

Country gdp analysis using pandas ,sns

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

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

```
Out[2]: '2.3.1'
```

```
In [12]: df = pd.read_csv(r"C:\Users\akiti\OneDrive\Desktop\data.csv")
```

```
In [14]: 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 [16]: id(df)
```

```
Out[16]: 2532139051456
```

```
In [18]: type(df)
```

```
Out[18]: pandas.core.frame.DataFrame
```

```
In [20]: len(df)
```

```
Out[20]: 195
```

```
In [22]: df.columns
```

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

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

```
Out[24]: 5
```

```
In [242...]: len(df)          # ✓ total rows
df.shape[0]
```

```
Out[242...]: 195
```

```
In [28]: df.shape
```

```
Out[28]: (195, 5)
```

```
In [30]: df.isnull()
```

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

Out[32]:

	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 [34]: `df.isnull().sum()# it tells about the missing values are present or not, column wise`

Out[34]:

In [35]: `df.head()`

Out[35]:

	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 [36]: `df.head`

```
Out[36]: <bound method NDFrame.head of
          CountryName  CountryCode  BirthRate  InternetUsers
          Aruba        ABW      10.244      78.9
          Afghanistan  AFG      35.253      5.9
          Angola       AGO      45.985     19.1
          Albania      ALB      12.877     57.2
          United Arab Emirates  ARE      11.044     88.0
          ...
          ...
          Yemen, Rep.   YEM      32.947     20.0
          South Africa  ZAF      20.850     46.5
          Congo, Dem. Rep.  COD      42.394      2.2
          Zambia        ZMB      40.471     15.4
          Zimbabwe      ZWE      35.715     18.5

          IncomeGroup
          High income
          Low income
          Upper middle income
          Upper middle income
          High income
          ...
          ...
          Lower middle income
          Upper middle income
          Low income
          Lower middle income
          Low income

[195 rows x 5 columns]>
```

In [37]: `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 [38]: `df.tail`

```
Out[38]: <bound method NDFrame.tail of
          CountryName  CountryCode  BirthRate  InternetUsers
          Aruba        ABW        10.244     78.9
          Afghanistan AFG        35.253     5.9
          Angola       AGO        45.985    19.1
          Albania      ALB        12.877    57.2
          United Arab Emirates ARE        11.044    88.0
          ...
          ...
          Yemen, Rep. YEM        32.947    20.0
          South Africa ZAF        20.850    46.5
          Congo, Dem. Rep. COD        42.394    2.2
          Zambia       ZMB        40.471    15.4
          Zimbabwe     ZWE        35.715    18.5

          IncomeGroup
          High income
          Low income
          Upper middle income
          Upper middle income
          High income
          ...
          ...
          Lower middle income
          Upper middle income
          Low income
          Lower middle income
          Low income

[195 rows x 5 columns]>
```

In [39]: `df.shape`

Out[39]: (195, 5)

In [40]: `df.dtypes`

```
Out[40]: CountryName      object
          CountryCode      object
          BirthRate        float64
          InternetUsers   float64
          IncomeGroup      object
          dtype: object
```

In [41]: `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
```

SLICING IN DATAFRAME

In [43]: `df[:]`

Out[43]:

	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 [44]: `df[:::-1]`

Out[44]:

	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 [45]:

df[:11]

Out[45]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9000	High income
1	Afghanistan	AFG	35.253	5.9000	Low income
2	Angola	AGO	45.985	19.1000	Upper middle income
3	Albania	ALB	12.877	57.2000	Upper middle income
4	United Arab Emirates	ARE	11.044	88.0000	High income
5	Argentina	ARG	17.716	59.9000	High income
6	Armenia	ARM	13.308	41.9000	Lower middle income
7	Antigua and Barbuda	ATG	16.447	63.4000	High income
8	Australia	AUS	13.200	83.0000	High income
9	Austria	AUT	9.400	80.6188	High income
10	Azerbaijan	AZE	18.300	58.7000	Upper middle income

In [46]:

df[0:200:50]

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.900000	High income
50	Ecuador	ECU	21.070	40.353684	Upper middle income
100	Libya	LBY	21.425	16.500000	Upper middle income
150	Sudan	SDN	33.477	22.700000	Lower middle income

In [47]: df['CountryName']

Out[47]:

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 [48]: df[['CountryName', 'CountryCode']]

	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 [49]: df[['CountryName', 'CountryCode', 'BirthRate']]

Out[49]:

	CountryName	CountryCode	BirthRate
0	Aruba	ABW	10.244
1	Afghanistan	AFG	35.253
2	Angola	AGO	45.985
3	Albania	ALB	12.877
4	United Arab Emirates	ARE	11.044
...
190	Yemen, Rep.	YEM	32.947
191	South Africa	ZAF	20.850
192	Congo, Dem. Rep.	COD	42.394
193	Zambia	ZMB	40.471
194	Zimbabwe	ZWE	35.715

195 rows × 3 columns

In [50]:

df.head(3)

Out[50]:

	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

descriptive statistics

In [52]:

df.describe()

Out[52]:

	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 [53]: df.describe().T # converts rows to columns and columns to rows

Out[53]:

	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 [55]: df_cat = df[['CountryName', 'CountryCode', 'IncomeGroup']]
df_cat

Out[55]:

	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 [56]: df_cat.describe()

Out[56]:

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

In [57]:

`df.describe()`

Out[57]:

	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 [58]:

`df.describe().all()`

Out[58]:

```
BirthRate      True
InternetUsers  True
dtype: bool
```

In [59]:

`df.describe(include='all')`

Out[59]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
count	195	195	195.000000	195.000000	195
unique	195	195	NaN	NaN	4
top	Aruba	ABW	NaN	NaN	High income
freq	1	1	NaN	NaN	67
mean	NaN	NaN	21.469928	42.076471	NaN
std	NaN	NaN	10.605467	29.030788	NaN
min	NaN	NaN	7.900000	0.900000	NaN
25%	NaN	NaN	12.120500	14.520000	NaN
50%	NaN	NaN	19.680000	41.000000	NaN
75%	NaN	NaN	29.759500	66.225000	NaN
max	NaN	NaN	49.661000	96.546800	NaN

In [60]:

df

Out[60]:

	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 [140...]

df.columns = ['a', 'b', 'c', 'd', 'e']

In [62]:

df.head(1)

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

```
In [142... df.columns = ['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers', 'IncomeGroup']
```

```
In [144... df.head(1)
```

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

```
In [146... df_categorical = df[['CountryName', 'CountryCode', 'IncomeGroup']]  
df_categorical.head()
```

	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 [148... df_categorical.describe()
```

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

```
In [150... ['CountryName', 'BirthRate']]
```

```
Out[150... ['CountryName', 'BirthRate']]
```

```
In [152... df[['CountryName', 'CountryCode', 'IncomeGroup']]
```

Out[152...]

	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 [154...]

df.head()

Out[154...]

	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 [156...]

#Mathematical operation
df.BirthRate * df.InternetUsers

Out[156...]

```
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 [158...]

```
#add a column
df['myCalc'] = df.BirthRate * df.InternetUsers
```

In [160...]

df

Out[160...]

	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 [162...]

df.columns

Out[162...]

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

In [164...]

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

In [166...]

df

Out[166...]

	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 [168...]

df.columns[3:4]

Out[168...]

Index(['InternetUsers'], dtype='object')

In [170...]

df.shape

Out[170...]

(195, 5)

In [172...]

df['InternetUsers']

Out[172...]

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 [174...]

df.InternetUsers<2

```
Out[174... 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 [176... Filter = df.InternetUsers<2
```

```
In [178... Filter
```

```
Out[178... 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 [180... df[Filter]
```

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

```
Out[182... 9
```

In [184...]

df

Out[184...]

	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 [186...]

df.BirthRate>40

Out[186...]

```

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 [188...]

Filter2 = df.BirthRate>40

In [190...]

Filter2

```
Out[190...]: 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 [192...]: df[Filter2]

	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 [194...]: Filter & Filter2

```
Out[194...]: 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 [196...]: df[Filter & Filter2]

Out[196...]

	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 [198...]

df

Out[198...]

	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 [200...]

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

Out[200...]

	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 [202...]

```
# hoe to get unique caterories
df.IncomeGroup.unique()
```

Out[202...]

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

In [204...]

```
df.IncomeGroup.nunique()
```

Out[204...]

4

In [206...]

```
# Introduction to seaborn. Seaborn is very powerfull visualization (STATISTIC VISUALIZATION)
import matplotlib.pyplot as plt # visulaization
import seaborn as sns # distribution visualion
# seaborn is used for advance visualization e.x --> distribution plot, line plot

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

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

In [208...]

```
df.head()
```

Out[208...]

	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 [210...]

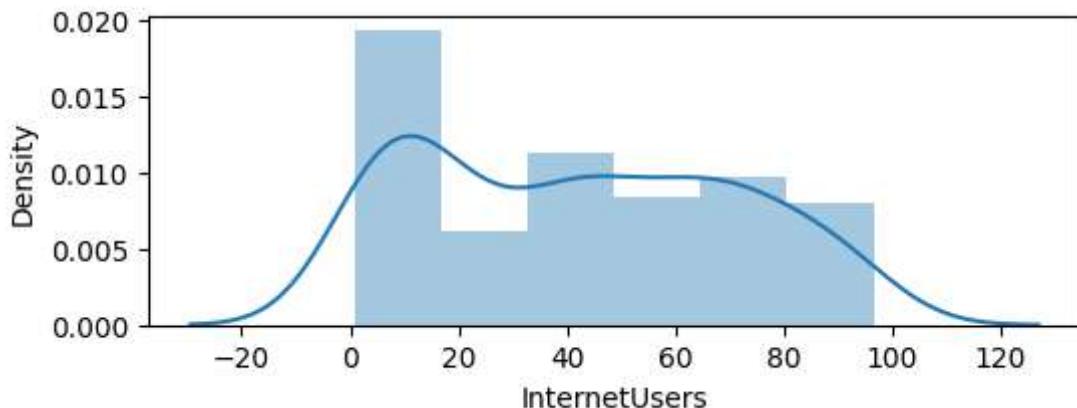
df["InternetUsers"]

Out[210...]

```
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
```

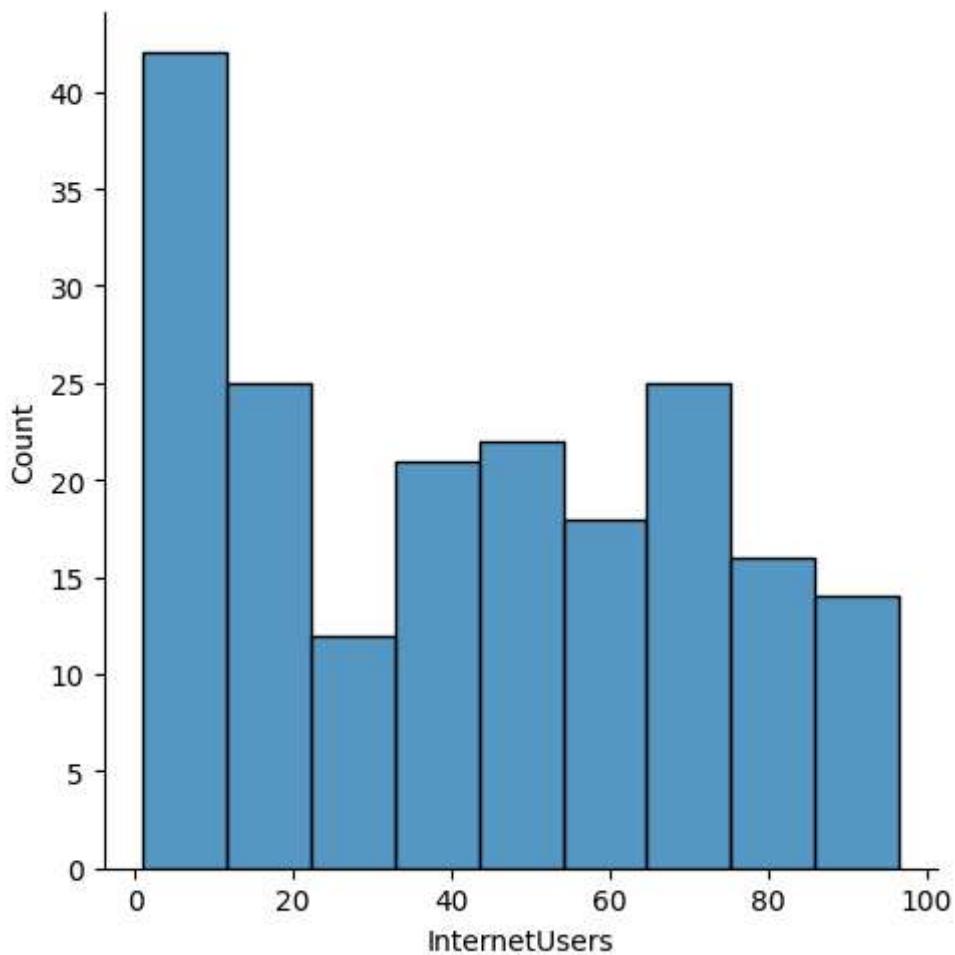
In [212...]

```
# Distibutions:
vis1 = sns.distplot(df["InternetUsers"])# UNIVARIATE ANALYSIS - statistics
plt.show(vis1)
```

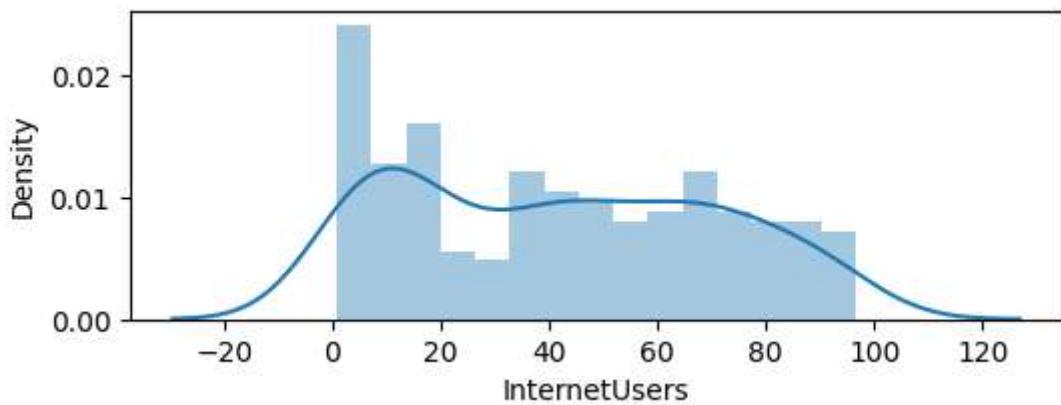


In [213...]

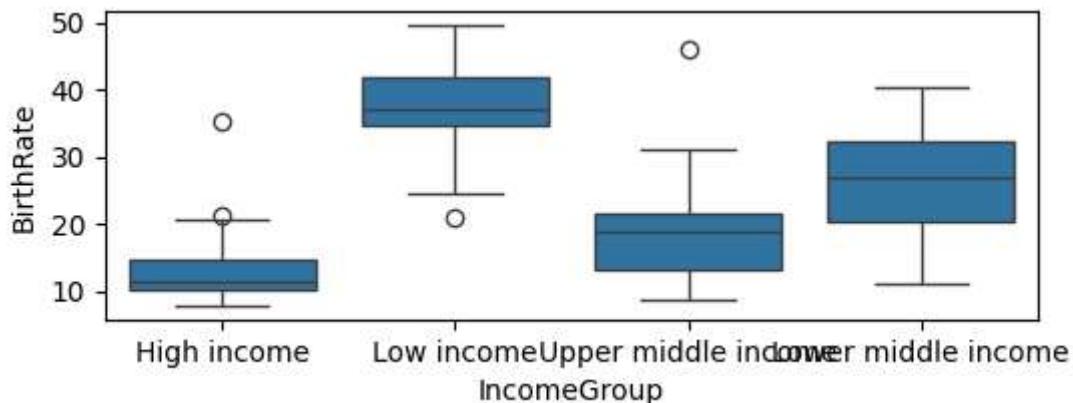
```
vis2 = sns.displot(df["InternetUsers"])
plt.show(vis2)
```



```
In [216]: vis3 = sns.distplot(df["InternetUsers"], bins = 15)  
plt.show(vis3)
```



```
In [218]: # Box plots:  
vis4 = sns.boxplot(data=df, x="IncomeGroup", y="BirthRate")  
plt.show(vis4)
```



In [220...]

df

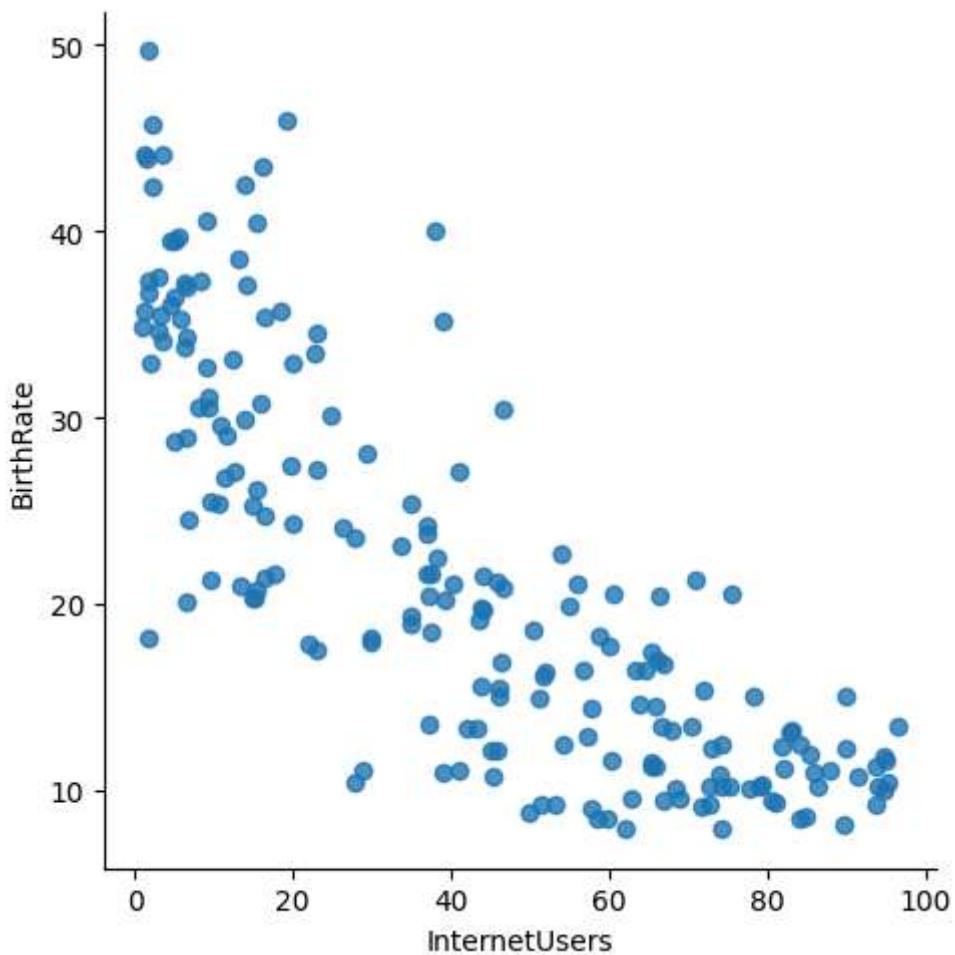
Out[220...]

	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

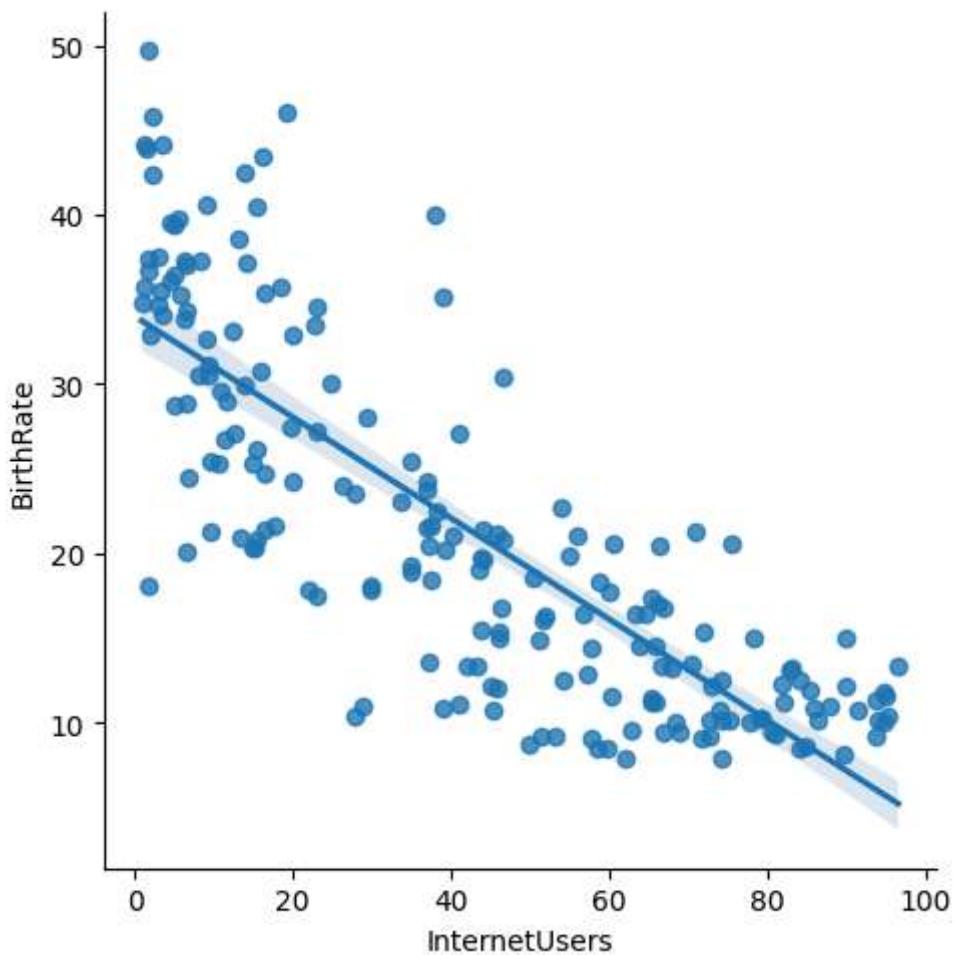
195 rows × 5 columns

In [222...]

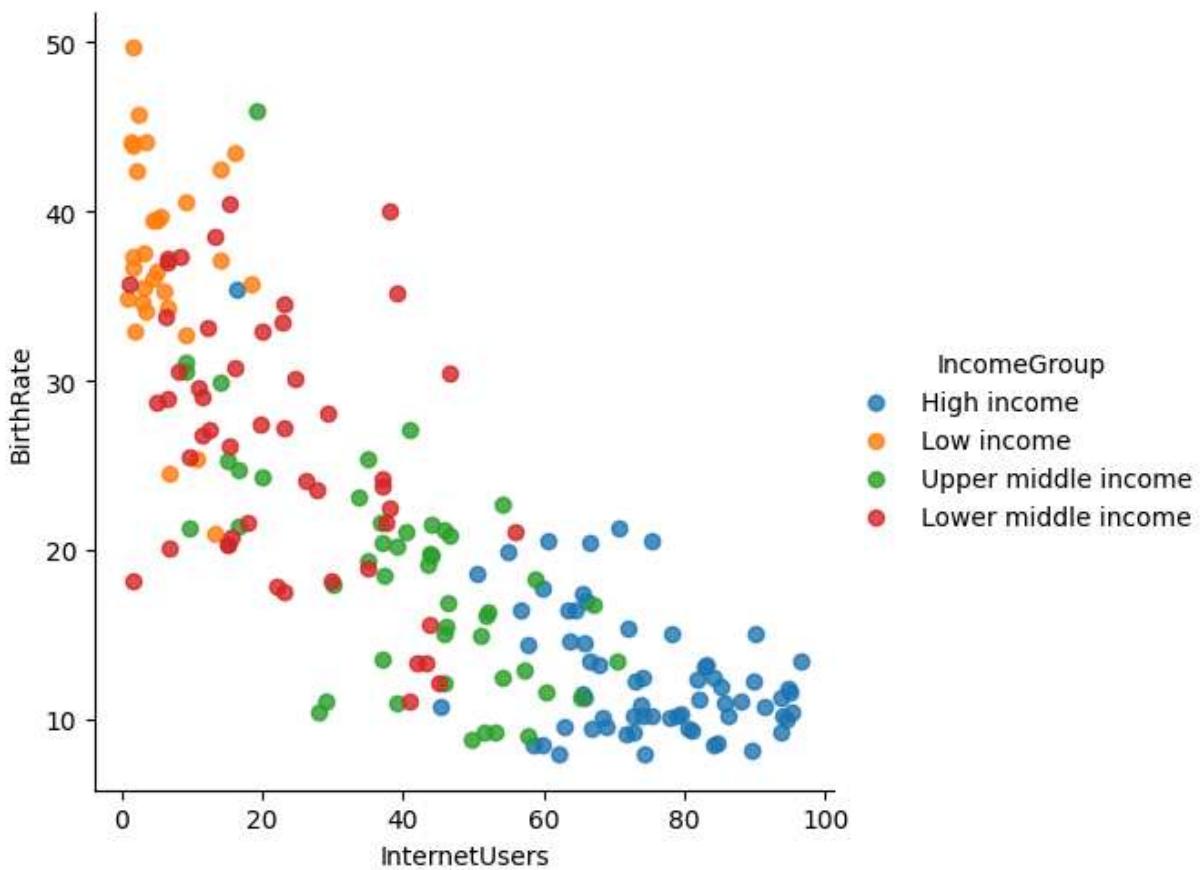
```
vis5 = sns.lmplot(data = df,x = "InternetUsers", y = "BirthRate",fit_reg = False)#
plt.show()
```



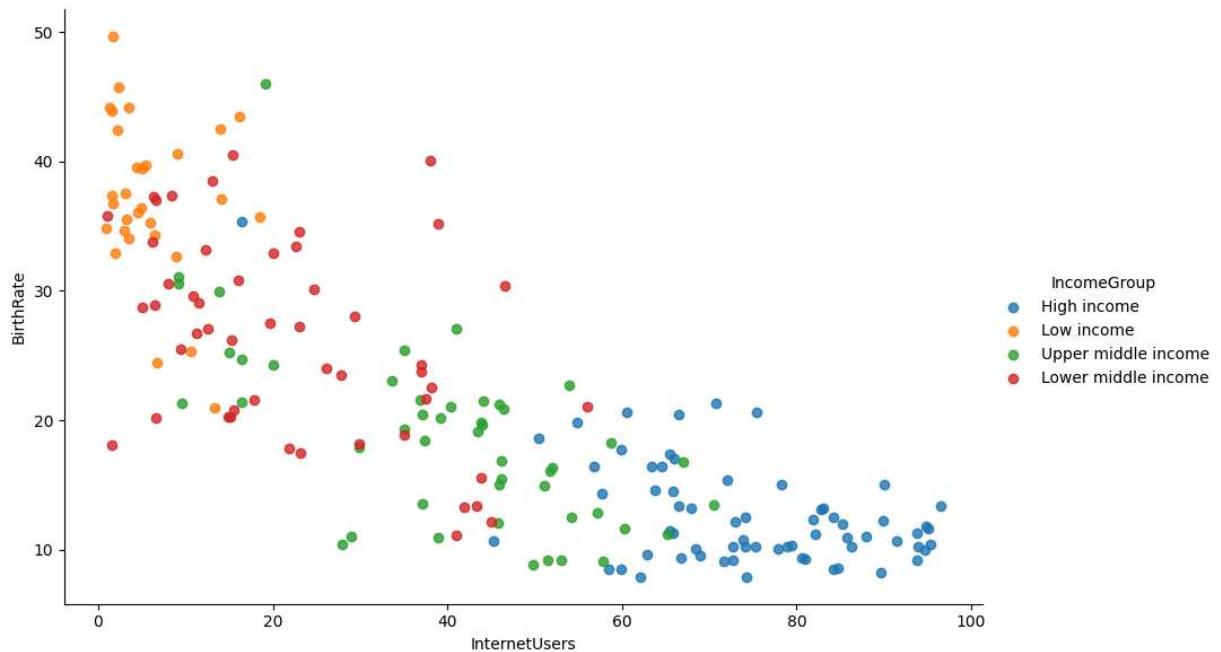
```
In [224]: vis6 = sns.lmplot(data = df,x = 'InternetUsers', y = 'BirthRate')  
plt.show()
```



```
In [226]: vis6 = sns.lmplot(data = df , x = 'InternetUsers' , y = 'BirthRate',fit_reg = False,  
plt.show()
```



```
In [233]: vis6 = sns.lmplot(data = df , x = 'InternetUsers' , y = 'BirthRate', fit_reg = False, plt.show()
```



In this section 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 8> filtering data frames 9> seaborn introduction

In []: