

시리즈 함수 다루기

csv 파일에 한글이 있는 경우 에러가 날 수 있다. 에러가 나는 경우 encoding 매개변수에 "euc-kr"을 추가한다.

In [2]:

```
import pandas as pd

df = pd.read_csv("person_info.csv", encoding='euc-kr')
df
```

Out[2]:

	이름	출생일	사망일	나이	직업
0	홍길동	1920-07-25	1958-04-16	37	회사원
1	이순신	1876-06-13	1937-10-16	61	군인
2	이홍길	1820-05-12	1910-08-13	90	의사
3	곽재우	1867-11-07	1934-07-04	66	연구원
4	신나라	1907-05-27	1964-04-14	56	연구원
5	오기영	1813-03-15	1858-06-16	45	연구원
6	기충현	1912-06-23	1954-06-07	41	회사원
7	심기한	1777-04-30	1855-02-23	77	백수

In [3]:

```
age = df['나이']
age
```

Out[3]:

```
0    37
1    61
2    90
3    66
4    56
5    45
6    41
7    77
Name: 나이, dtype: int64
```

시리즈의 평균 값 구하기

In [4]:

```
age.mean()
```

Out[4]:

```
59.125
```

시리즈의 중앙 값 구하기

In [5]:

```
age.median()
```

Out[5]:

58.5

시리즈의 최대 값 구하기

In [6]:

```
age.max()
```

Out[6]:

90

시리즈의 최소 값 구하기

In [7]:

```
age.min()
```

Out[7]:

37

시리즈의 표준편차 값 구하기

In [8]:

```
age.std()
```

Out[8]:

18.325918413937288

시리즈의 요약 통계량 값 구하기

In [9]:

```
age.describe()
```

Out[9]:

```
count      8.000000
mean       59.125000
std        18.325918
min        37.000000
25%        44.000000
50%        58.500000
75%        68.750000
max        90.000000
Name: 나이, dtype: float64
```

시리즈에서 중복된 값이 있다면 삭제하기

In [10]:

```
age.drop_duplicates()
```

Out[10]:

```
0    37
1    61
2    90
3    66
4    56
5    45
6    41
7    77
Name: 나이, dtype: int64
```

시리즈의 특정 값을 다른 값으로 바꾸기

inplace=True 값을 매개변수로 넣어주어야 저장이 된다

In [11]:

```
age.replace(37, 100)
```

Out[11]:

```
0    100
1     61
2     90
3     66
4     56
5     45
6     41
7     77
Name: 나이, dtype: int64
```

In [12]:

```
age
```

Out[12]:

```
0    37
1    61
2    90
3    66
4    56
5    45
6    41
7    77
Name: 나이, dtype: int64
```

In [13]:

```
age.replace(37, 100, inplace=True)
```

In [14]:

```
age
```

Out[14]:

```
0    100
1     61
2     90
3     66
4     56
5     45
6     41
7     77
Name: 나이, dtype: int64
```

시리즈에서 랜덤으로 샘플 값 추출하기

In [15]:

```
age.sample(5)
```

Out[15]:

```
2     90
7     77
0    100
1     61
5     45
Name: 나이, dtype: int64
```

시리즈 값 정렬하기

ascending=False를 매개변수에 넣어주면 내림차순으로 정렬 가능하다

In [16]:

```
age.sort_values()
```

Out[16]:

```
6    41
5    45
4    56
1    61
3    66
7    77
2    90
0   100
Name: 나이, dtype: int64
```

In [17]:

```
age.sort_values(ascending=False)
```

Out[17]:

```
0    100
2     90
7     77
3     66
1     61
4     56
5     45
6     41
Name: 나이, dtype: int64
```

시리즈를 데이터프레임으로 변환하기

In [18]:

```
age.to_frame()
```

Out[18]:

	나이
0	100
1	61
2	90
3	66
4	56
5	45
6	41
7	77

데이터 프레임 특정 열 기준으로 정렬하기

In [19]:

```
df.sort_values(by='나이')
```

Out [19]:

	이름	출생일	사망일	나이	직업
6	기충현	1912-06-23	1954-06-07	41	회사원
5	오기영	1813-03-15	1858-06-16	45	연구원
4	신나라	1907-05-27	1964-04-14	56	연구원
1	이순신	1876-06-13	1937-10-16	61	군인
3	곽재우	1867-11-07	1934-07-04	66	연구원
7	심기한	1777-04-30	1855-02-23	77	백수
2	이홍길	1820-05-12	1910-08-13	90	의사
0	홍길동	1920-07-25	1958-04-16	100	회사원