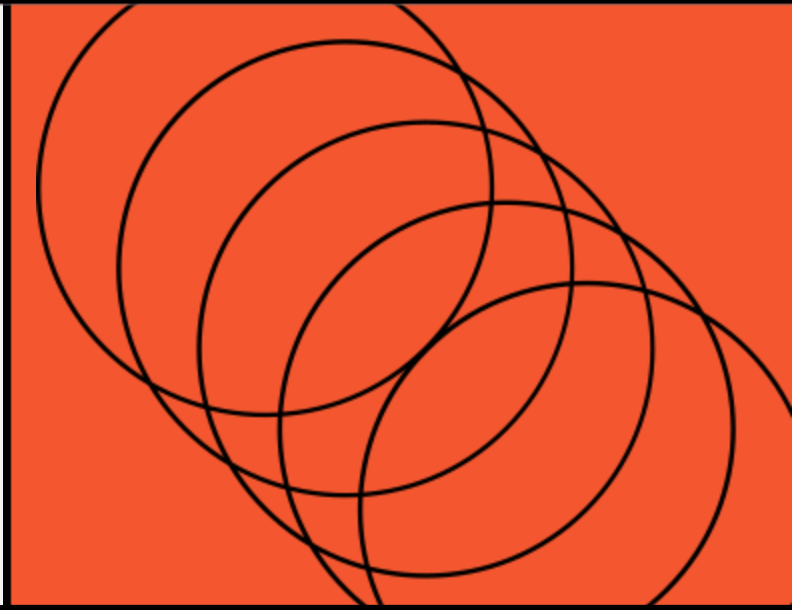


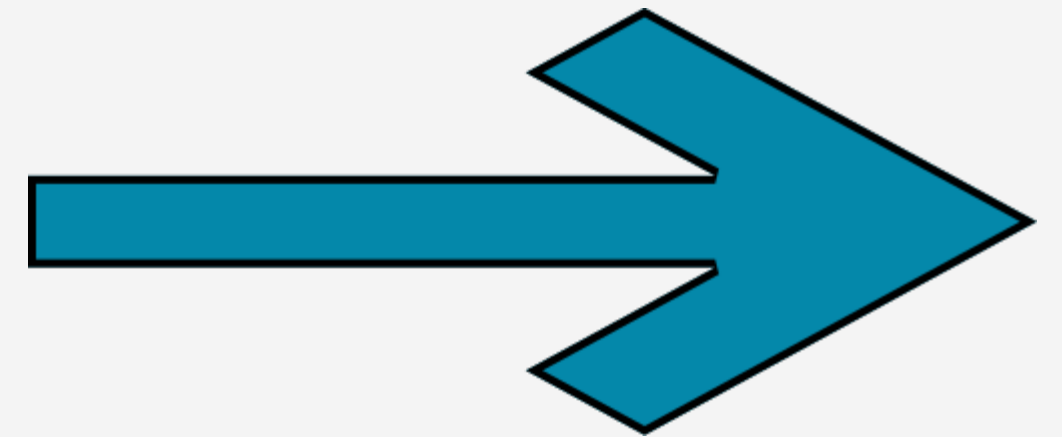
BITAMIN



11기 · FIRST SESSION

**NUMPY,
PANDAS**

교육부
최진우, 박소연



Numpy

Numpy(Numerical + Python)

- Numpy가 필요한 이유

데이터 수집, 데이터 정리, 시각화, 분석 등 할 일이 많아진다
이러한 과정들에서 계산 작업이 굉장히 복잡해질 수 있다.

이 때 **계산 작업**을 편하게 해줄 도구 중 하나



Numpy는 계산작업의 효율성을 높여주는 도구

Numpy 배열(array)

- Numpy 배열의 모양

```
array1 = np.array([1,3,4,7,8,10,11,13,15,21,23])  
array2 = np.array([[1,2,3,4],[5,6,7,8],[9,10,11,12]])
```



array1.shape

(11,)

array2.shape

(3, 4)

이들의 자료형은 어떻게 나올까요??!!

Numpy 배열(array)

- Numpy 배열의 자료형

```
array1  
array([ 1,  3,  4,  7,  8, 10, 11, 13, 15, 21, 23])  
  
type(array1) # 어떤 자료형인지 알아보자!  
numpy.ndarray
```

```
array1  
array([ 1,  3,  4,  7,  8, 10, 11, 13, 15, 21, 23])  
  
array2  
array([[ 1,  2,  3,  4],  
       [ 5,  6,  7,  8],  
       [ 9, 10, 11, 12]])
```

1차원 배열

2차원 배열

넘파이 배열은 파이썬의 리스트와 굉장히 비슷하다.

Numpy 배열(array)

- Numpy 배열과 파이썬 리스트

Numpy 배열

--> 파이썬의 리스트 개념과 비슷



리스트와 똑같이

인덱싱과 슬라이싱을 할 수 있다.

인덱싱과 슬라이싱

(0)	(1)	(2)	(3)
-----	-----	-----	-----

[1차원 벡터]

(0,0)	(0,1)	(0,2)	(0,3)
(1,0)	(1,1)	(1,2)	(1,3)
(2,0)	(2,1)	(2,2)	(2,3)
(3,0)	(3,1)	(3,2)	(3,3)

[2차원 배열]

파이썬 리스트를 인덱싱할 때와 굉장히 유사하다!!

인덱싱과 슬라이싱

- 인덱싱 예제

```
array1 = np.array([1,3,4,7,8,10,11,13,15,21,23])
```

```
array1[0] --> 1, array1[2] --> 4
```

```
array1[[1,3,4]] --> array([3, 7, 11])
```


인덱싱과 슬라이싱

- 인덱싱 예제

```
array2 = np.array([[1,2,3,4],[5,6,7,8],[9,10,11,12]])
```

```
array2[0] --> array([1,2,3,4]), array2[0,2] --> 3
```

```
array2[[1,0]] --> array([[5,6,7,8],[1,2,3,4]])
```

```
array2[[1,0],[0,2]] --> array([5,3])
```

인덱싱과 슬라이싱

- 슬라이싱 예제

```
array1 = np.array([1,3,4,7,8,10,11,13,15,21,23])
```

```
array1[1:4] --> array([3,4,7]), array1[:5] --> array([1,3,4,7,8])
```

```
array2 = np.array([[1,2,3,4],[5,6,7,8],[9,10,11,12]])
```

```
array2[0:2] --> array([[1,2,3,4],[5,6,7,8]])
```

Numpy 기본 연산

- Numpy 배열에서 +, *, % 실행하기

```
array3 = np.arange(10)
```

```
array3 --> array([0,1,2,3,4,5,6,7,8,9])
```

```
for i in range(len(array3)):
```

```
array[i] = 5 * array3[i]
```

```
list = [1,2,3,4]
```

```
list * 3 --> [1,2,3,4,1,2,3,4,1,2,3,4]
```

```
array3 * 5 --> array([0,5,10,15,20,25,30,35,40,45])
```

리스트와는 다르게 연산에 있어서 간단하고 뛰어난 성능

Numpy 다른 기능

- 통계 및 수학함수, 난수생성 등등

```
matrix = np.array([[1,2,3],[4,5,6],[7,8,9]])
```

```
np.sum(matrix) --> 45, matrix.mean() --> 5.0
```

```
np.random.random(10) -->
```

```
array([0.14990854, 0.68927404 ...])
```

```
np.random.randint(0,20,5) -->
```

```
array([13, 11, 18, 14, 13])
```

Numpy에도 다양한 기능이 있으니 활용해보자!!

파이썬 리스트 vs 넘파이 배열

- 상황에 따라서 사용하기



[10, 2023, '최진우', True]

여러가지 유형의 값들이 같이 들어갈 수 있음



[10, 2023, 27, 0253] ['Hello','Choi','Crush']

한 유형의 값들만 넣을 수 있음

리스트 --> 값을 추가하고 제거하는 일
넘파이 --> 수치 계산이 많고 복잡할 때, 행렬과 같은 다차원 배열일 때

코딩테스트 해설



코딩테스트 문제 1번

- 2023년 1월 29일, 자신의 이름, MBTI, 핸드폰 뒷번호를 리스트로 저장하고 그것을 Series 의 형태로 나타내시오

<답안 예시>

0 2023-01-29

1 홍길동

2 ESFP

3 0123

```
list_data = ['2023-01-29', '홍길동', 'ESFP', '0123']  
sr = pd.Series(list_data)  
print(sr)
```

```
0    2023-01-29  
1         홍길동  
2         ESFP  
3         0123
```

코딩테스트 문제 1번

딕셔너리를 사용하여 다음의 표를 만드시오

	국어	수학	영어	탐구
사람1	95	100	90	80
사람2	85	85	90	100
사람3	80	90	80	90



```
exam_data = {'국어' : [95,85,80], '수학' : [100,85,90], '영어' : [90,90,80], '탐구' : [80,100,90]}  
df = pd.DataFrame(exam_data, index = ['사람1', '사람2', '사람3'])  
df
```


코딩테스트 문제 1번

만들어진 표에서 탐구열을 삭제하세요

	국어	수학	영어	탐구
사람1	95	100	90	80
사람2	85	85	90	100
사람3	80	90	80	90



	국어	수학	영어
사람1	95	100	90
사람2	85	85	90
사람3	80	90	80

```
df.drop('탐구', axis = 1, inplace = True)  
df
```

코딩테스트 문제 1번

df에 국어, 수학, 영어 3과목의 평균 점수를 나타내는 '평균' 열을 추가하여 출력하여 주세요
(소수점 둘째자리까지)

	국어	수학	영어	평균
사람1	95	100	90	95.00
사람2	85	85	90	86.67
사람3	80	90	80	83.33

```
df['평균'] = round((df['국어'] + df['수학'] + df['영어']) / 3 , 2)  
df
```

코딩테스트 문제 1번

'평균'열에 나온 점수를 기준으로 등급으로 변환한 '등급'열을 추가하여 출력하세요.

	국어	수학	영어	평균	등급
사람1	95	100	90	95.00	A+
사람2	85	85	90	86.67	B+
사람3	80	90	80	83.33	B

```
grades = []
for i in df['평균']:
    if i >= 95:
        grades.append('A+')
    elif i >= 90:
        grades.append('A')
    elif i >= 85:
        grades.append('B+')
    else:
        grades.append('B')

df['등급'] = grades
df
```

코딩테스트 문제 2번

spotify.csv파일을 불러오고 data변수에 저장 후
맨 위 3개 행을 출력하세요.

	Country	Rank	Track_id	Streams	Track Name	Artist	URL	acousticness	danceability	energy	...
0	global	1.0	25sgk305KZfyuqVBQlahim	1166185736	Sweet but Psycho	Ava Max	https://open.spotify.com/track/25sgk305KZfyuqV...	0.0691	0.719	0.704	...
1	global	2.0	2Fxmhks0bxGSBdJ92vM42m	1052358787	bad guy	Billie Eilish	https://open.spotify.com/track/2Fxmhks0bxGSBdJ...	0.3280	0.701	0.425	...
2	global	3.0	6ocbgoVGwYJhOv1Ggl9NsF	789094044	7 rings	Ariana Grande	https://open.spotify.com/track/6ocbgoVGwYJhOv1...	0.5920	0.778	0.317	...

```
data = pd.read_csv('spotify.csv')  
data.head(3)
```

코딩테스트 문제 2번

data의 열 중에서 'Artist'열에서 고유값의 종류와 개수를 확인할 수 있는 함수를 쓰고 출력하세요.

```
Post Malone      45
Juice WRLD       31
Ariana Grande    28
Billie Eilish    27
Taylor Swift     27
..
Leona Lewis      1
Kylie Minogue    1
Omar Montes      1
Paloma Mami      1
Bee Gees         1
Name: Artist, Length: 488, dtype: int64
```

```
data['Artist'].value_counts()
```


코딩테스트 문제 2번

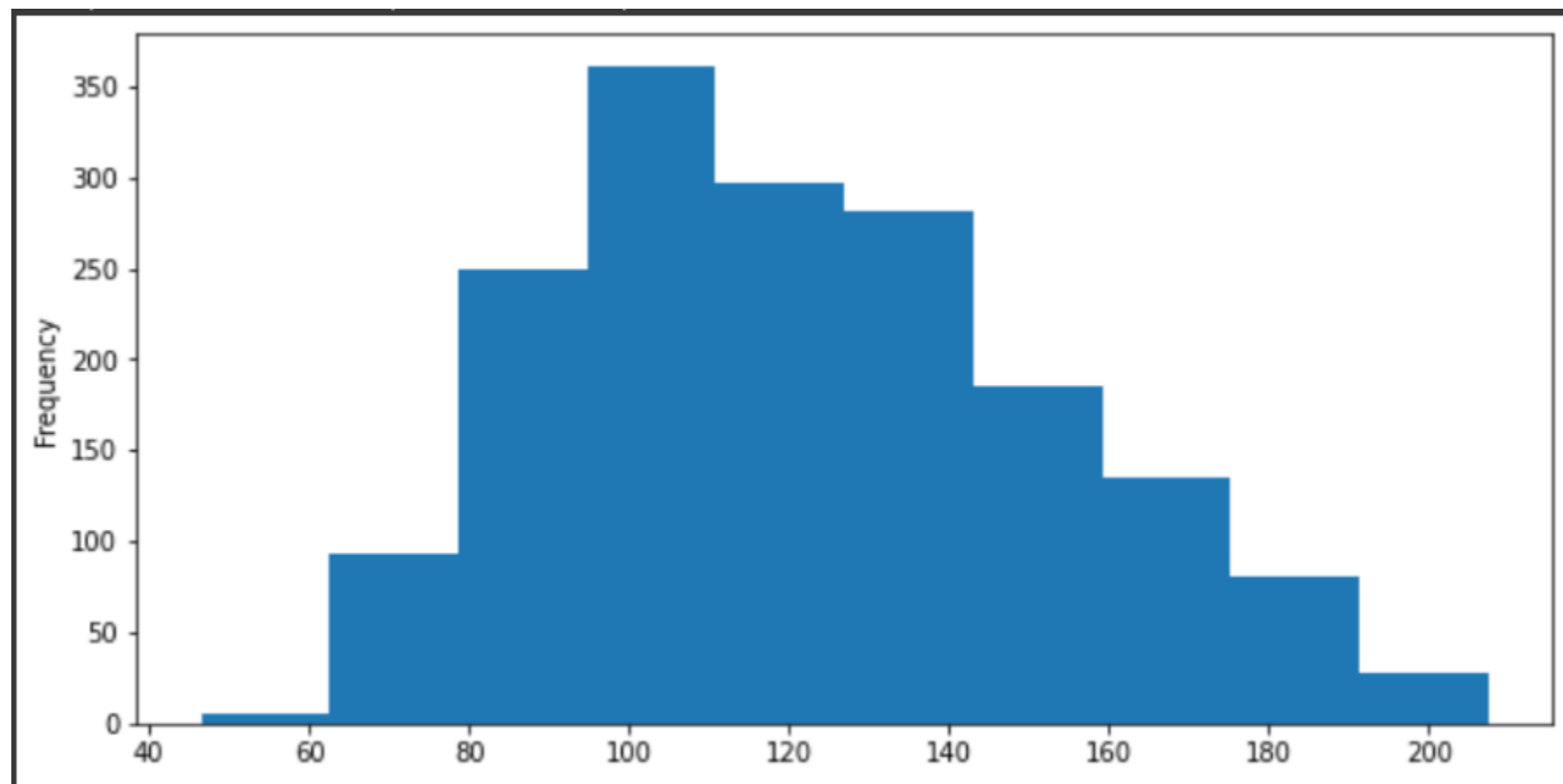
'Artist'열에서 아티스트가 BTS인 행들을 찾아 출력하세요.

	Country	Rank	Track_id	Streams	Track Name	Artist	URL	acousticness	danceability	energy	...
86	global	87.0	5KawlOMHjWeUjQtnuRs22c	261504425	Boy With Luv (feat. Halsey)	BTS	https://open.spotify.com/track/5KawlOMHjWeUjQt...	0.0923	0.645	0.862	...
464	global	465.0	314ZkcV7oLWG8yWE7LABvH	42873724	Make It Right (feat. Lauv)	BTS	https://open.spotify.com/track/314ZkcV7oLWG8yW...	0.0304	0.584	0.685	...
676	global	677.0	5hnbE5BF2e8BCk9OMR1UVC	18801822	Mikrokosmos	BTS	https://open.spotify.com/track/5hnbE5BF2e8BCk9...	0.1250	0.580	0.858	...
678	global	679.0	4c1WgUnHXq2LEncGHKgj1l	18648089	Dream Glow (BTS World Original Soundtrack) - P...	BTS	https://open.spotify.com/track/4c1WgUnHXq2LEnc...	0.0967	0.735	0.740	...
681	global	682.0	6KUE0HSSudAp9AOycaF6SH	18423901	Make It Right	BTS	https://open.spotify.com/track/6KUE0HSSudAp9AO...	0.0104	0.638	0.703	...

```
data[data['Artist'] == 'BTS']
```

코딩테스트 문제 2번

data의 열 중에서 'tempo'열에 대한 히스토그램을 그려주세요
(구간은 총 10개로 나누고 그림사이즈는 (10,5)로 설정해주세요)



```
data['tempo'].plot(kind='hist', bins = 10, figsize = (10,5))
```

코딩테스트 문제 2번

'Artist'가 BTS이면서 'tempo'가 평균 이상인 것들을 출력하세요
(열은 Rank, Track Name, Artist, tempo만 나오도록 하세요)

	Rank		Track Name	Artist	tempo
676	677.0		Mikrokosmos	BTS	174.039
678	679.0	Dream Glow (BTS World Original Soundtrack) - P..		BTS	141.948
710	711.0		HOME	BTS	142.991
725	726.0		Dionysus	BTS	176.084

```
mmm = data['tempo'].mean()  
data[(data['tempo'] >= mmm) & (data['Artist'] == 'BTS')][['Rank', 'Track Name', 'Artist', 'tempo']]
```