넘파이, 판다스 기본개념 익히기

파이썬 기본자료형과 컨테이너

기본자료형	• 숫자 : 정수형, 실수형 • 불리언(Booleans) : true, false • 문자열
컨테이너	• 리스트 • 딕셔너리 • 집합 • 튜플

Numpy

출처 : https://numpy.org/

- 고성능 과학(수치)계산을 위한 데이터분석(수치분석) 패키지
  - 벡터 산술연산
  - 다차원 배열 (ndarray)
  - 표준 수학 함수
  - 선형대수, 난수

Numpy 기술통계(descriptive statistics)

데이터의 개수 (count)	len(x) x = {1,2,3,4,5,....,N} ----> x = np.array([1,2,3,4,5,....N]) x의 갯수는 numpy에서 len(x)
평균 (average, mean)	np.mean(x)
분산 (variance)	데이터와 샘플 평균간의 거리의 제곱의 평균이다. 이 값이 작으면 데이터가 모여 있는 것이다. 반대로 이 값이 크면 흩어져 있는 것 np.var(x)

Numpy 기술통계(descriptive statistics)

표준 편차 (standard deviation)	분산의 제곱근 값이다. np.std(x)
최대값 (maximum)	데이터 중에서 가장 큰 값 np.max(x)
최소값 (minimum)	데이터 중에서 가장 작은 값 np.min(x)

Numpy 기술통계(descriptive statistics)

중앙값 (median)

데이터를 크기대로 정렬 하였을 때 가장 가운데 있는수를 의미한다. 만일 데이터의 갯수가 짝수인 경우에는 중앙의 두 수의 평균을 사용

np.median(x)

Numpy 기술통계(descriptive statistics)

사분위수 (quartile)

- 데이터를 크기대로 정렬하였을 때 1/4, 2/4, 3/4 위치에 있는 수를 말한다.
- 1사분위수, 2사분위수, 3사분위수라 한다.
- 데이터가 100개가 있다면 25번째 순서가 1사분위수 이다.
- 2사분위수는 중앙값과 같다.
- 위치를 1/100단위로 나눈 분위수(percentile)를 사용하기도 함.
- 1사분위수는 25% 백분위수와 같다.

Numpy 기술통계(descriptive statistics)

사분위수 (quartile)

```
np.percentile(x, 0)      #최솟값
np.percentile(x,25)      #1사분위수
np.percentile(x,50)      #2사분위수
np.percentile(x,75)      #3사분위수
np.percentile(x,100)     # 최댓값
```

sum

여러개의 수를 연속하여 더하는 연산

```
np.sum(x)
```

product

여러개의 수를 연속하여 곱하는 연산

```
np.product(x)
```

Numpy 배열

🎯 같은 종류의 데이터를 담는 자료형 (동일한 타입의 값)

rank

배열 차원 (ndim)

shape

각 차원 크기를
튜플로 리턴(행, 열)

dtype

배열에 저장된
자료를 알려주는
객체

Numpy 다차원배열

🎯 ndarray

다차원 배열 객체

array()

배열 생성

reshape()

배열 모양 변경

ones()

1이 들어있는 배열 생성

eye()

단위 행렬 생성

arange()

배열 생성 시 지정 값까지 자동 생성

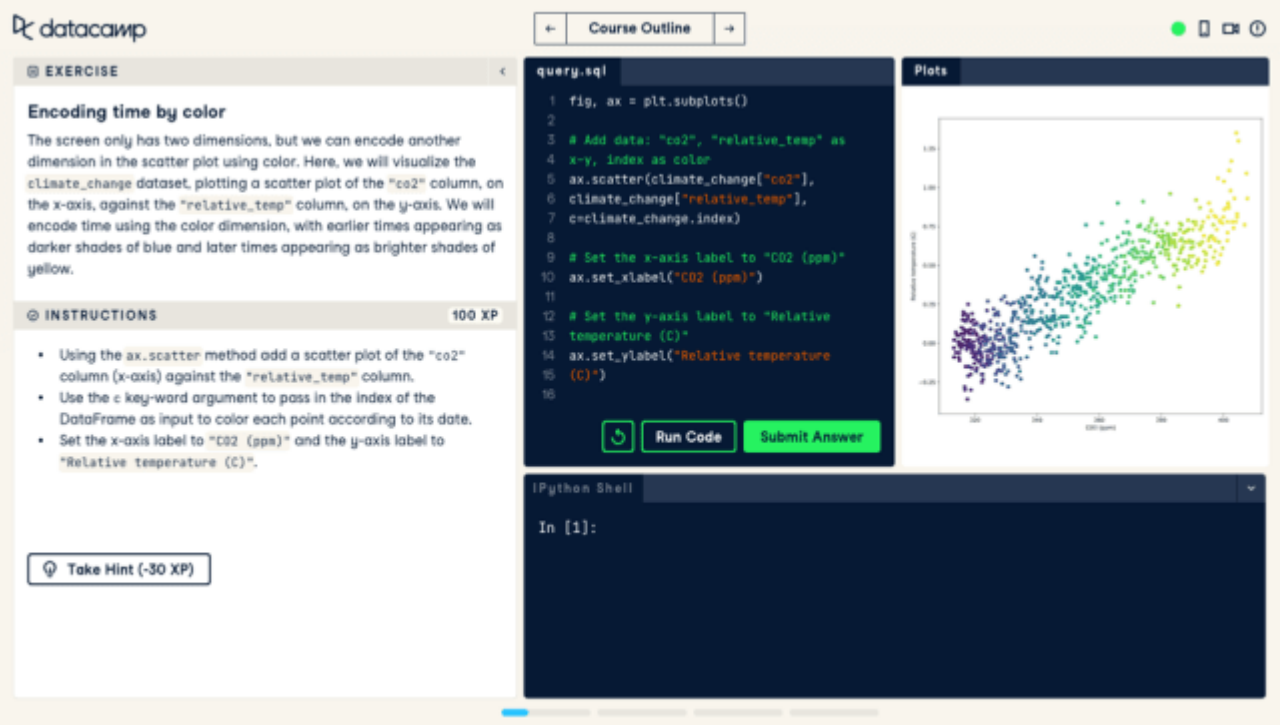
zeros()

0이 들어있는 배열 생성

full()

모든 값이 특정 상수인 배열을 생성

Numpy Cheet Sheet



출처 : https://www.datacamp.com/community/data-science-cheatsheets

Pandas

출처 : https://pandas.pydata.org/

- pandas는 데이터 조작 및 분석을 위해 Python 프로그래밍 언어로 작성된 소프트웨어 라이브러리
- CSV, Excel, JSON 등의 데이터를 읽고 원하는 데이터 형식으로 변환
- Series, Dataframe

Pandas 데이터 입출력 도구

File Format	Reader	Writer
CSV	read_csv	to_csv
JSON	read_json	to_json
HTML	read_html	to_html
local clipboard	read_clipboard	to_clipboard
MS Excel	read_excel	to_excel
HDF5 Format	read_hdf	to_hdf
SQL	read_sql	to_sql

출처 : http://pandas.pydata.org

Pandas Cheet Sheet

출처 : https://pandas.pydata.org/Pandas_Cheat_Sheet.pdf