

Python 데이터 분석과 이미지 처리

나동빈

Pandas의 활용

데이터 프레임의 마스킹

```
import numpy as np
import pandas as pd

df = pd.DataFrame(np.random.randint(1, 10, (2, 2)), index=[0, 1], columns=["A", "B"])
# 데이터 프레임 출력하기
print(df)
# 컬럼 A의 각 원소가 5보다 작거나 같은지 출력
print(df["A"] <= 5)
# 컬럼 A의 원소가 5보다 작고, 컬럼 B의 원소가 8보다 작은 행 추출
print(df.query("A <= 5 and B <= 8"))
```

Pandas의 활용

데이터 프레임의 개별 연산 ①

```
import pandas as pd

df = pd.DataFrame([[1, 2, 3, 4], [1, 2, 3, 4]], index=[0, 1], columns=["A", "B", "C", "D"])
print(df)
df = df.apply(lambda x: x + 1)
print(df)

def addOne(x):
    return x + 1

df = df.apply(addOne)
print(df)
```

Pandas의 활용

데이터 프레임의 개별 연산 ②

```
import pandas as pd

df = pd.DataFrame([
    ['Apple', 'Apple', 'Carrot', 'Banana'],
    ['Durian', 'Banana', 'Apple', 'Carrot']],
    index=[0, 1],
    columns=["A", "B", "C", "D"])
print(df)
df = df.replace({"Apple": "Airport"})
print(df)
```

Pandas의 활용

데이터 프레임의 그룹화 ①

```
import pandas as pd

df = pd.DataFrame([
    ['Apple', 7, 'Fruit'],
    ['Banana', 3, 'Fruit'],
    ['Beef', 5, 'Meal'],
    ['Kimchi', 4, 'Meal']],
    columns=["Name", "Frequency", "Type"])
print(df)
print(df.groupby(['Type']).sum())
```

Pandas의 활용

데이터 프레임의 그룹화 ②

```
import numpy as np
import pandas as pd

df = pd.DataFrame([
    ['Apple', 7, 5, 'Fruit'],
    ['Banana', 3, 6, 'Fruit'],
    ['Beef', 5, 2, 'Meal'],
    ['Kimchi', 4, 8, 'Meal']],
    columns=["Name", "Frequency", "Importance", "Type"])
print(df)
print(df.groupby(["Type"]).aggregate([min, max, np.average]))
```

Pandas의 활용

데이터 프레임의 그룹화 ③

```
import pandas as pd

df = pd.DataFrame([
    ['Apple', 7, 5, 'Fruit'],
    ['Banana', 3, 6, 'Fruit'],
    ['Beef', 5, 2, 'Meal'],
    ['Kimchi', 4, 8, 'Meal']],
    columns=["Name", "Frequency", "Importance", "Type"])

def my_filter(data):
    return data["Frequency"].mean() >= 5

df = df.groupby("Type").filter(my_filter)
print(df)
```

Pandas의 활용

데이터 프레임의 그룹화 ④

```
import pandas as pd

df = pd.DataFrame([
    ['Apple', 7, 5, 'Fruit'],
    ['Banana', 3, 6, 'Fruit'],
    ['Beef', 5, 2, 'Meal'],
    ['Kimchi', 4, 8, 'Meal']],
    columns=["Name", "Frequency", "Importance", "Type"])

df = df.groupby("Type").get_group("Fruit")
print(df)
```


Pandas의 활용

데이터 프레임의 그룹화 ⑤

```
import pandas as pd

df = pd.DataFrame([
    ['Apple', 7, 5, 'Fruit'],
    ['Banana', 3, 6, 'Fruit'],
    ['Beef', 5, 2, 'Meal'],
    ['Kimchi', 4, 8, 'Meal']],
    columns=["Name", "Frequency", "Importance", "Type"])

df["Gap"] = df.groupby("Type")["Frequency"].apply(lambda x: x - x.mean())
print(df)
```

Pandas의 활용

데이터 프레임의 다중화

```
import numpy as np
import pandas as pd

df = pd.DataFrame(
    np.random.randint(1, 10, (4, 4)),
    index=[['1차', '1차', '2차', '2차'], ['공격', '수비', '공격', '수비']],
    columns=['1회', '2회', '3회', '4회']
)

print(df)
print(df[["1회", "2회"]].loc["2차"])
```

Pandas의 활용

피벗 테이블의 기초

```
import numpy as np
import pandas as pd

df = pd.DataFrame([
    ['Apple', 7, 5, 'Fruit'],
    ['Banana', 3, 6, 'Fruit'],
    ['Coconut', 2, 6, 'Fruit'],
    ['Rice', 8, 2, 'Meal'],
    ['Beef', 5, 2, 'Meal'],
    ['Kimchi', 4, 8, 'Meal']],
    columns=["Name", "Frequency", "Importance", "Type"])

df = df.pivot_table(
    index="Importance", columns="Type", values="Frequency",
    aggfunc=np.max
)
print(df)
```