# Project on Country GDP Analysis

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

In [2]:
```
df = pd.read_csv(r'C:\Users\Gopi Reddy\Downloads\data.csv')
```

In [3]:
```
df
```

Out[3]:

| | 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 [12]:
```
len(df) # it gives a total length of dataframe
```

Out[12]:
```
195
```

In [13]:
```
df.shape # it gives a dimension
```

Out[13]:
```
(195, 5)
```

In [14]:
```
df.columns # it gives a column names
```

Out[14]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

In [7]:
```
type(df)
```

```
Out[7]: pandas.core.frame.DataFrame
```

```
In [8]: df
```

Out[8]:

|  | 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 [15]: pd.__version__ # it shows a which version of pandas
```

```
Out[15]: '2.2.2'
```

```
In [16]: df
```

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 |
| **...** | ... | ... | ... | ... | ... |
| **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 [28]:

```
df.info() # it gives a information about columns
```

```
<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 [18]:

```
df.columns
```

Out[18]:

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

In [19]:

```
len(df.columns) # It shows a columns length
```

Out[19]:

```
5
```

In [20]:

```
df.head() # it gives a first five rows
```

Out[20]:

| | 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 [21]: df.tail() # it gives a last fives rows
```

Out[21]:

| | 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 [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]:
```
df[::-1] # reversing the data (194 to 0)
```

Out[23]:

| | 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 [25]:
```
df[:5] # it gives upto five rows only
```

Out[25]:

| | 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 [26]:
```
df[6:] # it gives a output from 6th row to last row
```

Out[26]:

| | 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 [27]:

```
df[0:200:10] # from 0 to 200 it gives 10 step slicing
```

Out[27]:

| | 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 [30]:

```
df.describe() # it gives a descriptive statistics (print numerical data info onl
```

Out[30]:

| | 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 [44]:

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

Out[44]:

| | 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 [45]:
```
df.columns
```

Out[45]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

In [46]:
```
df.columns = ['a','b','c','d','e'] # you can change the attributes or column nam
```

In [47]:
```
df.head(1)
```

Out[47]:

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

In [48]:
```
df.columns
```

Out[48]:
```
Index(['a', 'b', 'c', 'd', 'e'], dtype='object')
```

In [49]:
```
df.columns = ['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers', 'Incom
```

In [50]:
```
df.head(1)
```

Out[50]:

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

In [51]:
```
df.columns
```

Out[51]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

In [52]:
```
df[['CountryName', 'CountryCode']]
```

Out[52]:

| | CountryName | CountryCode |
|---|---|---|
| **0** | Aruba | ABW |
| **1** | Afghanistan | AFG |
| **2** | Angola | AGO |
| **3** | Albania | ALB |
| **4** | United Arab Emirates | ARE |
| **...** | ... | ... |
| **190** | Yemen, Rep. | YEM |
| **191** | South Africa | ZAF |
| **192** | Congo, Dem. Rep. | COD |
| **193** | Zambia | ZMB |
| **194** | Zimbabwe | ZWE |

195 rows × 2 columns

In [53]:

```
df.isnull() # it will detects a missing value
```

Out[53]:

| | 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 [55]:

```
df.isnull().sum() # it shows a missing value information
```

Out[55]:

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

In [56]:

```
# Split the categorical data into numerical data
```

In [91]:

```
df.dtypes # it shows the datatypes
```

Out[91]:

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

In [92]:

```
df.columns
```

Out[92]:

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

In [93]:

```
df_categorical = df[['CountryName', 'CountryCode','IncomeGroup']]
df_categorical.head()
```

Out[93]:

| | 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 |

In [94]:

```
df.describe() #descriptive statistics - it prints numerical data only
```

Out[94]:

| | 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 [95]:

```
df_categorical.describe()
```

Out[95]:

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

In [96]:
```
df.columns
```

Out[96]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup', 'myCalc'],
      dtype='object')
```

In [97]:
```
df_num = df[['BirthRate', 'InternetUsers']]
df_num.head()
```

Out[97]:

| | 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 [98]:
```
df_num.describe() #descriptive statistics - it prints numerical data only
```

Out[98]:

| | 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 [99]:
```
df_num.describe().T  # transpose or T means convert column into rows
```

Out[99]:

| | 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 [100…

```
df.head()
```

Out[100…

| | 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 [101…

```
df['IncomeGroup']
```

Out[101…

```
0                High income
1                 Low income
2        Upper middle income
3        Upper middle income
4                High income
                ...
190      Lower middle income
191      Upper middle income
192               Low income
193      Lower middle income
194               Low income
Name: IncomeGroup, Length: 195, dtype: object
```

In [102…

```
df.columns
```

Out[102…

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

In [103…

```
['countryName','BirthRate']
```

Out[103…

```
['countryName', 'BirthRate']
```

In [104…

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

Out[104…

| | 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 [105…

```
df
```

Out[105…

| | 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 [106…

```
df[4:8]
```

Out[106…

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup | myCalc |
|---|---|---|---|---|---|---|
| **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 |
| **7** | Antigua and Barbuda | ATG | 16.447 | 63.4 | High income | 1042.7398 |

In [107…

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

Out[107…

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

In [108…

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

Out[108…

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

In [109…

```
df.head() # it will gives a first five rows
```

Out[109…

| | 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 [110…

```
df.BirthRate * df.InternetUsers  # Mathematical operations
```

Out[110…

```
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 [111…

```
df['myCalc'] = df.BirthRate * df.InternetUsers # Adding another new column
```

In [112…

```
df
```

Out[112…

| | 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 [113…

```
df.columns # myCalc column will be added
```

Out[113…

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

In [114…

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

Out[114…

```
6
```

In [115…

```
df = df.drop('myCalc',axis = 1) # deleting a myCalc column
```

In [116…

```
df.columns
```

Out[116…

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

In [117…

```
df.InternetUsers < 2
```

Out[117…

```
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 [118…

```
df
```

Out[118…

| | 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 [121…

```
Filter = df.InternetUsers < 2
```

In [122…

```
df[Filter]
```

Out[122…

| | 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 [123…

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

Out[123… 9

In [124…

```
df
```

Out[124…

| | 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 [127…

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

Out[127…

```
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 [128…

```
df[Filter2]
```

Out[128…

| | 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 [129…

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

Out[129…

```
12
```

In [130…

```
df[Filter2]
```

Out[130…

| | 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 [131…

```
Filter & Filter2
```

Out[131…

```
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 [132…

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

Out[132…

| | 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 [133…

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

Out[133…

| | 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 [134…

```
df.head()
```

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 |

In [135…

```
df
```

Out[135…

|   | 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 [138…

```
df[df.IncomeGroup == 'Low income'] # it shows all low income data
```

Out[138…

| | 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 [139…

```
df[df.IncomeGroup == 'High income'] # it shows all high income data
```

Out[139…

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

67 rows × 5 columns

In [140…

```
df.IncomeGroup.unique() # unique means it comes all low,high,upper middle,lower
```

Out[140…

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

In [141…

```
df.IncomeGroup.nunique() # It shows a length of IncomeGroup
```

Out[141…

```
4
```

In [142…

```
## matplotlib visualization
```

In [144…

```
import matplotlib.pyplot as plt
import seaborn as sns

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

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

In [145…

```
df.head()
```

Out[145…

| | 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 [146…

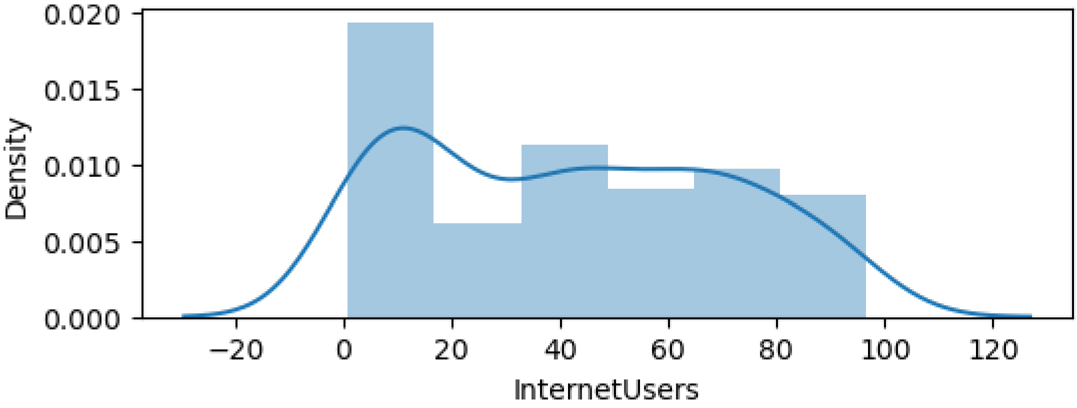
```
df['InternetUsers']
```

Out[146…

