

# PANDAS 데이터 필터링

## ◆ 필터링(Filtering)

- Series, DataFrame에서 **특정 조건식 만족하는 원소만 추출**
- 대표적인 방법 → **불린 인덱싱(Boolean Indexing)**

## ◆ 필터링(Filtering)

### ■ 불린 인덱싱(Boolean Indexing)

- 비교 연산자 or 논리 연산자 결과값 → True/False
- 값이 **bool** 타입으로만 구성된 **Series**
- 같은 **Index**가 **True**인 행/열만 추출 → 인덱스처럼 사용
- MASK라고도 함

## ◆ 필터링(Filtering)

### ■ 불린 인덱싱(Boolean Indexing)

연산결과					
False					
False					
True					
False					
True					
True					
False					

## ◆ 필터링(Filtering)

### ■ 불린 인덱싱(Boolean Indexing)

- 비교 연산자 →  $>$ ,  $>=$ ,  $<$ ,  $<=$ ,  $==$ ,  $!=$

- 논리 연산자 → (왼쪽 결과)  $\&$  (오른쪽 결과)

(왼쪽 결과)  $|$  (오른쪽 결과)

$\sim$  (조건식)

## ◆ 필터링(Filtering)

### ■ 불린 인덱싱(Boolean Indexing)

```
df = pd.DataFrame({'A': [1, 2, 3], 'B': [2, 2, 2]})
```

## - 특정 행의 원소만

```
result1 = df.iloc[0] >= 2
```

```
print(f'result1 == == \n{result1}', end='\n\n')
```

## - 0번, 1번 행의 원소만

```
result2 = df.iloc[0:2] >= 2
```

```
print(f'result2 == == \n{result2}', end='\n\n')
```

```
result1 == ==
```

```
A    False
```

```
B     True
```

```
Name: 0, dtype: bool
```

```
result2 == ==
```

```
      A      B
```

```
0  False  True
```

```
1   True  True
```

## ◆ 필터링(Filtering)

- 특정 값 가진 데이터 추출

```
DataFrame[열].isin( 추출값 리스트 )
```

- 특정 값 제외한 데이터 추출

```
~ DataFrame[열].isin( 추출값 리스트 )
```

## ◆ 필터링(Filtering)

### ■ 특정 값 가진 데이터 추출

```
import pandas as pd

df = pd.DataFrame({'A': [1, 2, 3], 'B': ['a', 'b', 'f']})

print('df.isin()' , df.isin([1, 3, 12, 'a']) , end='\n\n')

print('~ df.isin()' , ~df.isin([1, 3, 12, 'a']) , end='\n\n')
```

```
df.isin()
      A      B
0  True  True
1  False False
2  True  False
```

```
~ df.isin()
      A      B
0  False False
1   True  True
2  False  True
```



## ◆ 필터링(Filtering)

### ■ query( )

- 불리언 인덱싱으로 데이터 프레임의 각 행과 일치
- True 로 평가되는 행만 반환
- 특정 조건을 만족하는 결과를 추출

## ◆ 필터링(Filtering)

- query() 활용 특정 값 가진 데이터 추출

```
df = pd.DataFrame({'A': range(1, 6),  
                  'B': range(10, 0, -2),  
                  'C C': range(10, 5, -1)})
```

```
result1 = df.query('A > B')  
print(f'[result1]\n{result1}', end='\n\n')
```

[df]

	A	B	C C
0	1	10	10
1	2	8	9
2	3	6	8
3	4	4	7
4	5	2	6

[result1]

	A	B	C C
4	5	2	6

## ◆ 필터링(Filtering)

- query() 활용 특정 값 가진 데이터 추출

```
df = pd.DataFrame({ 'A': range(1, 6),  
                    'B': range(10, 0, -2),  
                    'C C': range(10, 5, -1)})
```

```
result2 = df[df.A > df.B]  
print(f'[result2]\n{result2}', end='\n\n')
```

```
result3 = df[df.B == df['C C']]  
print(f'[result3]\n{result3}', end='\n\n')
```

```
[result2]  
   A  B  C C  
4  5  2   6
```

```
[result3]  
   A  B  C C  
0  1 10 10
```