

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

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

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

```
In [3]: df=pd.read_csv(r'D:\DS and AI\PROJRCTS IN DATA_SCIENCE\data.csv')
```

```
In [4]: df
```

```
Out[4]:
```

	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 [5]: id(df)
```

```
Out[5]: 1229226812736
```

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

```
Out[6]: 195
```

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

```
Out[7]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
              'IncomeGroup'],
              dtype='object')
```

```
In [8]: len(df.columns)
```

Out[8]: 5

In [9]: `df.isnull()` *# here we are finding the missing values in my dataset*

Out[9]:

	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 [10]: `df.isna()` *# these is also same as the isnull()*

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()`

```
Out[11]: CountryName      0
CountryCode      0
BirthRate        0
InternetUsers     0
IncomeGroup       0
dtype: int64
```

```
In [12]: df.isna().sum()
```

```
Out[12]: CountryName      0
CountryCode      0
BirthRate        0
InternetUsers     0
IncomeGroup       0
dtype: int64
```

```
In [13]: df.head() # head()- it means it print the top five records
```

```
Out[13]:
```

	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 [14]: df.tail() # tail()- it means it print the bottom five records
```

```
Out[14]:
```

	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 [15]: df.info() # here this command give the information about dataset
```

```
<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 [16]: df[1:11]
```

Out[16]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
1	Afghanistan	AFG	35.253	5.9000	Low income
2	Angola	AGO	45.985	19.1000	Upper middle income
3	Albania	ALB	12.877	57.2000	Upper middle income
4	United Arab Emirates	ARE	11.044	88.0000	High income
5	Argentina	ARG	17.716	59.9000	High income
6	Armenia	ARM	13.308	41.9000	Lower middle income
7	Antigua and Barbuda	ATG	16.447	63.4000	High income
8	Australia	AUS	13.200	83.0000	High income
9	Austria	AUT	9.400	80.6188	High income
10	Azerbaijan	AZE	18.300	58.7000	Upper middle income

In [17]:

df[2:-1]

Out[17]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
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
6	Armenia	ARM	13.308	41.9	Lower middle income
...
189	Samoa	WSM	26.172	15.3	Lower middle 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

192 rows × 5 columns

In [18]: df[1:100:10]

Out[18]:

	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 [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

Out[20]:

	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 [22]: `df.head(4)`

Out[22]:

	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

In [24]: `df.describe()` *# here describe () is give you by default it give numerarical*

Out[24]:

	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 [25]: `df.head(1)`

Out[25]:

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

In [27]: `df['CountryName']`

Out[27]:

```

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 [29]: `df['CountryCode']`

```
Out[29]: 0      ABW
         1      AFG
         2      AGO
         3      ALB
         4      ARE
         ...
        190     YEM
        191     ZAF
        192     COD
        193     ZMB
        194     ZWE
        Name: CountryCode, Length: 195, dtype: object
```

```
In [31]: df[['CountryName', 'CountryCode']]
```

```
Out[31]:
```

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


Out[32]:

	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 [40]: `df_cat=df[['CountryName','CountryCode','IncomeGroup']] # here cat()-means catog`In [41]: `df_cat`

Out[41]:

	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 [42]: `print(len(df.columns))
print(len(df_cat.columns))`

5

3

```
In [43]: print((df.columns))
```

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

```
In [44]: print((df_cat.columns))
```

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

```
In [45]: df_cat.describe()
```

```
Out[45]:
```

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

```
In [47]: df_num=df[['BirthRate', 'InternetUsers']]  
df_num
```

```
Out[47]:
```

	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 [48]: df.info()
```

```
<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 [50]: df_cat.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 195 entries, 0 to 194
Data columns (total 3 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   CountryName     195 non-null   object
1   CountryCode     195 non-null   object
2   IncomeGroup     195 non-null   object
dtypes: object(3)
memory usage: 4.7+ KB
```

```
In [51]: df_num.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 195 entries, 0 to 194
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   BirthRate       195 non-null   float64
1   InternetUsers   195 non-null   float64
dtypes: float64(2)
memory usage: 3.2 KB
```

```
In [52]: print(df_num.info())
print(df_cat.info())
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 195 entries, 0 to 194
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   BirthRate       195 non-null   float64
1   InternetUsers   195 non-null   float64
dtypes: float64(2)
memory usage: 3.2 KB
None
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 195 entries, 0 to 194
Data columns (total 3 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   CountryName     195 non-null   object
1   CountryCode     195 non-null   object
2   IncomeGroup     195 non-null   object
dtypes: object(3)
memory usage: 4.7+ KB
None

```

```
In [54]: df.describe().transpose()  #it covert (rows to col)x(col-rows)
```

```
Out[54]:
```

	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 [59]: df.describe().T
```

```
Out[59]:
```

	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 [60]: df.columns
```

```
Out[60]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
               'IncomeGroup'],
              dtype='object')
```

```
In [62]: df.columns=['a','b','c','d','e']
```

```
In [63]: df.columns
```

```
Out[63]: Index(['a', 'b', 'c', 'd', 'e'], dtype='object')
```

```
In [64]: df.head(1)
```

```
Out[64]:
```

	a	b	c	d	e
0	Aruba	ABW	10.244	78.9	High income

```
In [65]: df.columns=['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
                    'IncomeGroup']
```

```
In [66]: df.columns
```

```
Out[66]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
                'IncomeGroup'],
              dtype='object')
```

```
In [67]: df.head(1)
```

```
Out[67]:
```

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

```
In [68]: df[['CountryCode', 'BirthRate', 'InternetUsers']][4:8] # subset
```

```
Out[68]:
```

	CountryCode	BirthRate	InternetUsers
4	ARE	11.044	88.0
5	ARG	17.716	59.9
6	ARM	13.308	41.9
7	ATG	16.447	63.4

```
In [69]: df[4:8][['CountryCode', 'BirthRate', 'InternetUsers']] # subset () another one
```

```
Out[69]:
```

	CountryCode	BirthRate	InternetUsers
4	ARE	11.044	88.0
5	ARG	17.716	59.9
6	ARM	13.308	41.9
7	ATG	16.447	63.4

```
In [70]: df.BirthRate*df.InternetUsers
```

```
Out[70]:
```

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 [71]: df.head(2)
```

Out[71]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
--	-------------	-------------	-----------	---------------	-------------

0	Aruba	ABW	10.244	78.9	High income
---	-------	-----	--------	------	-------------

1	Afghanistan	AFG	35.253	5.9	Low income
---	-------------	-----	--------	-----	------------

In [72]: `df['new column']=df.BirthRate*df.InternetUsers # here we are adding the new c`
`df.head(5)`

Out[72]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	new column
--	-------------	-------------	-----------	---------------	-------------	------------

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 [74]: `len(df.columns)`

Out[74]: 6

In [75]: `df=df.drop('new column',axis=1)`

In [77]: `df.head(5)`

Out[77]:

	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 [78]: `df`

Out[78]:

	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 [80]: `df.InternetUsers<2` *# here we are check the Internet users are <2*

Out[80]:

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 [82]: `df[df.InternetUsers<2]` *# here filtering the data set with df inside df is k*

Out[82]:

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

In [98]: `len(df[df.InternetUsers<2])`

Out[98]: 9

In [85]: `df[df.BirthRate<20]`

Out[85]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High 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
6	Armenia	ARM	13.308	41.9	Lower middle income
...
181	United States	USA	12.500	84.2	High income
183	St. Vincent and the Grenadines	VCT	16.306	52.0	Upper middle income
184	Venezuela, RB	VEN	19.842	54.9	High income
185	Virgin Islands (U.S.)	VIR	10.700	45.3	High income
186	Vietnam	VNM	15.537	43.9	Lower middle income

100 rows × 5 columns

In [99]: `len(df[df.BirthRate<20])`

Out[99]: 100


```
In [91]: Filter=df[df.InternetUsers<2]
Filter
```

```
Out[91]:
```

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

```
In [92]: Filter1=df[df.BirthRate>40]
Filter1
```

```
Out[92]:
```

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

```
In [95]: Filter=df.InternetUsers<2
```

```
In [96]: Filter1=df.BirthRate>40
```

```
In [97]: df[Filter&Filter1]
```

Out[97]:

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

In []: