

PANDAS

Analysing GDP growth

```
In [1]: import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```

```
In [2]: stats = pd.read_excel(r'/Users/sasidharbhagavatula/Desktop/data.xlsx')
```

```
In [3]: stats
```

```
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 [4]: len(stats)
```

```
Out[4]: 195
```

```
In [5]: stats.columns
```

```
Out[5]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
   'IncomeGroup'],
   dtype='object')
```

```
In [6]: len(stats.columns)
```

```
Out[6]: 5
```

```
In [7]: stats.shape
```

```
Out[7]: (195, 5)
```

```
In [8]: stats.isnull()
```

```
Out[8]:
```

	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 [9]: stats.isna()
```

Out[9]:

	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 [10]: `stats.isnull().sum()`

```
Out[10]: CountryName      0
          CountryCode     0
          BirthRate       0
          InternetUsers   0
          IncomeGroup     0
          dtype: int64
```

In [11]: `stats.dtypes`

```
Out[11]: CountryName      object
          CountryCode     object
          BirthRate        float64
          InternetUsers   float64
          IncomeGroup      object
          dtype: object
```

In [12]: `stats.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
```

```
In [13]: type(stats)
```

```
Out[13]: pandas.core.frame.DataFrame
```

```
In [14]: # top 5 rows
stats.head()
```

```
Out[14]:
```

	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 [15]: stats.tail()
```

```
Out[15]:
```

	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 [16]: stats.tail(1)
```

```
Out[16]:    CountryName  CountryCode  BirthRate  InternetUsers  IncomeGroup
194      Zimbabwe        ZWE     35.715       18.5  Low income
```

```
In [17]: stats.columns
```

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

```
In [18]: stats['CountryName']
```

```
Out[18]: 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 [19]: stats['BirthRate']
```

```
Out[19]: 0      10.244
1      35.253
2      45.985
3      12.877
4      11.044
...
190     32.947
191     20.850
192     42.394
193     40.471
194     35.715
Name: BirthRate, Length: 195, dtype: float64
```

```
In [20]: stats[['BirthRate', 'CountryName']]
```

```
Out[20]:
```

	BirthRate	CountryName
0	10.244	Aruba
1	35.253	Afghanistan
2	45.985	Angola
3	12.877	Albania
4	11.044	United Arab Emirates
...
190	32.947	Yemen, Rep.
191	20.850	South Africa
192	42.394	Congo, Dem. Rep.
193	40.471	Zambia
194	35.715	Zimbabwe

195 rows × 2 columns

```
In [21]: stats_numeric_data= stats[['BirthRate','InternetUsers']]
```

```
In [22]: stats_numeric_data.head()
```

```
Out[22]:
```

	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 [72]: stats_categorical_data= stats[['CountryName','CountryCode','InternetUsers']]
```

```
In [73]: stats_categorical_data.head()
```

Out[73]:

	CountryName	CountryCode	InternetUsers
0	Aruba	ABW	78.9
1	Afghanistan	AFG	5.9
2	Angola	AGO	19.1
3	Albania	ALB	57.2
4	United Arab Emirates	ARE	88.0

In [74]:

```
#total variables created
print(stats.shape)
print(stats_numeric_data.shape)
print(stats_categorical_data.shape)
```

```
(195, 5)
(195, 2)
(195, 3)
```

slicing in pandas

In [26]:

```
stats[:]
```

Out[26]:

	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 [27]: #slicing

In [28]: stats[:6]

Out[28]:

	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
5	Argentina	ARG	17.716	59.9	High income

In [29]: stats[3:]

Out [29]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
3	Albania	ALB	12.877	57.2	Upper middle income
4	United Arab Emirates	ARE	11.044	88.0	High income
5	Argentina	ARG	17.716	59.9	High income
6	Armenia	ARM	13.308	41.9	Lower middle income
7	Antigua and Barbuda	ATG	16.447	63.4	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

192 rows × 5 columns

In [30]: `stats[3:10]`

Out [30]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
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

In [31]: `stats[3:55:5]`

Out [31]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
3	Albania	ALB	12.877	57.2000	Upper middle income
8	Australia	AUS	13.200	83.0000	High income
13	Benin	BEN	36.440	4.9000	Low income
18	Bahamas, The	BHS	15.339	72.0000	High income
23	Bolivia	BOL	24.236	36.9400	Lower middle income
28	Botswana	BWA	25.267	15.0000	Upper middle income
33	China	CHN	12.100	45.8000	Upper middle income
38	Comoros	COM	34.326	6.5000	Low income
43	Cyprus	CYP	11.436	65.4548	High income
48	Dominican Republic	DOM	21.198	45.9000	Upper middle income
53	Spain	ESP	9.100	71.6350	High income

In [32]: `stats[::-1]`

Out[32]:

	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 [33]:

stats[-1::]

Out[33]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
194	Zimbabwe	ZWE	35.715	18.5	Low income

In [34]:

stats[0:200]

Out [34]:

	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 [35]: stats

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
...
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 [36]:

stats[:::-1]

Out [36]:

	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 [37]:

stats[:::-1:2]

Out [37]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income
2	Angola	AGO	45.985	19.1	Upper middle income
4	United Arab Emirates	ARE	11.044	88.0	High income
6	Armenia	ARM	13.308	41.9	Lower middle income
8	Australia	AUS	13.200	83.0	High income
...
184	Venezuela, RB	VEN	19.842	54.9	High income
186	Vietnam	VNM	15.537	43.9	Lower middle income
188	West Bank and Gaza	PSE	30.394	46.6	Lower middle income
190	Yemen, Rep.	YEM	32.947	20.0	Lower middle income
192	Congo, Dem. Rep.	COD	42.394	2.2	Low income

97 rows × 5 columns

In [38]:

stats[0:300]

Out[38]:

	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 [39]: `pd.__version__`

Out[39]: '2.2.3'

In []:

In [40]: `# 6th nov`In [41]: `stats.describe() # only numeric attributes`

```
Out[41]:      BirthRate  InternetUsers
```

	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 [69]: stats.describe().transpose()
```

```
Out[69]:
```

	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 [77]: stats_categorical_data.describe()
```

```
Out[77]:
```

	InternetUsers
count	195.000000
mean	42.076471
std	29.030788
min	0.900000
25%	14.520000
50%	41.000000
75%	66.225000
max	96.546800

```
In [90]: stats.head(2)
```

```
Out[90]:
```

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income
1	Afghanistan	AFG	35.253	5.9	Low income

```
In [91]: stats['calc'] = stats['BirthRate'] * stats['InternetUsers']
```

```
In [92]: stats['calc']
```

```
Out[92]: 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
Name: calc, Length: 195, dtype: float64
```

```
In [93]: stats.head()
```

```
Out[93]:   CountryName  CountryCode  BirthRate  InternetUsers  IncomeGroup      calc
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 [94]: stats=stats.drop('calc',axis=1) # drop the attribute from axis 1 column
```

```
In [95]: stats
```

Out[95]:

	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 [96]: `stats['InternetUsers'] < 2 # without filters`

Out[96]:

```
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 [98]: `stats[stats['InternetUsers'] < 2] # with filters`

Out[98]:

	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 [99]: `len(stats[stats['BirthRate']>40])` #no. of countries with birthrate > 4

Out[99]: 12

In [100...]: `stats[(stats.InternetUsers<2) & (stats.BirthRate>40)]` # and operation

Out[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

In [101...]: `len(stats[(stats.InternetUsers<2) & (stats.BirthRate>40)])`

Out[101...]: 3

In [102...]: `stats[stats.IncomeGroup == 'High income']`

Out[102...]

	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 [103...]

stats[stats.IncomeGroup == 'Low income'] #filtering

Out[103...]

	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 [104]: `stats.IncomeGroup.nunique() # no. of uninique entries in income group`

Out[104]: 4

In []:

In []:

In [68]: `stats.BirthRate.nunique() # no`

Out[68]: 177

In [79]: `stats.BirthRate.nunique()`

Out[79]: 177

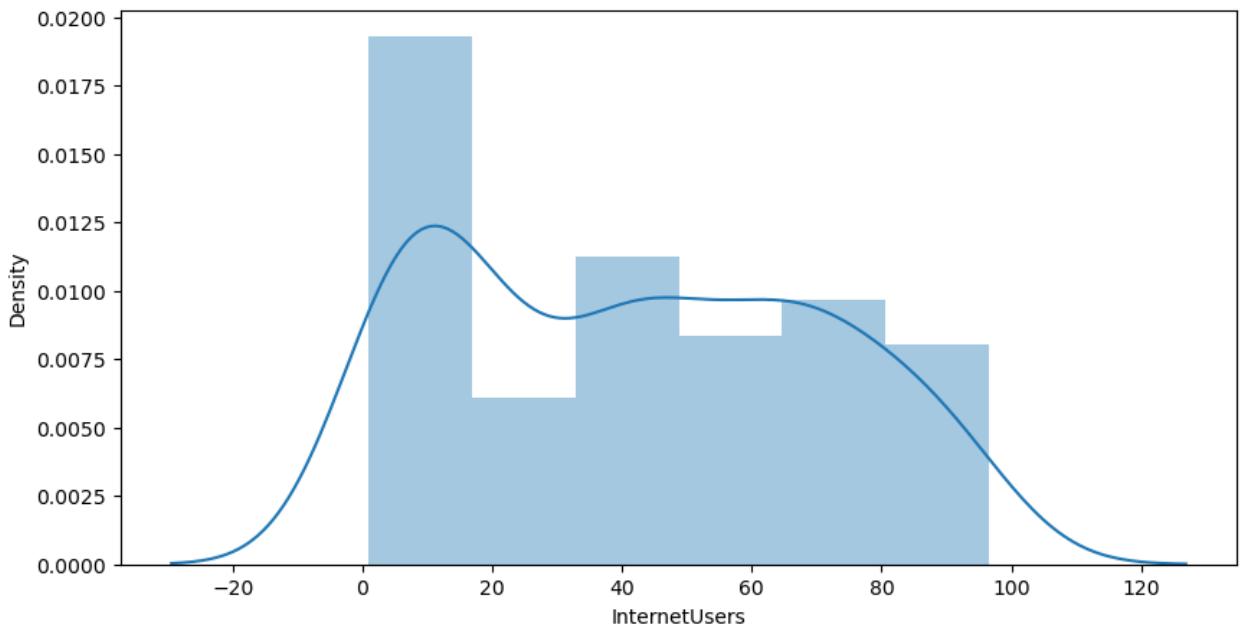
In [106]: `import pandas as pd
#import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib
plt.rcParams['figure.figsize']= 10,5
import warnings
warnings.filterwarnings('ignore')`

Using matplotlib backend: module://matplotlib_inline.backend_inline

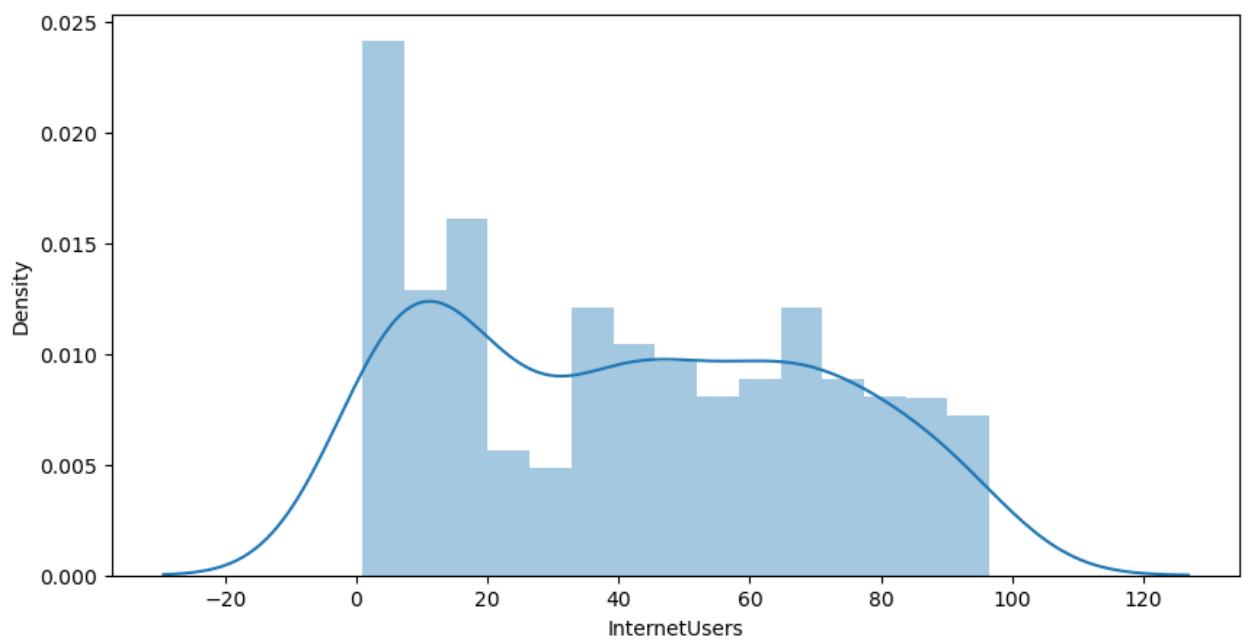
```
In [107]: vis1=plt.distplot(stats['InternetUsers'])
```

```
-  
AttributeError  
t)  
Cell In[107], line 1  
----> 1 vis1=plt.distplot(stats['InternetUsers'])  
  
AttributeError: module 'matplotlib.pyplot' has no attribute 'distplot'
```

```
In [108]: vis1=sns.distplot(stats['InternetUsers'])  
#plt.show()
```

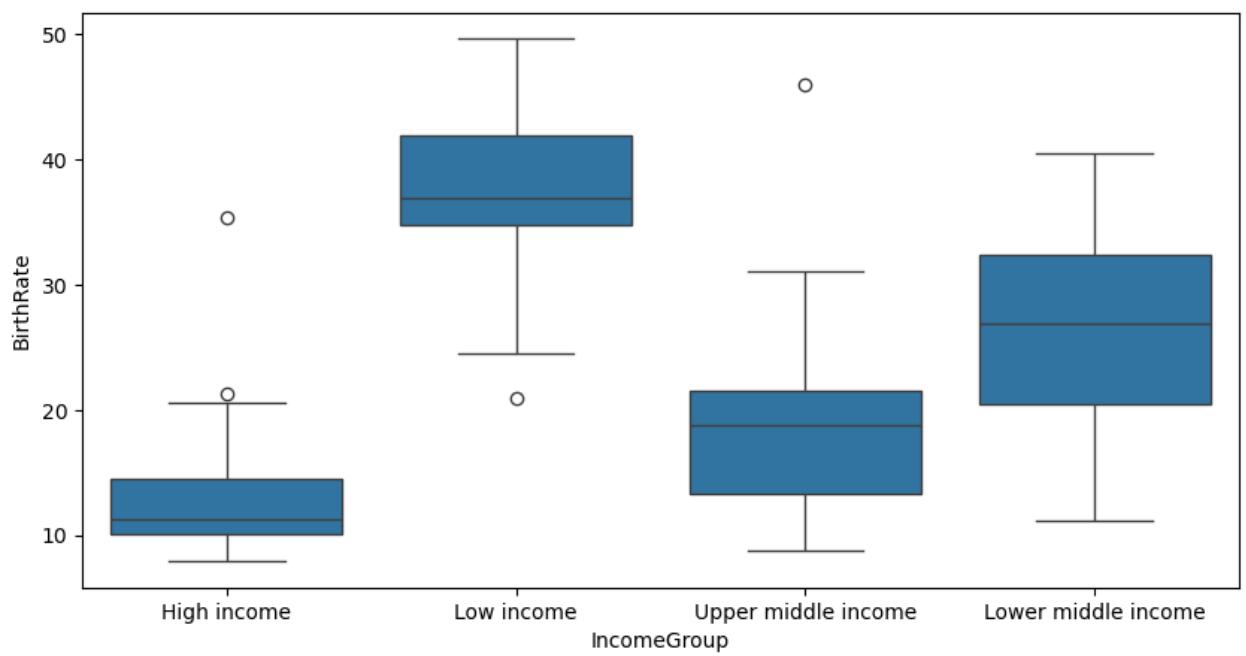


```
In [110]: vis3=sns.distplot(stats['InternetUsers'],bins=15) # bins=visualize in dif  
plt.show()
```



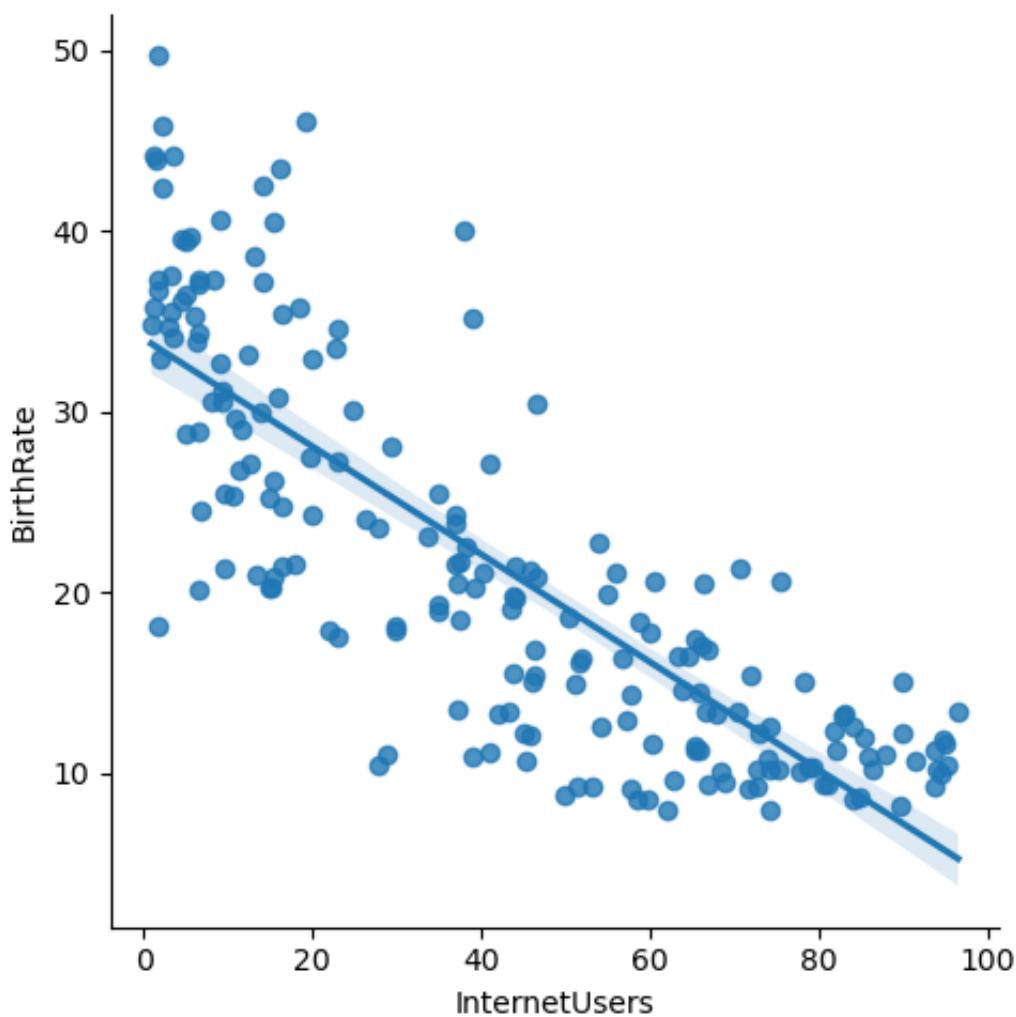
```
In [84]: #box plot for bivariate
```

```
In [111]: vis4=sns.boxplot(data=stats, x="IncomeGroup",y='BirthRate')
plt.show()
```

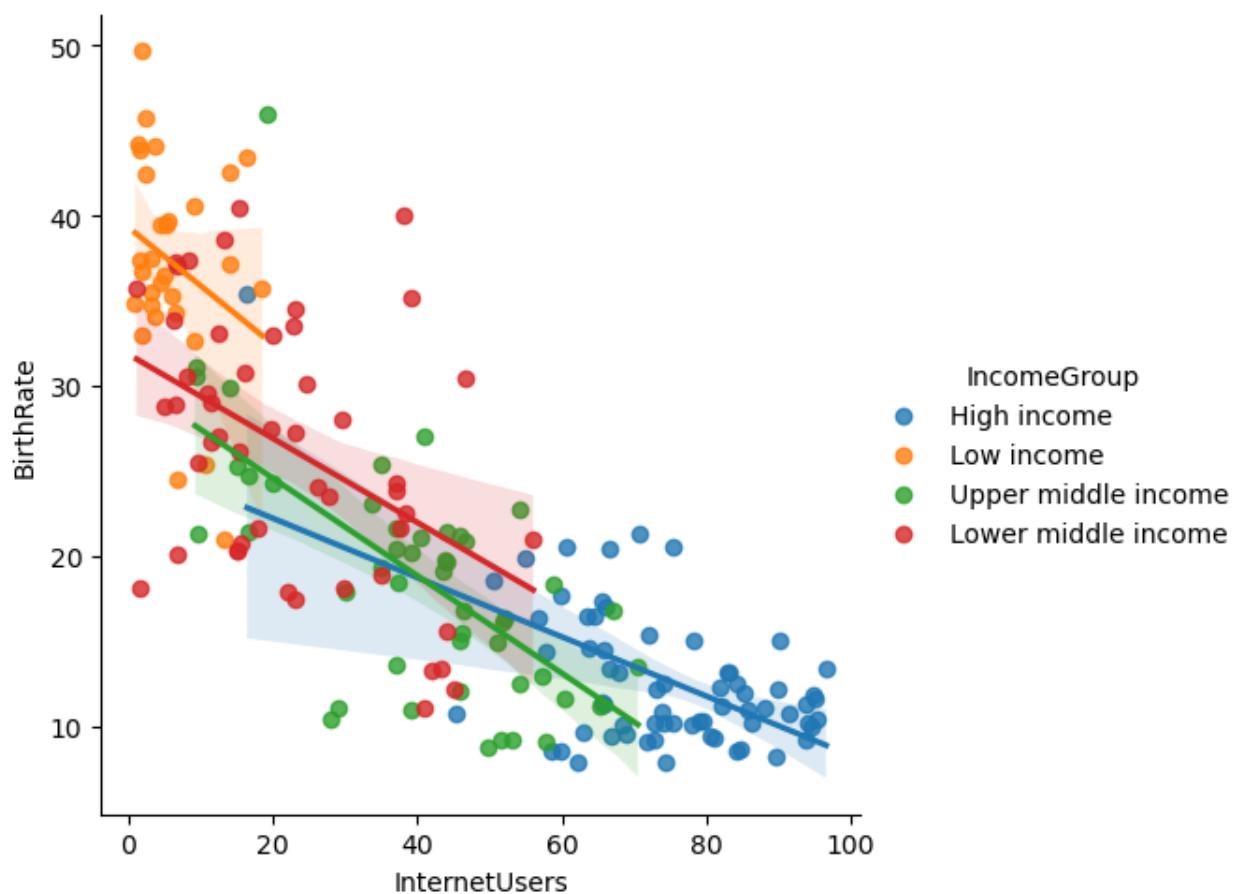


```
In [86]: # 3 outliers
```

```
In [87]: vis5=sns.lmplot(data=stats, x="InternetUsers",y='BirthRate')
plt.show()
```



```
In [118]: vis6=sns.lmplot(data=stats, x="InternetUsers",y='BirthRate',fit_reg=True,  
plt.show()
```



In []:

In []: