

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

```
In [2]: import pandas as pd
```

```
In [10]: pip install openpyxl
```

Requirement already satisfied: openpyxl in c:\users\santo\anaconda3\lib\site-packages (3.1.5)  
Requirement already satisfied: et-xmlfile in c:\users\santo\anaconda3\lib\site-packages (from openpyxl) (1.1.0)  
Note: you may need to restart the kernel to use updated packages.

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

stats = pd.read_excel(r"C:\Users\santo\OneDrive\Desktop\naresh it classroom\Project python\santosh data.xlsx")

stats
```

Out[1]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
<b>0</b>	Aruba	ABW	10.244	78.9	High income
<b>1</b>	Afghanistan	AFG	35.253	5.9	Low income
<b>2</b>	Angola	AGO	45.985	19.1	Upper middle income
<b>3</b>	Albania	ALB	12.877	57.2	Upper middle income
<b>4</b>	United Arab Emirates	ARE	11.044	88.0	High income
...	...	...	...	...	...
<b>190</b>	Yemen, Rep.	YEM	32.947	20.0	Lower middle income
<b>191</b>	South Africa	ZAF	20.850	46.5	Upper middle income
<b>192</b>	Congo, Dem. Rep.	COD	42.394	2.2	Low income
<b>193</b>	Zambia	ZMB	40.471	15.4	Lower middle income
<b>194</b>	Zimbabwe	ZWE	35.715	18.5	Low income

195 rows × 5 columns

In [41]: `len(stats)`

Out[41]: 195

In [43]: `stats.columns`

Out[43]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',  
'IncomeGroup'],  
dtype='object')

In [45]: `len(stats.columns)`

Out[45]: 5

In [2]: `stats.shape`

Out[2]: (195, 5)

In [3]: `stats.isnull()` *#To check the missing Value*

Out[3]:

	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 [4]: `stats.isna()`

Out[4]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
<b>0</b>	False	False	False	False	False
<b>1</b>	False	False	False	False	False
<b>2</b>	False	False	False	False	False
<b>3</b>	False	False	False	False	False
<b>4</b>	False	False	False	False	False
<b>...</b>	...	...	...	...	...
<b>190</b>	False	False	False	False	False
<b>191</b>	False	False	False	False	False
<b>192</b>	False	False	False	False	False
<b>193</b>	False	False	False	False	False
<b>194</b>	False	False	False	False	False

195 rows × 5 columns

In [5]: `stats.isnull().sum()`

```
Out[5]: CountryName    0
CountryCode    0
BirthRate    0
InternetUsers    0
IncomeGroup    0
dtype: int64
```

In [6]: `stats.dtypes`

```
Out[6]: CountryName    object
CountryCode    object
BirthRate    float64
InternetUsers    float64
IncomeGroup    object
dtype: object
```

```
In [7]: 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 [8]: type(stats)
```

```
Out[8]: pandas.core.frame.DataFrame
```

```
In [9]: stats.head()
```

```
Out[9]:
```

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

```
Out[10]:
```

	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 [11]: stats.head(2)
```

```
Out[11]:
```

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

```
In [12]: stats.tail(2)
```

```
Out[12]:
```

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
193	Zambia	ZMB	40.471	15.4	Lower middle income
194	Zimbabwe	ZWE	35.715	18.5	Low income

```
In [13]: # can you split
```

```
In [14]: stats.columns
```

```
Out[14]: Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
               'IncomeGroup'],
              dtype='object')
```

```
In [15]: stats['CountryName']
```

```
Out[15]: 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 [16]: stats['CountryCode']
```

```
Out[16]: 0    ABW
          1    AFG
          2    AGO
          3    ALB
          4    ARE
          ...
          190   YEM
          191   ZAF
          192   COD
          193   ZMB
          194   ZWE
          Name: CountryCode, Length: 195, dtype: object
```

```
In [17]: stats['BirthRate']
```

```
Out[17]: 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 [18]: stats['InternetUsers']
```

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



Out[19]:

	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 [20]: stats['IncomeGroup']

```

Out[20]: 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

```

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

```
Out[21]:
```

	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
...	...	...
190	32.947	20.0
191	20.850	46.5
192	42.394	2.2
193	40.471	15.4
194	35.715	18.5

195 rows × 2 columns

## lets split the dataset to numerical & categories a

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

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

```
In [23]: stats_numeric_data=stats[['BirthRate', 'InternetUsers']]
```

```
In [24]: stats_numeric_data.head()
```

Out[24]:

	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 [25]: stats\_numeric\_data=stats[['CountryName', 'CountryCode','IncomeGroup']]

In [26]: stats\_numeric\_data

Out[26]:

	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 [39]: import pandas as pd

stats = pd.read_excel(r"C:\Users\santo\OneDrive\Desktop\naresh it classroom\Project python\santosh data.xlsx")

# Numeric columns
stats_numeric_data = stats.select_dtypes(include=['number'])

# Categorical columns
stats_categorical_data = stats.select_dtypes(exclude=['number'])

print(stats.shape)
print(stats_numeric_data.shape)
print(stats_categorical_data.shape)

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

## slicing in pandas

```
In [28]: stats[:]
```

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
...	...	...	...	...	...
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 [29]: stats[:6]

Out[29]:

	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 [30]: stats[:3]
```

```
Out[30]:
```

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
<b>0</b>	Aruba	ABW	10.244	78.9	High income
<b>1</b>	Afghanistan	AFG	35.253	5.9	Low income
<b>2</b>	Angola	AGO	45.985	19.1	Upper middle income

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

```
Out[31]:
```

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
<b>3</b>	Albania	ALB	12.877	57.2	Upper middle income
<b>4</b>	United Arab Emirates	ARE	11.044	88.0	High income
<b>5</b>	Argentina	ARG	17.716	59.9	High income
<b>6</b>	Armenia	ARM	13.308	41.9	Lower middle income
<b>7</b>	Antigua and Barbuda	ATG	16.447	63.4	High income
...	...	...	...	...	...
<b>190</b>	Yemen, Rep.	YEM	32.947	20.0	Lower middle income
<b>191</b>	South Africa	ZAF	20.850	46.5	Upper middle income
<b>192</b>	Congo, Dem. Rep.	COD	42.394	2.2	Low income
<b>193</b>	Zambia	ZMB	40.471	15.4	Lower middle income
<b>194</b>	Zimbabwe	ZWE	35.715	18.5	Low income

192 rows × 5 columns

```
In [32]: stats[3:10]
```

Out[32]:

	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 [33]: stats[3:50:5]

Out[33]:

	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

In [34]: stats[::1]

Out[34]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
<b>194</b>	Zimbabwe	ZWE	35.715	18.5	Low income
<b>193</b>	Zambia	ZMB	40.471	15.4	Lower middle income
<b>192</b>	Congo, Dem. Rep.	COD	42.394	2.2	Low income
<b>191</b>	South Africa	ZAF	20.850	46.5	Upper middle income
<b>190</b>	Yemen, Rep.	YEM	32.947	20.0	Lower middle income
...	...	...	...	...	...
<b>4</b>	United Arab Emirates	ARE	11.044	88.0	High income
<b>3</b>	Albania	ALB	12.877	57.2	Upper middle income
<b>2</b>	Angola	AGO	45.985	19.1	Upper middle income
<b>1</b>	Afghanistan	AFG	35.253	5.9	Low income
<b>0</b>	Aruba	ABW	10.244	78.9	High income

195 rows × 5 columns

In [35]: `stats[:, -2]`



Out[35]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
<b>194</b>	Zimbabwe	ZWE	35.715	18.5	Low income
<b>192</b>	Congo, Dem. Rep.	COD	42.394	2.2	Low income
<b>190</b>	Yemen, Rep.	YEM	32.947	20.0	Lower middle income
<b>188</b>	West Bank and Gaza	PSE	30.394	46.6	Lower middle income
<b>186</b>	Vietnam	VNM	15.537	43.9	Lower middle income
...	...	...	...	...	...
<b>8</b>	Australia	AUS	13.200	83.0	High income
<b>6</b>	Armenia	ARM	13.308	41.9	Lower middle income
<b>4</b>	United Arab Emirates	ARE	11.044	88.0	High income
<b>2</b>	Angola	AGO	45.985	19.1	Upper middle income
<b>0</b>	Aruba	ABW	10.244	78.9	High income

98 rows × 5 columns

In [36]: stats[50:100]

Out[36]:

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
50	Ecuador	ECU	21.070	40.353684	Upper middle income
51	Egypt, Arab Rep.	EGY	28.032	29.400000	Lower middle income
52	Eritrea	ERI	34.800	0.900000	Low income
53	Spain	ESP	9.100	71.635000	High income
54	Estonia	EST	10.300	79.400000	High income
55	Ethiopia	ETH	32.925	1.900000	Low income
56	Finland	FIN	10.700	91.514400	High income
57	Fiji	FJI	20.463	37.100000	Upper middle income
58	France	FRA	12.300	81.919800	High income
59	Micronesia, Fed. Sts.	FSM	23.511	27.800000	Lower middle income
60	Gabon	GAB	30.555	9.200000	Upper middle income
61	United Kingdom	GBR	12.200	89.844100	High income
62	Georgia	GEO	13.332	43.300000	Lower middle income
63	Ghana	GHA	33.131	12.300000	Lower middle income
64	Guinea	GIN	37.337	1.600000	Low income
65	Gambia, The	GMB	42.525	14.000000	Low income
66	Guinea-Bissau	GNB	37.503	3.100000	Low income
67	Equatorial Guinea	GNQ	35.362	16.400000	High income
68	Greece	GRC	8.500	59.866300	High income
69	Grenada	GRD	19.334	35.000000	Upper middle income
70	Greenland	GRL	14.500	65.800000	High income
71	Guatemala	GTM	27.465	19.700000	Lower middle income

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
72	Guam	GUM	17.389	65.400000	High income
73	Guyana	GUY	18.885	35.000000	Lower middle income
74	Hong Kong SAR, China	HKG	7.900	74.200000	High income
75	Honduras	HND	21.593	17.800000	Lower middle income
76	Croatia	HRV	9.400	66.747600	High income
77	Haiti	HTI	25.345	10.600000	Low income
78	Hungary	HUN	9.200	72.643900	High income
79	Indonesia	IDN	20.297	14.940000	Lower middle income
80	India	IND	20.291	15.100000	Lower middle income
81	Ireland	IRL	15.000	78.247700	High income
82	Iran, Islamic Rep.	IRN	17.900	29.950000	Upper middle income
83	Iraq	IRQ	31.093	9.200000	Upper middle income
84	Iceland	ISL	13.400	96.546800	High income
85	Israel	ISR	21.300	70.800000	High income
86	Italy	ITA	8.500	58.459300	High income
87	Jamaica	JAM	13.540	37.100000	Upper middle income
88	Jordan	JOR	27.046	41.000000	Upper middle income
89	Japan	JPN	8.200	89.710000	High income
90	Kazakhstan	KAZ	22.730	54.000000	Upper middle income
91	Kenya	KEN	35.194	39.000000	Lower middle income
92	Kyrgyz Republic	KGZ	27.200	23.000000	Lower middle income
93	Cambodia	KHM	24.462	6.800000	Low income

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
94	Kiribati	KIR	29.044	11.500000	Lower middle income
95	Korea, Rep.	KOR	8.600	84.770000	High income
96	Kuwait	KWT	20.575	75.460000	High income
97	Lao PDR	LAO	27.051	12.500000	Lower middle income
98	Lebanon	LBN	13.426	70.500000	Upper middle income
99	Liberia	LBR	35.521	3.200000	Low income

In [37]: stats[0:200]

Out[37]:

	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 [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: