

PANDAS WITH SEABORN

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

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

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

```
In [3]: stats = pd.read_csv(r'C:\Users\Avinash\Downloads\data.csv')
stats
```

| | 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 [4]: len(stats)      # rows=195
```

```
Out[4]: 195
```

```
In [5]: stats.shape    # rows=195 and columns=5
```

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

```
In [6]: stats.columns  # get col info
```

```
Out[6]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
               'IncomeGroup'],
              dtype='object')
```

```
In [7]: len(stats.columns)    # no of columns
```

Out[7]: 5

In [8]: type(stats)

Out[8]: pandas.core.frame.DataFrame

In [9]: stats.info()

```
# provide quick summary of dataframe like data type of each column, memory usage ,n
<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 [10]: stats.columns

```
Out[10]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
               'IncomeGroup'],
              dtype='object')
```

In [11]: #5. top rows

```
#head()
stats.head()      # by default print top 5 rows cause required no. of rows are not s
```

| | 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 [12]: stats.head(2)

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

In [13]: stats.tail() #by default print last 5 rows

| | 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 [14]: `stats.tail(4)`

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|-----|------------------|-------------|-----------|---------------|---------------------|
| 191 | South Africa | ZAF | 20.850 | 46.5 | Upper middle income |
| 192 | Congo, Dem. Rep. | COD | 42.394 | 2.2 | Low income |
| 193 | Zambia | ZMB | 40.471 | 15.4 | Lower middle income |
| 194 | Zimbabwe | ZWE | 35.715 | 18.5 | Low income |

In [15]: `stats[::-1]` # give data in reverse order(Z-A)

| | 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 [16]: `stats[:5]` # give 1 to (5-1) rows #Slicing- rows sliced-(n-1) applied to high v

Out[16]:

| | 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 [17]:

stats[5:]

Out[17]:

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|-----|---------------------|-------------|-----------|---------------|---------------------|
| 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 |
| ... | ... | ... | ... | ... | ... |
| 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 |

190 rows × 5 columns

In [63]:

stats[::10]

Out[63]:

| | 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 [18]: stats[0:200:10] # rows between 0 to 200 with step count of 10

Out[18]:

| | 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 [19]: stats

Out[19]:

| | 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 [20]:

```
stats.describe()      # descriptive statistics # display only numerical value

# mean- avg, count- no of count, std- standard deviation, min-minimum, max -maximum
```

Out[20]:

| | 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 [21]:

```
stats.describe().transpose()      # row to col and col to row convert of stats.descr
```

Out[21]:

| | 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 |

column renaming as per the requirement # renaming the column name that is been readfrom csv file

```
In [22]: # column rename
stats.columns
```

```
Out[22]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
               'IncomeGroup'],
               dtype='object')
```

```
In [23]: stats.columns = ['a','b','c','d','e']
```

```
In [24]: stats.columns
```

```
Out[24]: Index(['a', 'b', 'c', 'd', 'e'], dtype='object')
```

```
In [25]: stats.head()
```

```
Out[25]:
```

| | 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 [26]: stats.columns =[ 'CountryName', 'CountryCode', 'BirthRate', 'InternetUsers', 'IncomeG
# given old names again
```

```
In [27]: stats.columns
```

```
Out[27]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
               'IncomeGroup'],
               dtype='object')
```

```
In [28]: stats.head()
```

```
Out[28]:
```

| | 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 [29]: stats[:] # display whole data; data slicing
```

Out[29]:

| | 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 [30]:

```
stats.isnull()      # also can write stats.isna()
# check if have any missing value  # False-no missing value
```

Out[30]:

| | 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 [31]:

```
stats.isna()
```

Out[31]:

| | 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 [32]: `stats.isnull().sum() # measure total missing value in columns`

Out[32]:

| | |
|---------------|---|
| CountryName | 0 |
| CountryCode | 0 |
| BirthRate | 0 |
| InternetUsers | 0 |
| IncomeGroup | 0 |
| dtype: int64 | |

In [33]: `stats.dtypes`

Out[33]:

| | |
|---------------|---------|
| CountryName | object |
| CountryCode | object |
| BirthRate | float64 |
| InternetUsers | float64 |
| IncomeGroup | object |
| dtype: object | |

In [34]: `stats.head(2)`

Out[34]:

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

In [35]: `stats['CountryName']`

```
Out[35]: 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 [36]: stats[['CountryName', 'CountryCode', 'IncomeGroup']]
```

| | 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 [37]: stats_categorical = stats[['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
                                'IncomeGroup']]
stats_categorical.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 [38]: `stats_categorical.describe()`

Out[38]:

| | 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 [39]: `stats.num = stats[['BirthRate', 'InternetUsers']]
stats.num.head()`

C:\Users\Avinash\AppData\Local\Temp\ipykernel_6756\4178692952.py:1: UserWarning: Pandas doesn't allow columns to be created via a new attribute name - see https://pandas.pydata.org/pandas-docs/stable/indexing.html#attribute-access
stats.num = stats[['BirthRate', 'InternetUsers']]

Out[39]:

| | 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 |

In [65]: `stats[4:8]`

Out[65]:

| | 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 [67]: `stats[4:8][['BirthRate', 'InternetUsers']]`

Out[67]: **BirthRate InternetUsers**

| | BirthRate | InternetUsers |
|----------|------------------|----------------------|
| 4 | 11.044 | 88.0 |
| 5 | 17.716 | 59.9 |
| 6 | 13.308 | 41.9 |
| 7 | 16.447 | 63.4 |

In [69]: `stats[['BirthRate', 'InternetUsers']][4:8]`

Out[69]: **BirthRate InternetUsers**

| | BirthRate | InternetUsers |
|----------|------------------|----------------------|
| 4 | 11.044 | 88.0 |
| 5 | 17.716 | 59.9 |
| 6 | 13.308 | 41.9 |
| 7 | 16.447 | 63.4 |

In [40]: `stats.num.describe() # use for descriptive statistics`

Out[40]: **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 [41]: `stats.BirthRate*stats.InternetUsers`

Out[41]:

| | |
|-----|----------|
| 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 [42]: stats['myCalc'] = stats.BirthRate*stats.InternetUsers
```

```
In [43]: stats
```

Out[43]:

| | 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 [44]: stats.columns
```

```
Out[44]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup', 'myCalc'],
      dtype='object')
```

```
In [45]: len(stats.columns)
```

```
Out[45]: 6
```

```
In [46]: stats = stats.drop('myCalc',axis = 1) #axis 1 = column; axis 0 = row
```

```
In [47]: stats
```

Out[47]:

| | 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 [48]: `stats.InternetUsers<2 # we are checking if given condition is correct`Out[48]: 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 [49]: `stats[stats.InternetUsers<2] # get actual values for which the condition is`

Out[49]:

| | 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 [50]:

`stats.BirthRate>40`

Out[50]:

```
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 [51]:

`stats[stats.BirthRate>40]`

Out[51]:

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

In [52]: stats[(stats.BirthRate>40) & (stats.InternetUsers<2)] # case where both conditions are met

Out[52]:

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

In [53]: stats.head()

Out[53]:

| | 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 [54]: stats[stats.IncomeGroup == 'Low income']

Out[54]:

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

```
In [55]: # get unique category
stats.IncomeGroup.unique()
```

```
Out[55]: array(['High income', 'Low income', 'Upper middle income',
       'Lower middle income'], dtype=object)
```

```
In [56]: # get unique value
stats.IncomeGroup.nunique()          # count no of unique values
```

```
Out[56]: 4
```

```
In [57]: # Introduction to seaborn # Seaborn is very powerfull visualization (static visual

import matplotlib.pyplot as plt # visualization
import seaborn as sns          # distribution visualization
#seaborn are used for advanced visualization e.x.--> distribution plot, line plot

%matplotlib inline
plt.rcParams['figure.figsize'] = 9,5

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

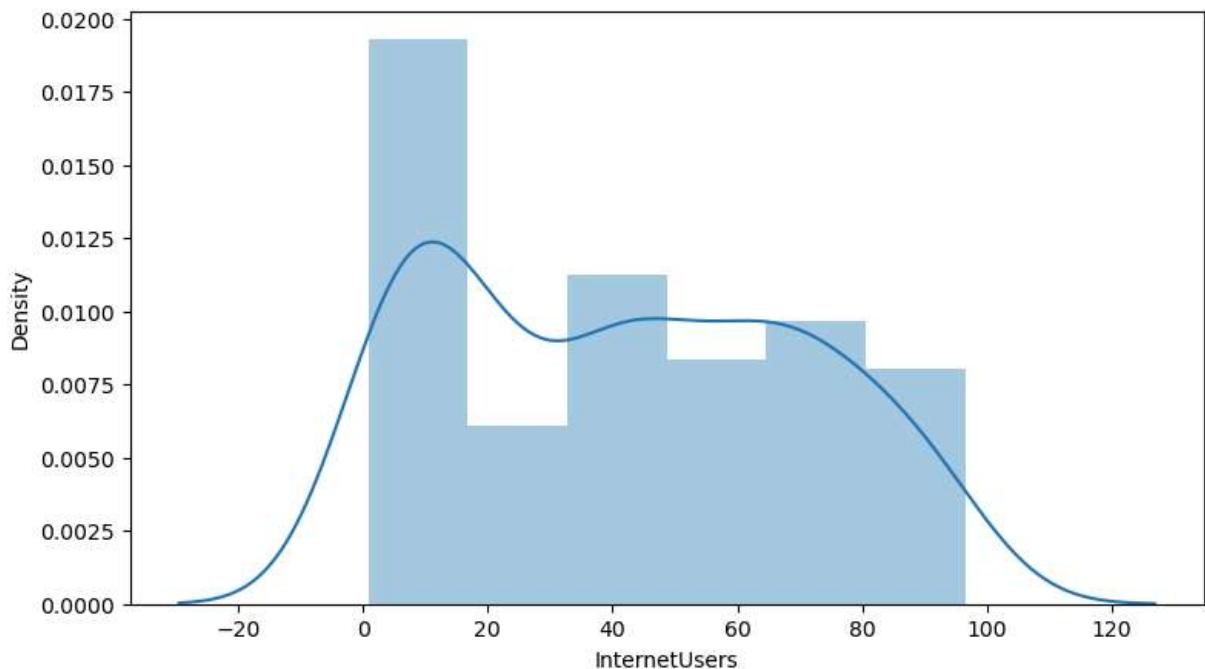
```
In [58]: stats.head()
```

| | 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 |

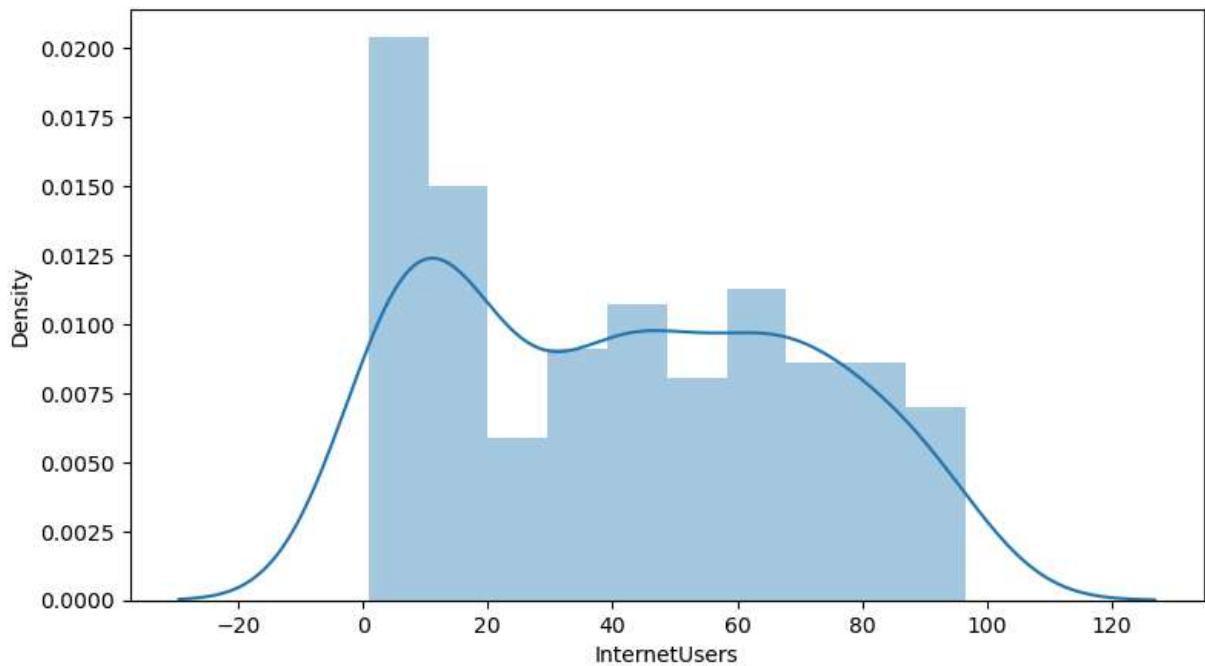
```
In [59]: stats["InternetUsers"]
```

```
Out[59]: 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
```

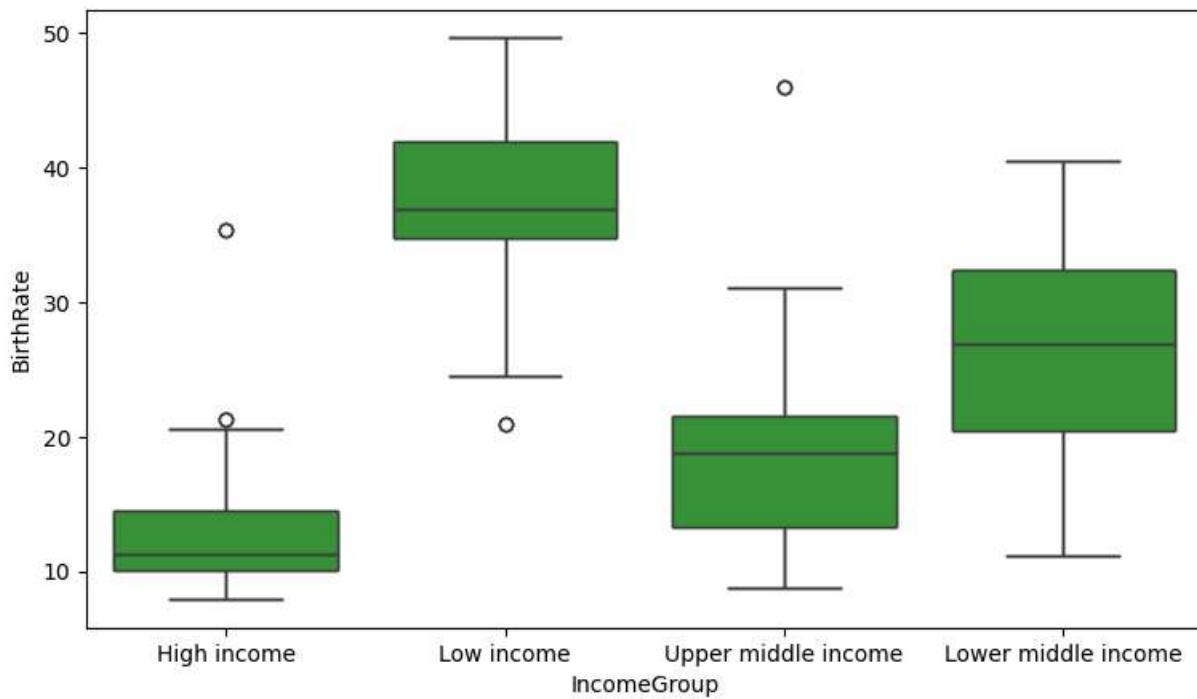
```
In [60]: #Distributions:
vis1 = sns.distplot(stats["InternetUsers"])      # Univariate analysis(plot graph us
plt.show()
```



```
In [61]: vis1 = sns.distplot(stats["InternetUsers"], bins=10) # get more bins to improve eng  
plt.show()
```

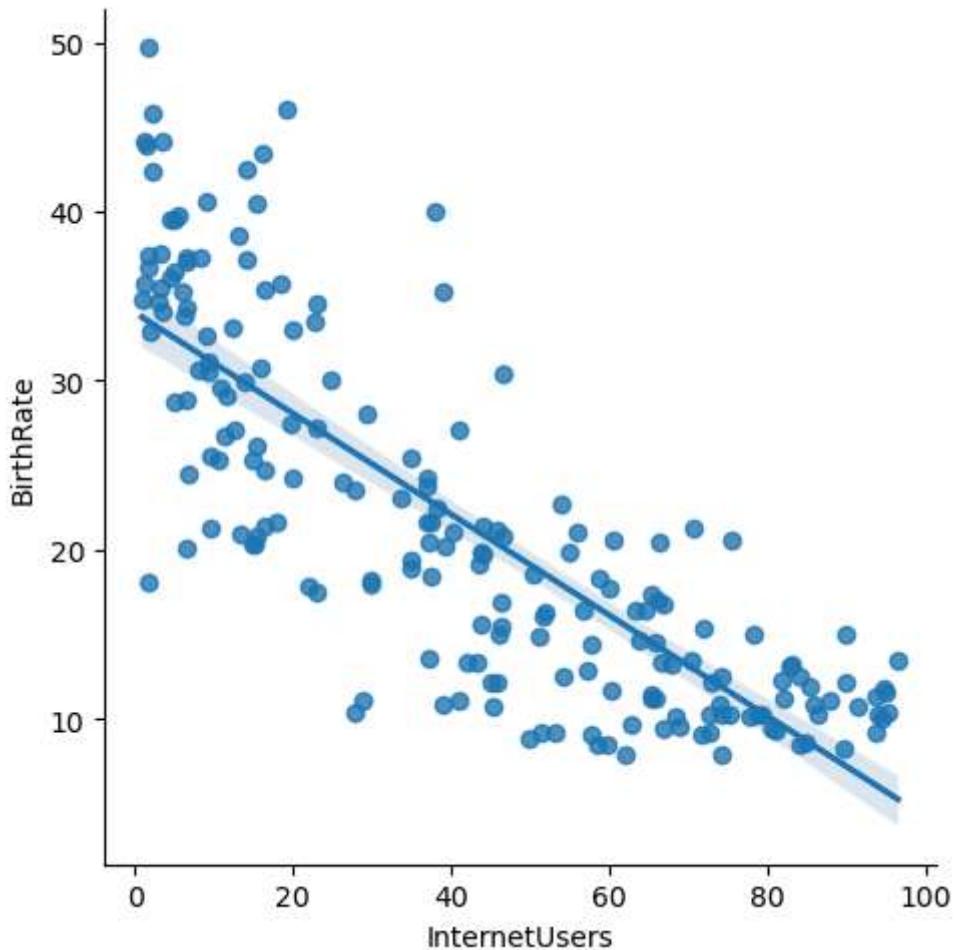


```
In [75]: #Box Plot:  
vis2 = sns.boxplot(data = stats, x = "IncomeGroup", y = "BirthRate") # BI-VARIATE  
plt.show()
```

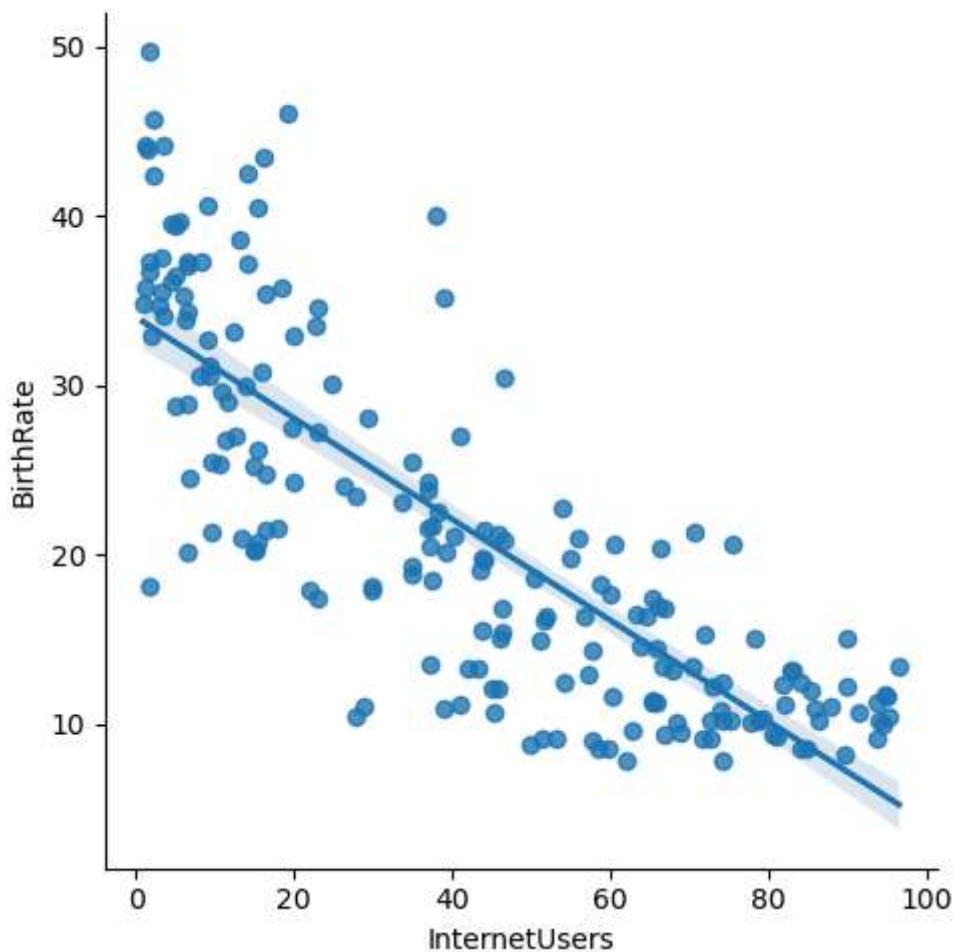


```
# refer to seaborn gallery # visualizing with seaborn
```