```
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
```

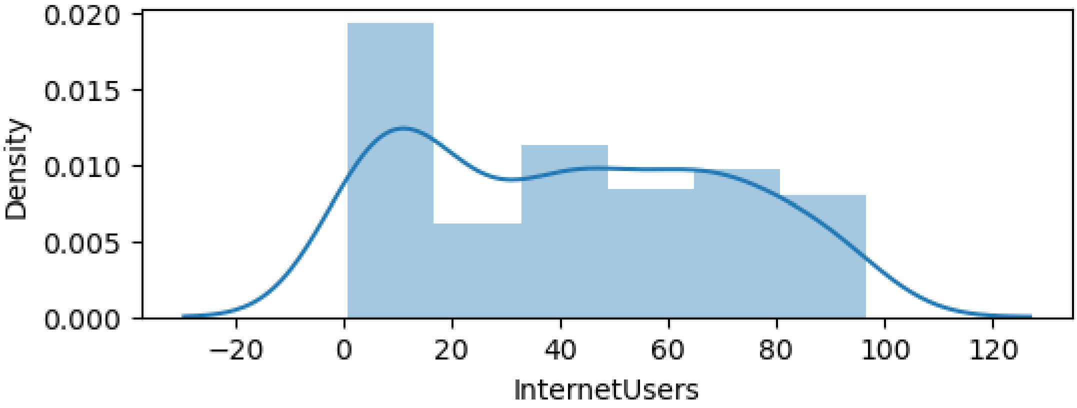
In [151…

```
vis1 = sns.distplot(df['InternetUsers']) # univariate analysis- plot the graph u
```



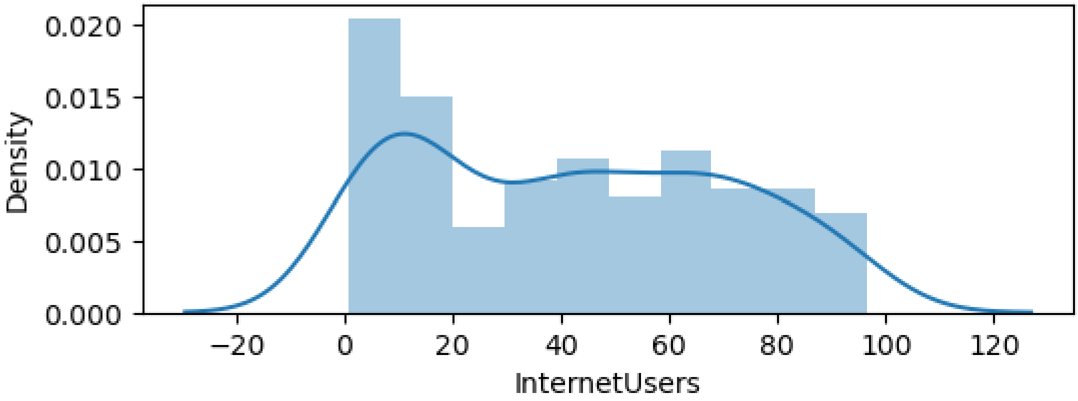
In [152…

```
vis2 = sns.distplot(df['InternetUsers'])
```



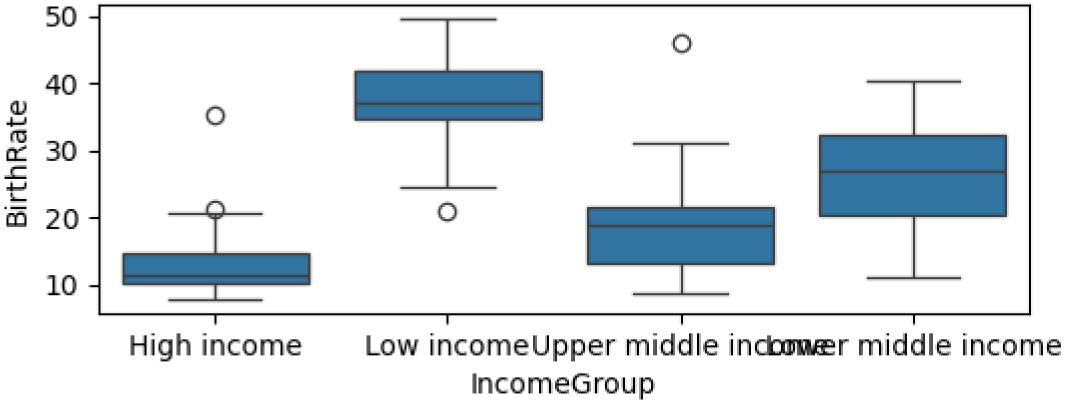
In [153…

```
vis3 = sns.distplot(df['InternetUsers'],bins = 10)
```



In [156…

```
# BOX PLOTS
vis4 = sns.boxplot(data = df, x = "IncomeGroup",y= 'BirthRate') # Bi-variate ana
```

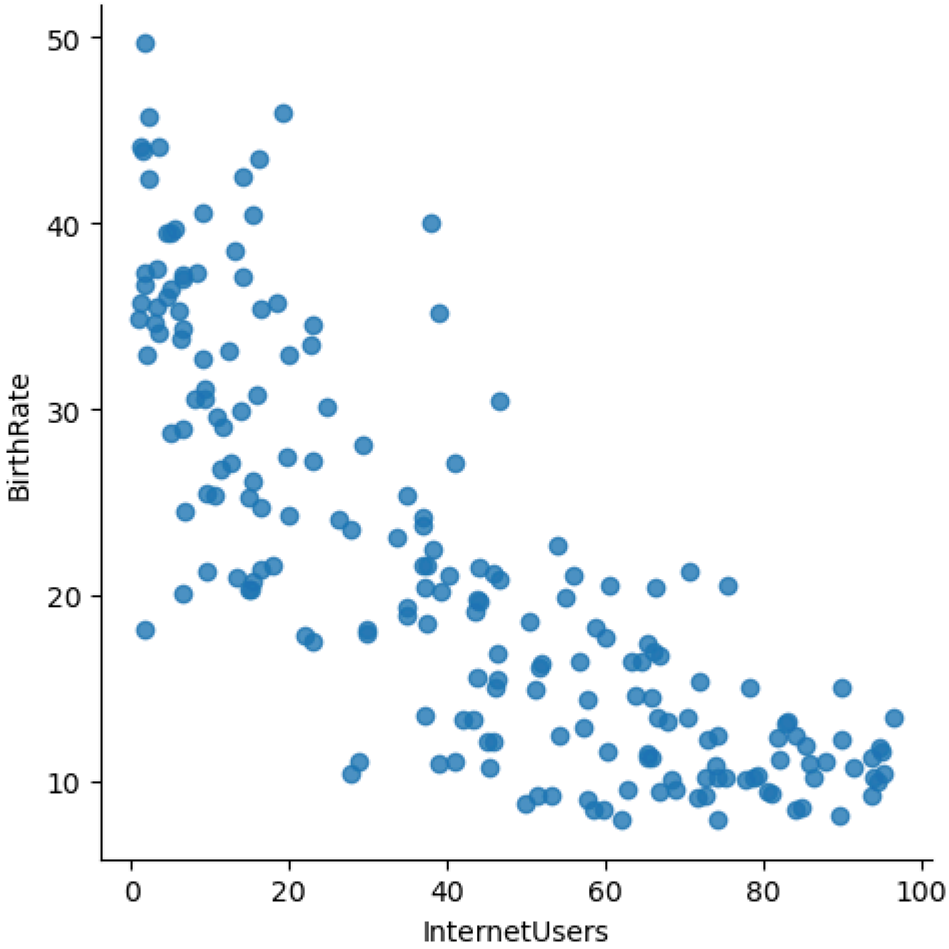


In [155…

```
# visualizing with seaborn
```

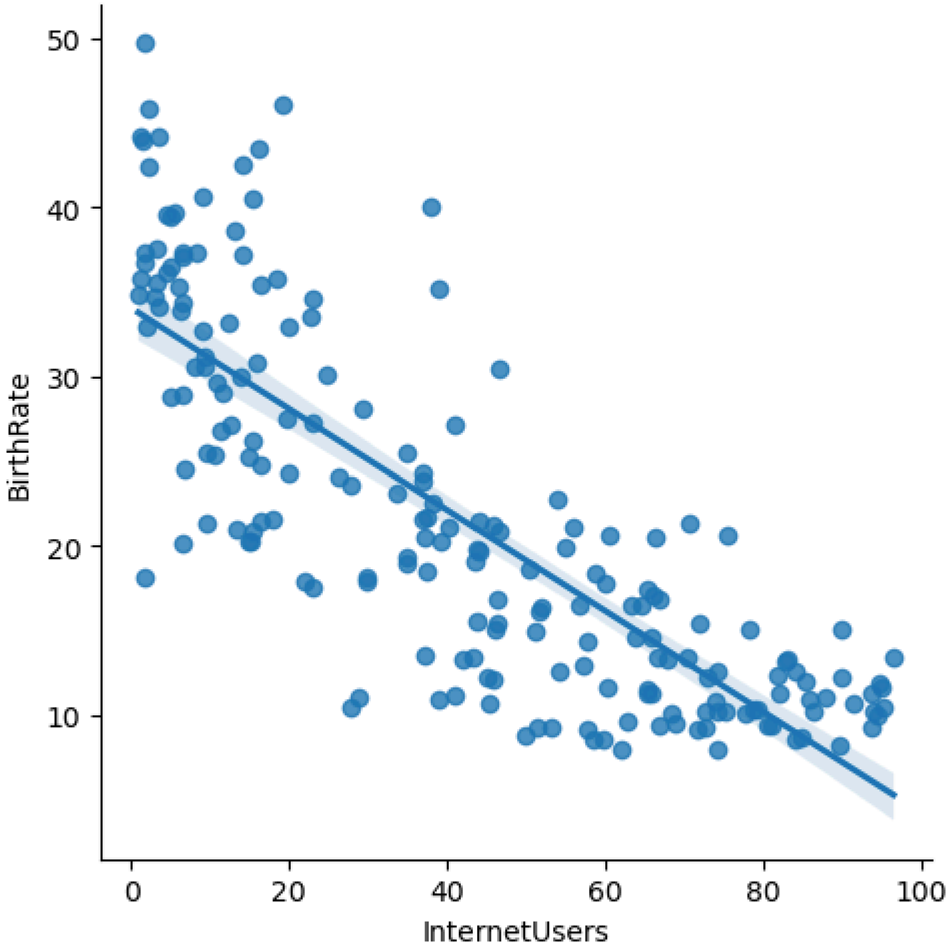
In [158…

```
vis5 = sns.lmplot(data = df,x = 'InternetUsers',y = 'BirthRate', fit_reg = False
```



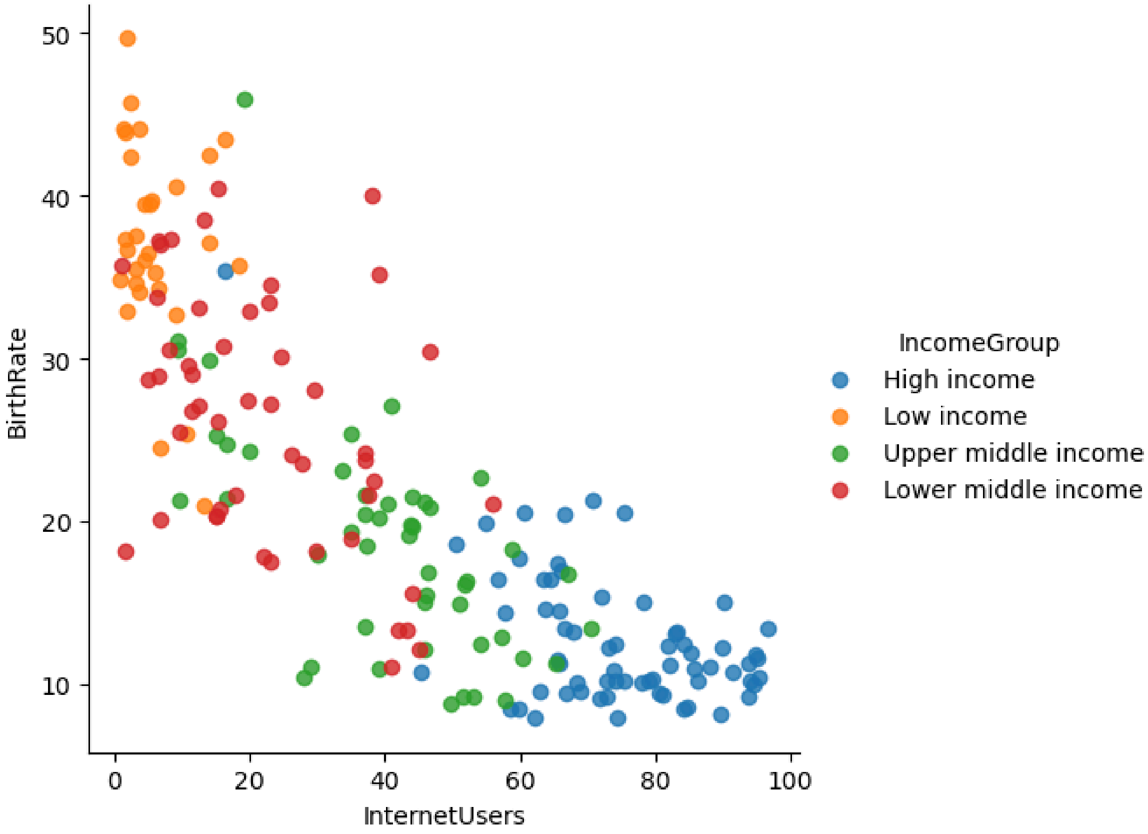
In [159…

```
vis6 = sns.lmplot(data = df,x = 'InternetUsers',y = 'BirthRate')
```



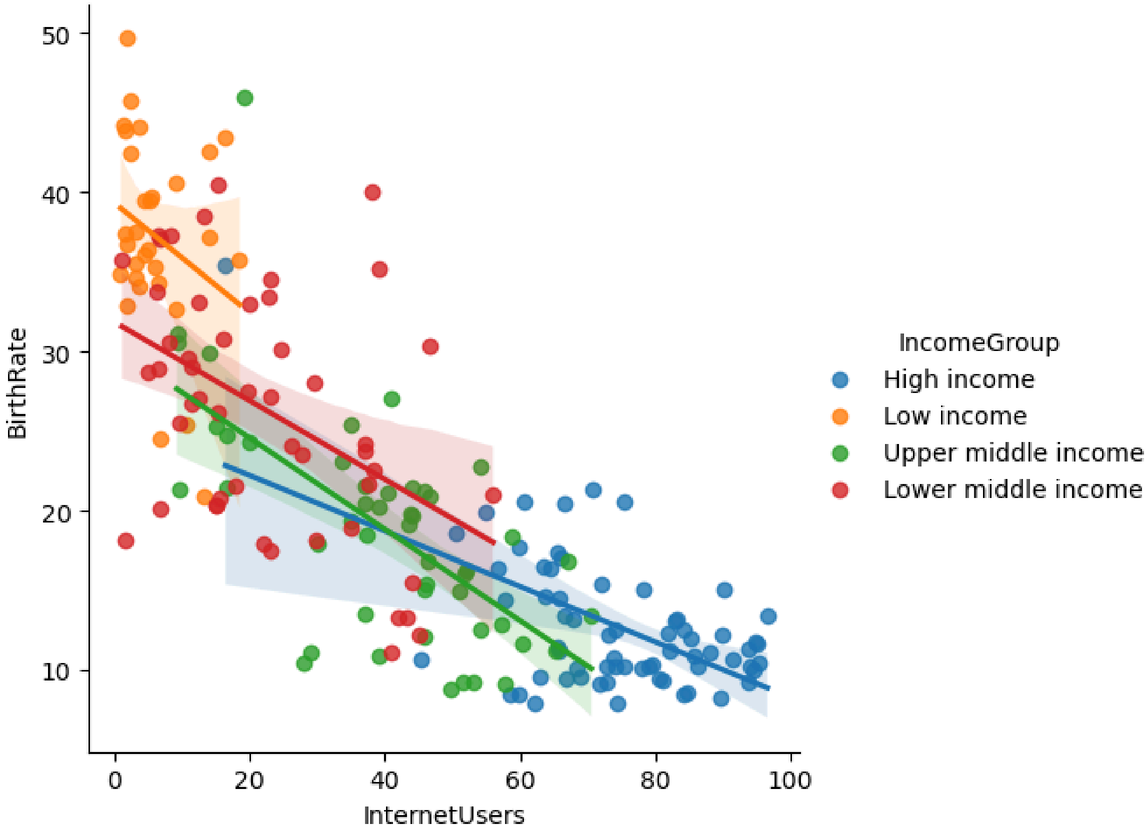
In [161…

```
vis7 = sns.lmplot(data = df,x = 'InternetUsers', y = 'BirthRate',
                  fit_reg = False,hue = 'IncomeGroup') # hue parameter for colou
```



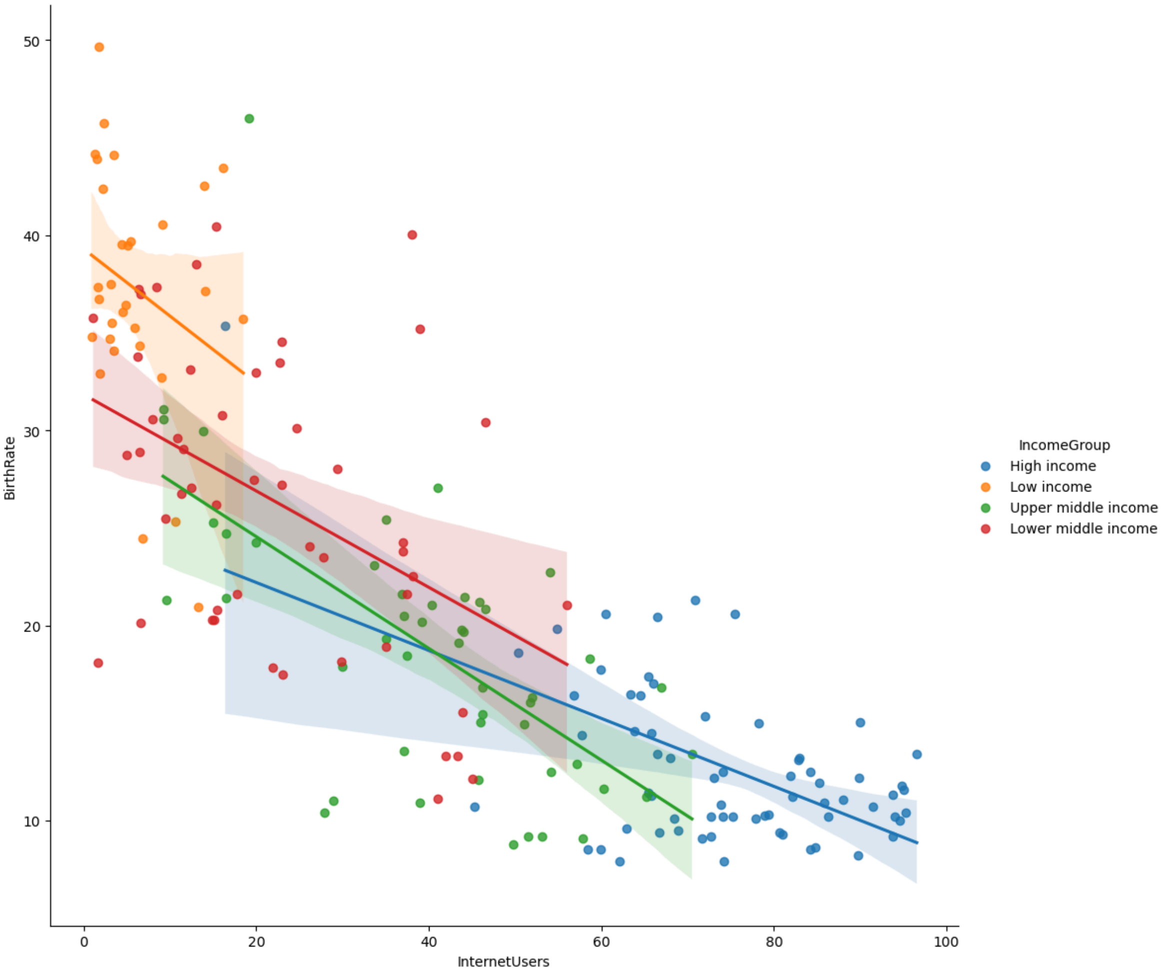
In [166…

```
vis8 = sns.lmplot(data = df,x = 'InternetUsers', y = 'BirthRate',
                  fit_reg = True,hue = 'IncomeGroup')
```



In [180…

```
vis8 = sns.lmplot(data = df,x = 'InternetUsers', y = 'BirthRate',
                  fit_reg = True,hue = 'IncomeGroup',height = 10)
plt.show(vis8)
```



# In this project we learned

1> importing data into python 2> Dataframe via panda 3> exploring datasets: head()tail()info()describe() 4> Renaming columns 5> subsetting dataframes 6> Basic operations with dataframe 7> filtering data frames 8> seaborn introduction

```
In [ ]:
```