

**Work on the give dataset..find out the insights from data
add your own analysis too**

Import Libraries

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

IMPORTING DATA

```
In [3]: d = pd.read_csv('as_2_data.csv')
```

Read data

```
In [4]: d
```

	Country Name	Country Code	Birth rate	Internet users	Income Group
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 [11]: d.shape # (rows, columns)
```

Out[11]: (195, 5)

```
In [13]: # 2. Number of rows
print('Number of rows in above file is :', len(d))
print('Number of rows in above file is :', d.shape[0])
```

Number of rows in above file is : 195
Number of rows in above file is : 195

```
In [10]: # 3. See columns
d.columns
```

Out[10]: Index(['Country Name', 'Country Code', 'Birth rate', 'Internet users',
'Income Group'],

```
dtype='object')
```

In [14]: # 4. Number of columns

```
print('Number of columns in above file is :', d.shape[1])
```

Number of columns in above file is : 5

In [15]: # 5. Top Rows

```
d.head() # Gives default 5 entries
```

Out[15]:

	Country Name	Country Code	Birth rate	Internet users	Income Group
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 [16]: # 6 Bottom rows

```
d.tail() # Gives default 5 entries
```

Out[16]:

	Country Name	Country Code	Birth rate	Internet users	Income Group
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 [17]: # 7. Information on the columns

```
d.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 195 entries, 0 to 194
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype  
 --- 
 0   Country Name    195 non-null   object  
 1   Country Code     195 non-null   object  
 2   Birth rate       195 non-null   float64 
 3   Internet users   195 non-null   float64 
 4   Income Group     195 non-null   object  
dtypes: float64(2), object(3)
memory usage: 7.7+ KB
```

In [21]: # 8. Get stats on the columns

```
d.describe()
```

Out[21]:

	Birthrate	Internet
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

	Birthrate	Internet
75%	29.759500	66.225000
max	49.661000	96.546800

Remove the space present in Column names of a Dataframe and update it

```
In [37]: d.columns = ['CountryName', 'CountryCode', 'Birthrate', 'Internetusers', 'IncomeGroup']
```

```
In [38]: d
```

	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

SUBSETTING DATAFRAMES IN PANDAS

Three Parts

1. Rows
2. Columns
3. Combine the two

Part 1: Subsetting Rows

```
In [39]: # Fetch row Starting at index 185 to the end
d.loc[185:] # by using Loc
```

	CountryName	CountryCode	Birthrate	Internetusers	IncomeGroup
185	Virgin Islands (U.S.)	VIR	10.700	45.3	High income
186	Vietnam	VNM	15.537	43.9	Lower middle income
187	Vanuatu	VUT	26.739	11.3	Lower middle income
188	West Bank and Gaza	PSE	30.394	46.6	Lower middle income
189	Samoa	WSM	26.172	15.3	Lower middle 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

	CountryName	CountryCode	Birthrate	Internetusers	IncomeGroup
194	Zimbabwe	ZWE	35.715	18.5	Low income

In [40]: `d[185:] # by using loc`

	CountryName	CountryCode	Birthrate	Internetusers	IncomeGroup
185	Virgin Islands (U.S.)	VIR	10.700	45.3	High income
186	Vietnam	VNM	15.537	43.9	Lower middle income
187	Vanuatu	VUT	26.739	11.3	Lower middle income
188	West Bank and Gaza	PSE	30.394	46.6	Lower middle income
189	Samoa	WSM	26.172	15.3	Lower middle 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

In [41]: `#fetch row from 100 to 150 and columns BirthRate CountryCode IncomeGroup-> InternetUsers
d.iloc[100:151, [2,1,4,3]] # using iloc`

	Birthrate	CountryCode	IncomeGroup	Internetusers
100	21.425	LBY	Upper middle income	16.5000
101	15.430	LCA	Upper middle income	46.2000
102	9.200	LIE	High income	93.8000
103	17.863	LKA	Lower middle income	21.9000
104	28.738	LSO	Lower middle income	5.0000
105	10.100	LTU	High income	68.4529
106	11.300	LUX	High income	93.7765
107	10.200	LVA	High income	75.2344
108	11.256	MAC	High income	65.8000
109	21.023	MAR	Lower middle income	56.0000
110	12.141	MDA	Lower middle income	45.0000
111	34.686	MDG	Low income	3.0000
112	21.447	MDV	Upper middle income	44.1000
113	19.104	MEX	Upper middle income	43.4600
114	11.222	MKD	Upper middle income	65.2400
115	44.138	MLI	Low income	3.5000
116	9.500	MLT	High income	68.9138
117	18.119	MMR	Lower middle income	1.6000
118	11.616	MNE	Upper middle income	60.3100
119	24.275	MNG	Upper middle income	20.0000
120	39.705	MOZ	Low income	5.4000
121	33.801	MRT	Lower middle income	6.2000
122	10.900	MUS	Upper middle income	39.0000

	Birthrate	CountryCode	IncomeGroup	Internetusers
123	39.459	MWI	Low income	5.0500
124	16.805	MYS	Upper middle income	66.9700
125	29.937	NAM	Upper middle income	13.9000
126	17.000	NCL	High income	66.0000
127	49.661	NER	Low income	1.7000
128	40.045	NGA	Lower middle income	38.0000
129	20.788	NIC	Lower middle income	15.5000
130	10.200	NLD	High income	93.9564
131	11.600	NOR	High income	95.0534
132	20.923	NPL	Low income	13.3000
133	13.120	NZL	High income	82.7800
134	20.419	OMN	High income	66.4500
135	29.582	PAK	Lower middle income	10.9000
136	19.680	PAN	Upper middle income	44.0300
137	20.198	PER	Upper middle income	39.2000
138	23.790	PHL	Lower middle income	37.0000
139	28.899	PNG	Lower middle income	6.5000
140	9.600	POL	High income	62.8492
141	10.800	PRI	High income	73.9000
142	7.900	PRT	High income	62.0956
143	21.588	PRY	Upper middle income	36.9000
144	16.393	PYF	High income	56.8000
145	11.940	QAT	High income	85.3000
146	8.800	ROU	Upper middle income	49.7645
147	13.200	RUS	High income	67.9700
148	32.689	RWA	Low income	9.0000
149	20.576	SAU	High income	60.5000
150	33.477	SDN	Lower middle income	22.7000

```
In [46]: d.loc[100:151, ['Birthrate', 'CountryCode', 'IncomeGroup', 'Internetusers']] # using Loc
```

	Birthrate	CountryCode	IncomeGroup	Internetusers
100	21.425	LBY	Upper middle income	16.5000
101	15.430	LCA	Upper middle income	46.2000
102	9.200	LIE	High income	93.8000
103	17.863	LKA	Lower middle income	21.9000
104	28.738	LSO	Lower middle income	5.0000
105	10.100	LTU	High income	68.4529
106	11.300	LUX	High income	93.7765
107	10.200	LVA	High income	75.2344
108	11.256	MAC	High income	65.8000
109	21.023	MAR	Lower middle income	56.0000
110	12.141	MDA	Lower middle income	45.0000

	Birthrate	CountryCode	IncomeGroup	Internetusers
111	34.686	MDG	Low income	3.0000
112	21.447	MDV	Upper middle income	44.1000
113	19.104	MEX	Upper middle income	43.4600
114	11.222	MKD	Upper middle income	65.2400
115	44.138	MLI	Low income	3.5000
116	9.500	MLT	High income	68.9138
117	18.119	MMR	Lower middle income	1.6000
118	11.616	MNE	Upper middle income	60.3100
119	24.275	MNG	Upper middle income	20.0000
120	39.705	MOZ	Low income	5.4000
121	33.801	MRT	Lower middle income	6.2000
122	10.900	MUS	Upper middle income	39.0000
123	39.459	MWI	Low income	5.0500
124	16.805	MYS	Upper middle income	66.9700
125	29.937	NAM	Upper middle income	13.9000
126	17.000	NCL	High income	66.0000
127	49.661	NER	Low income	1.7000
128	40.045	NGA	Lower middle income	38.0000
129	20.788	NIC	Lower middle income	15.5000
130	10.200	NLD	High income	93.9564
131	11.600	NOR	High income	95.0534
132	20.923	NPL	Low income	13.3000
133	13.120	NZL	High income	82.7800
134	20.419	OMN	High income	66.4500
135	29.582	PAK	Lower middle income	10.9000
136	19.680	PAN	Upper middle income	44.0300
137	20.198	PER	Upper middle income	39.2000
138	23.790	PHL	Lower middle income	37.0000
139	28.899	PNG	Lower middle income	6.5000
140	9.600	POL	High income	62.8492
141	10.800	PRI	High income	73.9000
142	7.900	PRT	High income	62.0956
143	21.588	PRY	Upper middle income	36.9000
144	16.393	PYF	High income	56.8000
145	11.940	QAT	High income	85.3000
146	8.800	ROU	Upper middle income	49.7645
147	13.200	RUS	High income	67.9700
148	32.689	RWA	Low income	9.0000
149	20.576	SAU	High income	60.5000
150	33.477	SDN	Lower middle income	22.7000
151	38.533	SEN	Lower middle income	13.1000

In [51]:

```
# Reverse the dataframe
d_reversed = d[::-1]
d_reversed
```

Out[51]:

	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 [59]:

```
# Get ONLY every 20th row: means row present at index 0-20-40-60
d[::-20]
```

Out[59]:

	CountryName	CountryCode	Birthrate	Internetusers	IncomeGroup
0	Aruba	ABW	10.244	78.9000	High income
20	Belarus	BLR	12.500	54.1700	Upper middle income
40	Costa Rica	CRI	15.022	45.9600	Upper middle income
60	Gabon	GAB	30.555	9.2000	Upper middle income
80	India	IND	20.291	15.1000	Lower middle income
100	Libya	LBY	21.425	16.5000	Upper middle income
120	Mozambique	MOZ	39.705	5.4000	Low income
140	Poland	POL	9.600	62.8492	High income
160	Suriname	SUR	18.455	37.4000	Upper middle income
180	Uruguay	URY	14.374	57.6900	High income

Part 2: Subsetting Columns

In [63]:

```
# Subset one column 'CountryName'
d[['CountryName']]
```

Out[63]:

	CountryName
0	Aruba
1	Afghanistan
2	Angola
3	Albania
4	United Arab Emirates
...	...
190	Yemen, Rep.
191	South Africa

CountryName	
192	Congo, Dem. Rep.
193	Zambia
194	Zimbabwe

195 rows × 1 columns

In [65]:

```
# Subset multiple columns 'CountryName', 'BirthRate'
d[['CountryName', 'Birthrate']]
```

Out[65]:

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

195 rows × 2 columns

BASIC OPERATIONS WITH A DATAFRAME

Mathematical Operations

In [60]:

```
d.head(2)
```

Out[60]:

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

In [72]:

```
# Mathematical Operations:
# do the multiplication of BirthRate & InternetUsers and create a new column to store it in dataframe
d['Birth_internet'] = np.multiply(d.Birthrate, d.Internetusers)
d
```

Out[72]:

	CountryName	CountryCode	Birthrate	Internetusers	IncomeGroup	Birth_internet
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
...
190	Yemen, Rep.	YEM	32.947	20.0	Lower middle income	658.9400

	CountryName	CountryCode	Birthrate	Internetusers	IncomeGroup	Birth_internet
191	South Africa	ZAF	20.850	46.5	Upper middle income	969.5250
192	Congo, Dem. Rep.	COD	42.394	2.2	Low income	93.2668
193	Zambia	ZMB	40.471	15.4	Lower middle income	623.2534
194	Zimbabwe	ZWE	35.715	18.5	Low income	660.7275

195 rows × 6 columns

```
In [73]: # Removing a column 'Birth_internet'
d.drop(columns='Birth_internet', inplace=True)
```

In [74]: d

	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

FILTERING DATAFRAMES

```
In [87]: #Filtering is about rows where 'InternetUsers' less than 2
d[d.Internetusers < 2]
```

	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 [84]: d.query('Internetusers < 2')      # using query
```

	CountryName	CountryCode	Birthrate	Internetusers	IncomeGroup
--	-------------	-------------	-----------	---------------	-------------

	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 [89]:

```
# Filter by birthrate greater than 40
d.query('Birthrate > 40')
```

Out[89]:

	CountryName	CountryCode	Birthrate	Internetusers	IncomeGroup
2	Angola	AGO	45.985	19.1	Upper middle income
11	Burundi	BDI	44.151	1.3	Low income
14	Burkina Faso	BFA	40.551	9.1	Low income
65	Gambia, The	GMB	42.525	14.0	Low income
115	Mali	MLI	44.138	3.5	Low income
127	Niger	NER	49.661	1.7	Low income
128	Nigeria	NGA	40.045	38.0	Lower middle income
156	Somalia	SOM	43.891	1.5	Low income
167	Chad	TCD	45.745	2.3	Low income
178	Uganda	UGA	43.474	16.2	Low income
192	Congo, Dem. Rep.	COD	42.394	2.2	Low income
193	Zambia	ZMB	40.471	15.4	Lower middle income

In [92]:

```
# Filter for high income countries
d.query('IncomeGroup == "High income"')
```

Out[92]:

	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 [101]: `d.groupby('IncomeGroup').get_group('High income') # using groupby`

	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 [93]: `# get unique categories of income group
d.IncomeGroup.unique()`

Out[93]: `array(['High income', 'Low income', 'Upper middle income',
'Lower middle income'], dtype=object)`

In [102]: `# Find out everything about Malta
d.groupby('CountryName').get_group('Malta')`

	CountryName	CountryCode	Birthrate	Internetusers	IncomeGroup
116	Malta	MLT	9.5	68.9138	High income