

```
import pandas as pd
```

```
pd.__version__
```

```
↗ '2.2.2'
```

```
df = pd.read_csv(r"C:\Users\telan\Downloads\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

```
id(df)
```

```
↗ 1951990983680
```

```
type(df)
```

```
↗ pandas.core.frame.DataFrame
```

```
len(df)
```

⇒ 195

df.columns

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

len(df.columns)

⇒ 5

df.shape

⇒ (195, 5)

df.isnull()

⇒

	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

df.isna()



	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

```
df.isnull().sum()
```



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

```
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

```
df.tail()
```



	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

df.dtypes



```
CountryName      object
CountryCode      object
BirthRate        float64
InternetUsers    float64
IncomeGroup      object
dtype: object
```

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
```

#

slicing in dataframe

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

```
df[:: -1]
```



	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

df[:11]



	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9000	High income
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

df[0:200:50]



	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.900000	High income
50	Ecuador	ECU	21.070	40.353684	Upper middle income
100	Libya	LBY	21.425	16.500000	Upper middle income
150	Sudan	SDN	33.477	22.700000	Lower middle income

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

```
df[['CountryName','CountryCode','BirthRate']] # To get through attributes two bracket car
```




	CountryName	CountryCode	BirthRate
0	Aruba	ABW	10.244
1	Afghanistan	AFG	35.253
2	Angola	AGO	45.985
3	Albania	ALB	12.877
4	United Arab Emirates	ARE	11.044
...
190	Yemen, Rep.	YEM	32.947
191	South Africa	ZAF	20.850
192	Congo, Dem. Rep.	COD	42.394
193	Zambia	ZMB	40.471
194	Zimbabwe	ZWE	35.715

195 rows × 3 columns

```
df.head(10)
```



	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9000	High income
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

```
# descriptive statistics
```

```
df.describe()
```



	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

```
df.describe().T # row to columns and column to row
```



	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

```
df.columns
```



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

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



	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

```
df_cat.describe()
```



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

```
# Renaming columns of a dataframes  
df.head(1)
```



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

```
df.columns
```

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

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

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

```
➡
```

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

```
df.columns
```

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

```
df.columns = ['CountryName', 'CountyCode', 'BirthRate', 'InternetUser', 'IncomeGroup']  
df.head(1)
```

```
➡
```

	CountryName	CountyCode	BirthRate	InternetUser	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income

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

```
➡
```

	CountryName	CountyCode	BirthRate	InternetUser	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income

```
df_categoical = df[['CountryName', 'CountyCode', 'IncomeGroup']]  
df_categoical.head()
```

```
➡
```

	CountryName	CountyCode	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

```
df_categorical.describe()
```



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

```
df_num = df[['CountryName','BirthRate',]]  
df_num.head()
```



	CountryName	BirthRate
0	Aruba	10.244
1	Afghanistan	35.253
2	Angola	45.985
3	Albania	12.877
4	United Arab Emirates	11.044

```
df_num = df[['CountryName','BirthRate','IncomeGroup']]  
df_num.head()
```



	CountryName	BirthRate	IncomeGroup
0	Aruba	10.244	High income
1	Afghanistan	35.253	Low income
2	Angola	45.985	Upper middle income
3	Albania	12.877	Upper middle income
4	United Arab Emirates	11.044	High income

```
#Mathematical operation
```

```
df.BirthRate * df.InternetUser
```

```

↔ 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

```

```

df['myCalc'] = df.BirthRate * df.InternetUser
df

```

```

↔

```

	CountryName	CountyCode	BirthRate	InternetUser	IncomeGroup	myCalc
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

```
df.columns
```

```

↔ Index(['CountryName', 'CountyCode', 'BirthRate', 'InternetUser',
        'IncomeGroup',
        'myCalc'],
        dtype='object')

```

```
len(df.columns)
```



6

df



	CountryName	CountyCode	BirthRate	InternetUser	IncomeGroup	myCalc
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

df



	CountryName	CountyCode	BirthRate	InternetUser	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

```
# Add a column
df['myCalc'] = df.BirthRate * df.InternetUser
df
```




	CountryName	CountyCode	BirthRate	InternetUser	IncomeGroup	myCalc
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

df.columns



```
Index(['CountryName', 'CountyCode', 'BirthRate', 'InternetUser',
      'IncomeGroup',
      'myCalc'],
      dtype='object')
```

len(df.columns)



6

```
df = df.drop('myCalc',axis = 1)
```

df.columns



```
Index(['CountryName', 'CountyCode', 'BirthRate', 'InternetUser',
      'IncomeGroup'],
      dtype='object')
```

df.shape



(195, 5)

```
df.InternetUser<2 # we are checking given condition if its corrct true or false
```

```
⇒ 0      False
   1      False
   2      False
   3      False
   4      False
   ...
  190     False
  191     False
  192     False
  193     False
  194     False
   Name: InternetUser, Length: 195, dtype: bool
```

```
Filter = df.InternetUser <2
Filter
```

```
⇒ 0      False
   1      False
   2      False
   3      False
   4      False
   ...
  190     False
  191     False
  192     False
  193     False
  194     False
   Name: InternetUser, Length: 195, dtype: bool
```

```
df[Filter]
```

```
⇒
```

	CountryName	CountyCode	BirthRate	InternetUser	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

```
len(df[Filter])
```

➡ 9

df

➡

	CountryName	CountyCode	BirthRate	InternetUser	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

df.BirthRate>40

➡

```
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
```

Filter2 = df.BirthRate>40

Filter2

➡

```
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

```

```
df[Filter2]
```



	CountryName	CountyCode	BirthRate	InternetUser	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

```
len(df[Filter2])
```



```
12
```

```
df[Filter2]
```



	CountryName	CountyCode	BirthRate	InternetUser	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

#Filter and Filter2
Filter & Filter2



```
0      False
1      False
2      False
3      False
4      False
...
190    False
191    False
192    False
193    False
194    False
Length: 195, dtype: bool
```

df[Filter & Filter2]



	CountryName	CountyCode	BirthRate	InternetUser	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

```
df[(df.BirthRate
```

```
df[df.IncomeGroup == 'High income']
```

↗

	CountryName	CountyCode	BirthRate	InternetUser	IncomeGroup
0	Aruba	ABW	10.244	78.90	High income
4	United Arab Emirates	ARE	11.044	88.00	High income
5	Argentina	ARG	17.716	59.90	High income
7	Antigua and Barbuda	ATG	16.447	63.40	High income
8	Australia	AUS	13.200	83.00	High income
...
174	Trinidad and Tobago	TTO	14.590	63.80	High income
180	Uruguay	URY	14.374	57.69	High income
181	United States	USA	12.500	84.20	High income
184	Venezuela, RB	VEN	19.842	54.90	High income
185	Virgin Islands (U.S.)	VIR	10.700	45.30	High income

67 rows × 5 columns

```
df.IncomeGroup.unique() # get categories
```

↗

```
array(['High income', 'Low income', 'Upper middle income',  
      'Lower middle income'], dtype=object)
```

```
df.IncomeGroup.nunique() # get number
```

↗ 4

```
# Intoduction to seaborn # seaborn is very powerfull visulization(Statistic visuliaizator  
import matplotlib.pyplot as plt # visulaization  
import seaborn as sns # distribution  
# seaborn are used for advance visualization e.x --> distribution plot, line polt
```

```
%matplotlib inline  
plt.rcParams['figure.figsize'] = 6,2
```

```
import warnings  
warnings.filterwarnings('ignore') # os error
```

```
df.head()
```



	CountryName	CountyCode	BirthRate	InternetUser	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income
1	Afghanistan	AFG	35.253	5.9	Low income