

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

DATA FRAME IN PYTHON AND HOW TO IMPORT THE DATASET

Pandas are very good package for dataframes & its perfect for dataset & very powerfull packages

```
In [385... # how to read the dataset

df = pd.read_csv(r'C:\Users\PC\Downloads\data.csv') # import the dataset
df
```

```
Out[385...

```

	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 [126... # explore the data in python
# 1. Full dataframe
# 2. how many rows & columns. you have tocheck the row because the no. of row sh

len(df) # 195 rows imported(this is for tracking later part)
```

```
Out[126... 195
```

```
In [128... df.shape
```

```
Out[128... (195, 5)
```

In [130...] `df.columns # see columns`

Out[130...] `Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
'IncomeGroup'],
 dtype='object')`

In [132...] `type(df)`

Out[132...] `pandas.core.frame.DataFrame`

In [134...] `df`

Out[134...]

	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 [158...] `#information of the column`

`df.info() # strings are called as object`

```
<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 [160... `len(df.columns)`

Out[160... 5

In [162... `df.head()`

Out[162...

	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 [164... `df.tail()`

Out[164...

	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 [166... `df.tail(2) # last 2 rows`

Out[166...

	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 [168... `df.head(2) # to 2 rows`

Out[168...

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

In [170... `df`

Out[170...

	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 [172...

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

Out[172...

	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 [174...

```
df[:,5]
```

Out[174...

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.900000	High income
5	Argentina	ARG	17.716	59.900000	High income
10	Azerbaijan	AZE	18.300	58.700000	Upper middle income
15	Bangladesh	BGD	20.142	6.630000	Lower middle income
20	Belarus	BLR	12.500	54.170000	Upper middle income
25	Barbados	BRB	12.188	73.000000	High income
30	Canada	CAN	10.900	85.800000	High income
35	Cameroon	CMR	37.236	6.400000	Lower middle income
40	Costa Rica	CRI	15.022	45.960000	Upper middle income
45	Germany	DEU	8.500	84.170000	High income
50	Ecuador	ECU	21.070	40.353684	Upper middle income
55	Ethiopia	ETH	32.925	1.900000	Low income
60	Gabon	GAB	30.555	9.200000	Upper middle income
65	Gambia, The	GMB	42.525	14.000000	Low income
70	Greenland	GRL	14.500	65.800000	High income
75	Honduras	HND	21.593	17.800000	Lower middle income
80	India	IND	20.291	15.100000	Lower middle income
85	Israel	ISR	21.300	70.800000	High income
90	Kazakhstan	KAZ	22.730	54.000000	Upper middle income
95	Korea, Rep.	KOR	8.600	84.770000	High income
100	Libya	LBY	21.425	16.500000	Upper middle income
105	Lithuania	LTU	10.100	68.452900	High income
110	Moldova	MDA	12.141	45.000000	Lower middle income
115	Mali	MLI	44.138	3.500000	Low income
120	Mozambique	MOZ	39.705	5.400000	Low income
125	Namibia	NAM	29.937	13.900000	Upper middle income
130	Netherlands	NLD	10.200	93.956400	High income
135	Pakistan	PAK	29.582	10.900000	Lower middle income
140	Poland	POL	9.600	62.849200	High income
145	Qatar	QAT	11.940	85.300000	High income
150	Sudan	SDN	33.477	22.700000	Lower middle income
155	El Salvador	SLV	17.476	23.109300	Lower middle income
160	Suriname	SUR	18.455	37.400000	Upper middle income

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
165	Seychelles	SYC	18.600	50.400000	High income
170	Tajikistan	TJK	30.792	16.000000	Lower middle income
175	Tunisia	TUN	19.800	43.800000	Upper middle income
180	Uruguay	URY	14.374	57.690000	High income
185	Virgin Islands (U.S.)	VIR	10.700	45.300000	High income
190	Yemen, Rep.	YEM	32.947	20.000000	Lower middle income

In [176...

```
df[6:]
```

Out[176...

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
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
...
190	Yemen, Rep.	YEM	32.947	20.0000	Lower middle income
191	South Africa	ZAF	20.850	46.5000	Upper middle income
192	Congo, Dem. Rep.	COD	42.394	2.2000	Low income
193	Zambia	ZMB	40.471	15.4000	Lower middle income
194	Zimbabwe	ZWE	35.715	18.5000	Low income

189 rows × 5 columns

In [178...

```
df[0:200:10]
```

Out[178...

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.900000	High income
10	Azerbaijan	AZE	18.300	58.700000	Upper middle income
20	Belarus	BLR	12.500	54.170000	Upper middle income
30	Canada	CAN	10.900	85.800000	High income
40	Costa Rica	CRI	15.022	45.960000	Upper middle income
50	Ecuador	ECU	21.070	40.353684	Upper middle income
60	Gabon	GAB	30.555	9.200000	Upper middle income
70	Greenland	GRL	14.500	65.800000	High income
80	India	IND	20.291	15.100000	Lower middle income
90	Kazakhstan	KAZ	22.730	54.000000	Upper middle income
100	Libya	LBY	21.425	16.500000	Upper middle income
110	Moldova	MDA	12.141	45.000000	Lower middle income
120	Mozambique	MOZ	39.705	5.400000	Low income
130	Netherlands	NLD	10.200	93.956400	High income
140	Poland	POL	9.600	62.849200	High income
150	Sudan	SDN	33.477	22.700000	Lower middle income
160	Suriname	SUR	18.455	37.400000	Upper middle income
170	Tajikistan	TJK	30.792	16.000000	Lower middle income
180	Uruguay	URY	14.374	57.690000	High income
190	Yemen, Rep.	YEM	32.947	20.000000	Lower middle income

In [180...

```
df.describe() #gives only numerical info/ it will work like statistics
```

Out[180...

	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 [182...

```
df.describe().transpose() # transpose convert column into rows
```


	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 [63]: `df.columns`

Out[63]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers', 'IncomeGroup'], dtype='object')

In [188... *# renaming columns of a dataframe*

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

	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 [186... `df.head(1)`

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

In [190... `df.columns`

Out[190... Index(['a', 'b', 'c', 'd', 'e'], dtype='object')

In [192... `df.columns = ['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers', 'IncomeGroup']`

In [198... `df.head(1)`

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

In [200... *# subsetting a dataframes in pandas*

```
#1 rows
#2. columns
#3. combine the two
```

In [202... *# rows:*

```
df[21:26] # how python know that only this is rows based on index
```

Out[202...

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
21	Belize	BLZ	23.092	33.60	Upper middle income
22	Bermuda	BMU	10.400	95.30	High income
23	Bolivia	BOL	24.236	36.94	Lower middle income
24	Brazil	BRA	14.931	51.04	Upper middle income
25	Barbados	BRB	12.188	73.00	High income

In [204...

```
df[:]
```

Out[204...

	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 [206...

```
df[:10]
```

Out[206...

	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

In [208...

df.head(10)

Out[208...

	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

In [210...

how to reverse the dataframe

df[::-1]

Out[210...

	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 [212...

df

Out[212...

	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 [214...

```
# get only every 20th row
df[::20]
```

Out[214...

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9000	High income
20	Belarus	BLR	12.500	54.1700	Upper middle income
40	Costa Rica	CRI	15.022	45.9600	Upper middle income
60	Gabon	GAB	30.555	9.2000	Upper middle income
80	India	IND	20.291	15.1000	Lower middle income
100	Libya	LBY	21.425	16.5000	Upper middle income
120	Mozambique	MOZ	39.705	5.4000	Low income
140	Poland	POL	9.600	62.8492	High income
160	Suriname	SUR	18.455	37.4000	Upper middle income
180	Uruguay	URY	14.374	57.6900	High income

In [216...

```
# columns:
df.columns
```

```
Out[216... Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
      'IncomeGroup'],
      dtype='object')
```

```
In [218... df.head() # top five rows
```

```
Out[218...
      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 [220... df['CountryName'].head()
```

```
Out[220...
0           Aruba
1  Afghanistan
2           Angola
3           Albania
4  United Arab Emirates
Name: CountryName, dtype: object
```

```
In [222... ['CountryName', 'BirthRate']
```

```
Out[222... ['CountryName', 'BirthRate']
```

```
In [224... df[['CountryName', 'BirthRate']].head()
```

```
Out[224...
      CountryName  BirthRate
0           Aruba    10.244
1  Afghanistan    35.253
2           Angola    45.985
3           Albania    12.877
4  United Arab Emirates    11.044
```

```
In [226... df.head()
```

```
Out[226...
      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 [228...

```
df['BirthRate']
```

Out[228...

```
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 [230...

```
# combine the two

df[4:8][['CountryName', 'BirthRate']]
```

Out[230...

	CountryName	BirthRate
4	United Arab Emirates	11.044
5	Argentina	17.716
6	Armenia	13.308
7	Antigua and Barbuda	16.447

In [232...

```
df[['CountryName', 'BirthRate']][4:8]
```

Out[232...

	CountryName	BirthRate
4	United Arab Emirates	11.044
5	Argentina	17.716
6	Armenia	13.308
7	Antigua and Barbuda	16.447

In [234...

```
df1 = df[['CountryName', 'BirthRate']]
```

In [236...

```
df1
```

Out[236...

	CountryName	BirthRate
0	Aruba	10.244
1	Afghanistan	35.253
2	Angola	45.985
3	Albania	12.877
4	United Arab Emirates	11.044
...
190	Yemen, Rep.	32.947
191	South Africa	20.850
192	Congo, Dem. Rep.	42.394
193	Zambia	40.471
194	Zimbabwe	35.715

195 rows × 2 columns

In [238...

df2 = df[4:8]

In [240...

df2

Out[240...

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
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
7	Antigua and Barbuda	ATG	16.447	63.4	High income

In [242...

```
# Basic operation of dataframe
df.head()
```

Out[242...

	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 [244...

df[['CountryName', 'BirthRate', 'InternetUsers']][4:8] # subset dataframe

Out[244...

	CountryName	BirthRate	InternetUsers
4	United Arab Emirates	11.044	88.0
5	Argentina	17.716	59.9
6	Armenia	13.308	41.9
7	Antigua and Barbuda	16.447	63.4

In [387...

```
df[['CountryName', 'BirthRate', 'InternetUsers']]
```

Out[387...

	CountryName	BirthRate	InternetUsers
0	Aruba	10.244	78.9
1	Afghanistan	35.253	5.9
2	Angola	45.985	19.1
3	Albania	12.877	57.2
4	United Arab Emirates	11.044	88.0
...
190	Yemen, Rep.	32.947	20.0
191	South Africa	20.850	46.5
192	Congo, Dem. Rep.	42.394	2.2
193	Zambia	40.471	15.4
194	Zimbabwe	35.715	18.5

195 rows × 3 columns

In [246...

```
df.head()
```

Out[246...

	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 [248...

```
# Mathematical operation::
df.BirthRate * df.InternetUsers
```

```
Out[248... 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 [252... # ADD a Column

df['myCalc'] = df.BirthRate * df.InternetUsers
```

```
In [254... df.head()
```

```
Out[254...   CountryName  CountryCode  BirthRate  InternetUsers  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  878.3135
   income
3      Albania        ALB      12.877         57.2  Upper middle  736.5644
   income
4  United Arab      ARE      11.044         88.0      High income  971.8720
   Emirates
```

```
In [256... # REMOVE A COLUMN

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

Out[256...

	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 [258...

`df.columns[2]`

Out[258...

'BirthRate'

In [266...

`df.InternetUsers < 2 # we are checking given condition if its correct true or False`

Out[266...

```

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 [268...

`Filter = df.InternetUsers < 2`

Filter

In [270...

`df[3:7]`

Out[270...

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	myCalc
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
5	Argentina	ARG	17.716	59.9	High income	1061.1884
6	Armenia	ARM	13.308	41.9	Lower middle income	557.6052

In [272...

df[30:40]

Out[272...

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	myCalc
30	Canada	CAN	10.900	85.80	High income	935.2200
31	Switzerland	CHE	10.200	86.34	High income	880.6680
32	Chile	CHL	13.385	66.50	High income	890.1025
33	China	CHN	12.100	45.80	Upper middle income	554.1800
34	Cote d'Ivoire	CIV	37.320	8.40	Lower middle income	313.4880
35	Cameroon	CMR	37.236	6.40	Lower middle income	238.3104
36	Congo, Rep.	COG	37.011	6.60	Lower middle income	244.2726
37	Colombia	COL	16.076	51.70	Upper middle income	831.1292
38	Comoros	COM	34.326	6.50	Low income	223.1190
39	Cabo Verde	CPV	21.625	37.50	Lower middle income	810.9375

In [274...

df[Filter] # it will take that row which are false

Out [274...

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	myCalc
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 [276...

```
df.BirthRate > 40
```

Out [276...

```
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 [282...

```
Filter2 = df.BirthRate > 40
```

Filter2

In [284...

```
df[Filter2]
```

Out[284...

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	myCalc
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 [286...

```
# Filter & Filter2

Filter & Filter2
```

Out[286...

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

In [288...

```
df[Filter & Filter2]
```

Out[288...

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	myCalc
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 [292...

```
df[(df.BirthRate > 40) & (df.InternetUsers < 2)]
```

Out[292...

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	myCalc
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 [294...

df.head()

Out[294...

	CountryName	CountryCode	BirthRate	InternetUsers	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

In [296...

df[df.IncomeGroup == 'Low income']

Out[296...

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	myCalc
1	Afghanistan	AFG	35.253	5.90	Low income	207.99270
11	Burundi	BDI	44.151	1.30	Low income	57.39630
13	Benin	BEN	36.440	4.90	Low income	178.55600
14	Burkina Faso	BFA	40.551	9.10	Low income	369.01410
29	Central African Republic	CAF	34.076	3.50	Low income	119.26600
38	Comoros	COM	34.326	6.50	Low income	223.11900
52	Eritrea	ERI	34.800	0.90	Low income	31.32000
55	Ethiopia	ETH	32.925	1.90	Low income	62.55750
64	Guinea	GIN	37.337	1.60	Low income	59.73920
65	Gambia, The	GMB	42.525	14.00	Low income	595.35000
66	Guinea-Bissau	GNB	37.503	3.10	Low income	116.25930
77	Haiti	HTI	25.345	10.60	Low income	268.65700
93	Cambodia	KHM	24.462	6.80	Low income	166.34160
99	Liberia	LBR	35.521	3.20	Low income	113.66720
111	Madagascar	MDG	34.686	3.00	Low income	104.05800
115	Mali	MLI	44.138	3.50	Low income	154.48300
120	Mozambique	MOZ	39.705	5.40	Low income	214.40700
123	Malawi	MWI	39.459	5.05	Low income	199.26795
127	Niger	NER	49.661	1.70	Low income	84.42370
132	Nepal	NPL	20.923	13.30	Low income	278.27590
148	Rwanda	RWA	32.689	9.00	Low income	294.20100
154	Sierra Leone	SLE	36.729	1.70	Low income	62.43930
156	Somalia	SOM	43.891	1.50	Low income	65.83650
158	South Sudan	SSD	37.126	14.10	Low income	523.47660
167	Chad	TCD	45.745	2.30	Low income	105.21350
168	Togo	TGO	36.080	4.50	Low income	162.36000
177	Tanzania	TZA	39.518	4.40	Low income	173.87920
178	Uganda	UGA	43.474	16.20	Low income	704.27880
192	Congo, Dem. Rep.	COD	42.394	2.20	Low income	93.26680
194	Zimbabwe	ZWE	35.715	18.50	Low income	660.72750

In [298... *# how to get the unique categories*

```
df.IncomeGroup.unique()
```

Out[298... array(['High income', 'Low income', 'Upper middle income',
'Lower middle income'], dtype=object)

In [300... *# introduction to seaborn*

seaborn is very poerfull visualization (STATISTIC VISUALIZATION) PACKAGE IN P

```
import matplotlib.pyplot as plt # visualization
import seaborn as sns # distribution visualization
```

```
%matplotlib inline
plt.rcParams['figure.figsize'] = 8,4
```

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

In [302... df.head()

	CountryName	CountryCode	BirthRate	InternetUsers	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

In [306... *# Distribution*

```
vis1 = sns.distplot(df["InternetUsers"])
plt.show()
```

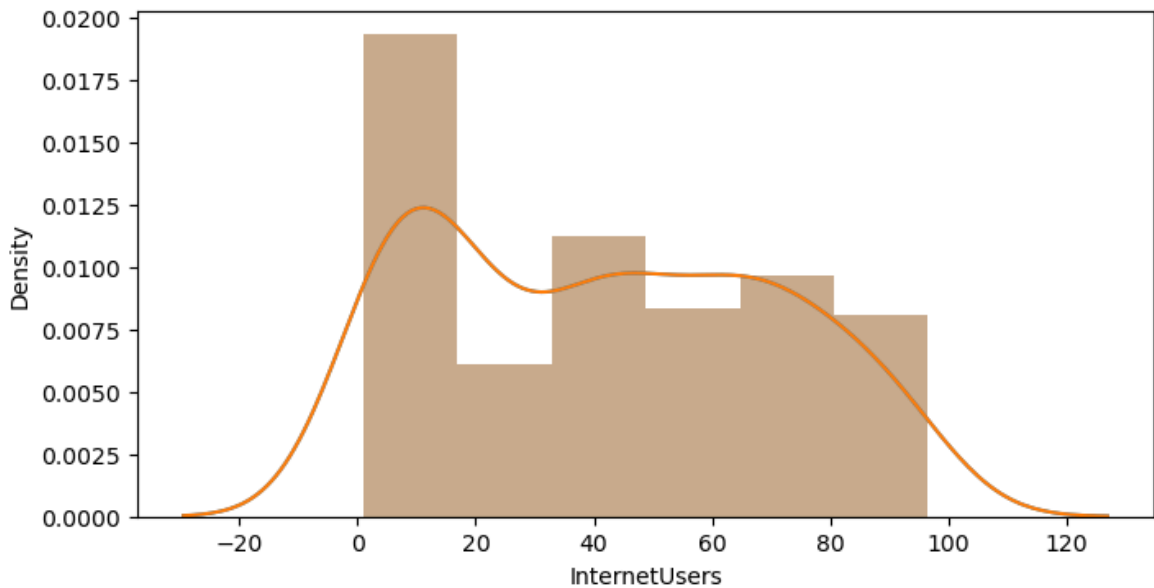
C:\Users\PC\AppData\Local\Temp\ipykernel_12404\1630422380.py:2: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
vis1 = sns.distplot(df["InternetUsers"])
```



```
In [310... vis1 = sns.distplot(df["InternetUsers"], bins=10)
plt.show()
```

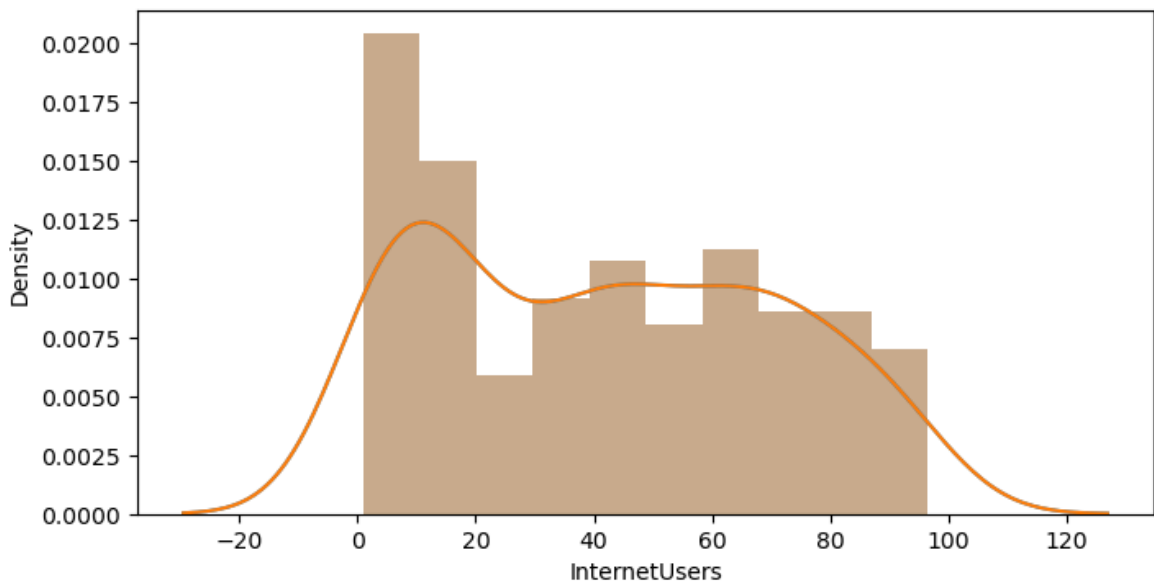
C:\Users\PC\AppData\Local\Temp\ipykernel_12404\2950760178.py:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

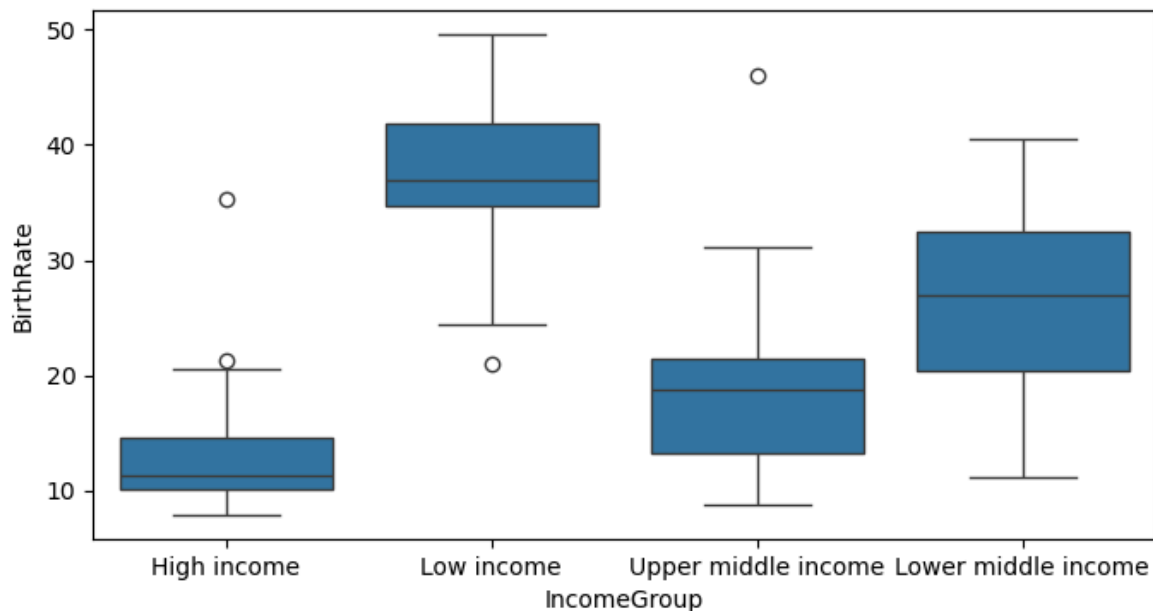
Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
vis1 = sns.distplot(df["InternetUsers"], bins=10)
```



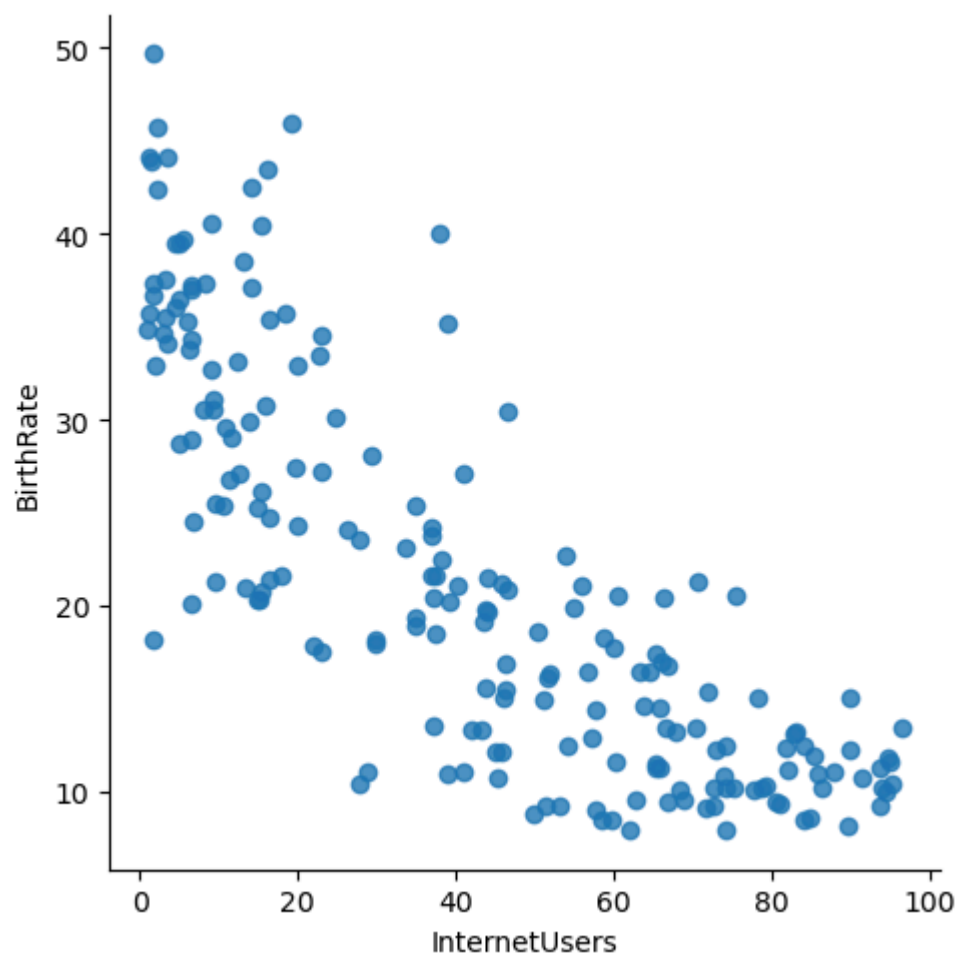
```
In [312... #BOX PLOTS::
vis2 = sns.boxplot(data = df, x = "IncomeGroup", y = 'BirthRate')
plt.show()
```

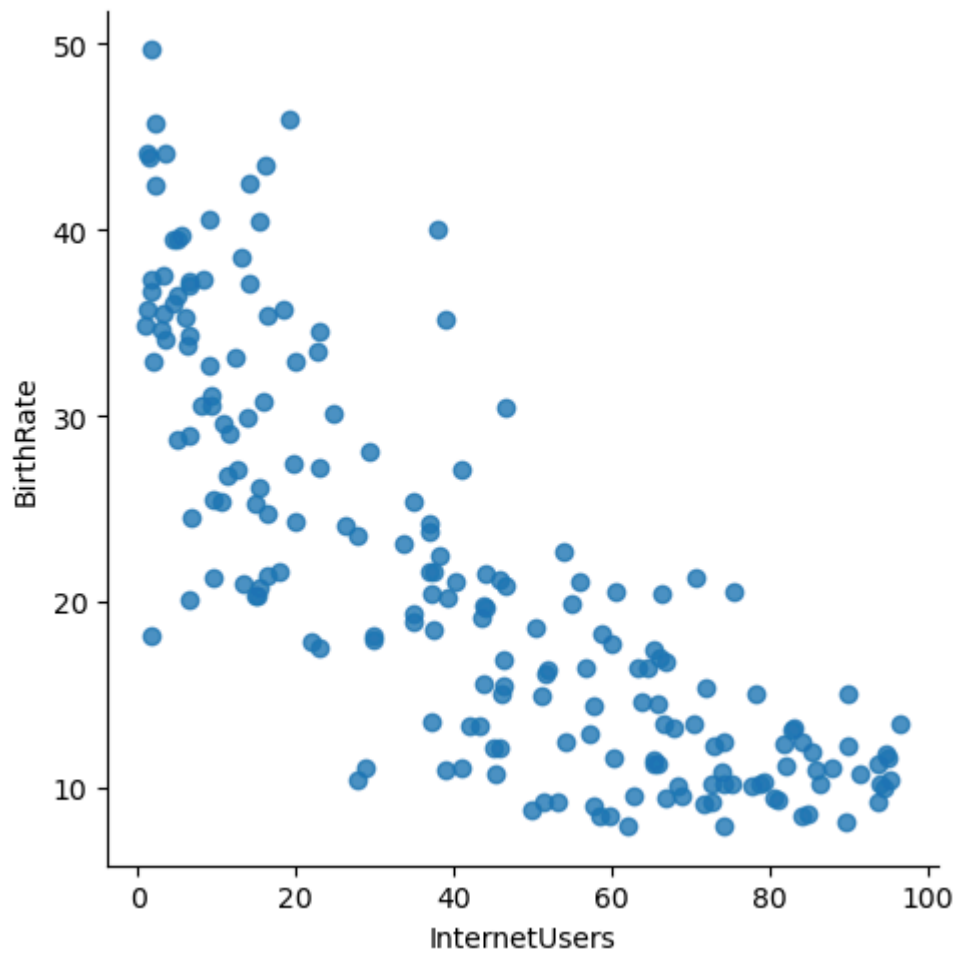


```
In [314... # refer to seaborn gallery
```

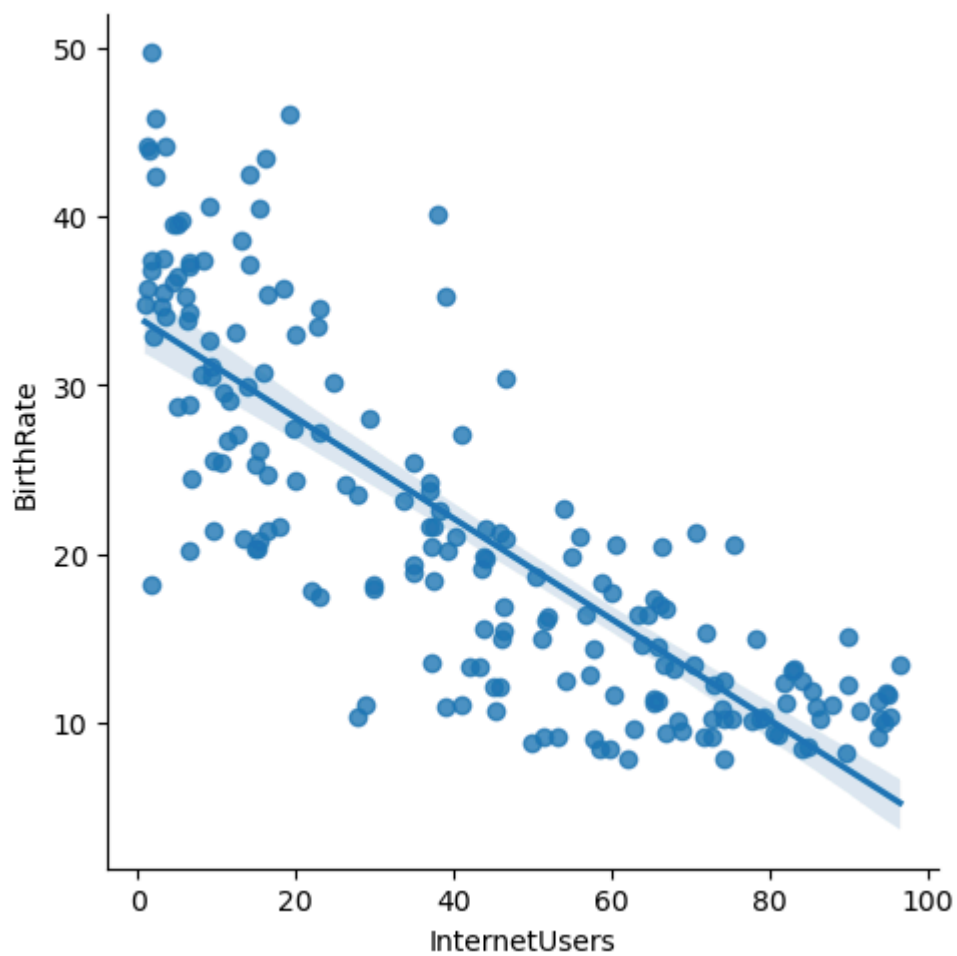
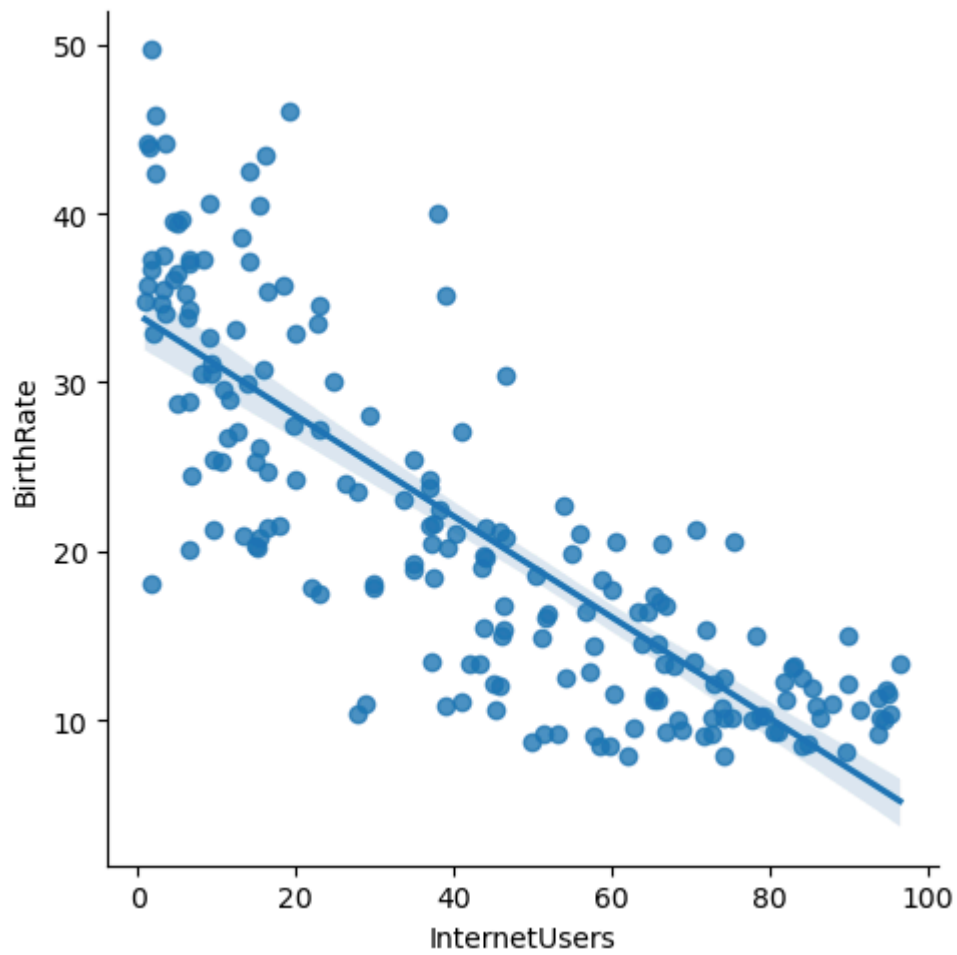
```
In [316... # visualization with seaborn
```

```
In [320... vis3 = sns.lmplot(data = df, x= 'InternetUsers', y = 'BirthRate', fit_reg= False  
plt.show()
```

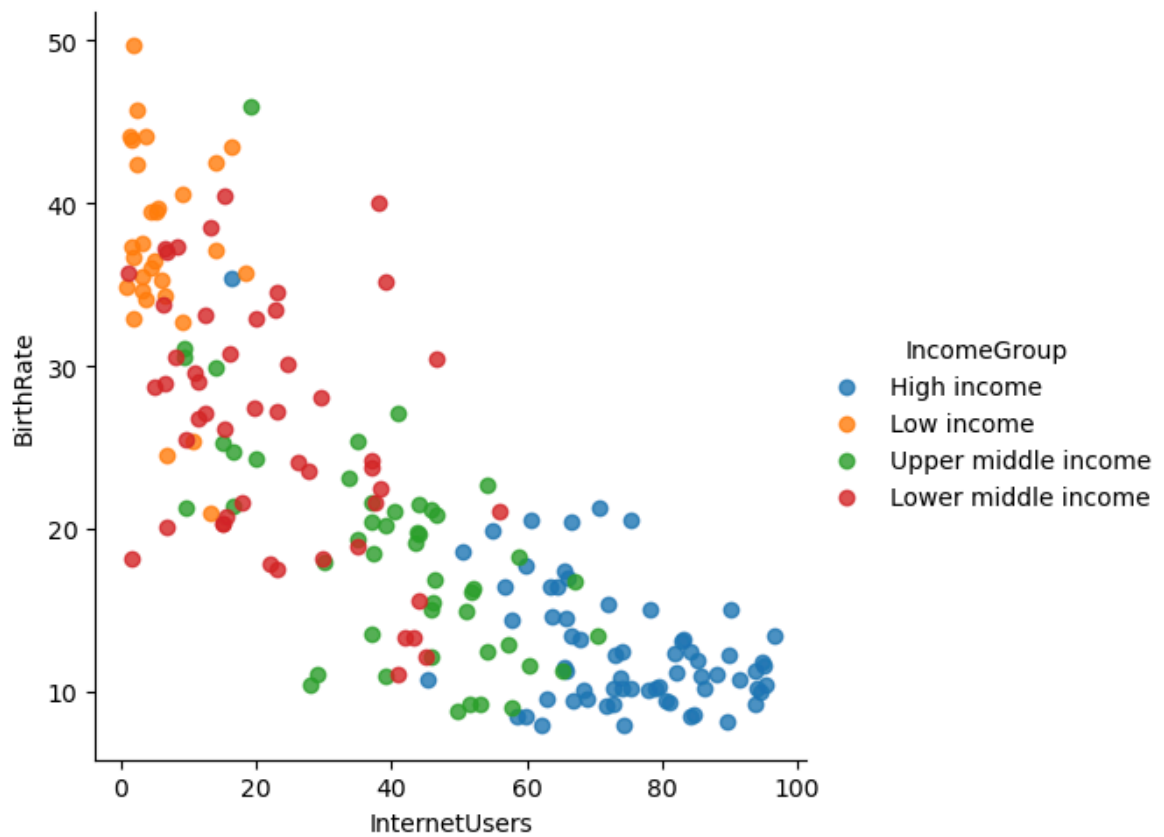




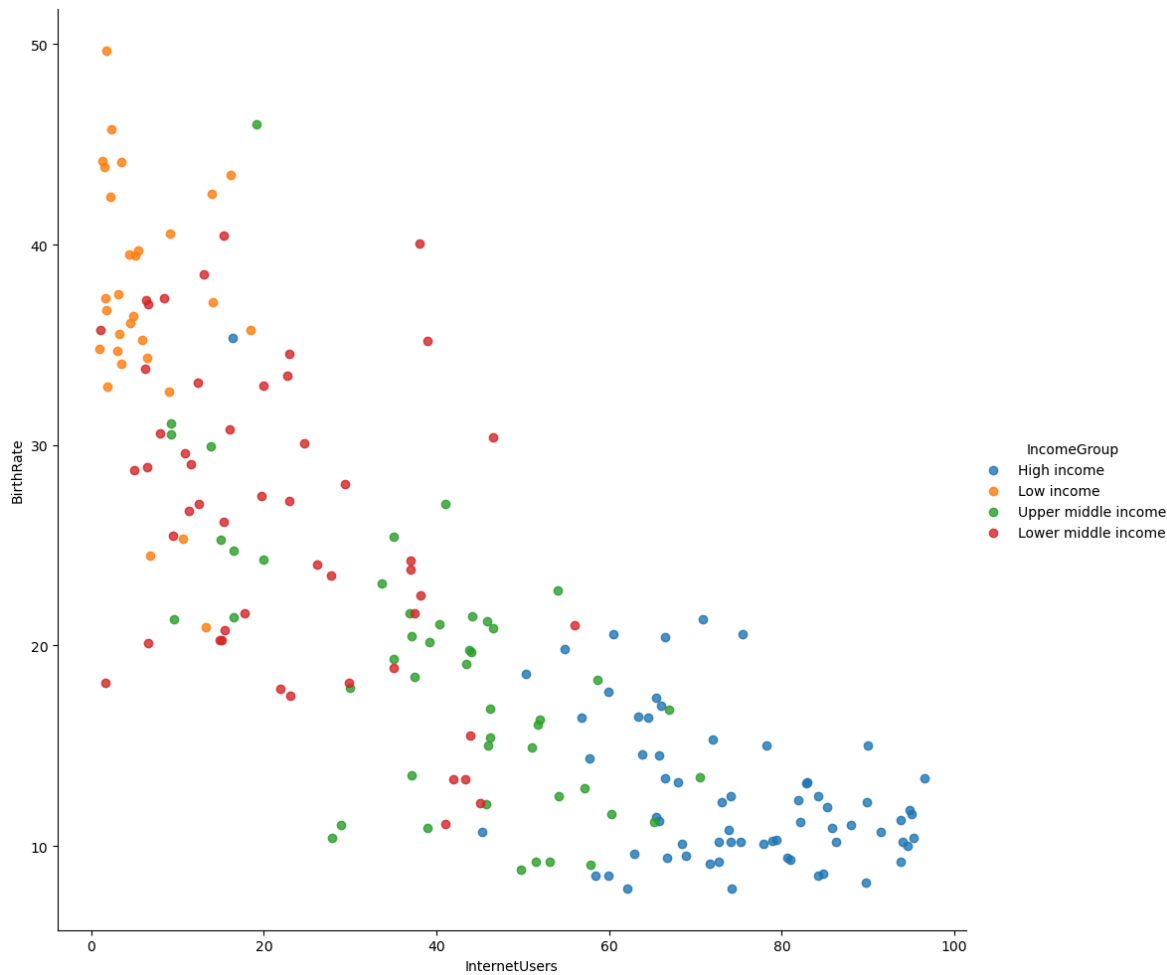
```
In [324... vis4= sns.lmplot(data = df, x= 'InternetUsers', y = 'BirthRate')  
plt.show()
```



```
In [328... vis5 = sns.lmplot(data = df, x= 'InternetUsers', y = 'BirthRate', fit_reg = False,
plt.show())
```



```
In [338... vis5 = sns.lmplot(data = df, x= 'InternetUsers', y = 'BirthRate', fit_reg = False,
plt.show())
```



In []:

In [340...

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

Out[340...

	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

In []:

In []:

```
In [344... df_categorical = df[['CountryName', 'CountryCode', 'BirthRate']]
df_categorical.head()
```

Out[344...]

	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

In [346...]

```
df.describe()
```

Out[346...]

	BirthRate	InternetUsers	myCalc
count	195.000000	195.000000	195.000000
mean	21.469928	42.076471	653.559009
std	10.605467	29.030788	351.553521
min	7.900000	0.900000	28.990400
25%	12.120500	14.520000	361.263300
50%	19.680000	41.000000	682.074300
75%	29.759500	66.225000	892.690170
max	49.661000	96.546800	1552.589500

In [348...]

```
df_categorical.describe()
```

Out[348...]

	BirthRate
count	195.000000
mean	21.469928
std	10.605467
min	7.900000
25%	12.120500
50%	19.680000
75%	29.759500
max	49.661000

In [350...

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

Out[350...

	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 [352...

```
df_num.describe()
```

Out[352...

	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 [354...

```
df.head()
```

Out [354...

	CountryName	CountryCode	BirthRate	InternetUsers	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

In [356...

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

Out [356...

```
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 [358...

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

Out [358...

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

In [360...

```
df['myCalc'] = df.BirthRate * df.InternetUsers
```

In [362...

```
df
```

Out[362...

	CountryName	CountryCode	BirthRate	InternetUsers	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

In [364...

```
df.columns
```

Out[364...

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

In [366...

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

Out[366...

6

In [368...

```
df
```

Out[368...

	CountryName	CountryCode	BirthRate	InternetUsers	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

In [370...

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

In [372...

```
df.columns
```

Out[372...

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

In [374...

```
df['InternetUsers']
```

Out[374...

```
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 [376...

```
df
```

Out[376...

	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 [378...

```
df.InternetUsers<2
```

Out[378...

```
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 [380...

```
df
```

Out[380...

	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 []:

In []:

In []:

In []: