Pandas Tasks

```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
```

[81] ✓ 0.0s Python

```
pd.__version__
```

[3] ✓ 0.0s Python

```
'2.3.1'
```

```
df = pd.read_csv(r'C:\Users\91965\FSDS\data.csv')
```

[4] ✓ 0.0s Python

```
df
```

[5] ✓ 0.0s Python

	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

95 rows × 5 columns

```
df.head()
```

[6] ✓ 0.0s Python

	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

```
len(df)#it will give you records
```

[7] ✓ 0.0s Python

```
195
```

```
id(df)
```

[8] ✓ 0.0s Python

```
1933066356112
```

```
df.shape #row and columns
```

[9] ✓ 0.0s Python

```
(195, 5)
```

```
df.columns #give you columns name
```

[10] ✓ 0.0s Python

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

```
len(df.columns)
```

[11] ✓ 0.0s Python

```
5
```

```
df.isnull()
```

[12] ✓ 0.0s Python

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGrou
0	False	False	False	False	Fals
1	False	False	False	False	Fals
2	False	False	False	False	Fals
3	False	False	False	False	Fals
4	False	False	False	False	Fals
...	...	...	...	...	
190	False	False	False	False	Fals
191	False	False	False	False	Fals
192	False	False	False	False	Fals
193	False	False	False	False	Fals
194	False	False	False	False	Fals

195 rows × 5 columns

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

[13] ✓ 0.0s Python

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

```
df.info()
```

[14] ✓ 0.0s Python

```
<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
```

```
len(df)
```

[15] ✓ 0.0s Python

```
195
```

```
len(df.columns)
```

[16] ✓ 0.0s Python

```
5
```

```
df.head(2)
```

[17] ✓ 0.0s Python

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

```
df.tail()
```

[18] ✓ 0.0s Python

	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

```
df.dtypes
```

[20] ✓ 0.0s Python

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

```
df.columns
```

[21] ✓ 0.0s Python

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

```
df['CountryName']
```

[22] ✓ 0.0s Python

```
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
```

```
df['CountryCode']
```

[23] ✓ 0.0s Python

```
0      ABW
1      AFG
2      AGO
3      ALB
4      ARE
      ...
190    YEM
191    ZAF
```

```
df['BirthRate']
```

[24] ✓ 0.0s Python

```
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
```

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

[25] ✓ 0.0s Python

```
0      78.9
1       5.9
2      19.1
3      57.2
4      88.0
       ...
190    20.0
```

```
df_cat = df[['CountryName', 'CountryCode','InternetUsers']]
df_cat
```

[26] ✓ 0.0s Python

	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
...	...	...	...
190	Yemen, Rep.	YEM	20.0
191	South Africa	ZAF	46.5
192	Congo, Dem. Rep.	COD	2.2
193	Zambia	ZMB	15.4
194	Zimbabwe	ZWE	18.5

195 rows × 3 columns

```
df_num = df[[ 'BirthRate', 'InternetUsers']]
df_num
```

[27] ✓ 0.0s Python

	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

```
print(df.shape) #5 col
print(df_cat.shape) # 3 categorial col
print(df_num.shape) #2 categorical col
```

[29] ✓ 0.0s Python

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

```
df[:]  # display entire dataset
```

[30] ✓ 0.0s Python

	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

```
df[:5]
```

[31] ✓ 0.0s Python

	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

```
df[5:]
```

[32] ✓ 0.0s Python

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
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
...	...	...	...	...	...
190	Yemen, Rep.	YEM	32.947	20.0000	Lower middle income
191	South Africa	ZAF	20.850	46.5000	Upper middle income
192	Congo, Dem. Rep.	COD	42.394	2.2000	Low income
193	Zambia	ZMB	40.471	15.4000	Lower middle income
194	Zimbabwe	ZWE	35.715	18.5000	Low income

90 rows × 5 columns

```
[33]  df[1:200:20]
      ✓ 0.0s                                                    Python
```

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
1	Afghanistan	AFG	35.253	5.9000	Low income
21	Belize	BLZ	23.092	33.6000	Upper middle income
41	Cuba	CUB	10.400	27.9300	Upper middle income
61	United Kingdom	GBR	12.200	89.8441	High income
81	Ireland	IRL	15.000	78.2477	High income
101	St. Lucia	LCA	15.430	46.2000	Upper middle income
121	Mauritania	MRT	33.801	6.2000	Lower middle income
141	Puerto Rico	PRI	10.800	73.9000	High income
161	Slovak Republic	SVK	10.100	77.8826	High income
181	United States	USA	12.500	84.2000	High income

```
[35]  df[::-1]
      ✓ 0.0s                                                    Python
```

	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

95 rows × 5 columns

```
df[::-5]
```

[36] ✓ 0.0s Python

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
194	Zimbabwe	ZWE	35.715	18.5000	Low income
189	Samoa	WSM	26.172	15.3000	Lower middle income
184	Venezuela, RB	VEN	19.842	54.9000	High income
179	Ukraine	UKR	11.100	41.0000	Lower middle income
174	Trinidad and Tobago	TTO	14.590	63.8000	High income
169	Thailand	THA	11.041	28.9400	Upper middle income
164	Swaziland	SWZ	30.093	24.7000	Lower middle income
159	Sao Tome and Principe	STP	34.537	23.0000	Lower middle income
154	Sierra Leone	SLE	36.729	1.7000	Low income
149	Saudi Arabia	SAU	20.576	60.5000	High income
144	French Polynesia	PYF	16.393	56.8000	High income
139	Papua New Guinea	PNG	28.899	6.5000	Lower middle income
64	Guinea	GIN	37.337	1.6000	Low income
59	Micronesia, Fed. Sts.	FSM	23.511	27.8000	Lower middle income
54	Estonia	EST	10.300	79.4000	High income
49	Algeria	DZA	24.738	16.5000	Upper middle income
44	Czech Republic	CZE	10.200	74.1104	High income
39	Cabo Verde	CPV	21.625	37.5000	Lower middle income
34	Cote d'Ivoire	CIV	37.320	8.4000	Lower middle income
29	Central African Republic	CAF	34.076	3.5000	Low income
24	Brazil	BRA	14.931	51.0400	Upper middle income
19	Bosnia and Herzegovina	BIH	9.062	57.7900	Upper middle income
14	Burkina Faso	BFA	40.551	9.1000	Low income
9	Austria	AUT	9.400	80.6188	High income
4	United Arab Emirates	ARE	11.044	88.0000	High income

```
#.describe - descriptive statistics
df.describe()
```

[37] ✓ 0.0s Python

	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

```
scribe().transpose() #row will convert the col and vice versa
```

[38] ✓ 0.0s Python

	count	mean	std	min	25%	50%	75%	max
Rate	195.0	21.469928	10.605467	7.9	12.1205	19.68	29.7595	49.6610
Jsers	195.0	42.076471	29.030788	0.9	14.5200	41.00	66.2250	96.5468

```
df_num.describe()
```

[39] ✓ 0.0s Python

	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

```
df_cat
```

[41] ✓ 0.0s Python

	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
...	...	...	...
190	Yemen, Rep.	YEM	20.0
191	South Africa	ZAF	46.5
192	Congo, Dem. Rep.	COD	2.2
193	Zambia	ZMB	15.4
194	Zimbabwe	ZWE	18.5

195 rows × 3 columns

```
df_cat.describe()
```

[43] ✓ 0.0s Python

	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

```
df.columns = ['a','b','c','d','e']
```

[46] ✓ 0.0s Python

```
df.head(1)
```

[47] ✓ 0.0s Python

	a	b	c	d	e
0	Aruba	ABW	10.244	78.9	High income

```
df.columns = ['CountryName', 'CountryCode', 'BirthRate', 'Interne
       'IncomeGroup']
```

[48] ✓ 0.0s Python

```
df.head(1)
```

[49] ✓ 0.0s Python

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income

```
df.BirthRate * df.InternetUsers
```

[50] ✓ 0.0s Python

```
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
Length: 195, dtype: float64
```

```
df.head(1)
```

[51] ✓ 0.0s Python

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income

```
df['myCalc'] = df.BirthRate * df.InternetUsers
```

[52] ✓ 0.0s Python

```
df.head(1)
```

[53] ✓ 0.0s Python

...tryName	CountryCode	BirthRate	InternetUsers	IncomeGroup	myCalc
Aruba	ABW	10.244	78.9	High income	808.2516

```
#xticks - x axis
#yticks - y axis
#axis - 0 means row
#axis -1 means col
df = df.drop('myCalc', axis=1)
```

[54] ✓ 0.0s Python

```
df
```

[55] ✓ 0.0s Python

	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

95 rows × 5 columns

```
df['InternetUsers'] < 2
```

[56] ✓ 0.0s Python

```
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
```

```
df[df['InternetUsers'] < 2]
```

[57] ✓ 0.0s Python

	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

```
len(df[df['InternetUsers'] < 2])
```

[58] ✓ 0.0s Python

```
9
```

```
filter1 = df[df['InternetUsers'] < 2]
filter1
```

[63] ✓ 0.0s Python

	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

```
filter2 = df[df['InternetUsers'] > 40]
filter2
```

[60] ✓ 0.0s Python

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGrou
0	Aruba	ABW	10.244	78.9	High incom
3	Albania	ALB	12.877	57.2	Upper middl incom
4	United Arab Emirates	ARE	11.044	88.0	High incom
5	Argentina	ARG	17.716	59.9	High incom
6	Armenia	ARM	13.308	41.9	Lower middl incom
...	...	...	...	...	
184	Venezuela, RB	VEN	19.842	54.9	High incom
185	Virgin Islands (U.S.)	VIR	10.700	45.3	High incom
186	Vietnam	VNM	15.537	43.9	Lower middl incom
188	West Bank and Gaza	PSE	30.394	46.6	Lower middl incom
191	South Africa	ZAF	20.850	46.5	Upper middl incom

```
df[(df.BirthRate > 40) & (df.InternetUsers < 2)]
```

[70] ✓ 0.0s Python

	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.head(1)
```

[71] ✓ 0.0s Python

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income

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

[72] ✓ 0.0s Python

	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

7 rows × 5 columns

```
df[df.IncomeGroup == 'Low income'].head()
```

[74] ✓ 0.0s Python

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
1	Afghanistan	AFG	35.253	5.9	Low income
11	Burundi	BDI	44.151	1.3	Low income
13	Benin	BEN	36.440	4.9	Low income
14	Burkina Faso	BFA	40.551	9.1	Low income
29	Central African Republic	CAF	34.076	3.5	Low income

```
df[df.IncomeGroup == 'High income'].head()
```

[75] ✓ 0.0s Python

	CountryName	CountryCode	BirthRate	InternetUsers	IncomeGroup
0	Aruba	ABW	10.244	78.9	High income
4	United Arab Emirates	ARE	11.044	88.0	High income
5	Argentina	ARG	17.716	59.9	High income
7	Antigua and Barbuda	ATG	16.447	63.4	High income
8	Australia	AUS	13.200	83.0	High income

```
df.IncomeGroup.unique() # return name of unique values
```

[76] ✓ 0.0s Python

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

```
df.IncomeGroup.nunique() #return count of unique value
```

[77] ✓ 0.0s Python

```
4
```

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

[80] ✓ 0.0s Python

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

[83] ✓ 0.0s Python

```
df["InternetUsers"]
```

[85] ✓ 0.0s Python

```
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
```

```
vis1 = sns.distplot(df["InternetUsers"]) #univariate analyasis
plt.show()
```

[84] ✓ 0.1s Python