```
In [93]: import pandas as pd  # USE FOR DATAFRAMES
```

```
In [94]: pd.__version__
```

```
Out[94]: '1.4.2'
```

```
In [95]: df = pd.read_csv('C:/Users/sidra/Downloads/data - data.csv')
```

```
In [96]: df
```

Out[96]:

	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 [97]: len(df)  # it will give you records
```

```
Out[97]: 195
```

```
In [98]: id(df)
```

```
Out[98]: 2404955096400
```

```
In [99]: df.shape  # rows * columns
```

```
Out[99]: (195, 5)
```

```
In [100]: df.columns
```

```
Out[100]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
            'IncomeGroup'],
          dtype='object')
```

```
In [101]: len(df.columns)
```

```
Out[101]: 5
```

```
In [102]: df.isnull()
```

Out[102]:

	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 [103]: df.isnull().sum()
```

```
Out[103]: CountryName      0
          CountryCode      0
          BirthRate        0
          InternetUsers    0
          IncomeGroup      0
          dtype: int64
```

```
In [104]: 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 [105]: len(df)
```

```
Out[105]: 195
```

```
In [106]: len(df.columns)
```

```
Out[106]: 5
```

```
In [107]: df.shape
```

Out[107]: (195, 5)

```
In [108]: df.head()
```

Out[108]:

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

Out[109]:

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

```
In [110]: df.tail()
```

Out[110]:

	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 [111]: df.tail(2)
```

Out[111]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
193	Zambia	ZMB	40.471	15.4	Lower middle income
194	Zimbabwe	ZWE	35.715	18.5	Low income

```
In [112]: df.dtypes
```

Out[112]: CountryName   object
          CountryCode   object
          BirthRate   float64
          InternetUsers   float64
          IncomeGroup   object
          dtype: object

```
In [113]: df.columns

Out[113]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
                  'IncomeGroup'],
                dtype='object')


In [114]: df['CountryName']

Out[114]: 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 [115]: df['BirthRate']

Out[115]: 0        10.244
          1        35.253
          2        45.985
          3        12.877
          4        11.044
                    ...
          190      32.947
          191      20.850
          192      42.394
          193      40.471
          194      35.715
          Name: BirthRate, Length: 195, dtype: float64


In [116]: df['InternetUsers']

Out[116]: 0        78.9
          1         5.9
          2        19.1
          3        57.2
          4        88.0
                   ...
          190      20.0
          191      46.5
          192       2.2
          193      15.4
          194      18.5
          Name: InternetUsers, Length: 195, dtype: float64
```

```
In [117]: df_cat = df[['CountryName', 'CountryCode', 'IncomeGroup']]
          df_cat
```

Out[117]:

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

Out[119]:

	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 [120]: print(df.shape)      # 5 columns
          print(df_cat.shape)  # 3 categorical columns
          print(df_num.shape)  # 2 numerical columns
```

```
(195, 5)
(195, 3)
(195, 2)
```

```python
In [121]: df[:]
```

Out[121]:

	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

```python
In [122]: df[:2]
```

Out[122]:

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

```python
In [123]: df[1:200:20]
```

Out[123]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
1	Afghanistan	AFG	35.253	5.9000	Low income
21	Belize	BLZ	23.092	33.6000	Upper middle income
41	Cuba	CUB	10.400	27.9300	Upper middle income
61	United Kingdom	GBR	12.200	89.8441	High income
81	Ireland	IRL	15.000	78.2477	High income
101	St. Lucia	LCA	15.430	46.2000	Upper middle income
121	Mauritania	MRT	33.801	6.2000	Lower middle income
141	Puerto Rico	PRI	10.800	73.9000	High income
161	Slovak Republic	SVK	10.100	77.8826	High income
181	United States	USA	12.500	84.2000	High income

```
In [124]: df[::-1]
```

Out[124]:

	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

```python
In [125]: df[::-5]
```

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
194	Zimbabwe	ZWE	35.715	18.5000	Low income
189	Samoa	WSM	26.172	15.3000	Lower middle income
184	Venezuela, RB	VEN	19.842	54.9000	High income
179	Ukraine	UKR	11.100	41.0000	Lower middle income
174	Trinidad and Tobago	TTO	14.590	63.8000	High income
169	Thailand	THA	11.041	28.9400	Upper middle income
164	Swaziland	SWZ	30.093	24.7000	Lower middle income
159	Sao Tome and Principe	STP	34.537	23.0000	Lower middle income
154	Sierra Leone	SLE	36.729	1.7000	Low income
149	Saudi Arabia	SAU	20.576	60.5000	High income
144	French Polynesia	PYF	16.393	56.8000	High income
139	Papua New Guinea	PNG	28.899	6.5000	Lower middle income
134	Oman	OMN	20.419	66.4500	High income
129	Nicaragua	NIC	20.788	15.5000	Lower middle income
124	Malaysia	MYS	16.805	66.9700	Upper middle income
119	Mongolia	MNG	24.275	20.0000	Upper middle income
114	Macedonia, FYR	MKD	11.222	65.2400	Upper middle income
109	Morocco	MAR	21.023	56.0000	Lower middle income
104	Lesotho	LSO	28.738	5.0000	Lower middle income
99	Liberia	LBR	35.521	3.2000	Low income
94	Kiribati	KIR	29.044	11.5000	Lower middle income
89	Japan	JPN	8.200	89.7100	High income
84	Iceland	ISL	13.400	96.5468	High income
79	Indonesia	IDN	20.297	14.9400	Lower middle income
74	Hong Kong SAR, China	HKG	7.900	74.2000	High income
69	Grenada	GRD	19.334	35.0000	Upper middle income
64	Guinea	GIN	37.337	1.6000	Low income
59	Micronesia, Fed. Sts.	FSM	23.511	27.8000	Lower middle income
54	Estonia	EST	10.300	79.4000	High income
49	Algeria	DZA	24.738	16.5000	Upper middle income
44	Czech Republic	CZE	10.200	74.1104	High income
39	Cabo Verde	CPV	21.625	37.5000	Lower middle income
34	Cote d'Ivoire	CIV	37.320	8.4000	Lower middle income
29	Central African Republic	CAF	34.076	3.5000	Low income
24	Brazil	BRA	14.931	51.0400	Upper middle income
19	Bosnia and Herzegovina	BIH	9.062	57.7900	Upper middle income
14	Burkina Faso	BFA	40.551	9.1000	Low income
9	Austria	AUT	9.400	80.6188	High income

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
4	United Arab Emirates	ARE	11.044	88.0000	High income

In [126]: `df.describe()`

Out[126]:

	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 [127]: `df.describe().transpose()`

Out[127]:

	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 [129]: `df_num.describe()`

Out[129]:

	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 [130]: `df_cat.describe()`

Out[130]:

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

```
In [131]: df.head(1)
```

Out[131]:

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

```
In [132]: df.columns = ['a','b','c','d','e']
```

```
In [133]: df.head(1)
```

Out[133]:

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

```
In [134]: df.columns = ['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
              'IncomeGroup']
```

```
In [135]: df.head(1)
```

Out[135]:

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

```
In [136]: df.BirthRate * df.InternetUsers
```

Out[136]:
```
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 [137]: df.head(1)
```

Out[137]:

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

```
In [138]: df['myCalc'] = df.BirthRate * df.InternetUsers
```

```
In [139]: df.head(1)
```

Out[139]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	myCalc
0	Aruba	ABW	10.244	78.9	High income	808.2516

```
In [140]: df = df.drop('myCalc',axis = 1)
```

```
In [141]: df
```

Out[141]:

	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 [142]: df
```

Out[142]:

	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 [143]: df['InternetUsers'] < 2
```

```
Out[143]: 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 [144]: df[df['InternetUsers'] < 2]
```

Out[144]:

	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 [145]: len(df[df['InternetUsers'] < 2])
```

```
Out[145]: 9
```

```
In [146]: filter = df[df['InternetUsers'] < 2]
          filter
```

Out[146]:

	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 [147]: filter1 = df[df['BirthRate'] < 40]
          filter1
```

Out[147]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income
1	Afghanistan	AFG	35.253	5.9	Low 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
...	...	...	...	...	...
188	West Bank and Gaza	PSE	30.394	46.6	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
194	Zimbabwe	ZWE	35.715	18.5	Low income

183 rows × 5 columns

```
In [148]: df(df'InternetUsers' < 2) & (df'BirthRate' < 40)
```

```
  Input In [148]
    df(df'InternetUsers' < 2) & (df'BirthRate' < 40)
         ^
SyntaxError: invalid syntax
```

```
In [149]: df.head(1)
```

Out[149]:

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

```
In [150]: df[df.IncomeGroup == 'Low income'].head()
```

Out[150]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
1	Afghanistan	AFG	35.253	5.9	Low income
11	Burundi	BDI	44.151	1.3	Low income
13	Benin	BEN	36.440	4.9	Low income
14	Burkina Faso	BFA	40.551	9.1	Low income
29	Central African Republic	CAF	34.076	3.5	Low income

```
In [151]: df[df.IncomeGroup == 'High income'].head()
```

Out[151]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income
4	United Arab Emirates	ARE	11.044	88.0	High income
5	Argentina	ARG	17.716	59.9	High income
7	Antigua and Barbuda	ATG	16.447	63.4	High income
8	Australia	AUS	13.200	83.0	High income

```
In [152]: df.IncomeGroup.unique()
```

```
Out[152]: array(['High income', 'Low income', 'Upper middle income',
                 'Lower middle income'], dtype=object)
```

```
In [153]: df.IncomeGroup.nunique()
```

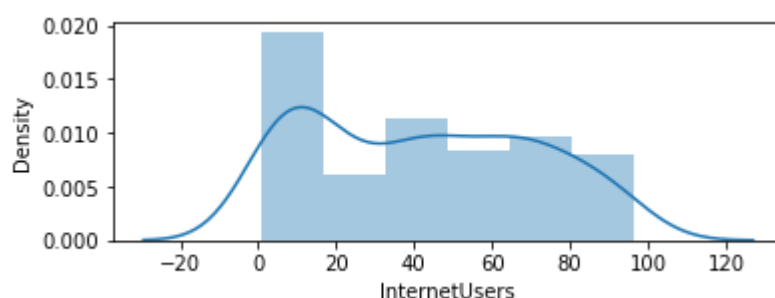```
Out[153]: 4
```

```
In [154]: import matplotlib.pyplot as plt  # visualization
          import seaborn as sns

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

          import warnings
          warnings.filterwarnings('ignore')
```

```
In [155]: df['InternetUsers']
```

```
Out[155]: 0        78.9
          1         5.9
          2        19.1
          3        57.2
          4        88.0
                   ...
          190      20.0
          191      46.5
          192       2.2
          193      15.4
          194      18.5
          Name: InternetUsers, Length: 195, dtype: float64
```

```
In [156]: vis1 = sns.distplot(df['InternetUsers'])
          plt.show()
```

In [ ]: