```python
import pandas as pd #importing pandas package

#reading a csv file
data=pd.read_csv(r"D:\Temp\Data Science & Artificial Intelligence\
Completed\Week-3\Class-16 on 07-02-2024 on Wednusday\Notes\7th\7th\
DataFrame_ Pandas\data.csv")

data
```

```
            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]
```

```python
len(data)# total 195 rows
```

```
195
```

```python
data.shape # 195 rows AND 5 columns
```

```
(195, 5)
```

```python
data.columns #Columns
```

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

```python
data.head()
```

```
          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
```

data.dtypes# *datatype of the variables*

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

data.info()# *total information*

```
<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
```

data.isnull().sum()# *no null values*

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

data.head()

```
      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
```

data.describe()

```
        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
```

```python
# importing supported libraries
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

top_20=data[['CountryName','BirthRate']].sort_values(by='BirthRate',as
cending=False).reset_index().drop('index',axis=1).head(20)

plt.rcParams['figure.figsize']=(10,5)
plt.bar(x=top_20.CountryName,height=top_20.BirthRate,width=0.8,ec='red
')
plt.xticks(ticks=range(len(top_20)),labels=top_20.CountryName,rotation
=45)
plt.tick_params(length=15,width=0.2)
plt.title('Top 20 Countries In BirthRate',size=25)
plt.xlabel('Country Name',size=15)
plt.ylabel('Birth Rate',size=15)
plt.show()
```

Top 20 Countries In BirthRate
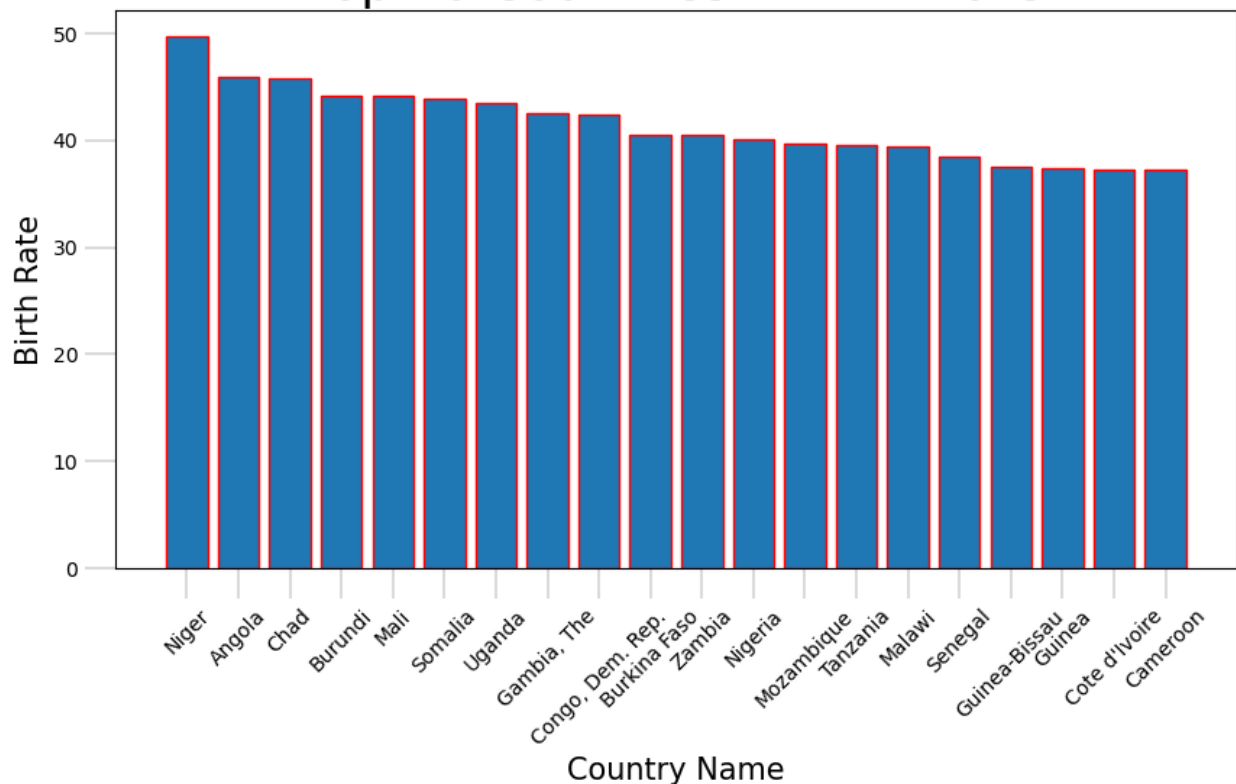

```
# Niger Country Having Top Birth Rate.

least_20=data[['CountryName','BirthRate']].sort_values(by='BirthRate',
ascending=False).reset_index().drop('index',axis=1).tail(20)

plt.bar(x=least_20.CountryName,height=least_20.BirthRate,width=0.8,ec=
'red')
plt.xticks(ticks=range(len(least_20)),labels=least_20.CountryName,rota
tion=80)
plt.tick_params(length=15,width=0.2)
plt.title('Least 20 Countries In BirthRate',size=25)
plt.xlabel('Country Name',size=15)
plt.ylabel('Birth Rate',size=15)
plt.show()
```
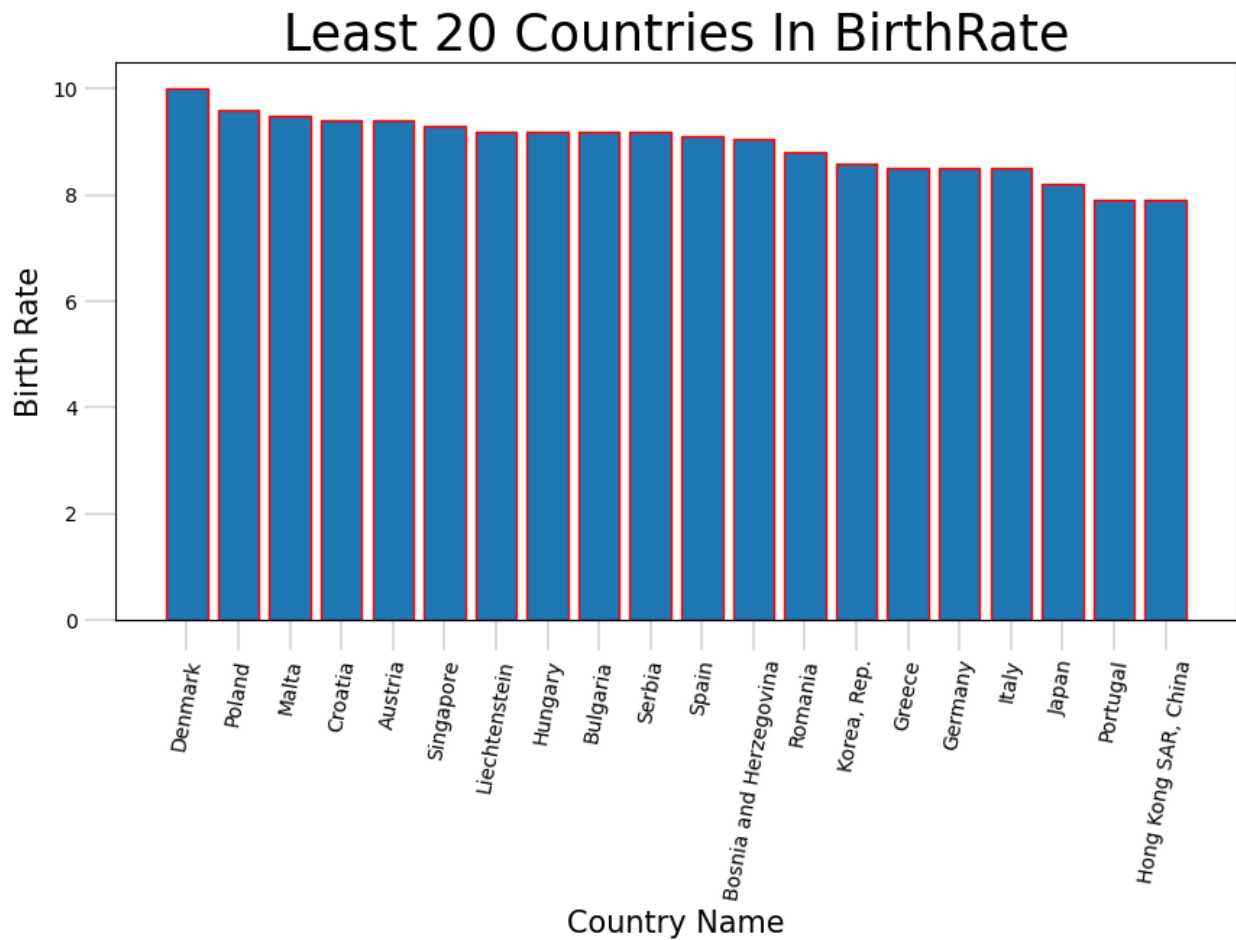
Least 20 Countries In BirthRate


```python
# According to the Hong Kong SAR, China having least Birth Rate.
sns.set_style('darkgrid')
sns.histplot(data=data.BirthRate,bins=10,kde=True,ec='white',alpha=1)
plt.xticks(ticks=range(0,len(data),5))
plt.title('Birth Rate Amoung The Countries',size=25)
plt.xlabel('BirthRate',size=20)
plt.ylabel('Count',size=20)
plt.xlim(0,60)
plt.show()
```

Birth Rate Amoung The Countries
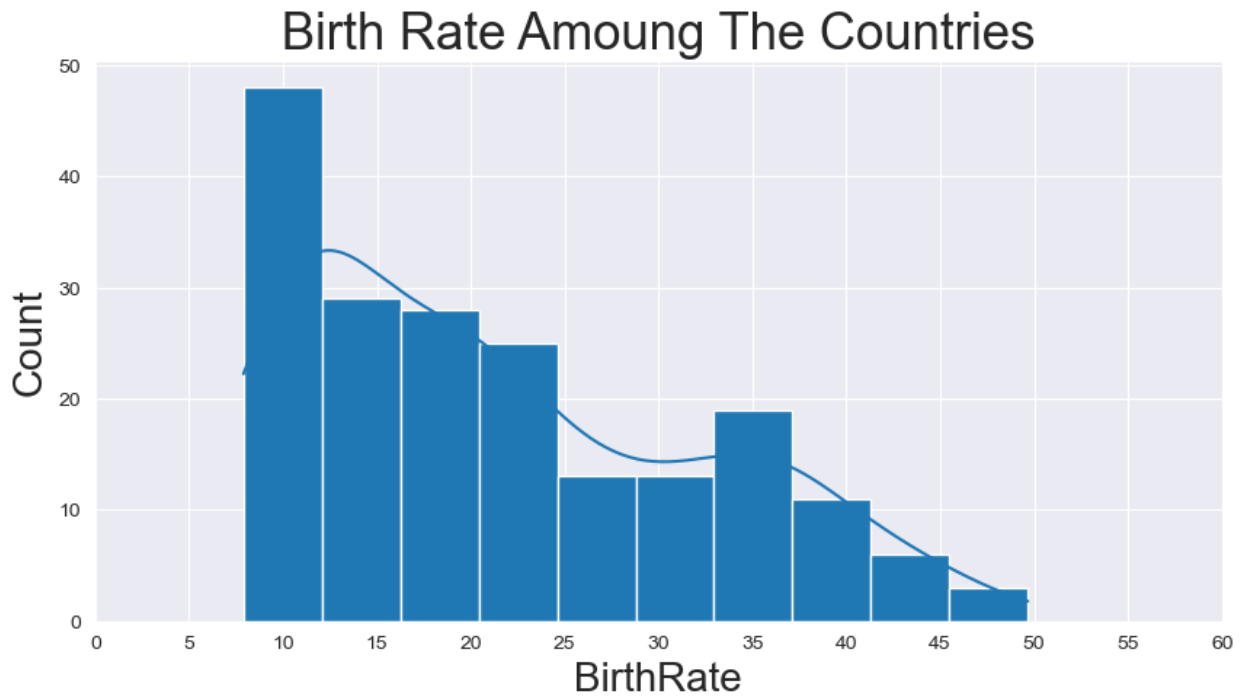
```python
# Amoung all countries nearly 48 countries having low birth rate.
data.head()
```

```
        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
```

```python
data.InternetUsers
```

```
0       78.9
1        5.9
2       19.1
3       57.2
4       88.0
        ...
190     20.0
191     46.5
192      2.2
```
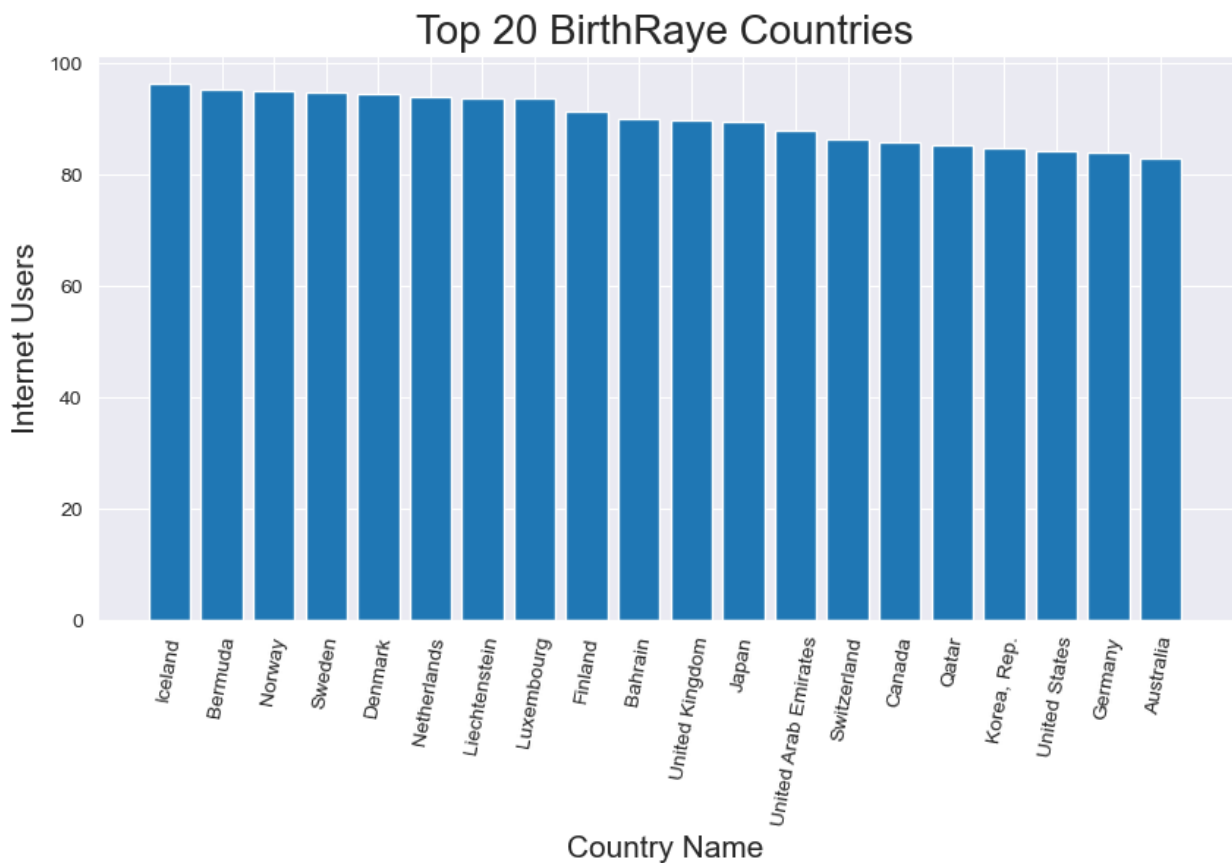
```
193     15.4
194     18.5
Name: InternetUsers, Length: 195, dtype: float64

top20_int_con=data[['CountryName','InternetUsers']].sort_values(by='In
ternetUsers',ascending=False).reset_index().drop('index',axis=1).head(
20)

plt.bar(x=top20_int_con.CountryName,height=top20_int_con.InternetUsers
)
plt.xticks(rotation=80)
plt.title('Top 20 BirthRaye Countries',size=20)
plt.xlabel('Country Name',size=15)
plt.ylabel('Internet Users',size=15)
plt.show()
```


Top 20 BirthRaye Countries

```
# Iceland having highest internet users amoung all the countries.

least20_int_con=data[['CountryName','InternetUsers']].sort_values(by='
InternetUsers',ascending=False).reset_index().drop('index',axis=1).tai
l(20)
```

```
plt.bar(x=least20_int_con.CountryName,height=least20_int_con.InternetU
sers)
plt.xticks(rotation=80)
plt.title('Least 20 Birthrate Countries',size=20)
plt.xlabel('Country Name',size=15)
plt.ylabel('Internet Users',size=15)
plt.show()
```
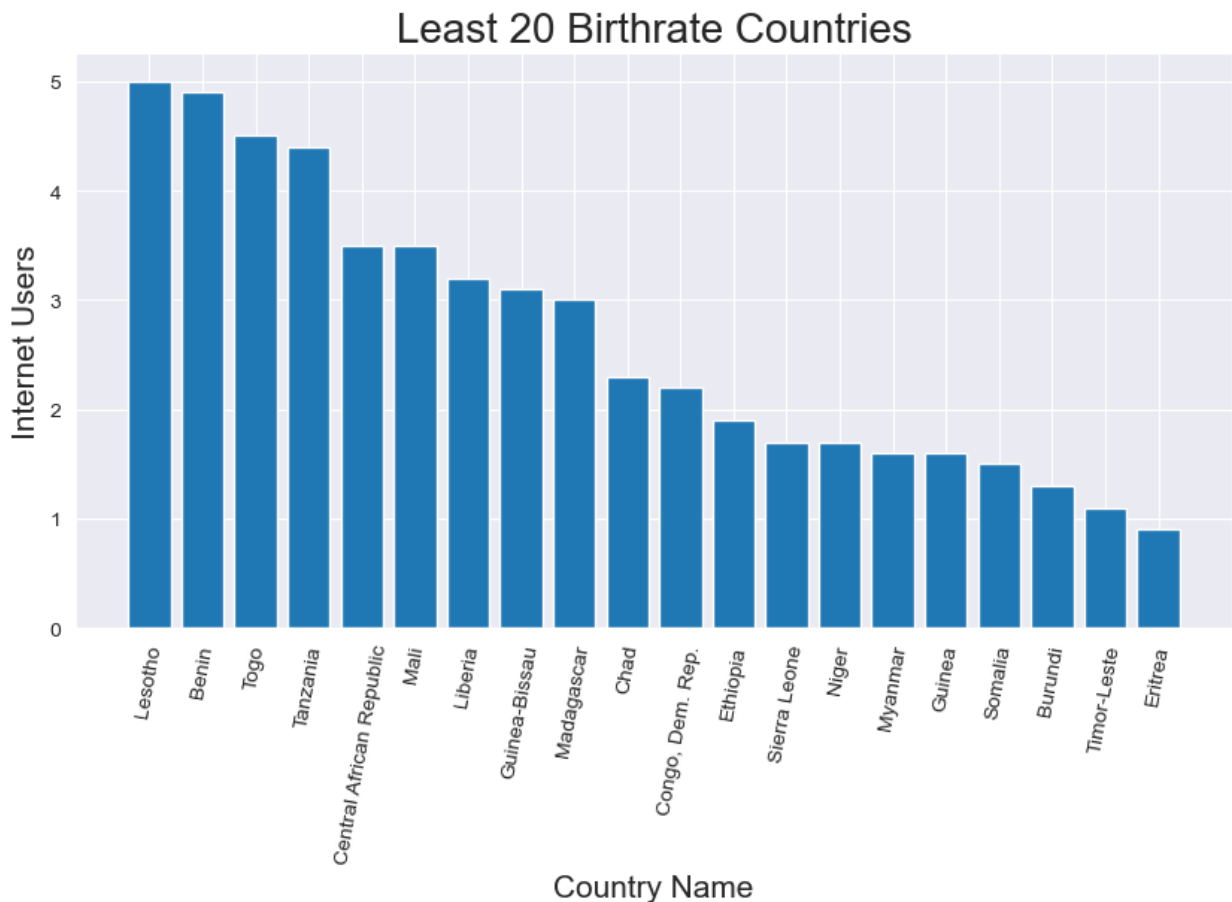


```
# Eritrea has least internet users amoung all the counties.

data.InternetUsers

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
```
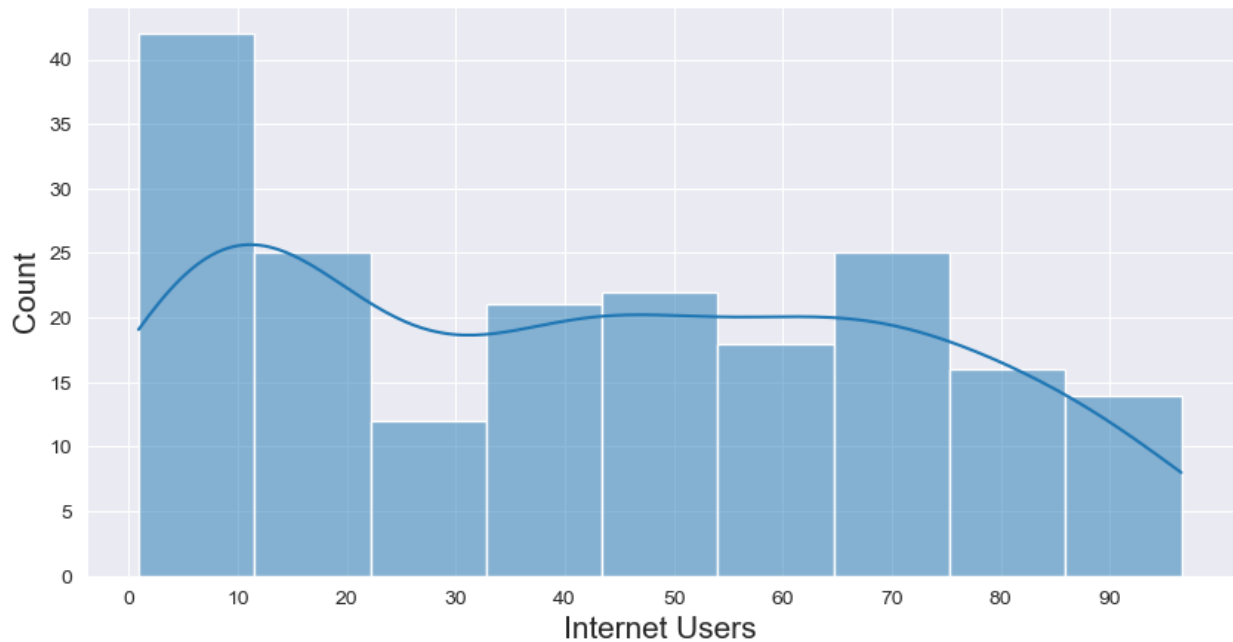
```
194       18.5
Name: InternetUsers, Length: 195, dtype: float64
```

```python
sns.histplot(data=data.InternetUsers,kde=True)
plt.xticks(ticks=range(0,100,10))
plt.xlabel('Internet Users',size=15)
plt.ylabel('Count',size=15)
plt.show()
```



```python
# Amoung all the countries, there are more countires having 0-10%
internet users in there respective countries.

top_20c=data.sort_values(by='BirthRate',ascending=False).head(20)

sns.set_style('ticks')

plt.plot(np.arange(len(top_20c)),top_20c.BirthRate,'-ro',label='Birth
Rate')
plt.plot(np.arange(len(top_20c)),top_20c.InternetUsers,'-
bo',label='Internet Users')
plt.xticks(ticks=range(len(top_20c)),labels=top_20c.CountryName,rotati
on=80)
plt.tick_params(length=5,color='red',axis='both')
plt.grid(which='both',linewidth=20,alpha=0.1,color='yellow')
plt.minorticks_on()
plt.title('Top 20 Countries Analysis',size=20,pad=5)
plt.legend(framealpha=1,edgecolor='red',shadow=True)
plt.show()
```

Top 20 Countries Analysis
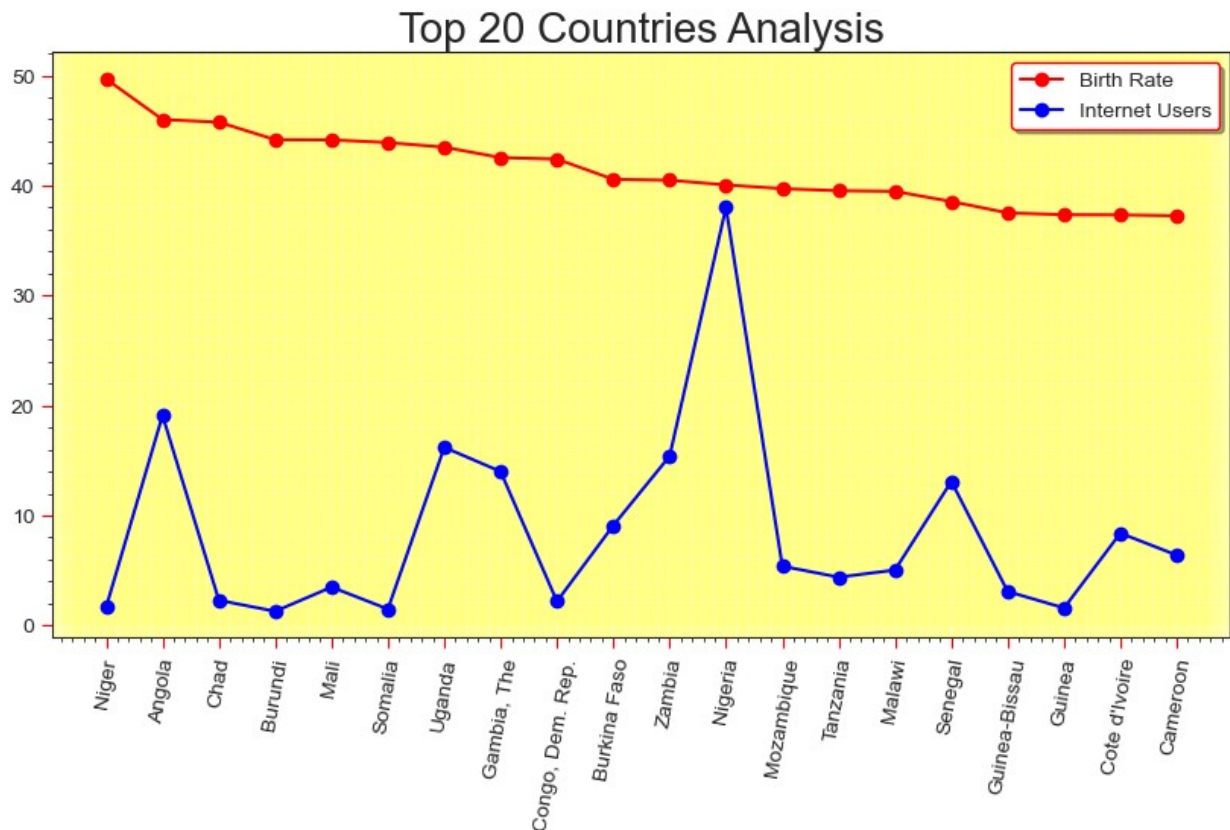

```python
# Accounding to the above plot, even though Niger having highest
birthrate but the internet users are less.
# In Nigeria, almost people are using the internet.

data.head()
```

```
        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
```
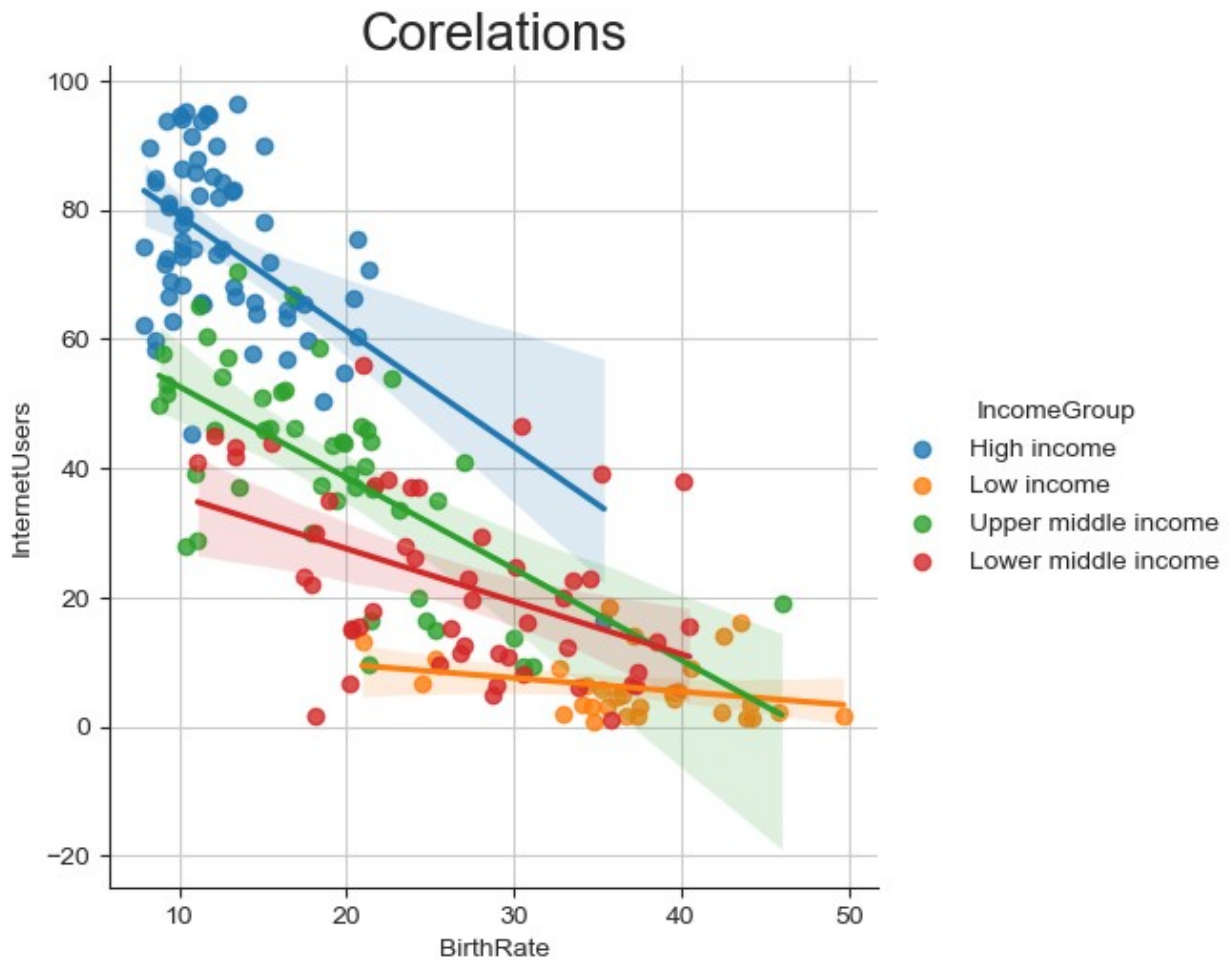
```python
sns.lmplot(data,x='BirthRate',y='InternetUsers')
plt.title('Corelations',size=20)
plt.grid()
plt.show()
```
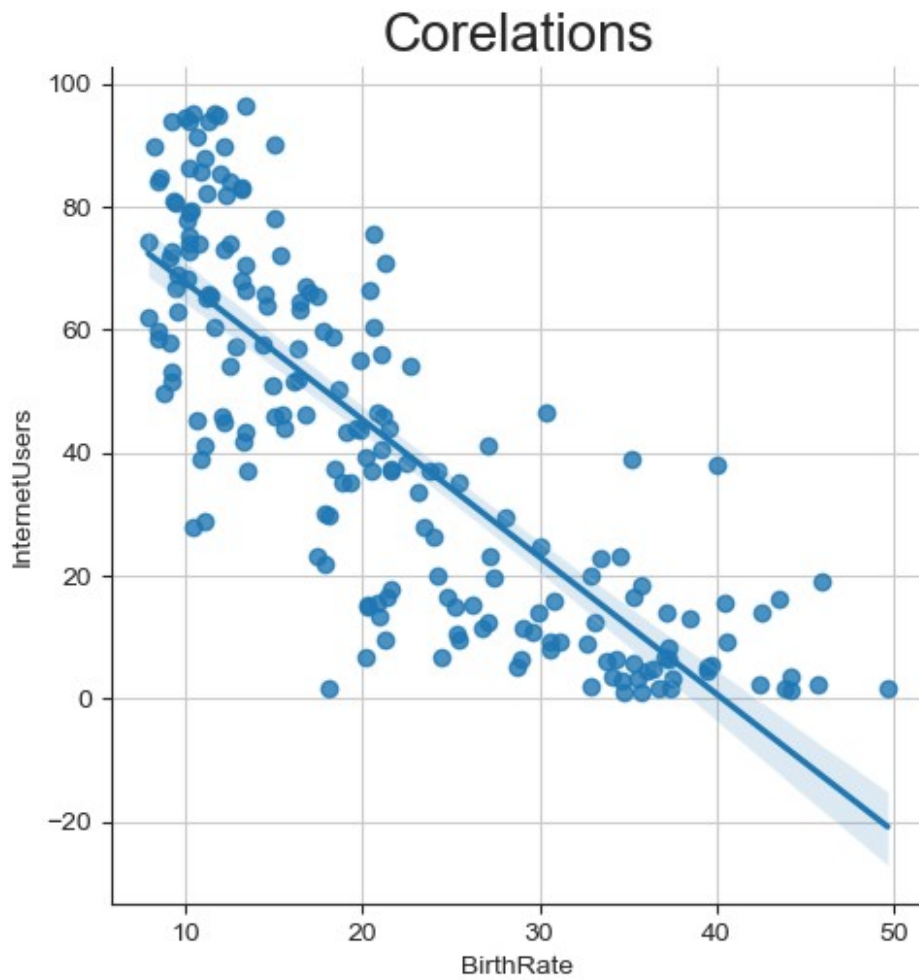
Corelations

```
sns.lmplot(data,x='BirthRate',y='InternetUsers',hue='IncomeGroup')
plt.title('Corelations',size=20)
plt.grid()
plt.show()
```

Corelations

```
sns.boxplot(data,x='IncomeGroup',y='BirthRate',hue='IncomeGroup')
plt.xlabel('Income Group',size=20,labelpad=20)
plt.ylabel('Birth Rate',size=20,labelpad=20)
plt.yticks(ticks=range(0,round(data.BirthRate.max())+10,5))
plt.xticks(fontsize=12)
plt.grid(which='minor',axis='y',alpha=1,linewidth=1,color='yellow')
plt.minorticks_on()
plt.show()
```

```
# From the above graph, in lower income group countris there is more
birthrate.
# we are having some outliers. if we focus on them, then we will get
more information.

data[(data.IncomeGroup=='High income')&(data.BirthRate>30)]# outlier

        CountryName CountryCode  BirthRate  InternetUsers
IncomeGroup
67  Equatorial Guinea        GNQ    35.362            16.4  High
income

# amoung all the highincome countries, Equatorial Guinea Country has
more income and more birth rate why?

data[(data.IncomeGroup=='Low income')&(data.BirthRate<24)]# outlier

    CountryName CountryCode  BirthRate  InternetUsers IncomeGroup
132       Nepal         NPL     20.923           13.3  Low income

# we found low income countries has ,more birthrate. eventhough Nepal
Country has low income but why low birth rate?

data[(data.IncomeGroup=='Upper middle income')&(data.BirthRate>40)]#
outlier

   CountryName CountryCode  BirthRate  InternetUsers
IncomeGroup
2       Angola         AGO     45.985           19.1  Upper middle
income
```
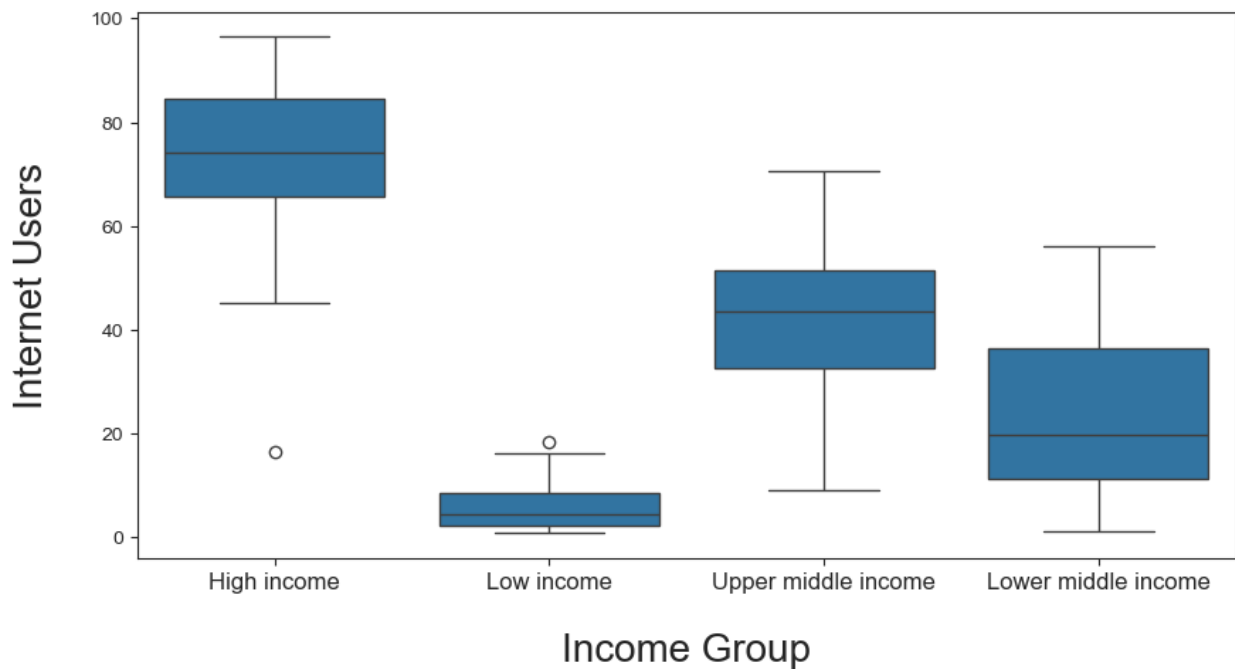
```python
# amoung all the countries in upper middle income group, why Angola
country has more birth rate?

sns.boxplot(data,x='IncomeGroup',y='InternetUsers')
plt.xlabel('Income Group',size=20,labelpad=20)
plt.ylabel('Internet Users',size=20,labelpad=20)
plt.xticks(fontsize=12)
plt.show()
```



```python
# From the above graph, where there is much internet usage there is
much income.

# That income is high because of maybe many factors like online
business...etc.
# Using technology in there respective works..

data[(data.IncomeGroup=='High
income')&(data.InternetUsers<20)]#Outlier
```

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
67	Equatorial Guinea	GNQ	35.362	16.4	High income

```python
# how Equatorial Guinea country in high income group with low internet
users?
# how they generating their income with less internet users?

data.head()
```

```
       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
```

```python
data[['BirthRate','InternetUsers']].corr()
```

```
               BirthRate  InternetUsers
BirthRate       1.000000      -0.815589
InternetUsers  -0.815589       1.000000
```

```python
data['IncomeGroup'].unique()
```

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

```python
data['Income grp num']=data['IncomeGroup'].map({'High income':4,'Upper
middle income':3,'Lower middle income':2,'Low income':1})
```

```python
data[['BirthRate','InternetUsers','Income grp num']].corr()
```

```
                BirthRate  InternetUsers  Income grp num
BirthRate        1.000000      -0.815589       -0.795319
InternetUsers   -0.815589       1.000000        0.869828
Income grp num  -0.795319       0.869828        1.000000
```

```python
# Accounding to our analysis InternetUsers And IncomeGroup are
corelated with eachother
```