

In [2]:

```
import pandas as pd

df = pd.read_csv('./data/birth_die.csv')
df
```

Out[2]:

	이름	주요경력	출생	사망
0	스티븐 호킹	이론 물리학자	1942-01-08	2018-03-14
1	마이클잭슨	가수	1958-08-29	2009-06-25
2	스티브잡스	CEO	1955-02-24	2011-10-05
3	로빈윌리엄스	배우	1951-07-21	2014-08-11
4	앨빈토폴러	미래학자	1928-10-04	2016-06-27

In [3]:

```
df.dtypes
```

Out[3]:

```
이름      object
주요경력  object
출생      object
사망      object
dtype: object
```

시계열 데이터

1. datetime 자료형으로 변환하기

```
pd.to_datetime(컬럼)
```

In [6]:

```
# 출생, 사망 컬럼을 datetime 자료형으로 변경하기
df['출생'] = pd.to_datetime(df.출생)
df['사망'] = pd.to_datetime(df.사망)
```

In [7]:

```
df
```

Out[7]:

	이름	주요경력	출생	사망
0	스티븐 호킹	이론 물리학자	1942-01-08	2018-03-14
1	마이클잭슨	가수	1958-08-29	2009-06-25
2	스티브잡스	CEO	1955-02-24	2011-10-05
3	로빈윌리엄스	배우	1951-07-21	2014-08-11
4	앨빈토폴러	미래학자	1928-10-04	2016-06-27

In [8]:

```
df.dtypes
```

Out[8]:

```
이름      object
주요경력  object
출생      datetime64[ns]
사망      datetime64[ns]
dtype: object
```

2. datetime 자료형에서 정보 추출하기

```
column.dt.year      // 연
column.dt.month     // 월
column.dt.day        // 일
column.dt.quarter    // 분기
```

In [10]:

```
df['출생']
```

Out[10]:

```
0    1942-01-08
1    1958-08-29
2    1955-02-24
3    1951-07-21
4    1928-10-04
Name: 출생, dtype: datetime64[ns]
```

In [9]:

```
# 출생 컬럼 0번 인덱스의 연도
df['출생'].dt.year[0]
```

Out[9]:

```
1942
```

In [11]:

```
df['출생'][0].year
```

Out[11]:

```
1942
```

In [12]:

```
df['출생'].dt.month
```

Out[12]:

```
0      1
1      8
2      2
3      7
4     10
Name: 출생, dtype: int64
```

In [13]:

```
df['출생'].dt.day
```

Out[13]:

```
0      8
1     29
2     24
3     21
4      4
Name: 출생, dtype: int64
```

In [16]:

```
# 분기 컬럼 만들기
df['출생분기'] = df['출생'].dt.quarter
df
```

Out[16]:

	이름	주요경력	출생	사망	출생분기
0	스티븐 호킹	이론 물리학자	1942-01-08	2018-03-14	1
1	마이클잭슨	가수	1958-08-29	2009-06-25	3
2	스티브잡스	CEO	1955-02-24	2011-10-05	1
3	로빈윌리엄스	배우	1951-07-21	2014-08-11	3
4	앨빈토폴러	미래학자	1928-10-04	2016-06-27	4

3. 날짜 계산하기

```
datetime2 - datetime1 = 일 수
datetime2.dt.year - datetime1.dt.year = 연도 수
```

In [18]:

```
# 생존일 수 컬럼 만들기
df['생존일수'] = df['사망'] - df['출생']
df
```

Out[18]:

	이름	주요경력	출생	사망	출생분기	생존일수
0	스티븐 호킹	이론 물리학자	1942-01-08	2018-03-14	1	27824 days
1	마이클잭슨	가수	1958-08-29	2009-06-25	3	18563 days
2	스티브잡스	CEO	1955-02-24	2011-10-05	1	20677 days
3	로빈윌리엄스	배우	1951-07-21	2014-08-11	3	23032 days
4	앨빈토폴러	미래학자	1928-10-04	2016-06-27	4	32043 days

In [20]:

```
# 생존 기간 컬럼 만들기
df['생존기간'] = df['사망'].dt.year - df['출생'].dt.year
df
```

Out[20]:

	이름	주요경력	출생	사망	출생분기	생존일수	생존기간
0	스티븐 호킹	이론 물리학자	1942-01-08	2018-03-14	1	27824 days	76
1	마이클잭슨	가수	1958-08-29	2009-06-25	3	18563 days	51
2	스티브잡스	CEO	1955-02-24	2011-10-05	1	20677 days	56
3	로빈윌리엄스	배우	1951-07-21	2014-08-11	3	23032 days	63
4	앨빈토폴러	미래학자	1928-10-04	2016-06-27	4	32043 days	88

4. 요일, 월 이름 추출하기

```
column.dt.strftime('%a')      // 요약요일명(Sun, Mon, ...,)
column.dt.strftime('%A')      // 긴요일명(Sunday, Monday, ...,)
column.dt.strftime('%w')      // 숫자요일(0: 일, 1 : 월, ...,)
column.dt.strftime('%b')      // 요약 월이름( Jan, Feb, ...,)
column.dt.strftime('%B')      // 긴 월이름( January, February, ...)
```

```
column.dt.strftime('%B') // 인플레이 (January, February, ...)
```

In [24]:

```
# 출생 요일 컬럼
df['출생요일'] = df['출생'].dt.strftime('%a')
df
```

Out[24]:

	이름	주요경력	출생	사망	출생분기	생존일수	생존기간	출생요일
0	스티븐 호킹	이론 물리학자	1942-01-08	2018-03-14	1	27824 days	76	Thu
1	마이클잭슨	가수	1958-08-29	2009-06-25	3	18563 days	51	Fri
2	스티브잡스	CEO	1955-02-24	2011-10-05	1	20677 days	56	Thu
3	로빈윌리엄스	배우	1951-07-21	2014-08-11	3	23032 days	63	Sat
4	앨빈토플러	미래학자	1928-10-04	2016-06-27	4	32043 days	88	Thu

In [29]:

```
# 출생월 컬럼 추가하기
df['출생월'] = df['출생'].dt.strftime('%b')
df
```

Out[29]:

	이름	주요경력	출생	사망	출생분기	생존일수	생존기간	출생요일	출생월
0	스티븐 호킹	이론 물리학자	1942-01-08	2018-03-14	1	27824 days	76	Thu	Jan
1	마이클잭슨	가수	1958-08-29	2009-06-25	3	18563 days	51	Fri	Aug
2	스티브잡스	CEO	1955-02-24	2011-10-05	1	20677 days	56	Thu	Feb
3	로빈윌리엄스	배우	1951-07-21	2014-08-11	3	23032 days	63	Sat	Jul
4	앨빈토플러	미래학자	1928-10-04	2016-06-27	4	32043 days	88	Thu	Oct

5. datetime 자료형을 인덱스로 사용하기

```
df.loc[연도] // 해당 연도의 데이터 추출
df.loc[연도-월] // 해당 연도-월의 데이터 추출
```

In [31]:

```
# 출생 컬럼을 인덱스로 만들기
df.index = df['출생']
df
```

Out[31]:

	이름	주요경력	출생	사망	출생분기	생존일수	생존기간	출생요일	출생월
출생									
1942-01-08	스티븐 호킹	이론 물리학자	1942-01-08	2018-03-14	1	27824 days	76	Thu	Jan
1958-08-29	마이클잭슨	가수	1958-08-29	2009-06-25	3	18563 days	51	Fri	Aug
1955-02-24	스티브잡스	CEO	1955-02-24	2011-10-05	1	20677 days	56	Thu	Feb
1951-07-21	로빈윌리엄스	배우	1951-07-21	2014-08-11	3	23032 days	63	Sat	Jul
1928-10-04	앨빈토플러	미래학자	1928-10-04	2016-06-27	4	32043 days	88	Thu	Oct

In [33]:

```
# 1955년에 출생한 데이터 추출하기
```

```
df.loc['1955']
```

Out[33]:

	이름	주요경력	출생	사망	출생분기	생존일수	생존기간	출생요일	출생월
출생									
1955-02-24	스티브잡스	CEO	1955-02-24	2011-10-05	1	20677 days	56	Thu	Feb

In [34]:

```
df.loc['1955-02']
```

Out[34]:

	이름	주요경력	출생	사망	출생분기	생존일수	생존기간	출생요일	출생월
출생									
1955-02-24	스티브잡스	CEO	1955-02-24	2011-10-05	1	20677 days	56	Thu	Feb

In [35]:

```
df.index.dtype
```

Out[35]:

dtype('<M8[ns]')