.745	2.3	Low income	105.2135
178	Uganda	UGA	43.474	16.2	Low income	704.2788
192	Congo, Dem. Rep.	COD	42.394	2.2	Low income	93.2668
193	Zambia	ZMB	40.471	15.4	Lower middle income	623.2534

In [73]:

```
filter1 = df.InternetUsers <2
filter2 = df.BirthRate >40
df[filter1 & filter2]
```

Out[73]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	newcol
11	Burundi	BDI	44.151	1.3	Low income	57.3963
127	Niger	NER	49.661	1.7	Low income	84.4237
156	Somalia	SOM	43.891	1.5	Low income	65.8365

In []: