

In [1]:

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

In [2]:

```
a = pd.DataFrame({'A':[10,30], 'B':[20,40]})
b = pd.DataFrame({'A':[10,30], 'C':[20,40]})
```

In [3]:

a

Out[3]:

	A	B
0	10	20
1	30	40

In [4]:

b

Out[4]:

	A	C
0	10	20
1	30	40

In []:

In [6]:

```
pd.concat([a, b])
```

Out[6]:

	A	B	C
0	10	20.0	NaN
1	30	40.0	NaN
0	10	NaN	20.0
1	30	NaN	40.0

In [7]:

```
pd.concat([a, b], axis=1)
```

Out[7]:

	A	B	A	C
0	10	20	10	20
1	30	40	30	40

In [8]:

```
pd.concat([a, b], axis=0, join='outer')
```

Out[8]:

	A	B	C
0	10	20.0	NaN
1	30	40.0	NaN
0	10	NaN	20.0
1	30	NaN	40.0

In [11]:

```
pd.concat([a, b], axis=0, join='inner')
```

Out[11]:

	A
0	10
1	30
0	10
1	30

pd.merge()

In [12]:

```
users = pd.DataFrame({'userid':[1, 2, 3], 'name':['A', 'B', 'C']})  
msgs = pd.DataFrame({'userid':[1, 1, 2], 'msg':['hello', 'bye', 'hi']})
```

In [13]:

```
users
```

Out[13]:

	userid	name
0	1	A
1	2	B
2	3	C

In [14]:

```
msgs
```

Out[14]:

	userid	msg
0	1	hello
1	1	bye
2	2	hi

In [15]:

```
pd.concat((users, msgs))
```

Out[15]:

	userid	name	msg
0	1	A	NaN
1	2	B	NaN
2	3	C	NaN
0	1	NaN	hello
1	1	NaN	bye
2	2	NaN	hi

In [16]:

```
pd.merge(users, msgs, on='userid')
```

Out[16]:

	userid	name	msg
0	1	A	hello
1	1	A	bye
2	2	B	hi

In [17]:

```
users.rename(columns={'userid' : 'id'}, inplace=True)
users
```

Out[17]:

	id	name
0	1	A
1	2	B
2	3	C

In [18]:

```
msgs
```

Out[18]:

	userid	msg
0	1	hello
1	1	bye
2	2	hi

In [19]:

```
pd.merge(users, msgs, left_on='id', right_on='userid')
```

Out[19]:

	id	name	userid	msg
0	1	A	1	hello
1	1	A	1	bye
2	2	B	2	hi

In [20]:

```
pd.merge(users, msgs, how='inner' , left_on='id', right_on='userid')
```

Out[20]:

	id	name	userid	msg
0	1	A	1	hello
1	1	A	1	bye
2	2	B	2	hi

In [22]:

```
users
```

Out[22]:

	id	name
0	1	A
1	2	B
2	3	C

In [23]:

```
msgs
```

Out[23]:

	userid	msg
0	1	hello
1	1	bye
2	2	hi

In [21]:

```
pd.merge(users, msgs, how='left' , left_on='id', right_on='userid')
```

Out[21]:

	id	name	userid	msg
0	1	A	1.0	hello
1	1	A	1.0	bye
2	2	B	2.0	hi
3	3	C	NaN	NaN

In [24]:

```
pd.merge(users, msgs, how='right' , left_on='id', right_on='userid')
```

Out[24]:

	id	name	userid	msg
0	1	A	1	hello
1	1	A	1	bye
2	2	B	2	hi

In [25]:

```
pd.merge(users, msgs, how='outer' , left_on='id', right_on='userid')
```

Out[25]:

	id	name	userid	msg
0	1	A	1.0	hello
1	1	A	1.0	bye
2	2	B	2.0	hi
3	3	C	NaN	NaN

In [30]:

```
users
```

Out[30]:

	id	name
0	1	A
1	2	B
2	3	C

In [32]:

```
users = users[['name', 'id']]  
users
```

Out[32]:

	name	id
0	A	1
1	B	2
2	C	3

In []:

In [34]:

```
df = pd.read_csv('gapminder.csv')
df.head()
```

Out[34]:

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

In [36]:

```
df.shape
```

Out[36]:

```
(1704, 6)
```

In [39]:

```
mask = df['lifeExp'] > 60
mask
```

Out[39]:

```
0      False
1      False
2      False
3      False
4      False
...
1699    True
1700    True
1701    False
1702    False
1703    False
Name: lifeExp, Length: 1704, dtype: bool
```

In [40]:

```
df[mask]
```

Out[40]:

	country	continent	year	lifeExp	population	gdpPerCap
14	Albania	Europe	1962	64.820	1728137	2312.888958
15	Albania	Europe	1967	66.220	1984060	2760.196931
16	Albania	Europe	1972	67.690	2263554	3313.422188
17	Albania	Europe	1977	68.930	2509048	3533.003910
18	Albania	Europe	1982	70.420	2780097	3630.880722
...
1678	Yemen, Rep.	Asia	2002	60.308	18701257	2234.820827
1679	Yemen, Rep.	Asia	2007	62.698	22211743	2280.769906
1698	Zimbabwe	Africa	1982	60.363	7636524	788.855041
1699	Zimbabwe	Africa	1987	62.351	9216418	706.157306
1700	Zimbabwe	Africa	1992	60.377	10704340	693.420786

877 rows × 6 columns

In [42]:

```
df[ df['lifeExp']>60 ][ 'country' ]
```

Out[42]:

```
14      Albania
15      Albania
16      Albania
17      Albania
18      Albania
...
1678  Yemen, Rep.
1679  Yemen, Rep.
1698      Zimbabwe
1699      Zimbabwe
1700      Zimbabwe
```

Name: country, Length: 877, dtype: object

In [53]:

```
df.loc[ df['lifeExp']>60 , ['country', 'lifeExp'] ]
```

Out[53]:

	country	lifeExp
14	Albania	64.820
15	Albania	66.220
16	Albania	67.690
17	Albania	68.930
18	Albania	70.420
...
1678	Yemen, Rep.	60.308
1679	Yemen, Rep.	62.698
1698	Zimbabwe	60.363
1699	Zimbabwe	62.351
1700	Zimbabwe	60.377

877 rows × 2 columns

In [55]:

```
df
```

Out[55]:

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

1704 rows × 6 columns

In [64]:

```
df[(df['country'] == 'Zimbabwe')]
```

Out[64]:

	country	continent	year	lifeExp	population	gdpPerCap
1692	Zimbabwe	Africa	1952	48.451	3080907	406.884115
1693	Zimbabwe	Africa	1957	50.469	3646340	518.764268
1694	Zimbabwe	Africa	1962	52.358	4277736	527.272182
1695	Zimbabwe	Africa	1967	53.995	4995432	569.795071
1696	Zimbabwe	Africa	1972	55.635	5861135	799.362176
1697	Zimbabwe	Africa	1977	57.674	6642107	685.587682
1698	Zimbabwe	Africa	1982	60.363	7636524	788.855041
1699	Zimbabwe	Africa	1987	62.351	9216418	706.157306
1700	Zimbabwe	Africa	1992	60.377	10704340	693.420786
1701	Zimbabwe	Africa	1997	46.809	11404948	792.449960
1702	Zimbabwe	Africa	2002	39.989	11926563	672.038623
1703	Zimbabwe	Africa	2007	43.487	12311143	469.709298

In [72]:

```
len(df[(df['country'] == 'Zimbabwe')])
```

Out[72]:

12

In [79]:

```
df[(df['lifeExp'] >30) & (df['lifeExp'] <40)]
```

Out[79]:

	country	continent	year	lifeExp	population	gdpPerCap
1	Afghanistan	Asia	1957	30.332	9240934	820.853030
2	Afghanistan	Asia	1962	31.997	10267083	853.100710
3	Afghanistan	Asia	1967	34.020	11537966	836.197138
4	Afghanistan	Asia	1972	36.088	13079460	739.981106
5	Afghanistan	Asia	1977	38.438	14880372	786.113360
...
1670	Yemen, Rep.	Asia	1962	35.180	6120081	825.623201
1671	Yemen, Rep.	Asia	1967	36.984	6740785	862.442146
1672	Yemen, Rep.	Asia	1972	39.848	7407075	1265.047031
1690	Zambia	Africa	2002	39.193	10595811	1071.613938
1702	Zimbabwe	Africa	2002	39.989	11926563	672.038623

121 rows × 6 columns

In []:

In [80]:

```
df[ (df['country'] == 'Kenya') | (df['country'] == 'Egypt') ]
```

Out[80]:

	country	continent	year	lifeExp	population	gdpPerCap
456	Egypt	Africa	1952	41.893	22223309	1418.822445
457	Egypt	Africa	1957	44.444	25009741	1458.915272
458	Egypt	Africa	1962	46.992	28173309	1693.335853
459	Egypt	Africa	1967	49.293	31681188	1814.880728
460	Egypt	Africa	1972	51.137	34807417	2024.008147
461	Egypt	Africa	1977	53.319	38783863	2785.493582
462	Egypt	Africa	1982	56.006	45681811	3503.729636
463	Egypt	Africa	1987	59.797	52799062	3885.460710
464	Egypt	Africa	1992	63.674	59402198	3794.755195
465	Egypt	Africa	1997	67.217	66134291	4173.181797
466	Egypt	Africa	2002	69.806	73312559	4754.604414
467	Egypt	Africa	2007	71.338	80264543	5581.180998
816	Kenya	Africa	1952	42.270	6464046	853.540919
817	Kenya	Africa	1957	44.686	7454779	944.438315
818	Kenya	Africa	1962	47.949	8678557	896.966373
819	Kenya	Africa	1967	50.654	10191512	1056.736457
820	Kenya	Africa	1972	53.559	12044785	1222.359968
821	Kenya	Africa	1977	56.155	14500404	1267.613204
822	Kenya	Africa	1982	58.766	17661452	1348.225791
823	Kenya	Africa	1987	59.339	21198082	1361.936856
824	Kenya	Africa	1992	59.285	25020539	1341.921721
825	Kenya	Africa	1997	54.407	28263827	1360.485021
826	Kenya	Africa	2002	50.992	31386842	1287.514732
827	Kenya	Africa	2007	54.110	35610177	1463.249282

In [86]:

```
not False
```

Out[86]:

True

In [89]:

```
df[~(df['country'] == 'Kenya')]
```

Out[89]:

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

1692 rows × 6 columns

In [90]:

```
df
```

Out[90]:

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

1704 rows × 6 columns

In [95]:

```
df[df['continent'] == 'Asia']['lifeExp'].mean()
```

Out[95]:

60.064903232323175

In [96]:

```
df[df['continent'] == 'Africa']['lifeExp'].mean()
```

Out[96]:

48.86533012820508

In [98]:

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

Out[98]:

5

In [99]:

```
df['continent'].unique()
```

Out[99]:

```
array(['Asia', 'Europe', 'Africa', 'Americas', 'Oceania'], dtype=object)
```

In [106]:

```
df['continent'].value_counts()
```

Out[106]:

```
Africa      624
Asia        396
Europe      360
Americas    300
Oceania      24
Name: continent, dtype: int64
```

In [107]:

```
# 21st century countries having less than 1000 gdp
```

In [109]:

```
df.loc[(df['year'] > 2000) & (df['gdpPerCap'] < 1000), 'country'].value_counts()
```

Out[109]:

Gambia	2
Rwanda	2
Togo	2
Liberia	2
Congo, Dem. Rep.	2
Burundi	2
Mozambique	2
Myanmar	2
Zimbabwe	2
Sierra Leone	2
Central African Republic	2
Guinea	2
Eritrea	2
Niger	2
Ethiopia	2
Guinea-Bissau	2
Somalia	2
Afghanistan	2
Malawi	2
Cambodia	1
Madagascar	1
Mali	1
Uganda	1
Comoros	1
Tanzania	1

Name: country, dtype: int64

In []:

In [110]:

```
string = "How much wood would a woodchuck chuck if a woodchuck could chuck wood"
```

In [115]:

```
pd.Series(string.split()).value_counts()
```

Out[115]:

woodchuck	2
wood	2
chuck	2
a	2
How	1
much	1
if	1
could	1
would	1

dtype: int64

In []:

Grouping

In [122]:

```
df.groupby(by='continent')['lifeExp'].mean()
```

Out[122]:

```
continent
Africa      48.865330
Americas    64.658737
Asia        60.064903
Europe      71.903686
Oceania     74.326208
Name: lifeExp, dtype: float64
```

In [123]:

```
df.groupby(by='continent')['lifeExp'].min()
```

Out[123]:

```
continent
Africa      23.599
Americas    37.579
Asia        28.801
Europe      43.585
Oceania     69.120
Name: lifeExp, dtype: float64
```

In [128]:

```
df.groupby(by='continent').mean()[['lifeExp', 'gdpPerCap']]
```

Out[128]:

	lifeExp	gdpPerCap
continent		
Africa	48.865330	2193.754578
Americas	64.658737	7136.110356
Asia	60.064903	7902.150428
Europe	71.903686	14469.475533
Oceania	74.326208	18621.609223

In [129]:

```
g = df.groupby(by='continent')
g
```

Out[129]:

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x7fede075abe0>
```

In [132]:

```
g.mean()
```

Out[132]:

	year	lifeExp	population	gdpPerCap
continent				
Africa	1979.5	48.865330	9.916003e+06	2193.754578
Americas	1979.5	64.658737	2.450479e+07	7136.110356
Asia	1979.5	60.064903	7.703872e+07	7902.150428
Europe	1979.5	71.903686	1.716976e+07	14469.475533
Oceania	1979.5	74.326208	8.874672e+06	18621.609223

In [131]:

```
g.get_group('Asia')
```

Out[131]:

	country	continent	year	lifeExp	population	gdpPerCap
0	Afghanistan	Asia	1952	28.801	8425333	779.445314
1	Afghanistan	Asia	1957	30.332	9240934	820.853030
2	Afghanistan	Asia	1962	31.997	10267083	853.100710
3	Afghanistan	Asia	1967	34.020	11537966	836.197138
4	Afghanistan	Asia	1972	36.088	13079460	739.981106
...
1675	Yemen, Rep.	Asia	1987	52.922	11219340	1971.741538
1676	Yemen, Rep.	Asia	1992	55.599	13367997	1879.496673
1677	Yemen, Rep.	Asia	1997	58.020	15826497	2117.484526
1678	Yemen, Rep.	Asia	2002	60.308	18701257	2234.820827
1679	Yemen, Rep.	Asia	2007	62.698	22211743	2280.769906

396 rows × 6 columns

In []:

In [142]:

```
temp = g.aggregate(func= [ 'min', 'max', 'mean', np.median ])[[ 'lifeExp', 'gdpPerCap' ]]
```

Out[142]:

continent	lifeExp				gdpF			
	min	max	mean	median	min	max	mean	n
Africa	23.599	76.442	48.865330	47.7920	241.165876	21951.21176	2193.754578	1192.7
Americas	37.579	80.653	64.658737	67.0480	1201.637154	42951.65309	7136.110356	5465.5
Asia	28.801	82.603	60.064903	61.7915	331.000000	113523.13290	7902.150428	2646.7
Europe	43.585	81.757	71.903686	72.2410	973.533195	49357.19017	14469.475533	12081.7
Oceania	69.120	81.235	74.326208	73.6650	10039.595640	34435.36744	18621.609223	17983.5

In [143]:

```
temp[ 'gdpPerCap' ]
```

Out[143]:

continent	min	max	mean	median
Africa	241.165876	21951.21176	2193.754578	1192.138217
Americas	1201.637154	42951.65309	7136.110356	5465.509853
Asia	331.000000	113523.13290	7902.150428	2646.786844
Europe	973.533195	49357.19017	14469.475533	12081.749115
Oceania	10039.595640	34435.36744	18621.609223	17983.303955

In [144]:

```
temp[ ( 'gdpPerCap', 'max' ) ]
```

Out[144]:

```
continent
Africa      21951.21176
Americas    42951.65309
Asia        113523.13290
Europe      49357.19017
Oceania     34435.36744
Name: (gdpPerCap, max), dtype: float64
```

In []:

In []:

In [147]:

```
df.groupby(by=['continent', 'year']).mean()[['lifeExp', 'gdpPerCap']]
```

Out[147]:

		lifeExp	gdpPerCap
continent	year		
Africa	1952	39.135500	1252.572466
	1957	41.266346	1385.236062
	1962	43.319442	1598.078825
	1967	45.334538	2050.363801
	1972	47.450942	2339.615674
	1977	49.580423	2585.938508
	1982	51.592865	2481.592960
	1987	53.344788	2282.668991
	1992	53.629577	2281.810333
	1997	53.598269	2378.759555
	2002	53.325231	2599.385159
	2007	54.806038	3089.032605
Americas	1952	53.279840	4079.062552
	1957	55.960280	4616.043733
	1962	58.398760	4901.541870
	1967	60.410920	5668.253496
	1972	62.394920	6491.334139
	1977	64.391560	7352.007126
	1982	66.228840	7506.737088
	1987	68.090720	7793.400261
	1992	69.568360	8044.934406
	1997	71.150480	8889.300863
	2002	72.422040	9287.677107
	2007	73.608120	11003.031625
Asia	1952	46.314394	5195.484004
	1957	49.318544	5787.732940
	1962	51.563223	5729.369625
	1967	54.663640	5971.173374
	1972	57.319269	8187.468699
	1977	59.610556	7791.314020
	1982	62.617939	7434.135157
	1987	64.851182	7608.226508
	1992	66.537212	8639.690248

		lifeExp	gdpPerCap
continent	year		
Europe	1997	68.020515	9834.093295
	2002	69.233879	10174.090397
	2007	70.728485	12473.026870
	1952	64.408500	5661.057435
	1957	66.703067	6963.012816
	1962	68.539233	8365.486814
	1967	69.737600	10143.823757
	1972	70.775033	12479.575246
	1977	71.937767	14283.979110
	1982	72.806400	15617.896551
	1987	73.642167	17214.310727
	1992	74.440100	17061.568084
	1997	75.505167	19076.781802
	2002	76.700600	21711.732422
	2007	77.648600	25054.481636
	1952	69.255000	10298.085650
	1957	70.295000	11598.522455
	1962	71.085000	12696.452430
	1967	71.310000	14495.021790
	1972	71.910000	16417.333380
Oceania	1977	72.855000	17283.957605
	1982	74.290000	18554.709840
	1987	75.320000	20448.040160
	1992	76.945000	20894.045885
	1997	78.190000	24024.175170
	2002	79.740000	26938.778040
	2007	80.719500	29810.188275

In [150]:

```
df.groupby(by=['year']).mean()[['lifeExp', 'gdpPerCap'] ]
```

Out[150]:

	lifeExp	gdpPerCap
year		
1952	49.057620	3725.276046
1957	51.507401	4299.408345
1962	53.609249	4725.812342
1967	55.678290	5483.653047
1972	57.647386	6770.082815
1977	59.570157	7313.166421
1982	61.533197	7518.901673
1987	63.212613	7900.920218
1992	64.160338	8158.608521
1997	65.014676	9090.175363
2002	65.694923	9917.848365
2007	67.007423	11680.071820

In [152]:

```
temp = df.groupby(by=['continent', 'year']).mean()[['lifeExp', 'gdpPerCap']]
temp
```

Out[152]:

		lifeExp	gdpPerCap
continent	year		
Africa	1952	39.135500	1252.572466
	1957	41.266346	1385.236062
	1962	43.319442	1598.078825
	1967	45.334538	2050.363801
	1972	47.450942	2339.615674
	1977	49.580423	2585.938508
	1982	51.592865	2481.592960
	1987	53.344788	2282.668991
	1992	53.629577	2281.810333
	1997	53.598269	2378.759555
	2002	53.325231	2599.385159
	2007	54.806038	3089.032605
Americas	1952	53.279840	4079.062552
	1957	55.960280	4616.043733
	1962	58.398760	4901.541870
	1967	60.410920	5668.253496
	1972	62.394920	6491.334139
	1977	64.391560	7352.007126
	1982	66.228840	7506.737088
	1987	68.090720	7793.400261
	1992	69.568360	8044.934406
	1997	71.150480	8889.300863
	2002	72.422040	9287.677107
	2007	73.608120	11003.031625
Asia	1952	46.314394	5195.484004
	1957	49.318544	5787.732940
	1962	51.563223	5729.369625
	1967	54.663640	5971.173374
	1972	57.319269	8187.468699
	1977	59.610556	7791.314020
	1982	62.617939	7434.135157
	1987	64.851182	7608.226508

		lifeExp	gdpPerCap
continent	year		
Europe	1992	66.537212	8639.690248
	1997	68.020515	9834.093295
	2002	69.233879	10174.090397
	2007	70.728485	12473.026870
	1952	64.408500	5661.057435
	1957	66.703067	6963.012816
	1962	68.539233	8365.486814
	1967	69.737600	10143.823757
	1972	70.775033	12479.575246
	1977	71.937767	14283.979110
	1982	72.806400	15617.896551
	1987	73.642167	17214.310727
	1992	74.440100	17061.568084
	1997	75.505167	19076.781802
	2002	76.700600	21711.732422
	2007	77.648600	25054.481636
	1952	69.255000	10298.085650
	1957	70.295000	11598.522455
	1962	71.085000	12696.452430
	1967	71.310000	14495.021790
Oceania	1972	71.910000	16417.333380
	1977	72.855000	17283.957605
	1982	74.290000	18554.709840
	1987	75.320000	20448.040160
	1992	76.945000	20894.045885
	1997	78.190000	24024.175170
	2002	79.740000	26938.778040
	2007	80.719500	29810.188275

In [156]:

```
temp.iloc[-1, :]
```

Out[156]:

```
lifeExp      80.719500
gdpPerCap    29810.188275
Name: (Oceania, 2007), dtype: float64
```

In [155]:

```
temp.loc[('Africa', 1957)]
```

Out[155]:

```
lifeExp      41.266346
gdpPerCap    1385.236062
Name: (Africa, 1957), dtype: float64
```

In []:

In [158]:

```
temp.reset_index()
```

Out[158]:

	continent	year	lifeExp	gdpPerCap
0	Africa	1952	39.135500	1252.572466
1	Africa	1957	41.266346	1385.236062
2	Africa	1962	43.319442	1598.078825
3	Africa	1967	45.334538	2050.363801
4	Africa	1972	47.450942	2339.615674
5	Africa	1977	49.580423	2585.938508
6	Africa	1982	51.592865	2481.592960
7	Africa	1987	53.344788	2282.668991
8	Africa	1992	53.629577	2281.810333
9	Africa	1997	53.598269	2378.759555

In [168]:

```
df.groupby(['continent', 'year']).aggregate({'year' : 'count', 'population' : 'sum'})
```

Out[168]:

		year	population
continent	year		
Africa	1952	52	237640501
	1957	52	264837738
	1962	52	296516865
	1967	52	335289489
	1972	52	379879541
	1977	52	433061021
	1982	52	499348587
	1987	52	574834110
	1992	52	659081517
	1997	52	743832984
	2002	52	833723916
	2007	52	929539692
Americas	1952	25	345152446
	1957	25	386953916
	1962	25	433270254
	1967	25	480746623
	1972	25	529384210
	1977	25	578067699
	1982	25	630290920
	1987	25	682753971
	1992	25	739274104
	1997	25	796900410
	2002	25	849772762
	2007	25	898871184
Asia	1952	33	1395357351
	1957	33	1562780599
	1962	33	1696357182
	1967	33	1905662900
	1972	33	2150972248
	1977	33	2384513556
	1982	33	2610135582
	1987	33	2871220762
	1992	33	3133292191

	year	population
continent	year	
Europe	1997	33 3383285500
	2002	33 3601802203
	2007	33 3811953827
	1952	30 418120846
	1957	30 437890351
	1962	30 460355155
	1967	30 481178958
	1972	30 500635059
	1977	30 517164531
	1982	30 531266901
	1987	30 543094160
	1992	30 558142797
	1997	30 568944148
	2002	30 578223869
	2007	30 586098529
Oceania	1952	2 10686006
	1957	2 11941976
	1962	2 13283518
	1967	2 14600414
	1972	2 16106100
	1977	2 17239000
	1982	2 18394850
	1987	2 19574415
	1992	2 20919651
	1997	2 22241430
	2002	2 23454829
	2007	2 24549947

In []:

In []:

In [172]:

```
def func(x):
    print(x)
    return x['lifeExp'].mean() < 50
```

In [173]:

```
df.groupby('continent').filter(func)
```

	country	continent	year	lifeExp	population	gdpPerCap
24	Algeria	Africa	1952	43.077	9279525	2449.008185
25	Algeria	Africa	1957	45.685	10270856	3013.976023
26	Algeria	Africa	1962	48.303	11000948	2550.816880
27	Algeria	Africa	1967	51.407	12760499	3246.991771
28	Algeria	Africa	1972	54.518	14760787	4182.663766
...
1699	Zimbabwe	Africa	1987	62.351	9216418	706.157306
1700	Zimbabwe	Africa	1992	60.377	10704340	693.420786
1701	Zimbabwe	Africa	1997	46.809	11404948	792.449960
1702	Zimbabwe	Africa	2002	39.989	11926563	672.038623
1703	Zimbabwe	Africa	2007	43.487	12311143	469.709298

[624 rows x 6 columns]

	country	continent	year	lifeExp	population	gdpPerCap
48	Argentina	Americas	1952	62.485	17876956	5911.315053
49	Argentina	Americas	1957	64.399	19610538	6856.856212
50	Argentina	Americas	1962	65.142	21283783	7133.166023
51	Argentina	Americas	1967	65.634	22934225	8052.953021
52	Argentina	Americas	1972	67.065	24779799	9443.038526
...
1639	Venezuela	Americas	1987	70.190	17910182	9883.584648
1640	Venezuela	Americas	1992	71.150	20265563	10733.926310
1641	Venezuela	Americas	1997	72.146	22374398	10165.495180
1642	Venezuela	Americas	2002	72.766	24287670	8605.047831
1643	Venezuela	Americas	2007	73.747	26084662	11415.805690

[300 rows x 6 columns]

	country	continent	year	lifeExp	population	gdpPerCap
0	Afghanistan	Asia	1952	28.801	8425333	779.445314
1	Afghanistan	Asia	1957	30.332	9240934	820.853030
2	Afghanistan	Asia	1962	31.997	10267083	853.100710
3	Afghanistan	Asia	1967	34.020	11537966	836.197138
4	Afghanistan	Asia	1972	36.088	13079460	739.981106
...
1675	Yemen, Rep.	Asia	1987	52.922	11219340	1971.741538
1676	Yemen, Rep.	Asia	1992	55.599	13367997	1879.496673
1677	Yemen, Rep.	Asia	1997	58.020	15826497	2117.484526
1678	Yemen, Rep.	Asia	2002	60.308	18701257	2234.820827
1679	Yemen, Rep.	Asia	2007	62.698	22211743	2280.769906

[396 rows x 6 columns]

	country	continent	year	lifeExp	population	gdpPerCa
p						
12	Albania	Europe	1952	55.230	1282697	1601.05613
6						
13	Albania	Europe	1957	59.280	1476505	1942.28424
4						
14	Albania	Europe	1962	64.820	1728137	2312.88895
8						
15	Albania	Europe	1967	66.220	1984060	2760.19693
1						
16	Albania	Europe	1972	67.690	2263554	3313.42218
8						
...
...						
1603	United Kingdom	Europe	1987	75.007	56981620	21664.78767

```

0
1604 United Kingdom Europe 1992 76.420 57866349 22705.09254
0
1605 United Kingdom Europe 1997 77.218 58808266 26074.53136
0
1606 United Kingdom Europe 2002 78.471 59912431 29478.99919
0
1607 United Kingdom Europe 2007 79.425 60776238 33203.26128
0

```

```
[360 rows x 6 columns]
```

	country	continent	year	lifeExp	population	gdpPerCap
60	Australia	Oceania	1952	69.120	8691212	10039.59564
61	Australia	Oceania	1957	70.330	9712569	10949.64959
62	Australia	Oceania	1962	70.930	10794968	12217.22686
63	Australia	Oceania	1967	71.100	11872264	14526.12465
64	Australia	Oceania	1972	71.930	13177000	16788.62948
65	Australia	Oceania	1977	73.490	14074100	18334.19751
66	Australia	Oceania	1982	74.740	15184200	19477.00928
67	Australia	Oceania	1987	76.320	16257249	21888.88903
68	Australia	Oceania	1992	77.560	17481977	23424.76683
69	Australia	Oceania	1997	78.830	18565243	26997.93657
70	Australia	Oceania	2002	80.370	19546792	30687.75473
71	Australia	Oceania	2007	81.235	20434176	34435.36744
1092	New Zealand	Oceania	1952	69.390	1994794	10556.57566
1093	New Zealand	Oceania	1957	70.260	2229407	12247.39532
1094	New Zealand	Oceania	1962	71.240	2488550	13175.67800
1095	New Zealand	Oceania	1967	71.520	2728150	14463.91893
1096	New Zealand	Oceania	1972	71.890	2929100	16046.03728
1097	New Zealand	Oceania	1977	72.220	3164900	16233.71770
1098	New Zealand	Oceania	1982	73.840	3210650	17632.41040
1099	New Zealand	Oceania	1987	74.320	3317166	19007.19129
1100	New Zealand	Oceania	1992	76.330	3437674	18363.32494
1101	New Zealand	Oceania	1997	77.550	3676187	21050.41377
1102	New Zealand	Oceania	2002	79.110	3908037	23189.80135
1103	New Zealand	Oceania	2007	80.204	4115771	25185.00911

```
Out[173]:
```

	country	continent	year	lifeExp	population	gdpPerCap
24	Algeria	Africa	1952	43.077	9279525	2449.008185
25	Algeria	Africa	1957	45.685	10270856	3013.976023
26	Algeria	Africa	1962	48.303	11000948	2550.816880
27	Algeria	Africa	1967	51.407	12760499	3246.991771
28	Algeria	Africa	1972	54.518	14760787	4182.663766
...
1699	Zimbabwe	Africa	1987	62.351	9216418	706.157306
1700	Zimbabwe	Africa	1992	60.377	10704340	693.420786
1701	Zimbabwe	Africa	1997	46.809	11404948	792.449960
1702	Zimbabwe	Africa	2002	39.989	11926563	672.038623
1703	Zimbabwe	Africa	2007	43.487	12311143	469.709298

```
624 rows x 6 columns
```

In []:

In [179]:

```
len('Afghanistan')
```

Out[179]:

11

Apply

In [186]:

```
def custom(x):  
    return len(x)
```

In [188]:

```
df['country_len'] = df['country'].apply(custom)
```

In [189]:

df

Out[189]:

	country	continent	year	lifeExp	population	gdpPerCap	country_len
0	Afghanistan	Asia	1952	28.801	8425333	779.445314	11
1	Afghanistan	Asia	1957	30.332	9240934	820.853030	11
2	Afghanistan	Asia	1962	31.997	10267083	853.100710	11
3	Afghanistan	Asia	1967	34.020	11537966	836.197138	11
4	Afghanistan	Asia	1972	36.088	13079460	739.981106	11
...
1699	Zimbabwe	Africa	1987	62.351	9216418	706.157306	8
1700	Zimbabwe	Africa	1992	60.377	10704340	693.420786	8
1701	Zimbabwe	Africa	1997	46.809	11404948	792.449960	8
1702	Zimbabwe	Africa	2002	39.989	11926563	672.038623	8
1703	Zimbabwe	Africa	2007	43.487	12311143	469.709298	8

1704 rows × 7 columns

In [194]:

```
df[['lifeExp', 'gdpPerCap']]
```

Out[194]:

	lifeExp	gdpPerCap
0	28.801	779.445314
1	30.332	820.853030
2	31.997	853.100710
3	34.020	836.197138
4	36.088	739.981106
...
1699	62.351	706.157306
1700	60.377	693.420786
1701	46.809	792.449960
1702	39.989	672.038623
1703	43.487	469.709298

1704 rows × 2 columns

In [192]:

```
df[['lifeExp', 'gdpPerCap']].apply(np.sum, axis=0)
```

Out[192]:

```
lifeExp      1.013444e+05
gdpPerCap    1.229492e+07
dtype: float64
```

In [193]:

```
df[['lifeExp', 'gdpPerCap']].apply(np.sum, axis=1)
```

Out[193]:

0	808.246315
1	851.185030
2	885.097710
3	870.217138
4	776.069106
...	...
1699	768.508306
1700	753.797786
1701	839.258960
1702	712.027623
1703	513.196298

Length: 1704, dtype: float64

In []:

Handling of Missing Values

In [197]:

```
type(None)
```

Out[197]:

NoneType

In [199]:

```
type(np.nan)
```

Out[199]:

float

In [211]:

```
None == np.nan
```

Out[211]:

False

In [200]:

```
np.nan == np.nan
```

Out[200]:

False

In [203]:

```
pd.isna(np.nan)
```

Out[203]:

True

In [205]:

```
ser = pd.Series([1, np.nan, 5, None])  
ser
```

Out[205]:

```
0    1.0  
1    NaN  
2    5.0  
3    NaN  
dtype: float64
```

In [213]:

```
pd.isna(ser).sum()
```

Out[213]:

2

In [210]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1704 entries, 0 to 1703
Data columns (total 7 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   country         1704 non-null   object
 1   continent       1704 non-null   object
 2   year            1704 non-null   int64
 3   lifeExp         1704 non-null   float64
 4   population      1704 non-null   int64
 5   gdpPerCap       1704 non-null   float64
 6   country_len     1704 non-null   int64
dtypes: float64(2), int64(3), object(2)
memory usage: 93.3+ KB
```

In [216]:

```
pd.isna(df).sum()
```

Out[216]:

```
country         0
continent        0
year            0
lifeExp         0
population       0
gdpPerCap       0
country_len     0
dtype: int64
```

In []:

In [217]:

```
disease = pd.read_csv('disease.csv')
disease.shape
```

Out[217]:

```
(122, 14)
```


In [218]:

```
disease.head(15)
```

Out[218]:

	Date	Day	Cases_Guinea	Cases_Liberia	Cases_Nigeria	Cases_Senegal	Cases_UnitedK
0	7/22/2014	122	2770.0	NaN	NaN	NaN	
1	7/21/2014	121	2771.0	NaN	NaN	NaN	
2	7/20/2014	120	2765.0	8165.0	NaN	NaN	
3	7/19/2014	119	NaN	8156.0	NaN	NaN	
4	7/18/2014	118	2730.0	8115.0	NaN	NaN	
5	7/17/2014	117	2706.0	8018.0	NaN	NaN	
6	7/16/2014	116	2695.0	NaN	NaN	NaN	
7	7/15/2014	115	2630.0	7977.0	NaN	NaN	
8	7/14/2014	114	2597.0	NaN	NaN	NaN	
9	7/13/2014	113	2571.0	7862.0	NaN	NaN	
10	7/12/2014	112	NaN	7830.0	NaN	NaN	
11	7/11/2014	111	2416.0	NaN	NaN	NaN	
12	7/10/2014	110	NaN	7797.0	NaN	NaN	
13	7/9/2014	109	2292.0	NaN	20.0	1.0	
14	7/8/2014	108	NaN	7719.0	NaN	NaN	

In [223]:

```
disease.isna().sum()
```

Out[223]:

```
Date          0
Day            0
Cases_Guinea   29
Cases_Liberia  39
Cases_Nigeria  84
Cases_Senegal  97
Cases_UnitedKingdom 104
Cases_Mali     110
Deaths_Guinea  30
Deaths_Liberia 41
Deaths_Nigeria 84
Deaths_Senegal 100
Deaths_UnitedKingdom 104
Deaths_Mali    110
dtype: int64
```

In []:

In [225]:

```
disease.drop(columns='Deaths_Mali', inplace = True)
```

In [226]:

```
disease.head()
```

Out[226]:

	Date	Day	Cases_Guinea	Cases_Liberia	Cases_Nigeria	Cases_Senegal	Cases_UnitedKii
0	7/22/2014	122	2770.0	NaN	NaN	NaN	
1	7/21/2014	121	2771.0	NaN	NaN	NaN	
2	7/20/2014	120	2765.0	8165.0	NaN	NaN	
3	7/19/2014	119	NaN	8156.0	NaN	NaN	
4	7/18/2014	118	2730.0	8115.0	NaN	NaN	

In [229]:

```
disease.dropna(axis=1, thresh=50)
```

Out[229]:

	Date	Day	Cases_Guinea	Cases_Liberia	Deaths_Guinea	Deaths_Liberia
0	7/22/2014	122	2770.0	NaN	1786.0	NaN
1	7/21/2014	121	2771.0	NaN	1781.0	NaN
2	7/20/2014	120	2765.0	8165.0	1767.0	3496.0
3	7/19/2014	119	NaN	8156.0	NaN	3496.0
4	7/18/2014	118	2730.0	8115.0	1739.0	3471.0
...
117	3/27/2014	5	103.0	8.0	66.0	6.0
118	3/26/2014	4	86.0	NaN	62.0	NaN
119	3/25/2014	3	86.0	NaN	60.0	NaN
120	3/24/2014	2	86.0	NaN	59.0	NaN
121	3/23/2014	1	49.0	NaN	29.0	NaN

122 rows × 6 columns

In []:

In [235]:

```
disease['Deaths_Guinea'].fillna(value=-1,)
```

Out[235]:

```
0      1786.0
1      1781.0
2      1767.0
3        -1.0
4      1739.0
...
117      66.0
118      62.0
119      60.0
120      59.0
121      29.0
Name: Deaths_Guinea, Length: 122, dtype: float64
```

In [237]:

```
m= disease['Deaths_Guinea'].mean()
m
```

Out[237]:

```
563.2391304347826
```

In [238]:

```
disease['Deaths_Guinea'].fillna(value=m)
```

Out[238]:

```
0      1786.00000
1      1781.00000
2      1767.00000
3      563.23913
4      1739.00000
...
117      66.00000
118      62.00000
119      60.00000
120      59.00000
121      29.00000
Name: Deaths_Guinea, Length: 122, dtype: float64
```

In [240]:

```
disease['Deaths_Guinea']
```

Out[240]:

```
0      1786.0
1      1781.0
2      1767.0
3         NaN
4      1739.0
...
117      66.0
118      62.0
119      60.0
120      59.0
121      29.0
Name: Deaths_Guinea, Length: 122, dtype: float64
```

In [241]:

```
disease['Deaths_Guinea'].fillna(method='bfill')
```

Out[241]:

```
0      1786.0
1      1781.0
2      1767.0
3      1739.0
4      1739.0
...
117      66.0
118      62.0
119      60.0
120      59.0
121      29.0
Name: Deaths_Guinea, Length: 122, dtype: float64
```

In []:

```
disease['Deaths_Guinea'].fillna(method='ffill')
```

In []:

In [244]:

```
disease['Deaths_Guinea'].replace(to_replace=np.nan, value = m)
```

Out[244]:

```
0      1786.00000
1      1781.00000
2      1767.00000
3        563.23913
4      1739.00000
...
117     66.00000
118     62.00000
119     60.00000
120     59.00000
121     29.00000
Name: Deaths_Guinea, Length: 122, dtype: float64
```

In []:

In [248]:

```
disease.fillna(method='ffill', inplace=True)
disease.fillna(method='bfill', inplace=True)
```

In [250]:

```
disease
```

Out[250]:

	Date	Day	Cases_Guinea	Cases_Liberia	Cases_Nigeria	Cases_Senegal	Cases_United
0	7/22/2014	122	2770.0	8165.0	20.0	1.0	
1	7/21/2014	121	2771.0	8165.0	20.0	1.0	
2	7/20/2014	120	2765.0	8165.0	20.0	1.0	
3	7/19/2014	119	2765.0	8156.0	20.0	1.0	
4	7/18/2014	118	2730.0	8115.0	20.0	1.0	
...	
117	3/27/2014	5	103.0	8.0	0.0	1.0	
118	3/26/2014	4	86.0	8.0	0.0	1.0	
119	3/25/2014	3	86.0	8.0	0.0	1.0	
120	3/24/2014	2	86.0	8.0	0.0	1.0	
121	3/23/2014	1	49.0	8.0	0.0	1.0	

122 rows × 13 columns

In []:

In []:

In []: