In [7]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [35]:
```
name='Jamil'
age=29
live='Kuakata'
print('my name is '+name)
print(name+' Is come from '+live)
print(name+' is naow ',age,'year old')
```

```
my name is Jamil
Jamil Is come from Kuakata
Jamil is naow  29 year old
```

In [40]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [12]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [22]:
```
Df=pd.read_csv('E:/gapminder.csv')
Df
```

Out[22]:

	country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	Afghanistan	1972	13079460	Asia	36.088	739.981106
...	...	...	...	...	...	...
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

1704 rows × 6 columns

In [26]:
```
print(Df.head)
```

```
<bound method NDFrame.head of            country  year  population continent  life_exp
gdp_cap
0     Afghanistan  1952     8425333      Asia    28.801  779.445314
1     Afghanistan  1957     9240934      Asia    30.332  820.853030
2     Afghanistan  1962    10267083      Asia    31.997  853.100710
3     Afghanistan  1967    11537966      Asia    34.020  836.197138
4     Afghanistan  1972    13079460      Asia    36.088  739.981106
...           ...   ...         ...       ...       ...         ...
1699     Zimbabwe  1987     9216418    Africa    62.351  706.157306
1700     Zimbabwe  1992    10704340    Africa    60.377  693.420786
1701     Zimbabwe  1997    11404948    Africa    46.809  792.449960
1702     Zimbabwe  2002    11926563    Africa    39.989  672.038623
1703     Zimbabwe  2007    12311143    Africa    43.487  469.709298

[1704 rows x 6 columns]>
```

In [27]:

```
print(Df.shape)
```

```
(1704, 6)
```

In [29]:

```
print(Df.columns)
```

```
Index(['country', 'year', 'population', 'continent', 'life_exp', 'gdp_cap'], dtype='o
bject')
```

In [31]:

```
print(Df.dtypes)
```

```
country        object
year            int64
population      int64
continent      object
life_exp      float64
gdp_cap       float64
dtype: object
```

In [ ]:

```
print(Df.info())
```

In [40]:

```
#Get the country column
country_Df=Df['country']

print(country_Df.head() )
```

```
0    Afghanistan
1    Afghanistan
2    Afghanistan
3    Afghanistan
4    Afghanistan
Name: country, dtype: object
```

In [45]:

```
print(country_Df.tail())
```

```
1699    Zimbabwe
1700    Zimbabwe
1701    Zimbabwe
1702    Zimbabwe
1703    Zimbabwe
Name: country, dtype: object
```

In [48]:

```
subset=Df[['country','continent','year']]
print(subset.head())
```

```
        country continent  year
0  Afghanistan      Asia  1952
1  Afghanistan      Asia  1957
2  Afghanistan      Asia  1962
3  Afghanistan      Asia  1967
4  Afghanistan      Asia  1972
```

In [49]:

```
print(subset.tail())
```

```
        country continent  year
1699  Zimbabwe    Africa  1987
1700  Zimbabwe    Africa  1992
1701  Zimbabwe    Africa  1997
1702  Zimbabwe    Africa  2002
1703  Zimbabwe    Africa  2007
```

In [50]:

```
print(Df.head())
```

```
        country  year  population continent  life_exp     gdp_cap
0  Afghanistan  1952     8425333      Asia    28.801  779.445314
1  Afghanistan  1957     9240934      Asia    30.332  820.853030
2  Afghanistan  1962    10267083      Asia    31.997  853.100710
3  Afghanistan  1967    11537966      Asia    34.020  836.197138
4  Afghanistan  1972    13079460      Asia    36.088  739.981106
```

In [52]:

```
Df.tail()
```

Out[52]:

	country	year	population	continent	life_exp	gdp_cap
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

In [57]:

```
Df.loc[0]
```

Out[57]:

```
country       Afghanistan
year                 1952
population        8425333
continent            Asia
life_exp           28.801
gdp_cap        779.445314
Name: 0, dtype: object
```

In [1]:

```
Df.loc[1702]
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[1], line 1
----> 1 Df.loc[1702]

NameError: name 'Df' is not defined
```

In [4]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [7]:

```
df=pd.read_csv('E:/gapminder_full.csv')
print(df.head())
```

```
       country  year  population continent  life_exp     gdp_cap
0  Afghanistan  1952     8425333      Asia    28.801  779.445314
1  Afghanistan  1957     9240934      Asia    30.332  820.853030
2  Afghanistan  1962    10267083      Asia    31.997  853.100710
3  Afghanistan  1967    11537966      Asia    34.020  836.197138
4  Afghanistan  1972    13079460      Asia    36.088  739.981106
```

In [10]:

```
df.tail()
```

Out[10]:

	country	year	population	continent	life_exp	gdp_cap
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

In [8]:

```
df.loc[0]
```

Out[8]:

```
country        Afghanistan
year                  1952
population         8425333
continent             Asia
life_exp            28.801
gdp_cap         779.445314
Name: 0, dtype: object
```

In [10]:

```
df.columns
```

Out[10]:

```
Index(['country', 'year', 'population', 'continent', 'life_exp', 'gdp_cap'], dtype='o
bject')
```

In [14]:

```
subset=df[['country', 'year', 'population', 'continent']]
subset.head()
```

Out[14]:

	country	year	population	continent
0	Afghanistan	1952	8425333	Asia
1	Afghanistan	1957	9240934	Asia
2	Afghanistan	1962	10267083	Asia
3	Afghanistan	1967	11537966	Asia
4	Afghanistan	1972	13079460	Asia

In [15]:

```
df.loc[99]
```

Out[15]:

```
country       Bangladesh
year                1967
population      62821884
continent           Asia
life_exp          43.453
gdp_cap       721.186086
Name: 99, dtype: object
```

In [18]:

```
print(df.tail(n=2))
```

```
       country  year  population continent  life_exp     gdp_cap
1702  Zimbabwe  2002    11926563    Africa    39.989  672.038623
1703  Zimbabwe  2007    12311143    Africa    43.487  469.709298
```

In [21]:

```
print(df.iloc[[1,99,999]])
```

```
         country  year  population continent  life_exp      gdp_cap
1    Afghanistan  1957     9240934      Asia    30.332   820.853030
99    Bangladesh  1967    62821884      Asia    43.453   721.186086
999     Mongolia  1967     1149500      Asia    51.253  1226.041130
```

In [22]:

```
print(df.iloc[1])
```

```
country       Afghanistan
year                 1957
population        9240934
continent            Asia
life_exp           30.332
gdp_cap         820.85303
Name: 1, dtype: object
```

In [24]:

```
print(df.iloc[-1])
```

```
country         Zimbabwe
year                2007
population      12311143
continent         Africa
life_exp          43.487
gdp_cap       469.709298
Name: 1703, dtype: object
```

In [25]:

```
print(df.tail(n=1))
```

```
       country  year  population continent  life_exp     gdp_cap
1703  Zimbabwe  2007    12311143    Africa    43.487  469.709298
```

In [32]:

```
subset=df.loc[:,['year','population']]
subset.head()
```

Out[32]:

	year	population
0	1952	8425333
1	1957	9240934
2	1962	10267083
3	1967	11537966
4	1972	13079460

subset columns with iloc

iloc will alow us to use integer, -1 will select the last columns

In [42]:

```
subset=df.iloc[:,[2,4,-1]]
subset.head()
```

Out[42]:

	population	life_exp	gdp_cap
0	8425333	28.801	779.445314
1	9240934	30.332	820.853030
2	10267083	31.997	853.100710
3	11537966	34.020	836.197138
4	13079460	36.088	739.981106

subseting column by range

creatte a range of integers from 0 to 4 inclusive

In [44]:

```
small_range=list(range(5))
small_range
```

Out[44]:

```
[0, 1, 2, 3, 4]
```

subset the dataframe with the range

In [49]:

```
subset=df.iloc[:,small_range]
subset.head()
```

Out[49]:

	country	year	population	continent	life_exp
0	Afghanistan	1952	8425333	Asia	28.801
1	Afghanistan	1957	9240934	Asia	30.332
2	Afghanistan	1962	10267083	Asia	31.997
3	Afghanistan	1967	11537966	Asia	34.020
4	Afghanistan	1972	13079460	Asia	36.088

subsetting rows and columns

using loc/iloc

In [50]:

```
df.loc[43,'country']
```

Out[50]:

```
'Angola'
```

In [51]:

```
df.iloc[43,0]
```

Out[51]:

```
'Angola'
```

subsetting multiple rows and column

get the 1st, 100th and 1000th rows

FROM 1th,4th and 6th column

In [57]:

```
df.iloc[[0,99,999],[0,3,5]]
```

Out[57]:

	country	continent	gdp_cap
0	Afghanistan	Asia	779.445314
99	Bangladesh	Asia	721.186086
999	Mongolia	Asia	1226.041130

In [65]:

```
df.loc[[0,99,999],['country','continent','gdp_cap']]
```

Out[65]:

	country	continent	gdp_cap
0	Afghanistan	Asia	779.445314
99	Bangladesh	Asia	721.186086
999	Mongolia	Asia	1226.041130

In [72]:

```
df.loc[10:13,['country','life_exp','gdp_cap']]
```

Out[72]:

	country	life_exp	gdp_cap
10	Afghanistan	42.129	726.734055
11	Afghanistan	43.828	974.580338
12	Albania	55.230	1601.056136
13	Albania	59.280	1942.284244

In [78]:

```
df.head(n=10)
```

Out[78]:

	country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	Afghanistan	1972	13079460	Asia	36.088	739.981106
5	Afghanistan	1977	14880372	Asia	38.438	786.113360
6	Afghanistan	1982	12881816	Asia	39.854	978.011439
7	Afghanistan	1987	13867957	Asia	40.822	852.395945
8	Afghanistan	1992	16317921	Asia	41.674	649.341395
9	Afghanistan	1997	22227415	Asia	41.763	635.341351

Group mean

for each year in our data,rhat was the average life expectancy?...To answare this question.... we need to split our data into parts by year..then we get the'life_exp' column and calculate the mean

In [79]:

```
df.groupby('year')['life_exp'].mean()
```

Out[79]:

```
year
1952    49.057620
1957    51.507401
1962    53.609249
1967    55.678290
1972    57.647386
1977    59.570157
1982    61.533197
1987    63.212613
1992    64.160338
1997    65.014676
2002    65.694923
2007    67.007423
Name: life_exp, dtype: float64
```

In [80]:

```
df.groupby('country')['life_exp'].mean()
```

Out[80]:

```
country
Afghanistan           37.478833
Albania               68.432917
Algeria               59.030167
Angola                37.883500
Argentina             69.060417
                        ...
Vietnam               57.479500
West Bank and Gaza    60.328667
Yemen, Rep.           46.780417
Zambia                45.996333
Zimbabwe              52.663167
Name: life_exp, Length: 142, dtype: float64
```

In [95]:

```
multi_group_var=df.groupby(['year','continent'])[['life_exp','gdp_cap']].mean()
multi_group_var.head(10)
```

Out[95]:

		life_exp	gdp_cap
year	**continent**		
1952	**Africa**	39.135500	1252.572466
	Americas	53.279840	4079.062552
	Asia	46.314394	5195.484004
	Europe	64.408500	5661.057435
	Oceania	69.255000	10298.085650
1957	**Africa**	41.266346	1385.236062
	Americas	55.960280	4616.043733
	Asia	49.318544	5787.732940
	Europe	66.703067	6963.012816
	Oceania	70.295000	11598.522455

"flatten" the dataframe use the rest_index method

In [91]:

```
flat=multi_group_var.reset_index()
flat.head(15)
```

Out[91]:

	year	continent	life_exp	gdp_cap
0	1952	Africa	39.135500	1252.572466
1	1952	Americas	53.279840	4079.062552
2	1952	Asia	46.314394	5195.484004
3	1952	Europe	64.408500	5661.057435
4	1952	Oceania	69.255000	10298.085650
5	1957	Africa	41.266346	1385.236062
6	1957	Americas	55.960280	4616.043733
7	1957	Asia	49.318544	5787.732940
8	1957	Europe	66.703067	6963.012816
9	1957	Oceania	70.295000	11598.522455
10	1962	Africa	43.319442	1598.078825
11	1962	Americas	58.398760	4901.541870
12	1962	Asia	51.563223	5729.369625
13	1962	Europe	68.539233	8365.486814
14	1962	Oceania	71.085000	12696.452430

Group frequency counts

use the nunique to geet counts of unique values on a Pandas series

In [96]:

```
df.groupby('continent')['country'].nunique()
```

Out[96]:

```
continent
Africa      52
Americas    25
Asia        33
Europe      30
Oceania      2
Name: country, dtype: int64
```

In [100…

```
global_yearly_life_expectancy =df.groupby('year')['life_exp'].mean()
global_yearly_life_expectancy
```

Out[100]:

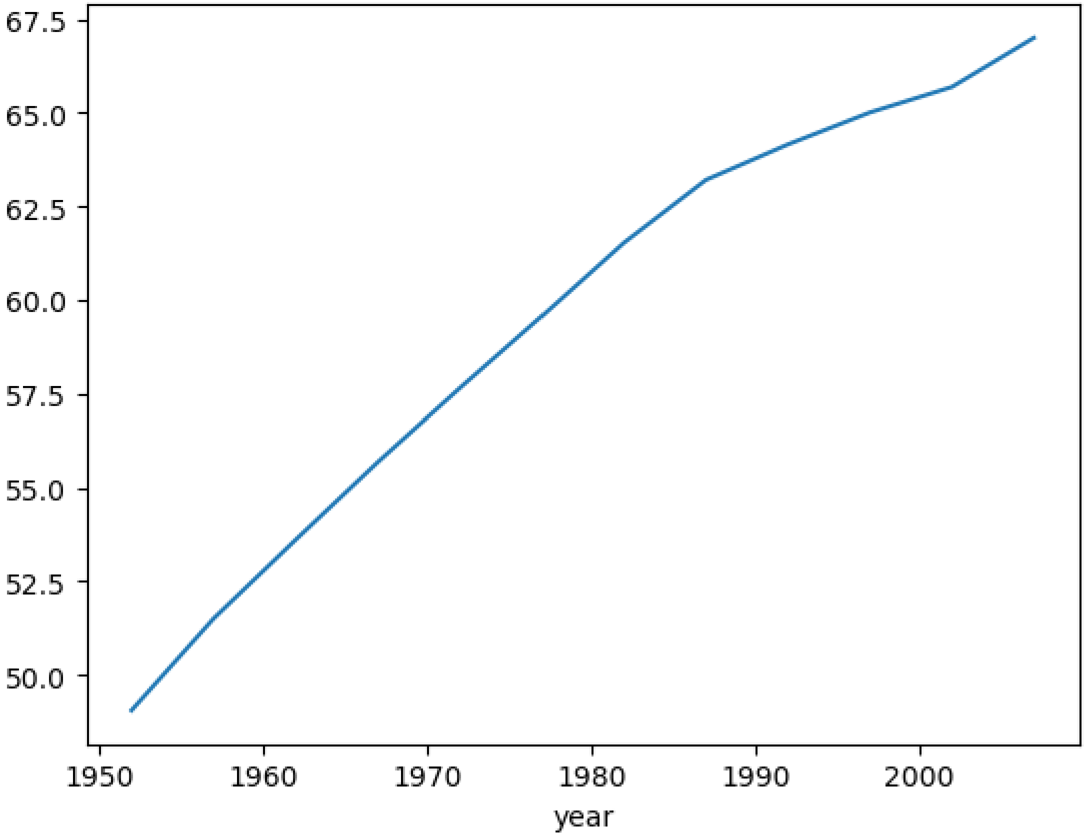
```
year
1952    49.057620
1957    51.507401
1962    53.609249
1967    55.678290
1972    57.647386
1977    59.570157
1982    61.533197
1987    63.212613
1992    64.160338
1997    65.014676
2002    65.694923
2007    67.007423
Name: life_exp, dtype: float64
```

In [101…

```
global_yearly_life_expectancy.plot()
```

Out[101]:

```
<Axes: xlabel='year'>
```


year