

# Python 데이터 분석과 이미지 처리

나동빈

# Pandas의 기본 사용법

## Pandas란?

- 데이터를 효과적으로 처리하고, 보여줄 수 있도록 도와주는 라이브러리입니다.
- Numpy와 함께 사용되어 다양한 연계적인 기능을 제공합니다.
- 인덱스(Index)에 따라 데이터를 나열하므로 사전(Dictionary) 자료형에 가깝습니다.
- 시리즈(Series)를 기본적인 자료형으로 사용합니다.
- 엑셀(Excel)과 거의 똑같습니다.

# Pandas의 기본 사용법

## Series란?

- 시리즈는 인덱스와 값으로 구성됩니다.

# Pandas의 기본 사용법

## 시리즈 사용해보기

```
import pandas as pd

array = pd.Series(['사과', '바나나', '당근'], index=['a', 'b', 'c'])
print(array)
print(array['a'])
```

# Pandas의 기본 사용법

## 시리즈 사용해보기

```
import pandas as pd

# Dict 자료형을 Series로 바꾸기
data = {
    'a': '사과',
    'b': '바나나',
    'c': '당근'
}

array = pd.Series(data)
print(array['a'])
```

# Pandas의 기본 사용법

## 데이터 프레임이란?

- 다수의 시리즈(Series)를 모아 처리하기 위한 목적으로 사용합니다.
- 표 형태로 데이터를 손쉽게 출력하고자 할 때 사용할 수 있습니다.

# Pandas의 기본 사용법

## 데이터 프레임 사용해보기

```
import pandas as pd

word_dict = {
    'Apple': '사과',
    'Banana': '바나나',
    'Carrot': '당근'
}
frequency_dict = {
    'Apple': 3,
    'Banana': 5,
    'Carrot': 7
}

word = pd.Series(word_dict)
frequency = pd.Series(frequency_dict)
summary = pd.DataFrame({
    'word': word,
    'frequency': frequency
})

print(summary)
```

# Pandas의 기본 사용법

## 시리즈의 연산

- 시리즈를 서로 연산하여 새로운 시리즈를 만들 수 있습니다.

|     | 가격   |   | 개수  |    | 최종 가격 |
|-----|------|---|-----|----|-------|
| 사과  | 300원 | + | 사과  | 3개 | 900원  |
| 바나나 | 400원 |   | 바나나 | 2개 | 800원  |
| 당근  | 350원 |   | 당근  | 3개 | 1050원 |



# Pandas의 기본 사용법

## 시리즈의 연산

```
import pandas as pd

word_dict = {
    'Apple': '사과',
    'Banana': '바나나',
    'Carrot': '당근'
}

frequency_dict = {
    'Apple': 3,
    'Banana': 5,
    'Carrot': 7
}

importance_dict = {
    'Apple': 3,
    'Banana': 2,
    'Carrot': 1
}
```

```
word = pd.Series(word_dict)
frequency = pd.Series(frequency_dict)
importance = pd.Series(importance_dict)

summary = pd.DataFrame({
    'word': word,
    'frequency': frequency,
    'importance': importance
})

score = summary['frequency'] * summary['importance']
summary['score'] = score

print(summary)
```

# Pandas의 기본 사용법

## 데이터 프레임의 슬라이싱

```
import pandas as pd

word_dict = {
    'Apple': '사과',
    'Banana': '바나나',
    'Carrot': '당근',
    'Durian': '두리안'
}

frequency_dict = {
    'Apple': 3,
    'Banana': 5,
    'Carrot': 7,
    'Durian': 2
}

importance_dict = {
    'Apple': 3,
    'Banana': 2,
    'Carrot': 1,
    'Durian': 1
}
```

```
word = pd.Series(word_dict)
frequency = pd.Series(frequency_dict)
importance = pd.Series(importance_dict)

summary = pd.DataFrame({
    'word': word,
    'frequency': frequency,
    'importance': importance
})

print(summary)

# 이름을 기준으로 슬라이싱
print(summary.loc['Banana':'Carrot', 'importance':])

# 인덱스를 기준으로 슬라이싱
print(summary.iloc[1:3, 2:])
```

# Pandas의 기본 사용법

## 데이터 프레임의 연산

```
import pandas as pd

word_dict = {
    'Apple': '사과',
    'Banana': '바나나',
    'Carrot': '당근',
    'Durian': '두리안'
}

frequency_dict = {
    'Apple': 3,
    'Banana': 5,
    'Carrot': 7,
    'Durian': 2
}

importance_dict = {
    'Apple': 3,
    'Banana': 2,
    'Carrot': 1,
    'Durian': 1
}
```

```
word = pd.Series(word_dict)
frequency = pd.Series(frequency_dict)
importance = pd.Series(importance_dict)

summary = pd.DataFrame({
    'word': word,
    'frequency': frequency,
    'importance': importance
})

print(summary)

summary.loc['Apple', 'importance'] = 5 # 데이터의 변경
summary.loc['Elderberry'] = ['엘더베리', 5, 3] # 새 데이터 삽입

print(summary)
```

# Pandas의 기본 사용법

## 엑셀로 내보내기/불러오기

```
import pandas as pd

word_dict = {
    'Apple': '사과',
    'Banana': '바나나',
    'Carrot': '당근'
}

frequency_dict = {
    'Apple': 3,
    'Banana': 5,
    'Carrot': 7
}
```

```
word = pd.Series(word_dict)
frequency = pd.Series(frequency_dict)

summary = pd.DataFrame({
    'word': word,
    'frequency': frequency
})

summary.to_csv("summary.csv", encoding="utf-8-sig")
saved = pd.read_csv("summary.csv", index_col=0)
print(saved)
```