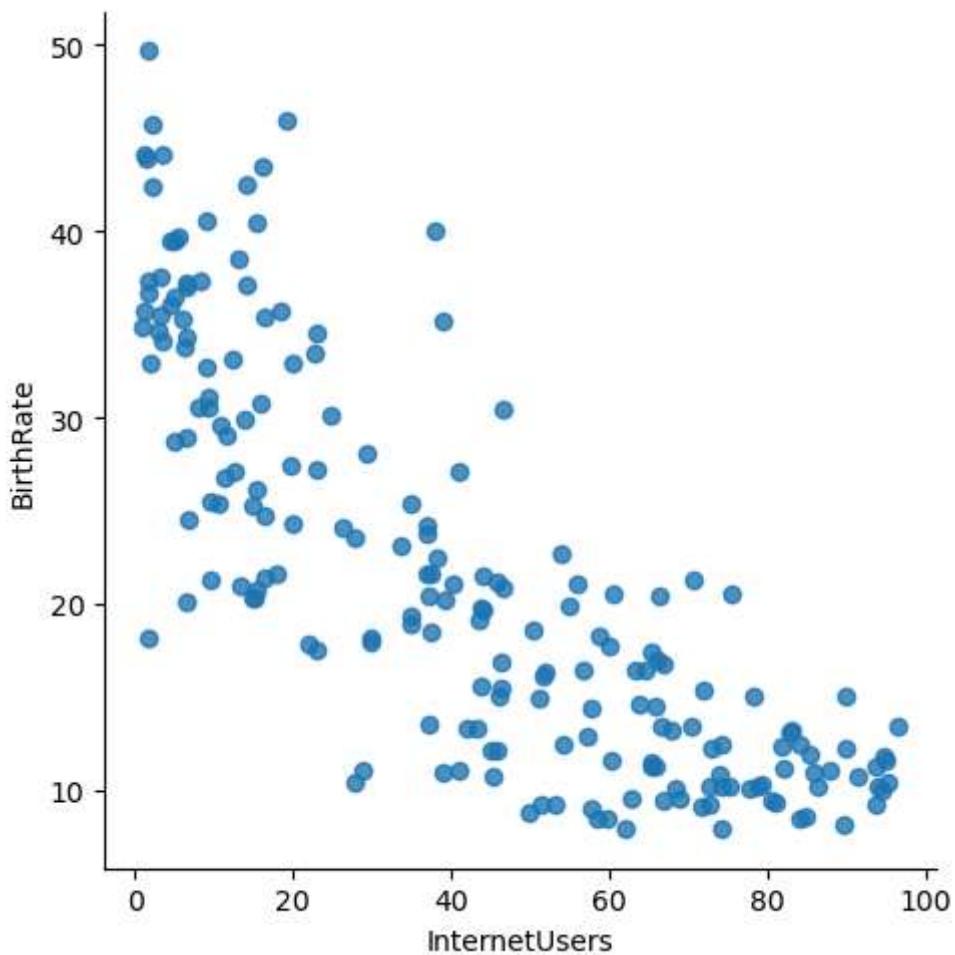
```
In [106]: vis4 = sns.lmplot(data = stats, x='InternetUsers', y='BirthRate') # Lm-Linear model
plt.show()
```



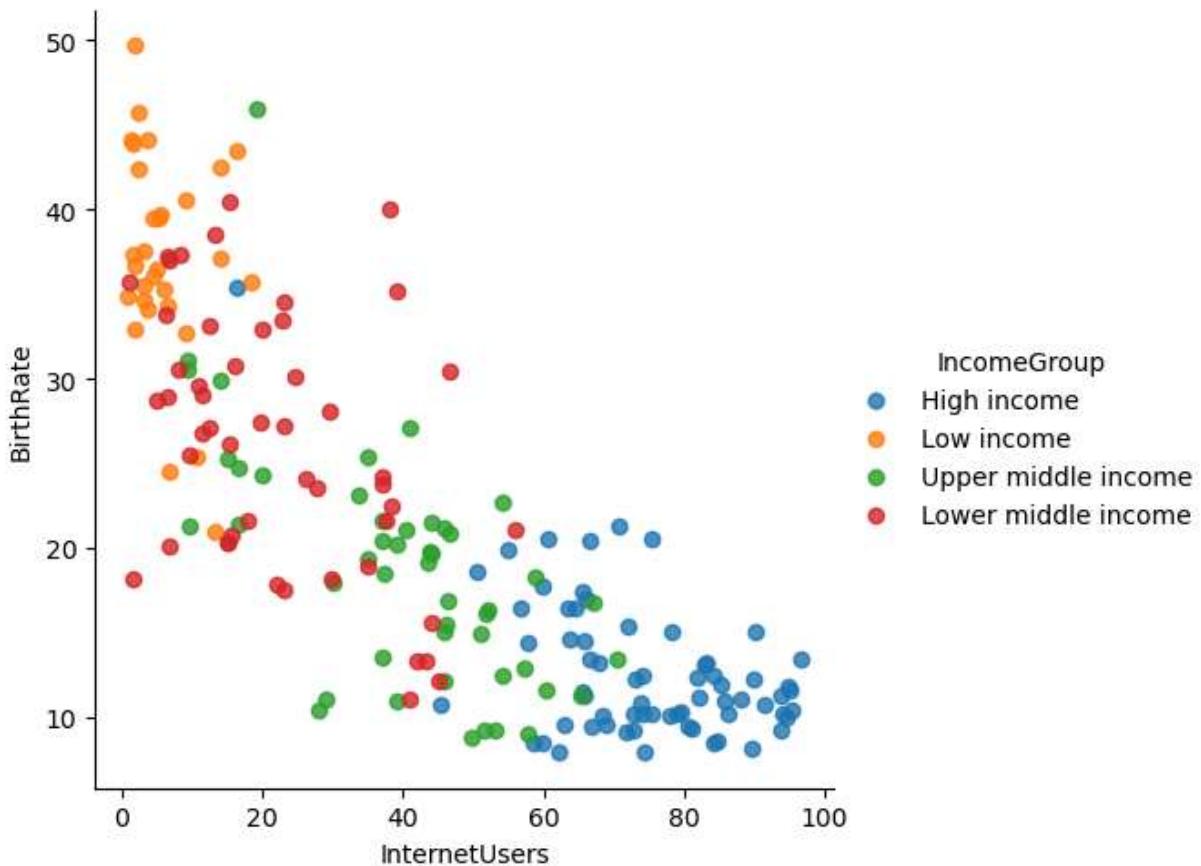
```
In [108... vis3 = sns.lmplot(data = stats, x='InternetUsers', y='BirthRate', fit_reg=True)  
plt.show()
```



```
In [110... vis31 = sns.lmplot(data = stats, x='InternetUsers', y='BirthRate', fit_reg=False)  
plt.show()
```



```
In [112]: vis5 = sns.lmplot(data = stats, x='InternetUsers', y='BirthRate', fit_reg=False, hue="# hu")  
plt.show()
```



```
In [114]: vis5 = sns.lmplot(data = stats, x='InternetUsers', y='BirthRate', fit_reg=False, hue='IncomeGroup', size=7)
plt.show() # hu
```

```
-----  
TypeError                                     Traceback (most recent call last)  
Cell In[114], line 1  
----> 1 vis5 = sns.lmplot(data = stats, x='InternetUsers', y='BirthRate', fit_reg=False,  
  1se, hue='IncomeGroup', size=7)  
    2 plt.show()  
  
TypeError: lmplot() got an unexpected keyword argument 'size'
```