```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
import pandas as pd

df=pd.read_csv("/content/drive/MyDrive/FSDS @Kodi Senapati/data.csv")

df.columns
```

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

```
df.shape
```

```
(195, 5)
```

```
df.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

```
# transpose the output of describe

df.describe().transpose()
```

	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

```
print(df)
```

```
             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]
```

```
df.isnull().sum()
```

	0
CountryName	0
CountryCode	0
BirthRate	0
InternetUsers	0
IncomeGroup	0

dtype: int64

```
filter=df.InternetUsers<2
```

```
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

```
#filter the data frame with condition BirthRate>40 and InternetUsers<2
df[(df.BirthRate>40)&(df.InternetUsers<2)]
```

	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

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

	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

```
#Get unique values of an column
print(df.IncomeGroup.unique()) #prints values
print(df.IncomeGroup.nunique()) #prints number of values
```

```
['High income' 'Low income' 'Upper middle income' 'Lower middle income']
4
```

```
# Introduction to seaborn(Statistic Visulaization- random distribution)

import matplotlib.pyplot as plt
import seaborn as sns

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

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

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

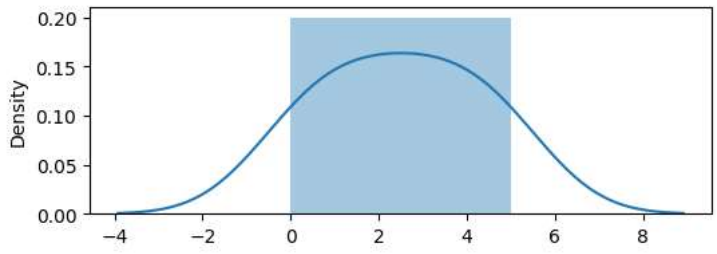
	InternetUsers
0	78.9
1	5.9
2	19.1
3	57.2
4	88.0
...	...
190	20.0
191	46.5
192	2.2
193	15.4
194	18.5

195 rows × 1 columns

dtype: float64

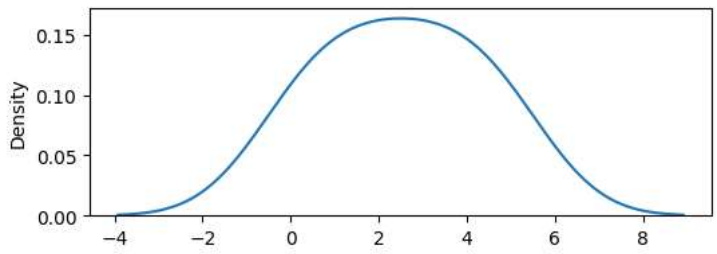
```
#Distplot stands for distribution plot,
#it takes as input an array and plots a curve corresponding to the distribution of points in the array.

sns.distplot([0, 1, 2, 3, 4, 5])
```

```
<Axes: ylabel='Density'>
```



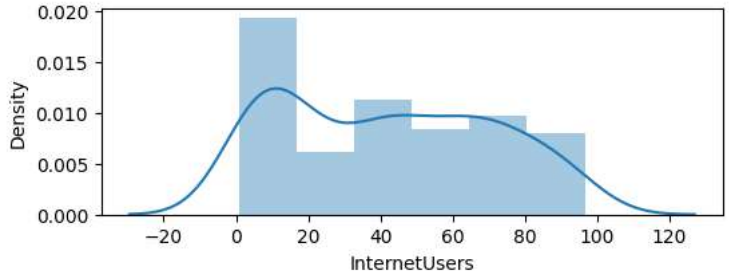

```
#without histogram
sns.distplot([0, 1, 2, 3, 4, 5],hist=False)
```

```
<Axes: ylabel='Density'>
```



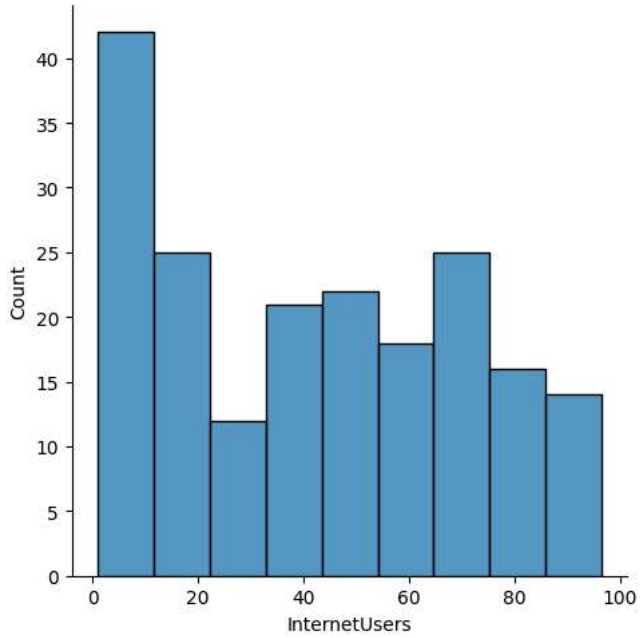

```
#distplot
sns.distplot(df['InternetUsers']) #Univariate analysis
```

```
<Axes: xlabel='InternetUsers', ylabel='Density'>
```



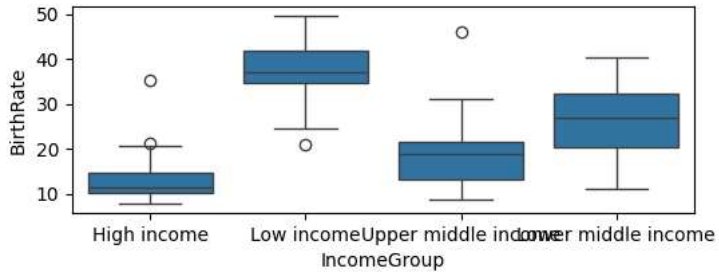

```
#displot
sns.displot(df['InternetUsers']) #Univariate analysis
```

```
<seaborn.axisgrid.FacetGrid at 0x7d9c1815b7f0>
```



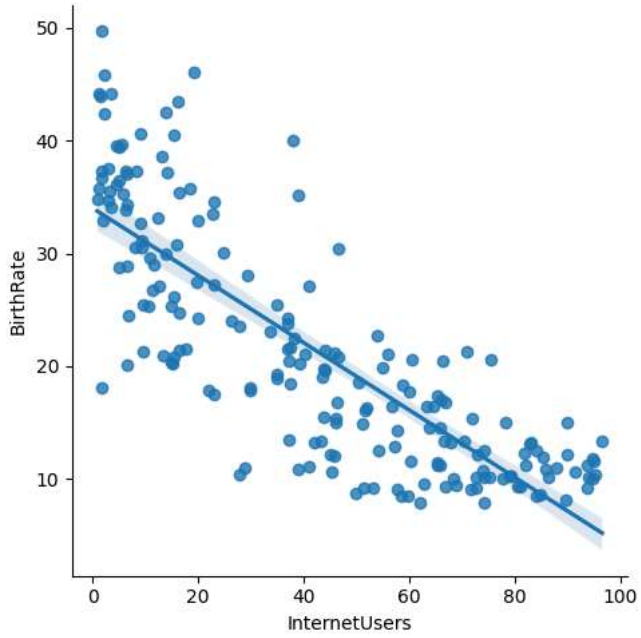

```
#box plot
sns.boxplot(data=df,x='IncomeGroup',y='BirthRate') #bivariate analysis
```

```
<Axes: xlabel='IncomeGroup', ylabel='BirthRate'>
```



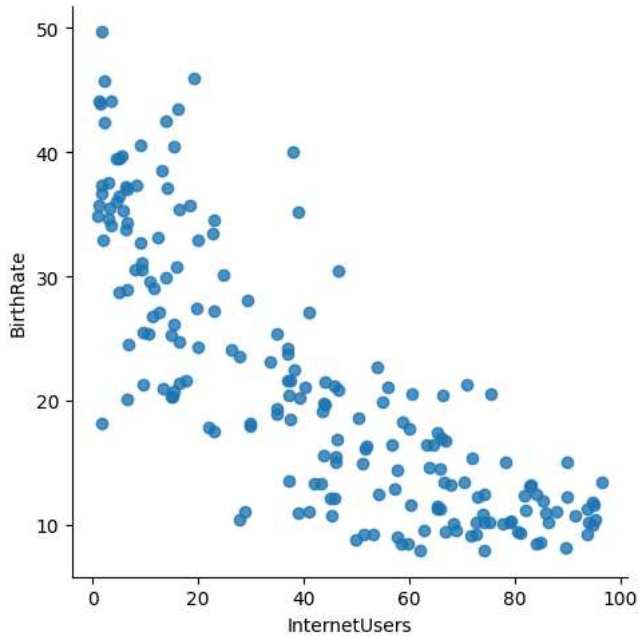

```
#lmplot-linear model plot
sns.lmplot(data=df,x='InternetUsers',y='BirthRate')
```

```
<seaborn.axisgrid.FacetGrid at 0x7d9c1a3fca60>
```



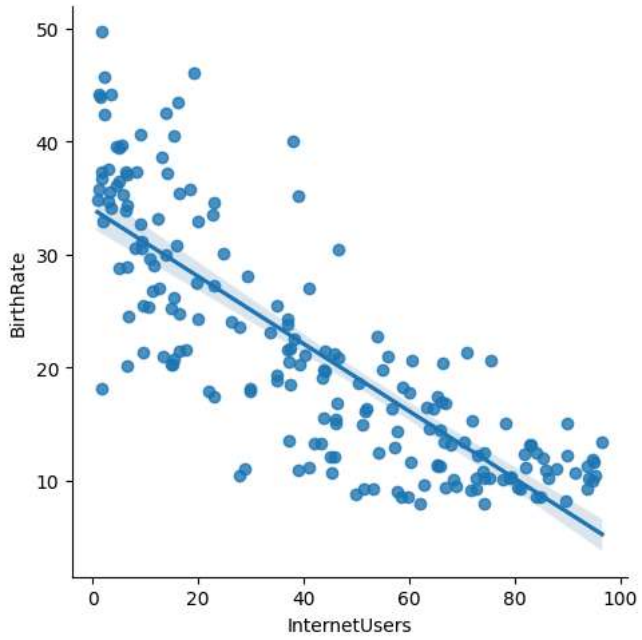

```
sns.lmplot(data=df,x='InternetUsers',y='BirthRate',fit_reg=False)
```

```
<seaborn.axisgrid.FacetGrid at 0x7d9c18158d90>
```



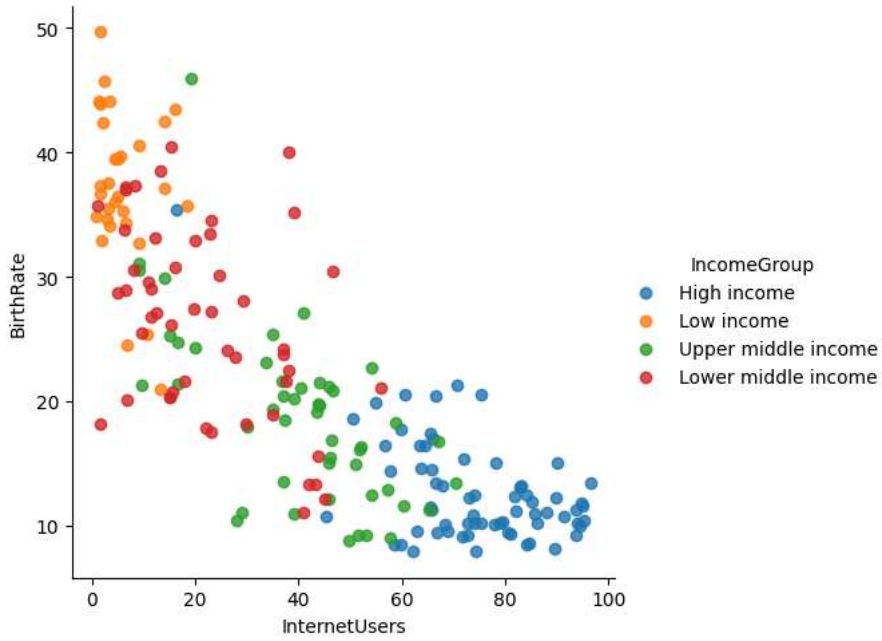
```
sns.lmplot(data=df,x='InternetUsers',y='BirthRate',fit_reg=True)
```

```
<seaborn.axisgrid.FacetGrid at 0x7d9c17fac9a0>
```




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

```
<seaborn.axisgrid.FacetGrid at 0x7d9c18159180>
```




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

```
<seaborn.axisgrid.FacetGrid at 0x7d9c163a09a0>
```