

```
In [15]: import numpy as np  
np.__version__
```

```
Out[15]: '2.2.6'
```

```
In [16]: import pandas as pd
```

```
In [20]: pd.__version__
```

```
Out[20]: '2.2.3'
```

```
In [21]: df = pd.read_csv(r'Downloads\data.csv')
```

```
In [22]: df
```

```
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
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 [23]: id(df)
```

```
Out[23]: 1740802700880
```

```
In [24]: len(df)
```

```
Out[24]: 195
```

```
In [25]: df.columns
```

```
Out[25]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
       dtype='object')
```

```
In [26]: len(df.columns)
```

```
Out[26]: 5
```

```
In [27]: df.isnull()
```

```
Out[27]:
```

	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 [28]: df.isna()
```

```
Out[28]:
```

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

```
Out[29]: CountryName      0  
CountryCode      0  
BirthRate        0  
InternetUsers    0  
IncomeGroup       0  
dtype: int64
```

```
In [30]: df.isna().sum()
```

```
Out[30]: CountryName      0  
CountryCode      0  
BirthRate        0  
InternetUsers    0  
IncomeGroup       0  
dtype: int64
```

```
In [31]: 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 [32]: 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

```
In [33]: 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 [34]: 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 [35]: df[1:10]
```

	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

```
In [36]: 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

```
In [37]: df[1:100:10]
```

Out[37]:

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

Out[38]:

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

In [39]: `df.describe()`

Out[39]:

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

Out[40]:

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

```
In [41]: df[ 'CountryName' ]
```

```
Out[41]: 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 [42]: df[ 'CountryCode' ]
```

```
Out[42]: 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 [43]: df[[ 'CountryName', 'CountryCode', 'IncomeGroup']]
```

Out[43]:

	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 [44]: df[['CountryName', 'CountryCode', 'IncomeGroup']]

Out[44]:

	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 [45]: df\_cat= df[['CountryName', 'CountryCode', 'IncomeGroup']]  
df\_cat

Out[45]:

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

195 rows × 3 columns

In [46]:

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

5  
3

In [47]:

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

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

In [48]:

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

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

In [49]:

```
df_cat.describe()
```

Out[49]:

	<b>CountryName</b>	<b>CountryCode</b>	<b>IncomeGroup</b>
<b>count</b>	195	195	195
<b>unique</b>	195	195	4
<b>top</b>	Aruba	ABW	High income
<b>freq</b>	1	1	67

In [50]:

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

```
Out[50]:
```

	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 [51]: 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 [52]: 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 [53]: 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 [54]: `df.describe()`

Out[54]:

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

In [55]: `df.describe().transpose()`

Out[55]:

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

In [56]: `df.describe().T`

Out[56]:

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

In [57]: `df.columns`

Out[57]:

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

In [58]: `df.columns = ['a', 'b', 'c', 'd', 'e']`

In [59]: `df.head(1)`

```
Out[59]:    a      b      c      d      e
```

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

```
In [60]: df.columns = ['CountryName', 'CountryCode', 'BirthRate', 'Internetuser', 'IncomeGro
```

```
In [61]: df[['CountryCode', 'BirthRate', 'Internetuser']][4:8] #Subset
```

```
Out[61]:    CountryCode BirthRate Internetuser
```

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 [62]: df[4:8][['CountryCode', 'BirthRate', 'Internetuser']]
```

```
Out[62]:    CountryCode BirthRate Internetuser
```

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 [63]: df.columns
```

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

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

```
Out[64]:    CountryName  CountryCode  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

```
In [65]: df.BirthRate * df.Internetuser
```

```
Out[65]: 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 [66]: df.head(2)
```

```
Out[66]:   CountryName  CountryCode  BirthRate  Internetuser  IncomeGroup
0          Aruba        ABW      10.244       78.9  High income
1  Afghanistan        AFG      35.253       5.9  Low income
```

```
In [67]: df['Newcolumns']=df.BirthRate * df.Internetuser
```

```
In [68]: df.head(5)
```

```
Out[68]:   CountryName  CountryCode  BirthRate  Internetuser  IncomeGroup  Newcolumns
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 [69]: len(df.columns)
```

```
Out[69]: 6
```

```
In [70]: df= df.drop('Newcolumns',axis = 1)
```

```
In [71]: df.head(1)
```

```
Out[71]:   CountryName  CountryCode  BirthRate  Internetuser  IncomeGroup
0          Aruba        ABW      10.244       78.9  High income
```

```
In [72]: df
```

Out[72]:

	CountryName	CountryCode	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

In [73]:

```
df[df.Internetuser<2]
```

Out[73]:

	CountryName	CountryCode	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

In [74]:

```
df[df.Internetuser>2]
```

Out[74]:

	CountryName	CountryCode	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

In [75]: `len(df[df.Internetuser<2])`

Out[75]: 9

In [76]: `df.BirthRate>40`

Out[76]:

```
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 [77]: `low_educat = df[df.Internetuser<2]`  
`low_educat`

Out[77]:

	CountryName	CountryCode	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

In [78]:

```
low_ InternetUser_country = df[df.Internetuser<2]
low_ InternetUser_country
```

Out[78]:

	CountryName	CountryCode	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

In [79]:

```
high_birth_rate = df[df.BirthRate>40]
high_birth_rate
```

Out[79]:

	CountryName	CountryCode	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

In [80]: `Filter1 = df.Internetuser < 2`

In [81]: `Filter2 = df.BirthRate > 40`

In [82]: `df[Filter1 & Filter2]`

Out[82]:

	CountryName	CountryCode	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

In [83]: `df_num`

```
Out[83]:
```

	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 [84]: df_cat
```

```
Out[84]:
```

	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 [86]: df
```

Out[86]:

	CountryName	CountryCode	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

In [91]: df[df.IncomeGroup == 'high income']

Out[91]:

	CountryName	CountryCode	BirthRate	Internetuser	IncomeGroup
--	-------------	-------------	-----------	--------------	-------------

In [92]: df

Out[92]:

	CountryName	CountryCode	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

```
In [93]: df[df.IncomeGroup == 'low income']
```

```
Out[93]: CountryName  CountryCode  BirthRate  Internetuser  IncomeGroup
```

```
In [94]: df.IncomeGroup.unique()
```

```
Out[94]: array(['High income', 'Low income', 'Upper middle income',
       'Lower middle income'], dtype=object)
```

```
In [95]: df.IncomeGroup.nunique()
```

```
Out[95]: 4
```

```
In [98]: import matplotlib.pyplot as plt # visualization
import seaborn as sns    # stats visualization , advanced visualization

%matplotlib inline
# plot the graph in the Line
plt.rcParams ['figure.figsize'] = 6,2 # rcpParam param comes from plt Librar where

import warnings
warnings.filterwarnings('ignore') # whenever os will update ignore the os error us
```

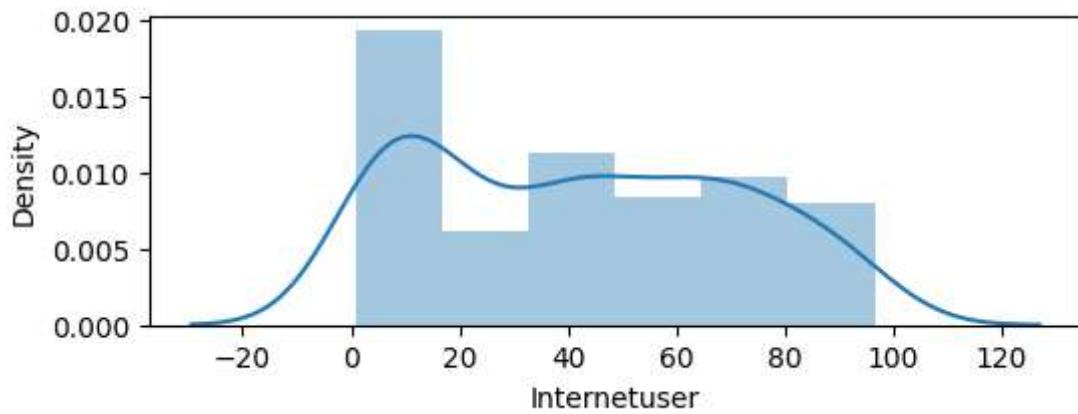
```
In [99]: df.columns
```

```
Out[99]: Index(['CountryName', 'CountryCode', 'BirthRate', 'Internetuser',
       'IncomeGroup'],
       dtype='object')
```

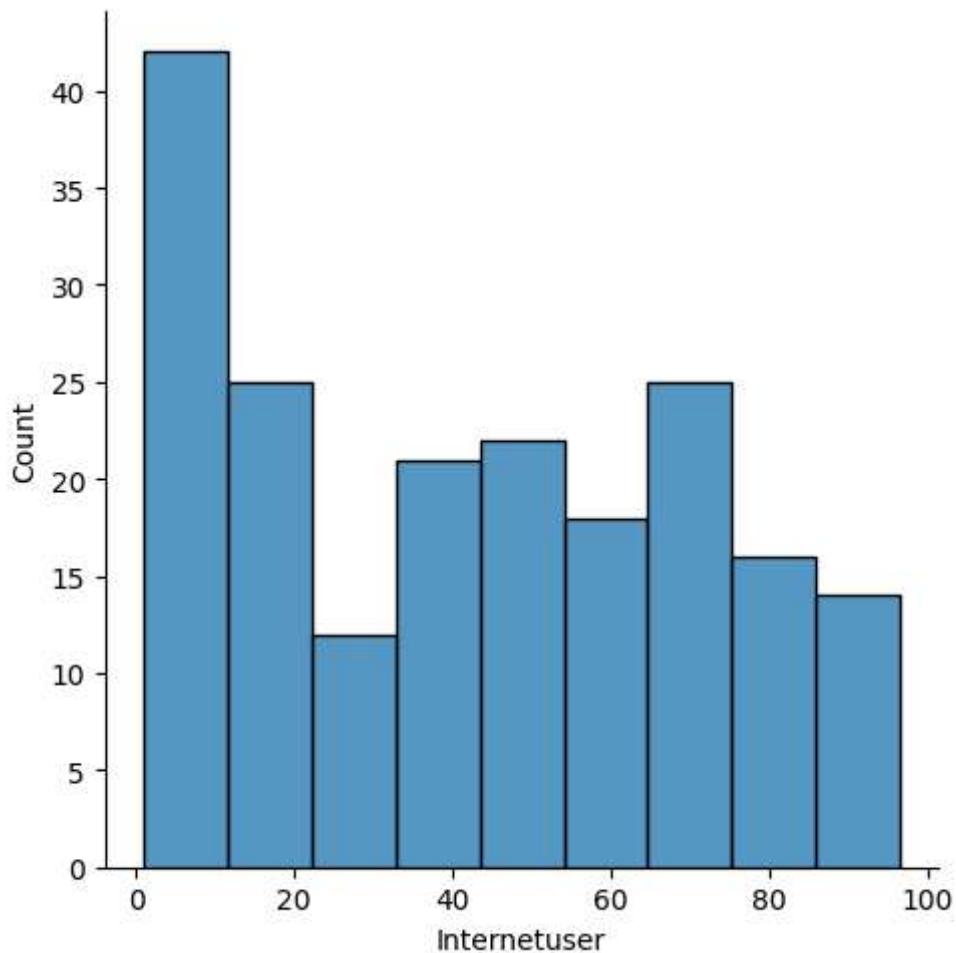
```
In [102...]: df['Internetuser']
```

```
Out[102...]: 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: Internetuser, Length: 195, dtype: float64
```

```
In [104...]: vis1 = sns.distplot(df["Internetuser"])
plt.show(vis1)      #univariate analysis--`plot the graph using 1 variable is cal
```

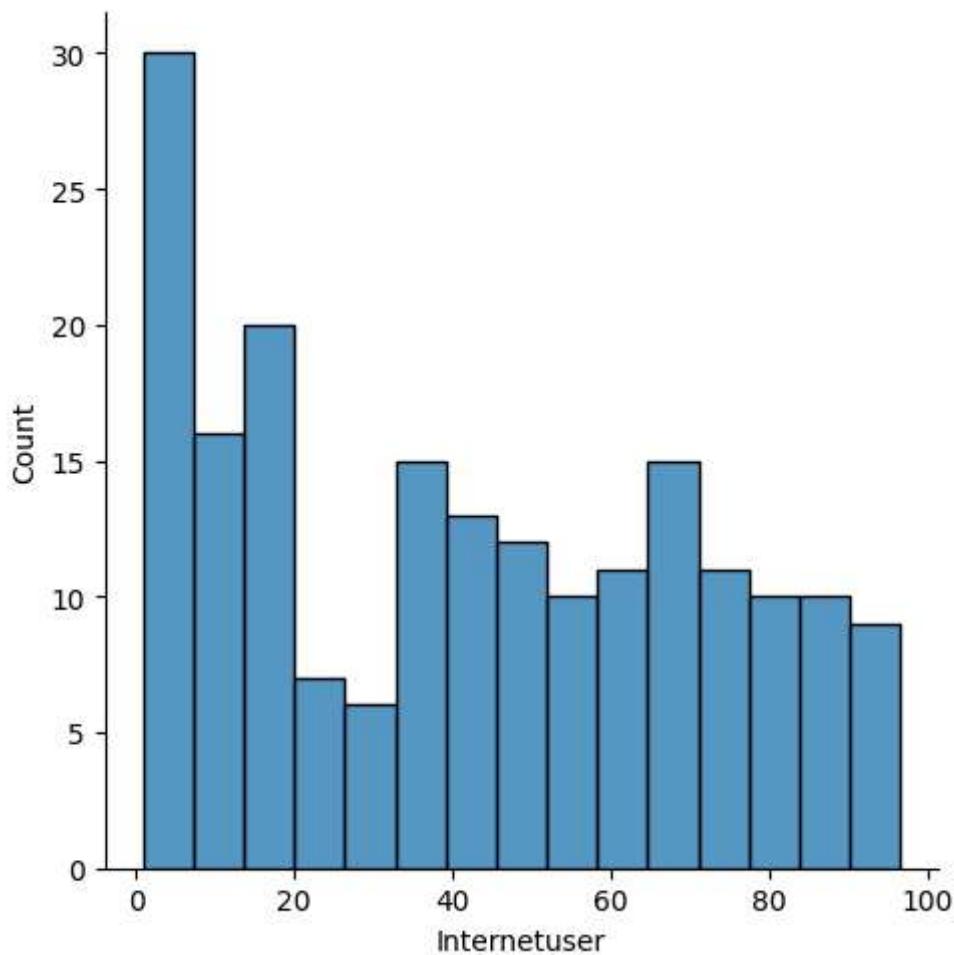


```
In [128]: vis2 = sns.displot(df["Internetuser"])
plt.show(vis2)
```

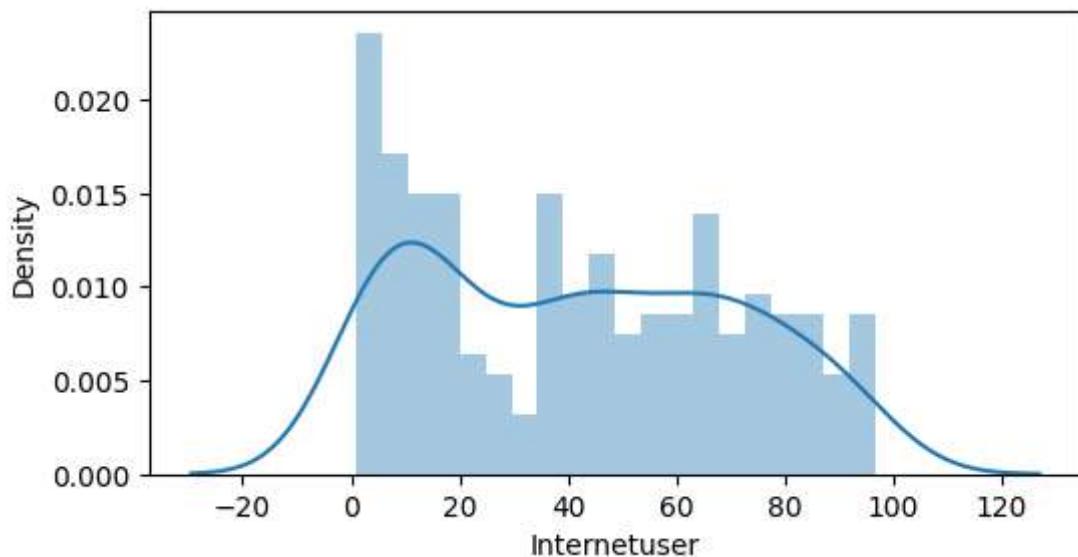


```
In [129]: # univariate analysis--- plot the graph using 1 variable
```

```
In [132]: vis3=sns.displot(df["Internetuser"], bins=15)
plt.show(vis3)
```

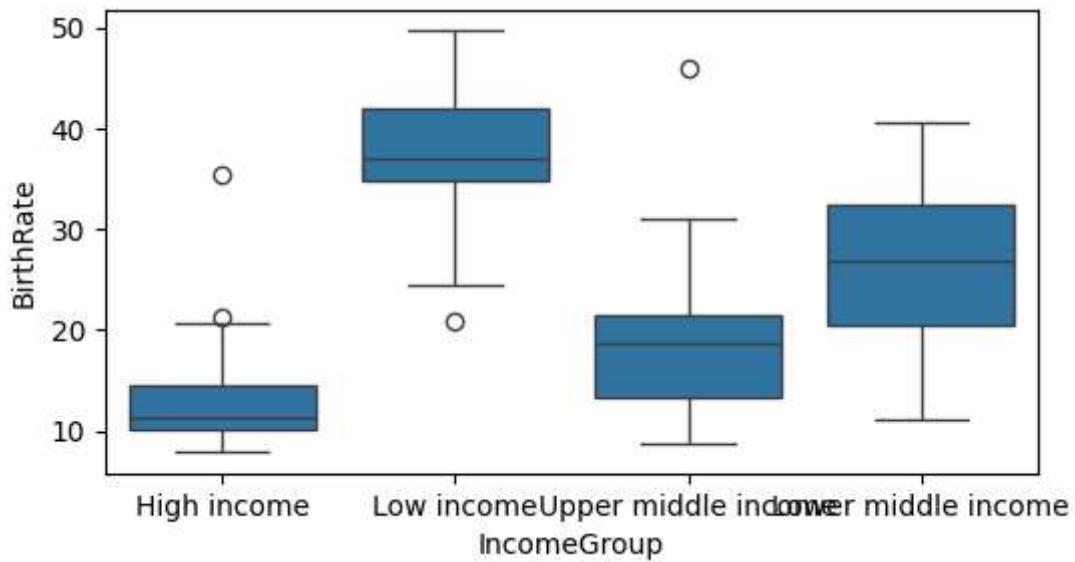


```
In [133]: vis0=sns.distplot(df["Internetuser"], bins=20)  
plt.show(vis0)
```



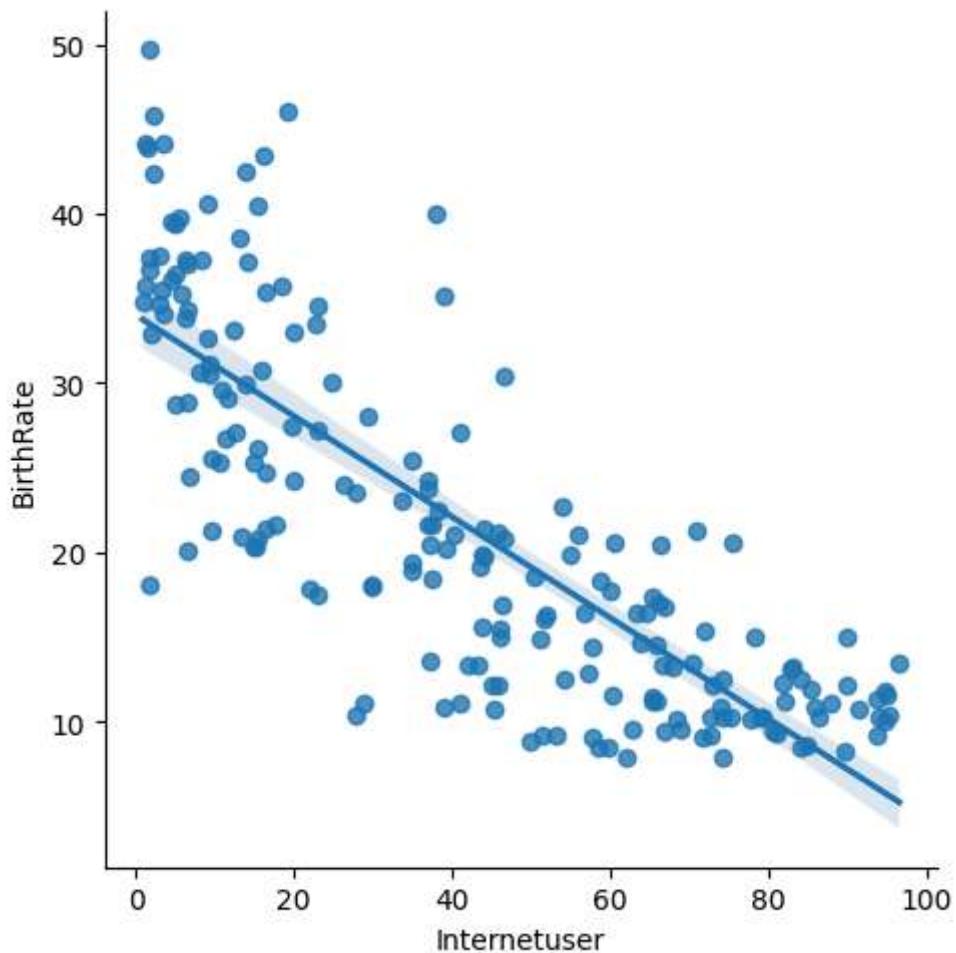
```
In [134]: plt.rcParams['figure.figsize'] = 6,3
```

```
In [135]: vis4=sns.boxplot(data=df,x="IncomeGroup",y='BirthRate')  
plt.show(vis4) #bivariate analysis=plot the graph using 2 variable is called univa
```



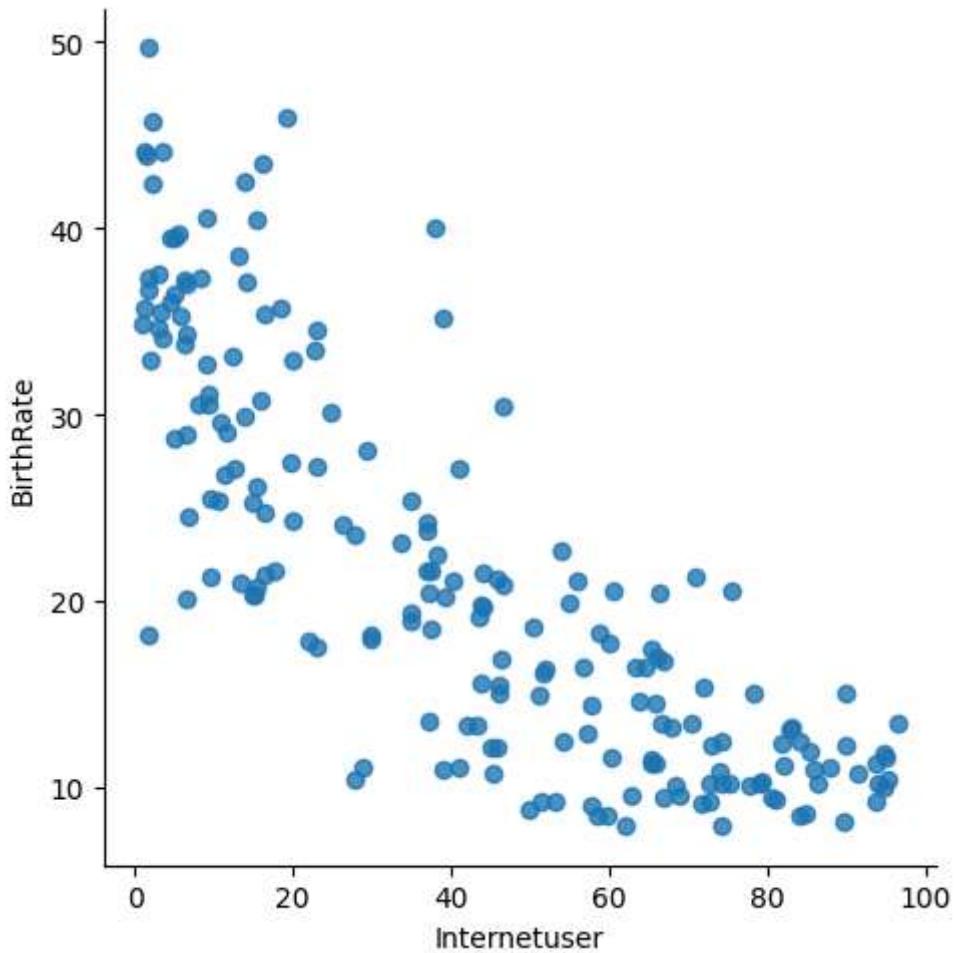
```
In [136]: #outlier=an statistic is the datapoint which is far from other observation  
#ALSO called as anomaly detection
```

```
In [138]: vis5=sns.lmplot(data=df,x='Internetuser', y='BirthRate')  
plt.show(vis5)
```

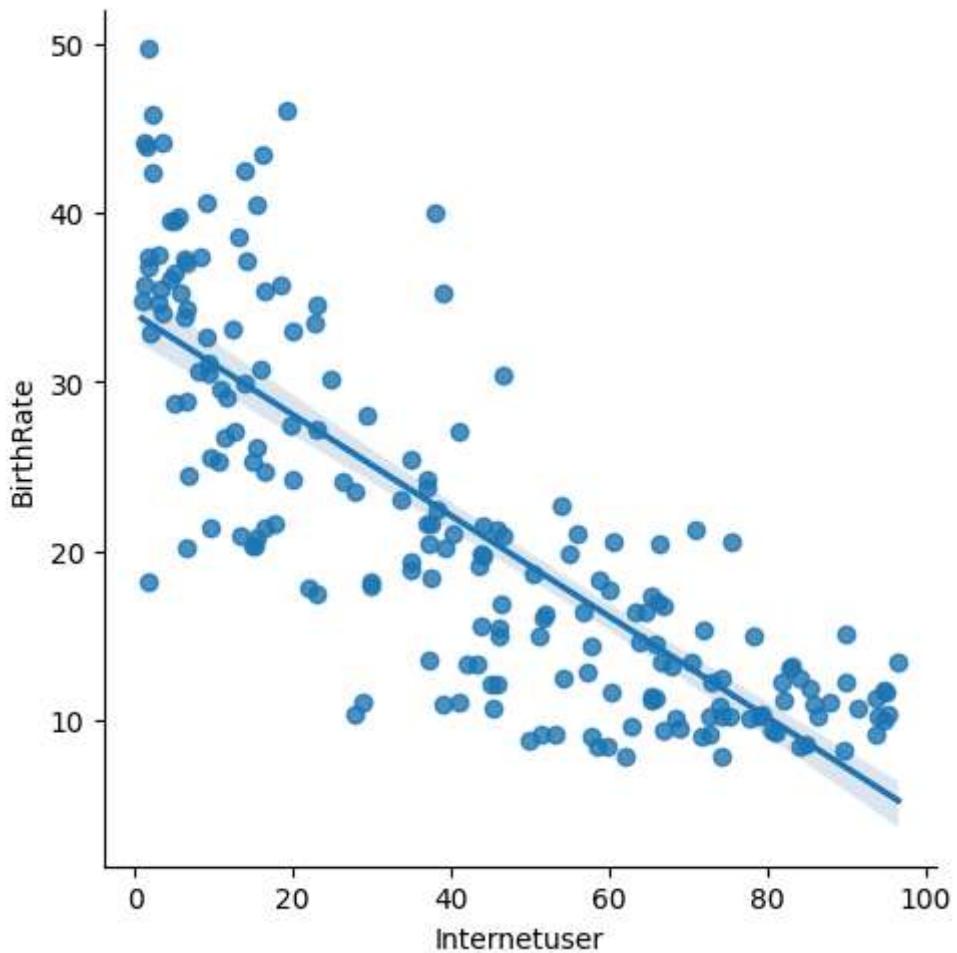


```
In [139]: vis5=sns.lmplot(data=df,x='Internetuser', y='BirthRate' ,fit_reg=False)
```

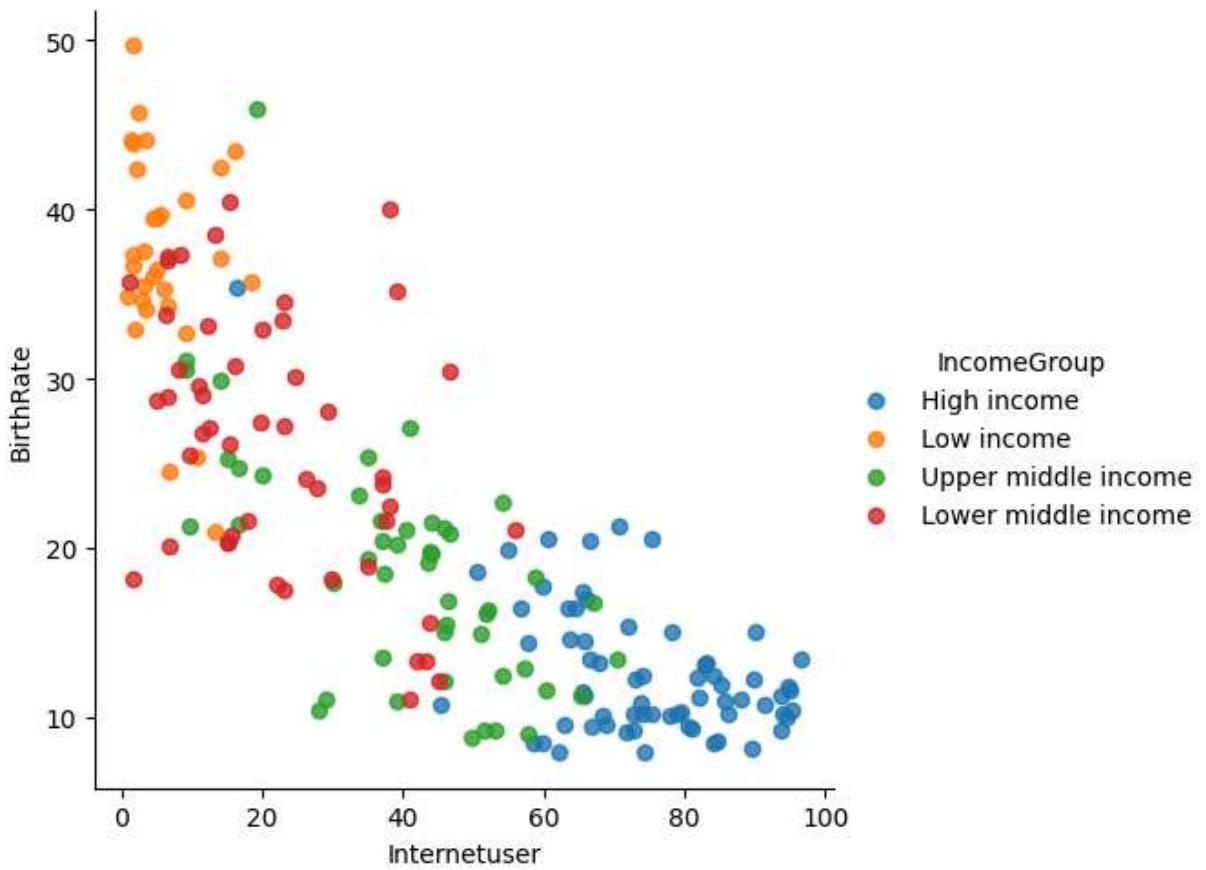
```
plt.show(vis5)
```



```
In [140]: vis5=sns.lmplot(data=df,x='Internetuser', y='BirthRate' ,fit_reg=True)  
plt.show(vis5)
```



```
In [141]: vis5=sns.lmplot(data=df,x='Internetuser', y='BirthRate' ,fit_reg=False, hue='Income')
plt.show(vis5)
```



```
In [142]: vis5=sns.lmplot(data=df,x='Internetuser', y='BirthRate' ,fit_reg=True, hue='IncomeG  
plt.show(vis5)
```

