intro to data science course

In [1]:

```
import os
os.getcwd()
```

Out[1]:

```
'C:\\Users\\LILENDAR'
```

In [2]:

```
os.chdir('C:/Users/LILENDAR/Desktop/Python Directory')
```

WEEK 3 PRACTICE

In [9]:

```
import pandas as pd
import numpy as np
```

In [27]:

```
data=pd.read_excel('Energy.xlsx')
```

In [148]:

```
type(data.columns)
```

Out[148]:

```
pandas.core.indexes.base.Index
```

In [149]:

```
type(data)
```

Out[149]:

```
pandas.core.frame.DataFrame
```

In [150]:

```
data.head()
```

Out[150]:

	Unnamed: 0	Energy Supply	Energy Supply per capita	Renewable Electricity Production	Energy Supp
0	NaN	Petajoules	Gigajoules	%	PetajoulesPetajoulesPetajoulesPetajoulesPetajo..
1	NaN	NaN	NaN	NaN	NaN
2	Afghanistan	321	10	78.6693	321000000
3	Albania	102	35	100	102000000
4	Algeria	1959	51	0.55101	1959000000

In [29]:

```
data['Energy Supp']=data['Energy Supply']*1000000
```

In [30]:

```
data
```

Out[30]:

	Unnamed: 0	Energy Supply	Energy Supply per capita	Renewable Electricity Production	Energy Su
0	NaN	Petajoules	Gigajoules	%	PetajoulesPetajoulesPetajoulesPetajoulesPetaj
1	NaN	NaN	NaN	NaN	N
2	Afghanistan	321	10	78.6693	321000
3	Albania	102	35	100	102000
4	Algeria	1959	51	0.55101	1959000
...	...	...	...	...	
224	Viet Nam	2554	28	45.3215	2554000
225	Wallis and Futuna Islands	0	26	0	
226	Yemen	344	13	0	344000
227	Zambia	400	26	99.7147	400000
228	Zimbabwe	480	32	52.5361	480000

229 rows × 5 columns

In [22]:

```
```

Out[22]:

	Unnamed: 0	Energy Supply	Energy Supply per capita	Renewable Electricity Production	Energy Supp
2	Afghanistan	321	10	78.6693	321000000
3	Albania	102	35	100	102000000
4	Algeria	1959	51	0.55101	1959000000
6	Andorra	9	121	88.6957	9000000
7	Angola	642	27	70.9091	642000000
...	...	...	...	...	...
224	Viet Nam	2554	28	45.3215	2554000000
225	Wallis and Futuna Islands	0	26	0	0
226	Yemen	344	13	0	344000000
227	Zambia	400	26	99.7147	400000000
228	Zimbabwe	480	32	52.5361	480000000

226 rows × 5 columns

In [31]:

```
data2=data.drop('Energy Supply',axis=1)
```

In [32]:

```
data2.head()
```

Out[32]:

	Unnamed: 0	Energy Supply per capita	Renewable Electricity Production	Energy Supp
0	NaN	Gigajoules	%	PetajoulesPetajoulesPetajoulesPetajoulesPetajo...
1	NaN	NaN	NaN	NaN
2	Afghanistan	10	78.6693	321000000
3	Albania	35	100	102000000
4	Algeria	51	0.55101	1959000000

In [33]:

```
data2.drop([0,1],axis=0)
```

Out[33]:

	Unnamed: 0	Energy Supply per capita	Renewable Electricity Production	Energy Supp
2	Afghanistan	10	78.6693	321000000
3	Albania	35	100	102000000
4	Algeria	51	0.55101	1959000000
5	American Samoa	NaN	0.641026	NaN
6	Andorra	121	88.6957	9000000
...	...	...	...	...
224	Viet Nam	28	45.3215	2554000000
225	Wallis and Futuna Islands	26	0	0
226	Yemen	13	0	344000000
227	Zambia	26	99.7147	400000000
228	Zimbabwe	32	52.5361	480000000

227 rows × 4 columns

In [35]:

```
energy=data2
```

In [115]:

```
energy['Country']=energy['Unnamed: 0']
```

In [113]:

```
energy.head()
energy=energy.drop([0,1],axis=0)
```

In [120]:

```
energy.head()
#energy=energy.drop('Unnamed: 0',axis=1)
energy= energy.set_index('Country')
```

In [ ]:

```

```

In [ ]:

```

```

In [48]:

```
gdp=pd.read_excel('gdp.xls')
```

In [50]:

```
gdp.head()
gdp=gdp.drop([0,1],axis=0)
```

In [51]:

```
gdp.head()
```

Out[51]:

	Data Source	World Development Indicators	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unn
2	Country Name	Country Code	Indicator Name	Indicator Code	2.006000e+03	2.007000e+03	2.008
3	Aruba	ABW	GDP (current US$)	NY.GDP.MKTP.CD	2.424581e+09	2.615084e+09	2.745
4	Afghanistan	AFG	GDP (current US$)	NY.GDP.MKTP.CD	6.971286e+09	9.747880e+09	1.010
5	Angola	AGO	GDP (current US$)	NY.GDP.MKTP.CD	5.238101e+10	6.526645e+10	8.853
6	Albania	ALB	GDP (current US$)	NY.GDP.MKTP.CD	8.896074e+09	1.067732e+10	1.288

In [55]:

```
gdp=gdp.set_index(['Data Source'])
```

In [75]:

```
#gdp=gdp.set_column(['Country Name'])
gdp.head()
gdpp=gdp.drop(['World Development Indicators','Unnamed: 2','Unnamed: 3'],axis=1)
gdpp.head()
#gdpp.iloc(0)
```

Out[75]:

Data Source	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnar
Country Name	2.006000e+03	2.007000e+03	2.008000e+03	2.009000e+03	2.010000e+03	2.01100
Aruba	2.424581e+09	2.615084e+09	2.745251e+09	2.498883e+09	2.390503e+09	2.54972
Afghanistan	6.971286e+09	9.747880e+09	1.010923e+10	1.243909e+10	1.585657e+10	1.78042
Angola	5.238101e+10	6.526645e+10	8.853861e+10	7.030716e+10	8.379950e+10	1.11789
Albania	8.896074e+09	1.067732e+10	1.288135e+10	1.204422e+10	1.192693e+10	1.28907

In [93]:

```
header_row = 0

gdpp.columns = gdpp.iloc[header_row]
gdpp
```

Out[93]:

Country Name / Data Source	2006.0	2007.0	2008.0	2009.0	2010.0	2
Country Name	2.006000e+03	2.007000e+03	2.008000e+03	2.009000e+03	2.010000e+03	2.01100
Aruba	2.424581e+09	2.615084e+09	2.745251e+09	2.498883e+09	2.390503e+09	2.54972
Afghanistan	6.971286e+09	9.747880e+09	1.010923e+10	1.243909e+10	1.585657e+10	1.78042
Angola	5.238101e+10	6.526645e+10	8.853861e+10	7.030716e+10	8.379950e+10	1.11789
Albania	8.896074e+09	1.067732e+10	1.288135e+10	1.204422e+10	1.192693e+10	1.28907
...	...	...	...	...	...	
Kosovo	3.846820e+09	4.655899e+09	5.687418e+09	5.653793e+09	5.835874e+09	6.70169
Yemen, Rep.	1.906198e+10	2.165053e+10	2.691085e+10	2.513027e+10	3.090675e+10	3.27264
South Africa	2.716385e+11	2.994155e+11	2.867698e+11	2.959365e+11	3.753494e+11	4.16418
Zambia	1.275686e+10	1.405696e+10	1.791086e+10	1.532834e+10	2.026556e+10	2.34595
Zimbabwe	5.443896e+09	5.291950e+09	4.415703e+09	9.665793e+09	1.204166e+10	1.41019

265 rows × 10 columns

In [104]:

```
#a=gdpp.iloc[0]
#gdpp.drop(['Country Name'],axis=0)
gdpp.index
```

Out[104]:

```
Index(['Aruba', 'Afghanistan', 'Angola', 'Albania', 'Andorra', 'Arab Worl
d',
       'United Arab Emirates', 'Argentina', 'Armenia', 'American Samoa',
       ...
       'Virgin Islands (U.S.)', 'Vietnam', 'Vanuatu', 'World', 'Samoa',
       'Kosovo', 'Yemen, Rep.', 'South Africa', 'Zambia', 'Zimbabwe'],
      dtype='object', name='Data Source', length=264)
```

In [ ]:

```
```

In [ ]:

```
```

In [43]:

```
scimen=pd.read_excel('ScimEn.xlsx')
```

In [44]:

```
scimen.head()
```

Out[44]:

	Rank	Country	Region	Documents	Citable documents	Citations	Self-citations	Citations per document	H index
0	1	China	Asiatic Region	235126	233883	1909601	1306438	8.12	224
1	2	United States	Northern America	157811	154288	1940563	639345	12.30	333
2	3	Japan	Asiatic Region	46032	45559	436961	109968	9.49	181
3	4	India	Asiatic Region	39893	38848	368175	123446	9.23	171
4	5	United Kingdom	Western Europe	38873	37780	536378	100038	13.80	208

In [106]:

```
data2=scimen[0:15]
data2.head()
```

Out[106]:

	Rank	Country	Region	Documents	Citable documents	Citations	Self-citations	Citations per document	H index
0	1	China	Asiatic Region	235126	233883	1909601	1306438	8.12	224
1	2	United States	Northern America	157811	154288	1940563	639345	12.30	333
2	3	Japan	Asiatic Region	46032	45559	436961	109968	9.49	181
3	4	India	Asiatic Region	39893	38848	368175	123446	9.23	171
4	5	United Kingdom	Western Europe	38873	37780	536378	100038	13.80	208

In [135]:

```
#data2=data2.set_index('Country')
#data2=data2.drop('Region',axis=1)
data2.shape
```

Out[135]:

```
(15, 7)
```

In [136]:

```
gdpp.shape
```

Out[136]:

```
(264, 10)
```

In [137]:

```
energy.shape
```

Out[137]:

```
(227, 3)
```

In [139]:

```
data3=pd.merge(data2, gdpp, how='left', left_index=True, right_index=True)
```

In [141]:

```
data4=pd.merge(data3, energy, how='left', left_index=True, right_index=True)
```

In [142]:

```
data4.shape
```

Out[142]:

```
(15, 20)
```

In [144]:

```
def answer_one():

    return data4
```

In [146]:

```
data4
```

Out[146]:

Country	Rank	Documents	Citable documents	Citations	Self-citations	Citations per document	H index	200
China	1	235126	233883	1909601	1306438	8.12	224	2.752132e+
United States	2	157811	154288	1940563	639345	12.30	333	1.381461e+
Japan	3	46032	45559	436961	109968	9.49	181	4.530377e+
India	4	39893	38848	368175	123446	9.23	171	9.402599e+
United Kingdom	5	38873	37780	536378	100038	13.80	208	2.713750e+
Germany	6	32935	32227	367356	78265	11.15	186	2.992197e+
Russian Federation	7	31880	31664	91906	42529	2.88	84	9.899305e+
Canada	8	29633	29011	491467	88987	16.59	210	1.315415e+
Italy	9	23725	22819	312631	76446	13.18	155	1.947920e+
South Korea	10	23451	23166	279709	53110	11.93	146	Na
France	11	22429	21913	300015	58151	13.38	169	2.318594e+
Iran	12	19371	19085	242250	77012	12.51	127	Na
Spain	13	18882	18529	312632	56968	16.56	163	1.259344e+
Australia	14	18077	17661	263733	44799	14.59	161	7.460542e+
Brazil	15	18024	17745	152380	37112	8.45	121	1.107640e+

WEEK 4 PRACTICE

In [ ]:

```

```