

```
1 from google.colab import files
2 uploaded = files.upload() #로컬 PC에서 데이터 불러오기
```

파일 선택 선택된 파일 없음

Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving gapminder.tsv to gapminder.tsv

```
1 import pandas as pd
2 import numpy as np
3 df = pd.read_csv('gapminder.tsv', sep='Wt') #파일을 df라는 변수에 저장 (sep='Wt') : 탭 형태로 구분
4 df.head() #데이터를 0부터 9번까지 확인
5 # df.shape #변수명.shape : (row)행, (column)열 개수 출력
6
```

	country	continent	year	lifeExp	pop	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

```
1 df.info() #해당 데이터 프레임의 정보 출력 (Dtype = object는 문자열을 의미한다.)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1704 entries, 0 to 1703
Data columns (total 6 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   pop         1704 non-null   int64
5   gdpPercap   1704 non-null   float64
dtypes: float64(2), int64(2), object(2)
memory usage: 80.0+ KB
```

▼ 데이터 추출하기

- 열 단위 데이터 추출: [열이름]
- 열 한개만 추출 : 시리즈
- 열 두개 이상이면 : 데이터 프레임
- 순서를 표시하는 인덱스(라벨)이 자동으로 붙는다.

```
1 s2= pd.Series(['a', 'b', 1, 2])
2 s2
```

3 # object는 문자열을 의미

```
0    a
1    b
2    1
3    2
dtype: object
```

```
1 index_data = ['2018-5-7', '2019-5-7', '2020-5-7', '2021-5-7']
```

```
2 s3 = pd.Series([200, np.nan, 200, 208], index=index_data) #np.nan은 NULL을 의미 : 데이터 값이 없음
```

```
3 #인덱스를 직접 지정해 줄 수 있고, 값과 인덱스가 1:1로 대응되어야 한다.
```

```
4 s3
```

```
2018-5-7    200.0
2019-5-7      NaN
2020-5-7    200.0
2021-5-7    208.0
dtype: float64
```

```
1 df['country'] # 'country' 열 만 출력
```

```
0    Afghanistan
1    Afghanistan
2    Afghanistan
3    Afghanistan
4    Afghanistan
...
1699    Zimbabwe
1700    Zimbabwe
1701    Zimbabwe
1702    Zimbabwe
1703    Zimbabwe
Name: country, Length: 1704, dtype: object
```

```
1 subset = df[['country', 'continent', 'year']] #3가지 열의 값만 출력
```

```
2 subset
```

	country	continent	year
0	Afghanistan	Asia	1952

```
1 print(type(df))
```

```
<class 'pandas.core.frame.DataFrame'>
```

3	Afghanistan	Asia	1967
----------	-------------	------	------

▼ 행 단위 데이터 추출

- `.loc[]`은 인덱스를 기준으로 행 데이터를 추출
- `.iloc[]`은 데이터 순서를 의미하는 행 번호를 기준으로 행 데이터 추출

```
1 type(subset)
```

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

1703	Zimbabwe	Africa	2007
-------------	----------	--------	------

```
1 df.loc[0] #인덱스 0번의 데이터 출력
```

country	Afghanistan
continent	Asia
year	1952
lifeExp	28.801
pop	8425333
gdpPercap	779.445

Name: 0, dtype: object

```
1 df.iloc[-1] #마지막 행 정보를 보여준다
```

country	Zimbabwe
continent	Africa
year	2007
lifeExp	43.487
pop	12311143
gdpPercap	469.709

Name: 1703, dtype: object

```
1 df.loc[[0,5,99]] # 입력한 인덱스의 데이터만 출력
```

	country	continent	year	lifeExp	pop	gdpPercap
0	Afghanistan	Asia	1952	28.801	8425333	779.445314
5	Afghanistan	Asia	1977	38.438	14880372	786.113360
99	Bangladesh	Asia	1967	43.453	62821884	721.186086

```
1 df.iloc[[0,5,99]]
```

	country	continent	year	lifeExp	pop	gdpPer cap
0	Afghanistan	Asia	1952	28.801	8425333	779.445314
5	Afghanistan	Asia	1977	38.438	14880372	786.113360

```
1 # subset = df.loc[:, [2,4]] 에러
2 # subset=df.iloc[:, ['year', 'pop']] 에러
3 subset = df.iloc[:, [2,4]]
4 subset
```

	year	pop
0	1952	8425333
1	1957	9240934
2	1962	10267083
3	1967	11537966
4	1972	13079460
...
1699	1987	9216418
1700	1992	10704340
1701	1997	11404948
1702	2002	11926563
1703	2007	12311143

1704 rows × 2 columns

[퀴즈1] 0,10,100번 행 데이터를 추출하는데 열은 country, lifeExp 정보만 추출하기

```
1 subset = df.loc[[0,10,100], ['country', 'lifeExp']] #행, 열 순서
2 subset
```

	country	lifeExp
0	Afghanistan	28.801
10	Afghanistan	42.129
100	Bangladesh	45.252

```
1 df.groupby('year')['lifeExp'].mean() #연도별로 lifeExp의 평균 구하기
2 df.groupby('country')['pop'].mean() #국가별로 pop의 평균 구하기
```

country	
Afghanistan	1.582372e+07
Albania	2.580249e+06
Algeria	1.987541e+07
Angola	7.309390e+06

```
Argentina      2.860224e+07
...
Vietnam        5.456857e+07
West Bank and Gaza  1.848606e+06
Yemen, Rep.    1.084319e+07
Zambia         6.353805e+06
Zimbabwe       7.641966e+06
Name: pop, Length: 142, dtype: float64
```

[퀴즈2] lifeExp, gdpPercap 열의 평균값을 연도, 지역별로 그룹화하여 한번에 계산하기

```
1 df.groupby(['year', 'continent'])['lifeExp', 'gdpPercap'].mean()
2 # df.groupby('continent')['gdpPercap'].mean()
```



```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Indexing with
    ""Entry point for launching an IPython kernel.
```

		lifeExp	gdpPerCap
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
1962	Africa	43.319442	1598.078825
	Americas	58.398760	4901.541870
	Asia	51.563223	5729.369625
	Europe	68.539233	8365.486814
	Oceania	71.085000	12696.452430
1967	Africa	45.334538	2050.363801
	Americas	60.410920	5668.253496
	Asia	54.663640	5971.173374
	Europe	69.737600	10143.823757
	Oceania	71.310000	14495.021790
1972	Africa	47.450942	2339.615674
	Americas	62.394920	6491.334139
	Asia	57.319269	8187.468699
	Europe	70.775033	12479.575246
	Oceania	71.910000	16417.333380
1977	Africa	49.580423	2585.938508
	Americas	64.391560	7352.007126
	Asia	59.610556	7791.314020
	Europe	71.937767	14283.979110
	Oceania	72.855000	17283.957605

1982	Africa	51.592865	2481.592960
	Americas	66.228840	7506.737088
	Asia	62.617939	7434.135157
	Europe	72.806400	15617.896551
	Oceania	74.290000	18554.709840
1987	Africa	53.344788	2282.668991
	Americas	66.228840	7506.737088
	Asia	62.617939	7434.135157
	Europe	72.806400	15617.896551
	Oceania	74.290000	18554.709840
1992	Africa	53.629577	2281.810333
	Americas	69.568360	8044.934406
	Asia	66.537212	8639.690248
	Europe	74.440100	17061.568084
	Oceania	76.945000	20894.045885
1997	Africa	53.598269	2378.759555
	Americas	71.150480	8889.300863
	Asia	68.020515	9834.093295
	Europe	75.505167	19076.781802
	Oceania	78.190000	24024.175170
2002	Africa	53.325231	2599.385159

✓ 0초 오전 10:09에 완료됨

