69x0hkkur

March 24, 2025

[6]:
```
import pandas as pd
```

[8]:
```
pd.__version__
```

[8]:
```
'2.2.2'
```

[10]:
```
df=pd.read_csv(r"C:\Users\Dhanwantari Devre\OneDrive\Engg\HYD_CLASS\data.csv")
```

[12]:
```
df
```

[12]:

	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 x 5 columns]

[14]:
```
len(df)
```

[14]:
```
195
```

[16]:
```
df.shape
```

[16]:
```
(195, 5)
```

[17]:
```
df.columns
```

[17]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

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

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

[19]:
```
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
```

[20]:
```
len(df.columns)
```

[20]:
```
5
```

[21]:
```
df.head()
```

[21]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	
1	Afghanistan	AFG	35.253	5.9	
2	Angola	AGO	45.985	19.1	
3	Albania	ALB	12.877	57.2	
4	United Arab Emirates	ARE	11.044	88.0	

```
0          High income
1           Low income
2  Upper middle income
3  Upper middle income
4          High income
```

[22]:
```
df.tail()
```

[22]:
```
           CountryName CountryCode  BirthRate  InternetUsers  \
190       Yemen, Rep.          YEM     32.947           20.0
191      South Africa          ZAF     20.850           46.5
192  Congo, Dem. Rep.          COD     42.394            2.2
193            Zambia          ZMB     40.471           15.4
194          Zimbabwe          ZWE     35.715           18.5

             IncomeGroup
190  Lower middle income
191  Upper middle income
192           Low income
193  Lower middle income
194           Low income
```

[23]:
```
df.info
```

[23]:
```
<bound method DataFrame.info of              CountryName CountryCode  BirthRate
InternetUsers  \
0                  Aruba         ABW     10.244           78.9
1            Afghanistan         AFG     35.253            5.9
2                 Angola         AGO     45.985           19.1
3                Albania         ALB     12.877           57.2
4    United Arab Emirates         ARE     11.044           88.0
..                    …         …       …                …
190          Yemen, Rep.         YEM     32.947           20.0
191         South Africa         ZAF     20.850           46.5
192     Congo, Dem. Rep.         COD     42.394            2.2
193               Zambia         ZMB     40.471           15.4
194             Zimbabwe         ZWE     35.715           18.5

             IncomeGroup
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

[195 rows x 5 columns]>
```

[24]:
```
df.columns
```

[24]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

[25]:
```
len(df.columns)
```

[25]:
```
5
```

[26]:
```
df.tail(2)
```

[26]:
```
    CountryName CountryCode  BirthRate  InternetUsers          IncomeGroup
193      Zambia         ZMB     40.471           15.4  Lower middle income
194    Zimbabwe         ZWE     35.715           18.5           Low income
```

[27]:
```
df.head(2)
```

[27]:
```
   CountryName CountryCode  BirthRate  InternetUsers  IncomeGroup
0        Aruba         ABW     10.244           78.9  High income
1  Afghanistan         AFG     35.253            5.9   Low income
```

[28]:
```
df[::-1]
```

[28]:
```
              CountryName CountryCode  BirthRate  InternetUsers  \
194              Zimbabwe         ZWE     35.715           18.5
193                Zambia         ZMB     40.471           15.4
192      Congo, Dem. Rep.         COD     42.394            2.2
191          South Africa         ZAF     20.850           46.5
190          Yemen, Rep.          YEM     32.947           20.0
..                    …          …          …              …
4    United Arab Emirates         ARE     11.044           88.0
3                 Albania         ALB     12.877           57.2
2                  Angola         AGO     45.985           19.1
1            Afghanistan          AFG     35.253            5.9
0                  Aruba          ABW     10.244           78.9

             IncomeGroup
194           Low income
193  Lower middle income
192           Low income
```

```
191  Upper middle income
190  Lower middle income
..                   …
4             High income
3    Upper middle income
2    Upper middle income
1              Low income
0             High income

[195 rows x 5 columns]
```

[29]:
```
df[:5]
```

[29]:
```
             CountryName CountryCode  BirthRate  InternetUsers  \
0                  Aruba         ABW     10.244           78.9
1            Afghanistan         AFG     35.253            5.9
2                 Angola         AGO     45.985           19.1
3                Albania         ALB     12.877           57.2
4  United Arab Emirates         ARE     11.044           88.0

           IncomeGroup
0         High income
1          Low income
2  Upper middle income
3  Upper middle income
4         High income
```

[30]:
```
df[6:]
```

[30]:
```
              CountryName CountryCode  BirthRate  InternetUsers  \
6                 Armenia         ARM     13.308        41.9000
7    Antigua and Barbuda          ATG     16.447        63.4000
8               Australia         AUS     13.200        83.0000
9                 Austria         AUT      9.400        80.6188
10             Azerbaijan         AZE     18.300        58.7000
..                    …         …          …             …
190          Yemen, Rep.         YEM     32.947        20.0000
191         South Africa         ZAF     20.850        46.5000
192     Congo, Dem. Rep.         COD     42.394         2.2000
193               Zambia         ZMB     40.471        15.4000
194             Zimbabwe         ZWE     35.715        18.5000

              IncomeGroup
6    Lower middle income
7             High income
8             High income
9             High income
```

```
10   Upper middle income
..                   …
190  Lower middle income
191  Upper middle income
192           Low income
193  Lower middle income
194           Low income

[189 rows x 5 columns]
```

[46]:
```
df[0:200:10]
```

[46]:

	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

[48]:
```
df[0:200:50]
```

[48]:

	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

[49]:
```
df.describe() #descriptive analysis
```

[49]:

	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
```

[50]:
```
df.describe().transpose()
```

[50]:
```
                count       mean        std  min      25%    50%      75%  \
BirthRate       195.0  21.469928  10.605467  7.9  12.1205  19.68  29.7595
InternetUsers   195.0  42.076471  29.030788  0.9  14.5200  41.00  66.2250

                    max
BirthRate       49.6610
InternetUsers   96.5468
```

[51]:
```
df.describe().T
```

[51]:
```
                count       mean        std  min      25%    50%      75%  \
BirthRate       195.0  21.469928  10.605467  7.9  12.1205  19.68  29.7595
InternetUsers   195.0  42.076471  29.030788  0.9  14.5200  41.00  66.2250

                    max
BirthRate       49.6610
InternetUsers   96.5468
```

[56]:
```
df.columns
```

[56]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

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

[60]:
```
df.columns
```

[60]:
```
Index(['a', 'b', 'c', 'd', 'e'], dtype='object')
```

[62]:
```
df.head(1)
```

[62]:
```
        a    b       c     d            e
0  Aruba  ABW  10.244  78.9  High income
```

[64]:
```
df.columns = ['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup']
```

[66]:
```
df.head(1)
```

[66]:
```
  CountryName CountryCode  BirthRate  InternetUsers  IncomeGroup
0       Aruba         ABW     10.244           78.9  High income
```

[68]:
```
df.dtypes
```

[68]:
```
CountryName       object
CountryCode       object
BirthRate        float64
InternetUsers    float64
IncomeGroup       object
dtype: object
```

[70]:
```
df[4:8][['CountryName','BirthRate']]
```

[70]:
```
             CountryName  BirthRate
4  United Arab Emirates     11.044
5             Argentina     17.716
6               Armenia     13.308
7   Antigua and Barbuda     16.447
```

[72]:
```
df[['CountryName','BirthRate']][4:8]
```

[72]:
```
             CountryName  BirthRate
4  United Arab Emirates     11.044
5             Argentina     17.716
6               Armenia     13.308
7   Antigua and Barbuda     16.447
```

[74]:
```
df['myCalc']=df.BirthRate * df.InternetUsers
```

[75]:
```
df
```

[75]:
```
               CountryName CountryCode  BirthRate  InternetUsers  \
0                    Aruba         ABW     10.244           78.9
1              Afghanistan         AFG     35.253            5.9
2                   Angola         AGO     45.985           19.1
3                  Albania         ALB     12.877           57.2
4    United Arab Emirates         ARE     11.044           88.0
..                     …         …       …              …
190           Yemen, Rep.         YEM     32.947           20.0
191          South Africa         ZAF     20.850           46.5
192      Congo, Dem. Rep.         COD     42.394            2.2
193                Zambia         ZMB     40.471           15.4
194              Zimbabwe         ZWE     35.715           18.5
```

```
              IncomeGroup    myCalc
0             High income  808.2516
1              Low income  207.9927
2    Upper middle income  878.3135
3    Upper middle income  736.5644
4             High income  971.8720
..                    …       …
190  Lower middle income  658.9400
191  Upper middle income  969.5250
192           Low income   93.2668
193  Lower middle income  623.2534
194           Low income  660.7275

[195 rows x 6 columns]
```

[76]:
```
df.columns
```

[76]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup', 'myCalc'],
      dtype='object')
```

[77]:
```
len(df.columns)
```

[77]:
```
6
```

[78]:
```
df=df.drop('myCalc',axis=1)
```

[84]:
```
df.columns
```

[84]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

[86]:
```
df.InternetUsers<2
```

[86]:
```
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
```

[98]:
```
df.InternetUsers
```

[98]:
```
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
```

[90]:
```
Filter=df.InternetUsers<2
```

[90]:
```
df[Filter]
```

[90]:

	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

[92]:
```
Filter2= df.BirthRate>40
```

[94]:
```
Filter2
```

[94]:
```
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
```

[96]:
```
Filter & Filter2
```

[96]:
```
0      False
1      False
2      False
3      False
4      False
       …
190    False
191    False
192    False
193    False
194    False
Length: 195, dtype: bool
```

[98]:
```
df[Filter & Filter2]
```

[98]:

	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

[108]:
```
df[df.IncomeGroup == 'High income']
```

[108]:

	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 x 5 columns]

[110]:
```
df.IncomeGroup.nunique()
```

[110]:
```
4
```

[114]:
```
#for unique categories
df.IncomeGroup.unique()
```

[114]:
```
array(['High income', 'Low income', 'Upper middle income',
       'Lower middle income'], dtype=object)
```

[100]:
```
df[(df.BirthRate>40) & (df.InternetUsers<2)]
```

[100]:

	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

[104]:
```
df.head()
```

[104]:

	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

[108]:
```
df[df.IncomeGroup=='Low income']
```

[108]:

	CountryName	CountryCode	BirthRate	InternetUsers
1	Afghanistan	AFG	35.253	5.90
11	Burundi	BDI	44.151	1.30
13	Benin	BEN	36.440	4.90
14	Burkina Faso	BFA	40.551	9.10
29	Central African Republic	CAF	34.076	3.50
38	Comoros	COM	34.326	6.50
52	Eritrea	ERI	34.800	0.90
55	Ethiopia	ETH	32.925	1.90
64	Guinea	GIN	37.337	1.60
65	Gambia, The	GMB	42.525	14.00
66	Guinea-Bissau	GNB	37.503	3.10
77	Haiti	HTI	25.345	10.60
93	Cambodia	KHM	24.462	6.80
99	Liberia	LBR	35.521	3.20
111	Madagascar	MDG	34.686	3.00
115	Mali	MLI	44.138	3.50
120	Mozambique	MOZ	39.705	5.40
123	Malawi	MWI	39.459	5.05

```
127                    Niger         NER     49.661          1.70
132                    Nepal         NPL     20.923         13.30
148                   Rwanda         RWA     32.689          9.00
154             Sierra Leone         SLE     36.729          1.70
156                  Somalia         SOM     43.891          1.50
158              South Sudan         SSD     37.126         14.10
167                     Chad         TCD     45.745          2.30
168                     Togo         TGO     36.080          4.50
177                 Tanzania         TZA     39.518          4.40
178                   Uganda         UGA     43.474         16.20
192         Congo, Dem. Rep.         COD     42.394          2.20
194                 Zimbabwe         ZWE     35.715         18.50
```

```
    IncomeGroup
1    Low income
11   Low income
13   Low income
14   Low income
29   Low income
38   Low income
52   Low income
55   Low income
64   Low income
65   Low income
66   Low income
77   Low income
93   Low income
99   Low income
111  Low income
115  Low income
120  Low income
123  Low income
127  Low income
132  Low income
148  Low income
154  Low income
156  Low income
158  Low income
167  Low income
168  Low income
177  Low income
178  Low income
192  Low income
194  Low income
```

[110]:
```
df.IncomeGroup.unique
```

[110]:
```
<bound method Series.unique of 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>
```

[112]:
```
df.IncomeGroup.unique()
```

[112]:
```
array(['High income', 'Low income', 'Upper middle income',
       'Lower middle income'], dtype=object)
```

0.0.1 Seaborn Intro

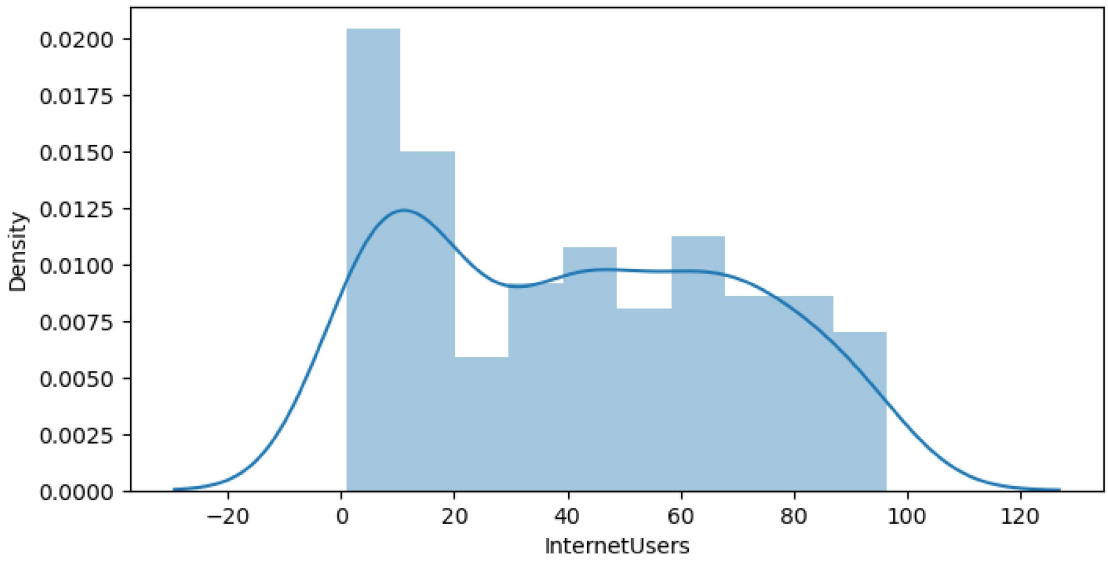
[131]:
```
import matplotlib.pyplot as plt # visualization
import seaborn as sns # distribution visualtion

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

[208]:
```
df.head()
```

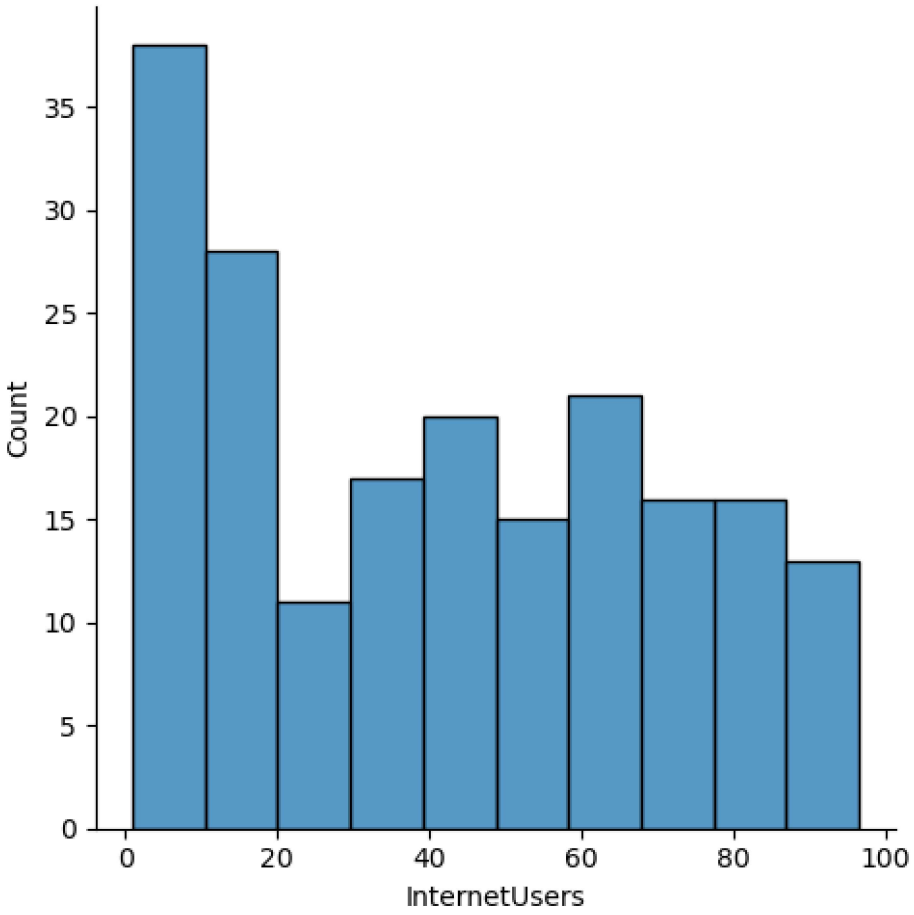
[208]:

	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

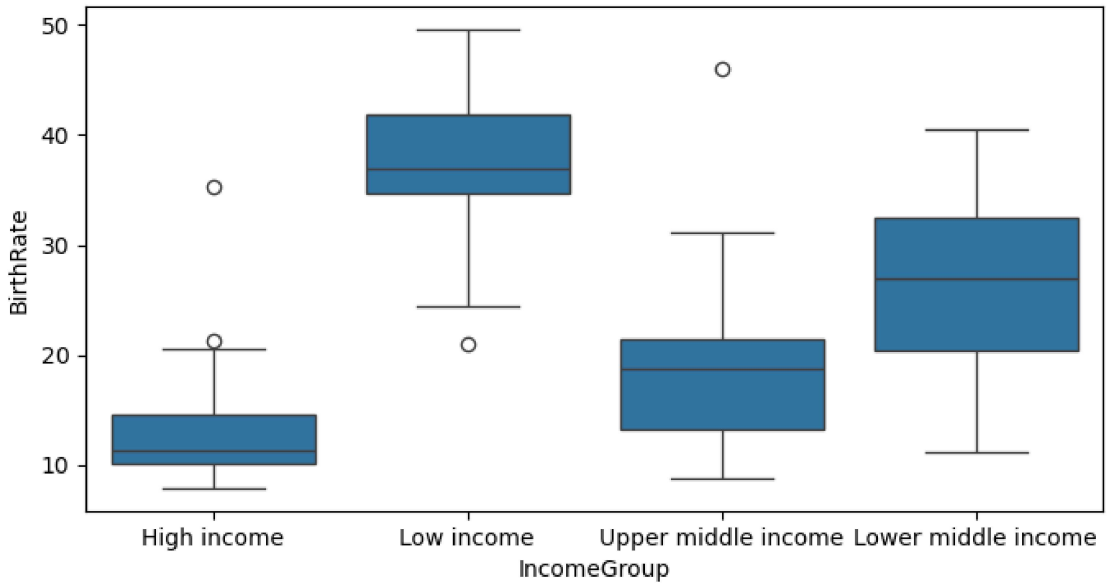
[210]:
```
vis1=sns.distplot(df["InternetUsers"],bins=10)
plt.show()
```



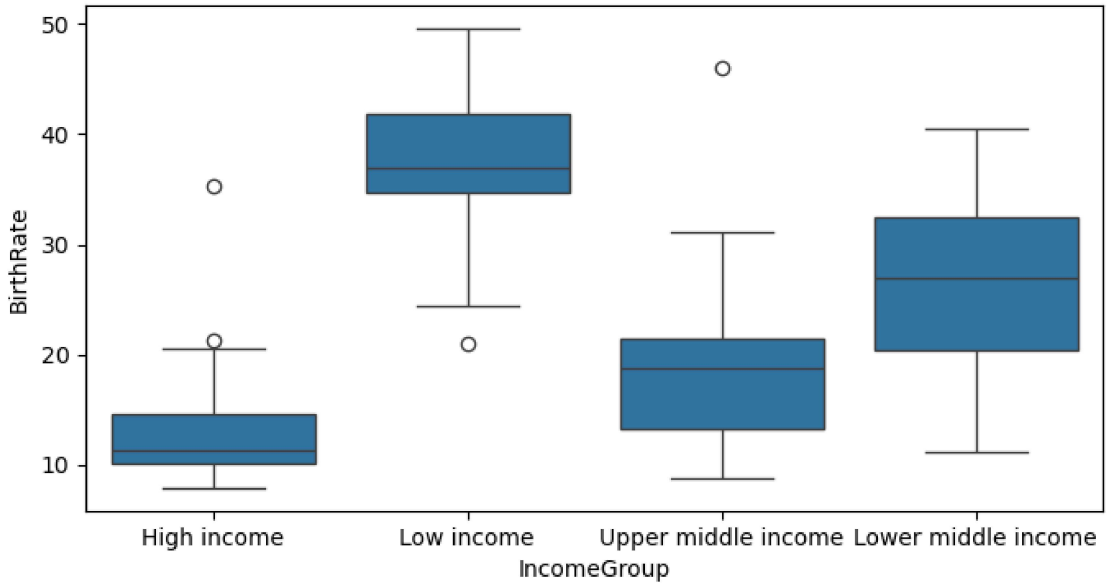
[211]:
```
vis1=sns.displot(df["InternetUsers"],bins=10)
plt.show()
```



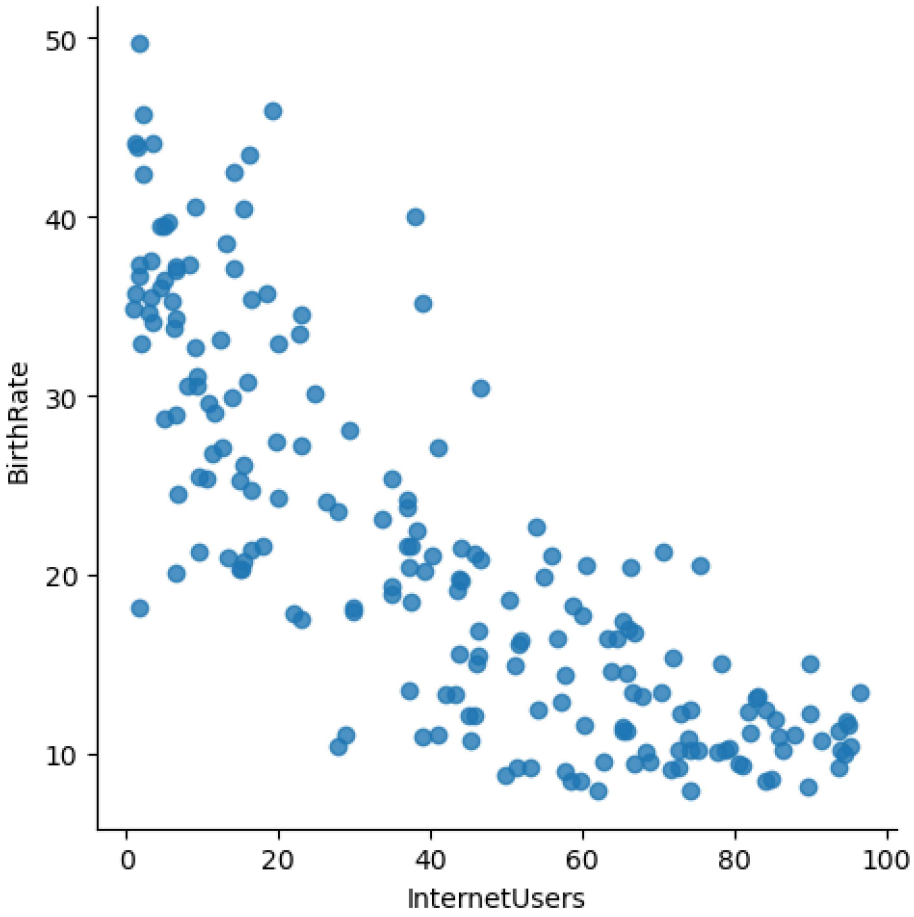
[212]:
```
vis2 = sns.boxplot(data = df, x="IncomeGroup", y='BirthRate')
plt.show()
```



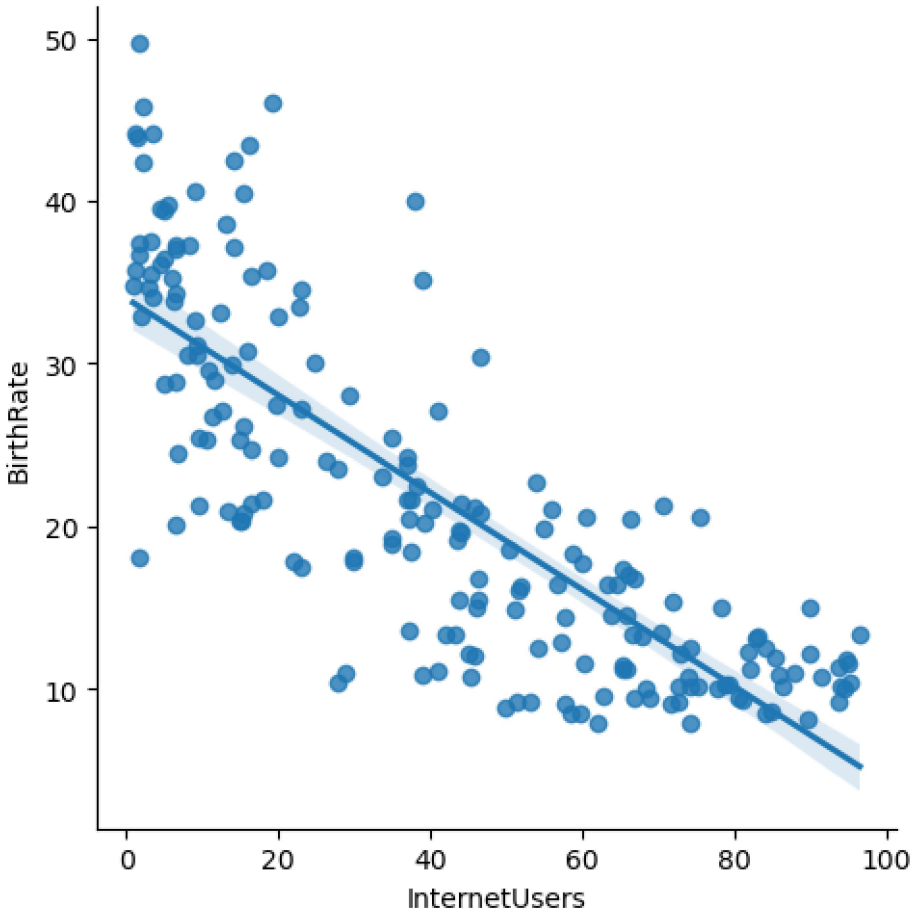
[213]:
```
vis2 = sns.boxplot(data = df, x="IncomeGroup", y='BirthRate')
plt.show()
```



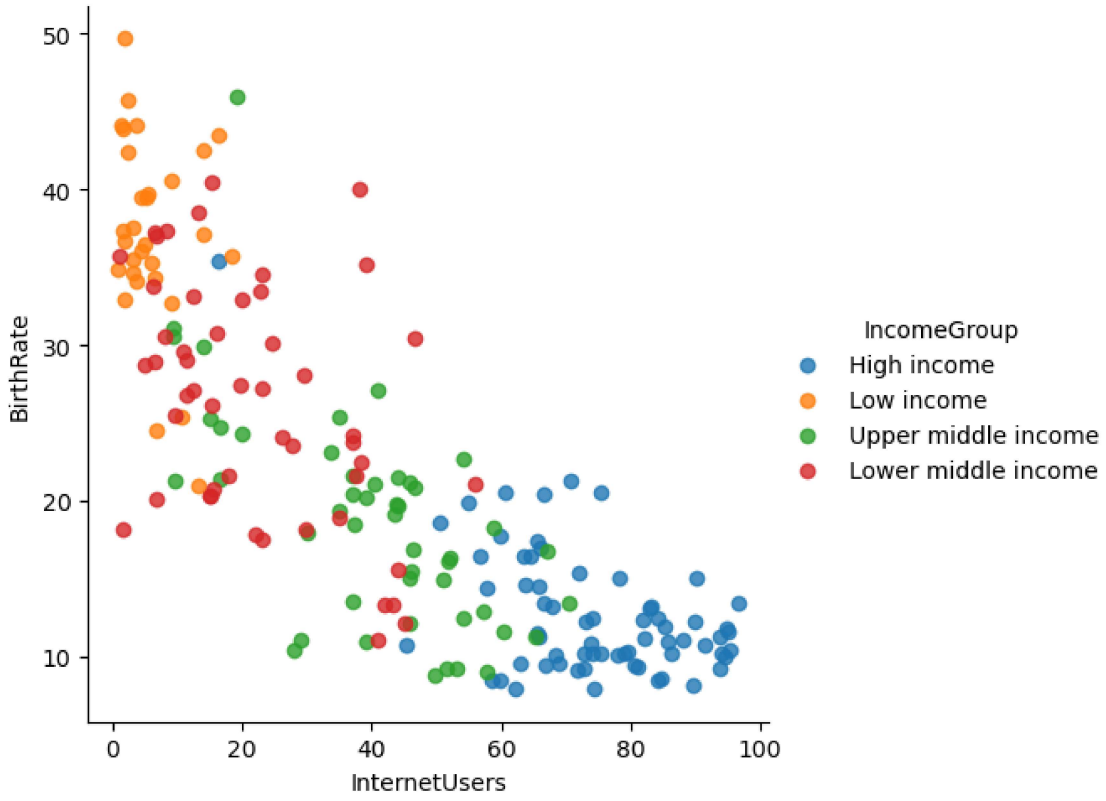
[214]:
```
vis3 = sns.lmplot(data = df,x = 'InternetUsers', y = 'BirthRate', fit_reg =␣
↪False) #lm - linear model
plt.show()
```



[215]:
```
vis4 = sns.lmplot(data = df,x = 'InternetUsers', y = 'BirthRate')
plt.show()
```



[216]:
```
vis5 = sns.lmplot(data = df,x = 'InternetUsers', y = 'BirthRate',fit_reg =␣
↪False,hue = 'IncomeGroup')
plt.show()
```



[219]:
```
vis5 = sns.lmplot(data = df,x = 'InternetUsers', y = 'BirthRate',fit_reg =␣
↪False,hue='IncomeGroup', size=5)
plt.show()
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Cell In[219], line 1
----> 1 vis5 = sns.lmplot(data = df,x = 'InternetUsers', y = 'BirthRate',fit_re␣
↪= False,hue='IncomeGroup', size=5)
      2 plt.show()

TypeError: lmplot() got an unexpected keyword argument 'size'
```

[ ]: